Regression and Multimodal Learning to Aid Diagnosis in Ophthalmology and Histopathology

Santiago Toledo Cortés

Universidad Nacional de Colombia
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2023

Regression and Multimodal Learning to Aid Diagnosis in Ophthalmology and Histopathology

Santiago Toledo Cortés

Submitted to the Engineering School of the Universidad Nacional de Colombia, in fulfillment of
the requirements for the degree of:
Doctor of Engineering
Systems and Computer Engineering

Advisor:
Fabio A. González Ph.D.

Research area:
Machine Learning and Intelligent Systems

Research Group:
MindLab Research Group

Universidad Nacional de Colombia
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2023

Progress; it is extraordinary to witness the
intellectual progress of humanity.

Acknowledgements

To my wife María, to my parents Nestor and Astrid, to my brother Sergio, and to my family.

Thanks to the advice of Professor Fabio, who taught me the craft of research. Thanks to the advice of my colleagues at the MindLab, who taught me the details of many of the things I needed, and others that I didn't even know I needed. To Victor Contreras, Andrés Pérez, Melissa de la Pava, Diego Useche, Glenn Amaya, Sebastián Neiza, Óscar Perdomo, Andrés Rosso and Alejandro Gallego. Thanks to Professor Henning Müller at the MedGIFT group. Thanks to Minciencias and Google Research funding.

Thank you for your patience and time.

Abstract

Regression and Multimodal Learning to Aid Diagnosis in Ophthalmology and Histopathology

The main contribution of this thesis is the development of probabilistic machine learning models to support disease diagnosis from medical data sources. We show how a probabilistic approach offers great versatility in exploiting all available information about the target task. Based on the mathematical formalism of quantum mechanics, we develop and apply machine learning models that allow us to handle the flow of information using density matrices in different ways. We develop mechanisms that can naturally encode not only categorical but also ordinal information, and can also merge different data modalities. Furthermore, we show that the proposed models are naturally interpretable, which allows and facilitates their use in sensitive domains such as health applications. In particular, our models are tested in the diagnosis of several eye diseases and prostate cancer. First, we show the effectiveness and benefit of using regression models in the diagnosis of eye diseases of genetic origin. We then demonstrate the importance of including disease grading information and performing discrete regression to improve the performance of the binary diagnosis of diabetic retinopathy and prostate cancer. We show that a probabilistic interpretation of the results provides information on the uncertainty of the models, which can also be used in training processes. Finally, the proposed framework allows us to encode information using kernel functions, which in turn allows us to naturally introduce flexible information fusion mechanisms and thus to address multimodal tasks. Overall, we show that incorporating ordinal and multimodal information using probabilistic kernel-based frameworks allows learning better data representations, which improves the performance of the models and provides them with a higher level of interpretability.

Keywords:
Deep Learning, Histopathology, Kernel Methods, Medical Image Analysis, Multimodal Learning, Ophthalmology, Ordinal Regression, Probabilistic Models, Quantum Machine Learning.

Resumen

Regresión y aprendizaje multimodal como ayuda al diagnóstico en oftalmología e histopatología

La principal contribución de esta tesis es el desarrollo de modelos probabilísticos de aprendizaje de máquina para apoyar el diagnóstico de enfermedades a partir de información médica. Mostramos cómo un enfoque probabilístico ofrece una gran versatilidad al momento de aprovechar toda la información disponible sobre la tarea objetivo. Basándonos en el formalismo matemático de la mecánica cuántica, desarrollamos y aplicamos modelos de aprendizaje que nos permiten manejar el flujo de información utilizando matrices de densidad de diferentes maneras. Desarrollamos mecanismos que pueden codificar de forma natural no sólo información categórica, sino también ordinal, y que también pueden fusionar distintas modalidades de información. Además, demostramos que los modelos propuestos son naturalmente interpretables, lo que permite y facilita su aplicación en dominios sensibles como las aplicaciones médicas. Precisamente, en este trabajo probamos nuestros modelos en tareas específicas de diagnóstico de enfermedades oculares y cáncer de próstata. En primer lugar, mostramos la eficacia y el beneficio de usar modelos de regresión en el diagnóstico de enfermedades oculares de origen genético. A continuación, demostramos la importancia de incluir información sobre el estadio de las enfermedades y realizar una regresión discreta para mejorar el rendimiento del diagnóstico binario de la retinopatía diabética y el cáncer de próstata. Demostramos que la interpretación probabilística de los resultados proporciona información sobre la incertidumbre de los modelos, que puede utilizarse también en los procesos de entrenamiento. Por último, los modelos propuestos nos permiten codificar la información mediante funciones kernel, que a su vez nos permiten introducir de forma natural mecanismos de fusión de información, flexibles y versátiles, y con estos abordar tareas multimodales. En conjunto, demostramos que la incorporación de información ordinal y multimodal mediante modelos probabilísticos basados en funciones de kernel permite aprender mejores representaciones de los datos, lo que mejora el rendimiento de los modelos y les proporciona un mayor nivel de interpretabilidad.

Palabras clave:
Aprendizaje Profundo, Análisis de Imágenes Médicas, Aprendizaje de Máquina Cuántico, Aprendizaje Multimodal, Histopatología, Métodos de Kernel, Modelos Probabilísticos, Oftalmología, Regresión Ordinal.

Esta tesis de doctorado se sustentó el 7 de diciembre de 2023 a las 9:00 a.m., y fue evaluada por los siguientes jurados:

Lola Xiomara Bautista Rozo, Ph.D.
Profesor Asistente
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander

Francisco Albeiro Gómez Jaramillo, Ph.D.
Profesor Asociado
Departamento de Física, Estadística y Matemáticas
Universidad Nacional de Colombia

Luis Fernando Niño Vásquez, Ph.D.
Profesor Titular
Departamento de Ingeniería de Sistemas e Industrial
Universidad Nacional de Colombia

Henning Müller, Ph.D.
Profesor Titular
Departamento de Radiología e Informática Médica
Universidad de Ginebra, Suiza

Content

List of Figures

List of Tables

1. Introduction

1.1. Motivation

To diagnose certain eye diseases, it is necessary to carry out medical examinations and diagnostic imaging such as fundus photography or tomography. Common eye diseases include diabetic retinopathy (DR), diabetic macular edema (DME), glaucoma and age-related macular degeneration (AMD). However, the acquisition and manual reading of these images and examinations require very expensive equipment and high levels of professional training [112]. The situation is similar for prostate cancer (PCa). PCa is currently the second most common cancer among men in America [134]. Early detection allows for more treatment options and a higher chance of success, but while there are several methods of initial screening, a definitive diagnosis of PCa can only be made with a prostate biopsy, and this diagnosis usually occurs at an advanced stage of the disease [62].

The need for specialised personnel and equipment, coupled with the lack of healthcare in developing countries, results in a large number of undiagnosed patients. According to the World Health Organization, more than 2 billion people are visually impaired or blind, of whom at least 1 billion have visual impairment that could have been prevented or is still untreated [164]. Regarding cancer, in the United States alone it is estimated that around 30% of PCa patients are not diagnosed and treated in time [42]. To make matters worse, these figures are expected to worsen due to factors such as the ageing population and current lifestyles [62].

Deep learning models such as deep convolutional neural networks (CNNs) have been shown to be effective in analysing images and visual content of all kinds: from X-rays to diagnose osteoporosis [84] to MRIs to diagnose brain diseases [175]. In the case of eye diseases, many automatic analysis mechanisms have been developed to detect retinal lesions, diabetic retinopathy, glaucoma, etc, using fundus images, angiography and tomography [151]. In the case of histopathology image analysis, there are also many automatic diagnostic models that even follow standard medical-inspired pre-processing steps, such as performing an analysis based on small patches extracted from the whole slide image [67].

However, although physicians have established different stages on a progressive scale for many diseases, and although this information is usually available, binary categorical labels are generally preferred for training these models [93]. But the progress of a disease is not as simple. The degeneration associated to any disease is not a discrete jump between *healthy* and *ill*, but a progressive continuum [93]. Therefore, the stages or grades given by physicians are an attempt to discretize continuous behavior. Although not completely accurate, this information is useful in the generation of automated systems if a model with appropriate descriptive power is used.

As mentioned above, CNNs represent the state of the art for automated medical image analysis. However, treating the task of classifying medical datas as a simple categorical problem with conventional neural networks has two major drawbacks: first, the ordinal information of the grades is not taken into account in the training process, and second, the predictions of the models, usually subject to a softmax activation function, cannot be interpreted as probability distributions [154]. This brings us to another very important issue: the interpretability of the models. It is already well known that a trained model ends up being a black box, producing results without much explanation. This is an obstacle to real implementations of models that are intended to support diagnosis of any disease [141].

In addition, much of the work on automated support for the diagnosis of ophthalmological diseases and prostate cancer has focused on image data, but other types of modalities, such as text or clinical data, are a potential source of information. One way to exploit all these possible sources of information, and also to achieve interpretability, is through probabilistic models capable of capturing and combining the patterns and tinformation structures present in ordinal labels and multimodal sources. Precisely, this research focuses on the development of algorithms capable of processing ordinal and multimodal information using probabilistic schemes. The main research question motivating this research is: *What is the effect of including ordinal information and other data modalities in the development of probabilistic algorithms to support the diagnosis of ocular diseases and histopathology?*

1.2. Problem Statement

Many diseases are perfectly treatable when detected on time. Among these diseases, diabetic retinopathy or glaucoma, are one of the leading causes of blindness worldwide. For its part, prostate cancer is the second most common cancer in men worldwide [134]. However, in developing countries, the lack of trained personnel, and the low coverage of health systems, mean that approximately 40% of affected patients do not have access to specialized medical equipment that can give a timely diagnosis [170] [162] [100]. Diagnostic methods for the detection of eye diseases includes color photographs of the posterior pole of the eye, optical coherence tomography and some others. For prostate cancer, while there are several methods of initial screening, a concrete diagnosis can only be made with a prostate biopsy. In any case, the diagnosis requires interpretation by a specialist, being susceptible to inter-observer variability and not always available in health systems [103] [142].

Computerized assistance methods have been developed to make the diagnostic process faster and more affordable. These methods should be able to jointly analyze all of the available information for a patient when supporting the process of diagnosis of a disease. Several studies applied to the diagnosis and treatment of other diseases, such as Alzheimer's and cancer [175] [76], have shown that the development of ordinal regression and multimodal methods, capable of jointly analyzing multiple types of information, leads to better results in terms of precision and even interpretability [5].

But in the area of eye diseases and histopathology there is still much to explore. This is partly due to the lack of ordinal regression models and the lack of multimodal databases [105]. Most of the learning models developed to support the diagnosis of eye diseases and histopathology have focused on analyzing information from a single modality in a binary (healthy/unhealthy) schema, leaving aside ordinal disease-grading and complementary information from medical records [5].

This project aims to design and evaluate the performance of regression, ordinal regression and multimodal learning methods for supporting the diagnosis of eye diseases and prostate cancer. Mechanisms for learning representations should be explored, that allows the capture of information from different modalities and different grades of the same disease, in an expressive and compact way, and furthermore that allows fusion mechanisms that exploit the intrinsic correlation between them. A satisfactory solution to this general challenge requires answering some particular research questions:

- How to effectively and efficiently learn ordinal representations for different data modalities related with the diagnosis of eye diseases and histopathology?
- How to design ordinal and multimodal representations for data modalities related with the diagnosis of eye diseases and prostate cancer?
- How regression-based and multimodal approaches affects the performance of automatic models for the diagnosis of eye diseases and prostate cancer?

1.3. Objectives

1.4. Main goal

To design, implement and evaluate models based on deep learning to support the detection and diagnosis of eye diseases and prostate cancer using ordinal information and multiple information sources.

1.5. Specific objectives

- To design a strategy for learning representations for different data modalities or ordinal data structures, associated with the diagnosis of ophthalmological and histopathological diseases.

- To design a strategy for summarization and fusion mechanisms for different data modalities associated with the diagnosis of ophthalmological and histopathological diseases.

- To design methods for the automatic analysis of different data modalities to support the diagnosis of ophthalmological and histopathological diseases.

- To systematically evaluate the performance of the proposed methods.

1.6. Contributions

The main contributions of this thesis are listed below. Overall, we have developed several diagnostic support models capable of exploiting ordinal relationships and multiple data modalities from the degenerative disease process itself, finally resulting in probabilistic models that allow for interpretability and versatility.

- ***Toledo-Cortes, S.**, Dubis, A. M., González, F.A., Müller, H. Deep Density Estimation for Cone Counting and Diagnosis of Genetic Eye Diseases From Adaptive Optics Scanning Light Ophthalmoscope Images. In: Translational Vision Science Technology 12 (2023), 11, Nr. 11, S. 25–25. – ISSN 2164–2591*

  In this paper we show the importance of cellular pattern recognition in the diagnosis of genetic eye diseases. Diseases such as retinitis pigmentosa affect the cellular pattern of the retina and ultimately lead to partial or total loss of vision. Adaptive optics scanning light ophthalmoscope images allow us to assess retinal damage at the cellular level in living patients, identifying changes before macroscopic lesions become apparent. Using techniques originally developed for segmentation tasks, we have developed a regression model to estimate cell density in Adaptive Optics Scanning Light Ophthalmoscope images. We show that incorporating density estimation in advance improves the subsequent performance of disease classification models.

- ***Toledo-Cortés S.**, de la Pava M., Perdomo O., González F.A. Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification. In: Ophthalmic Medical Image Analysis. OMIA 2020. Lecture Notes in Computer Science, vol 12069. Springer, Cham., 2020, S. 206–215*

  In this work we developed a hybrid model that integrates Gaussian processes with deep learning models for image analysis. We address the problem of diagnosing diabetic retinopathy by training the model as a regressor using the five grades of DR. We show that this approach allows to improve the traditional binary approach and to perform an uncertainty analysis of the model, where the confidence of the predictions is understood as the associated variance, given that each prediction of the probabilistic model is a probability distribution.

- ***Toledo-Cortés S.***, *Useche D.H., Müller H., González F.A. Grading diabetic retinopathy and prostate cancer diagnostic images with deep quantum ordinal regression. In: Computers in Biology and Medicine 145 (2022), S. 105472. – ISSN 0010–4825*

  In this paper, we retained the probabilistic regression approach for diagnostic models but developed quantum-inspired models for them. In addition to developing a family of algorithms based on quantum measurements using density matrices and random features, we showed that performing ordinal regression on the task of diagnosing prostate cancer from histopathology images improved the summation process (necessary for patch-based analysis), the binary diagnosis, and also provided the model with a first level of interpretability by calculating the uncertainty of the predictions. We also extended the ideas to the diagnosis of diabetic retinopathy. The results were robust and again gave better results in the final binary task.

- ***Toledo-Cortés S.*** *González F.A. Multimodal Quantum Kernel Fusion for Glaucoma Diagnosis. In preparation to be submitted.*

  Finally, the previous proposed models based on density matrix processing can be formulated in terms of kernel functions to represent the data. This allows us to create highly versatile fusion mechanisms with which we address multimodal problems applied to prostate cancer and glaucoma diagnosis.

Software

- **Toledo-Cortés S.**, Dubis, A. M., González, F.A., Müller, H. (2022). Implementation of CODE and CODED for AOSLO image analysis.
  `https://github.com/stoledoc/AOSLO-CNN_Diagnosis_Counting`

- **Toledo-Cortés S.**, de la Pava M., Perdomo O., González F.A. (2020). Implementation of Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis.
  `https://github.com/stoledoc/DLGP-DR-Diagnosis`

- **Toledo-Cortés S.**, Useche D.H., Müller H., González F.A. (2021). Implementation of Deep Quantum Ordinal Regressor for Prostate Cancer and Diabetic Retinopathy Grading.
  `https://github.com/stoledoc/DQOR`

Awards

The proposal of this thesis was awarded by the Latin America Reseach Awards - Google LARA 2019. The development of this work was partially funded by Google Research. `https://latam.googleblog.com/2019/11/celebramos-los-ganadores-de-la-septima-edicion-de-LARA-programa-becas-investigacion-latinoamerica.html`

1.6.1. Other Contributions

During the development of this work, other publications, software, oral presentations, posters, and master's thesis advising were made on topics related to the focus of the research, as well as on other independent topics:

Papers

- **Toledo-Cortés S.**, Castellanos-Martinez I.Y., Gonzalez F.A. (2019) Large Scale Learning Techniques for Least Squares Support Vector Machines. In: Vera-Rodriguez R., Fierrez J., Morales A. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2018. Lecture Notes in Computer Science, vol 11401. Springer, Cham. `https://doi.org/10.1007/978-3-030-13469-3_1`

- Perdomo-Charry, O., Pérez-Pérez, A., de-la-Pava-Rodríguez, M., Ríos-Calixto, H., Arias-Vanegas, V., Lara-Ramírez, J., **Toledo-Cortés S.**, Camargo-Mendoza, J., Rodríguez-Alvira, F., González-Osorio, F. (2020). SOPHIA: System for Ophthalmic Image Acquisition, Transmission, and Intelligent Analysis. Revista Facultad De Ingeniería, 29(54), e11769. `https://doi.org/10.19053/01211129.v29.n54.2020.11769`

- Perdomo O.J., **Toledo-Cortés S.**, Orjuela A., González F.A. (2021) What You Need to Know About Artificial Intelligence: Technical Introduction. In: Ichhpujani P., Thakur S. (eds) Artificial Intelligence and Ophthalmology. Current Practices in Ophthalmology. Springer, Singapore. `https://doi.org/10.1007/978-981-16-0634-2_2`

- **Toledo-Cortés S.**, Useche, D.H., Gonzalez, F.A. (2021). Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression. `arXiv:2103.03188`

- Gonzalez, F.A., Gallego, A., **Toledo-Cortés S.**, Vargas-Calderon, V. (2022). Learning with Density Matrices and Random Features. Quantum Mach. Intell. 4, 23. `https://doi.org/10.1007/s42484-022-00079-9`

- **Toledo-Cortés S.**, Lara, J.S., Zambrano, A., González F.A., Rosero García, J. Characterization of Electricity Demand Based on Energy Consumption Data from Colombia (2023). International Journal of Electrical and Computer Engineering (IJECE), 13(5), 4798-4809. `https://doi.org/10.11591/ijece.v13i5.pp4798-4809`

- **Toledo-Cortés S.**, Zambrano, A., González F.A., Rosero García, J. Short-term Forecasting of Power Demand Kernel-based Methods for Short-term Forecasting of Power Demand. Submitted to Journal of Electrical and Computer Engineering.

- **Toledo-Cortés S.**, Serna J. E., González F.A., Romero E. Cognitive Impairment Inference in Parkinson Disease Patients from Spatio-Temporal Gait Assessment Using Machine Learning. To be submitted at journal.

Coadvisor

- Amaya-Cruz, G. H., (2023). Análisis de calibración en modelos de aprendizaje de máquina cuántico. Master's Thesis, Universidad Nacional de Colombia.

Posters

- **Toledo-Cortés S.**, Castellanos-Martinez I.Y., Gonzalez F.A. (2018) Large Scale Learning Techniques for Least Squares Support Vector Machines. The 23rd Iberoamerican Congress on Pattern Recognition (CIARP), Madrid, Spain.

- **Toledo-Cortés S.**, Castellanos-Martinez I.Y., Gonzalez F.A. (2019) Large Scale Learning Techniques for Least Squares Support Vector Machines. LatinX Workshop at the 33rd Neural Information Processing Systems (NeurIPS), Vancouver, Canada.

- **Toledo-Cortés S.**, de la Pava M., Perdomo O., González F.A. (2020) Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification. 23rd International Conference of Medical Image Computing and Computer Assisted Intervention (MICCAI). Virtual.

Presentations

- **Toledo-Cortés S.**, Castellanos-Martinez I.Y., Gonzalez F.A. (2018) Evaluating Kernel Approximation Thechniques for Large-Scale Learning. MAPI1, Universidad Nacional de Colombia, Bogotá, Colombia.

- **Toledo-Cortés S.**, Castellanos-Martinez I.Y., Gonzalez F.A. (2018) Large Scale Learning Techniques for Least Squares Support Vector Machines. The 23rd Iberoamerican Congress on Pattern Recognition (CIARP), Madrid, Spain.

- **Toledo-Cortés S.**, de la Pava M., Perdomo O., González F.A. (2020) Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification. 23rd International Conference of Medical Image Computing and Computer Assisted Intervention (MICCAI). Virtual.

1.6.2. Thesis Organization

The remaining chapters of the thesis are organized as follows:

- ***Chapter 2: Background and Related Work***. This chapter discusses previous work related to regression and multimodal methods applied to the analysis of medical data, specifically those related to eye disease and prostate cancer.

- ***Chapter 3: Deep Regression for Cone Density Estimation and Genetic Ophthalmic Diseases Diagnosis***. This chapter presents two new regression models based on deep learning for automatic cell counting in AOSLO images and for the detection of genetic eye diseases such as retinitis pigmentosa and Stargardt's disease.

- ***Chapter 4: Deep Probabilistic Regression for Diabetic Retinopathy Grading*** . This chapter presents a new hybrid model that combines deep learning with Gaussian processes to grade diabetic retinopathy from eye fundus images.

- ***Chapter 5: Quantum Measurements***. This chapter presents the framework for the upcoming methods. The core behind the learning models based on quantum measurement is explained here.

- ***Chapter 6: Deep Quantum Measurement Regression for Diabetic Retinopathy and Prostate Cancer Grading***. In this chapter we present the Quantum Measurement Regressor, a deep probabilistic model that allows for ordinal regression. Here we show its implementation in the task of grading diabetic retinopathy and prostate from medical images.

- ***Chapter 7: Multimodal Deep Kernel Quantum Measurement for Prostate Cancer and Glaucoma Diagnosis***.This chapter presents the Multimodal Kernel Quantum Measurement Unit, a modular processing unit that uses kernel functions as information encoders, allowing multimodal implementations. Here we show its effectiveness for prostate cancer and glaucoma diagnosis.

- ***Chapter 8: Conclusions***. The final chapter presents the main conclusions and discussions of this dissertation, summarising the main contributions and highlighting the main findings. Some future research directions are also presented and discussed.

2. Background and Related Work

This chapter introduces the fundamental background necessary for understanding and developing the following topics of the thesis. We begin with a general overview of the diseases we will focus on and give an overview of the state of the art in regression and multimodal machine learning models applied to diagnostic support.

2.1. Medical Background

2.1.1. Ophthalmological Diseases

Several medical tests are used to diagnose and monitor the progress of various diseases. An examination of the ocular fundus allows visual examination of the retina (the light-sensitive tissue at the back of the eye) and is one of the most common (although not definitive) tests for the following conditions:

Diabetic Retinopathy (RD)

Diabetic retinopathy (DR) is a progressive, chronic eye disease that is one of the most common complications of diabetes. It affects the blood vessels in the retina, the light-sensitive tissue at the back of the eye. In people with diabetes, prolonged exposure to high blood sugar levels can damage the blood vessels, leading to various abnormalities in the retina. These lesions include microaneurysms (MAs), haemorrhages and dark spot-like lesions known as exudates. The early stages of DR can be asymptomatic, but as the disease progresses, symptoms such as blurred vision, floaters and even vision loss may occur. DR is usually diagnosed by a comprehensive eye examination that includes visual acuity testing, dilated eye exams and imaging techniques such as eye fundus imaging or optical coherence tomography (OCT) [162].

Glaucoma

Glaucoma is an increase in the internal pressure of the eyeball, mostly caused by obstruction of the drainage pathways of the eye's internal fluids [113]. As the pressure increases, the head of the optic nerve suffers a compression that causes progressive irreversible damage (death of nerve cells), leading to total blindness. The detection of glaucoma begins with the measurement of intraocular pressure with a tonometer. Although this is not definitive, it does give a clue as to when other tests should be carried out. In the fundus image, the area associated with the optic nerve, known as the optic disc (OD), is analysed by looking for patterns in the morphological parameters provided by the image.

Retinitis Pigmentosa (RP) and Stargardt Disease (SGD)

These are two forms of inherited degenerative eye disease that primarily affect the retina. RP is characterised by a progressive deterioration of the photoreceptor cells [22], leading to a gradual loss of peripheral vision and, in advanced stages, to total blindness. It often presents with symptoms such as night blindness and reduced visual field. Stargardt's disease, the most common form of inherited juvenile macular degeneration, primarily affects the macula, the central area of the retina responsible for sharp, straight ahead vision [150]. It typically manifests itself in childhood or adolescence, resulting in loss of central vision but preservation of peripheral vision. Although there is currently no cure for either condition [32] [64], ongoing research is focusing on therapies such as gene therapy, stem cell treatment and artificial retinal implants to mitigate their effects. They are diagnosed by a combination of clinical examination, symptom history and specialised tests.

2.1.2. Prostate Cancer (PCa)

Prostate cancer (PCa) is one of the most common cancers in men in America and is also one of the leading causes of cancer death among men of all races and Hispanic origin populations [41]. The risk of prostate cancer increases with age, and it's most common after age 50. In many cases, prostate cancer is a slow-growing cancer. It occurs when abnormal cells start to grow uncontrollably in the prostate, a small walnut-shaped gland that produces semen in men. While some types of prostate cancer grow slowly and may require little or no treatment, others can be aggressive and spread quickly [136].

Diagnosing prostate cancer usually starts with screening tests [45]. The two most commonly used tests for prostate cancer are the prostate-specific antigen (PSA) blood test and the digital rectal examination (DRE). Elevated levels of PSA can indicate the presence of cancer, but it's not definitive because other conditions, such as an enlarged or inflamed prostate, can also raise PSA levels. The DRE involves the doctor manually examining the prostate for any abnormalities. If these tests suggest the possibility of cancer, a biopsy will be done. A biopsy involves taking a sample of tissue from the prostate and examining it under a microscope. Based on the microscopic appearance of the tissue, the aggressiveness of prostate cancer is then measured using the Gleason grading system [90]. It helps pathologists classify prostate cancer cells into different grades, or patterns, to provide information about the tumour's behaviour and potential to spread.

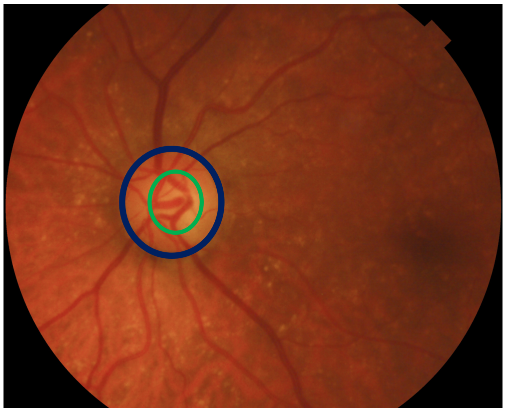


Figure 2-1.: Eye fundus image example from the PAPILA dataset [78]. This is an RGB photograph of the inner posterior part of the eye, showing an overview of the retinal tissue. Dark blue and green circles indicate the optic disc (OD) and optic cup (OC) zones, respectively. In diabetic retinopathy, lesions can occur throughout the retina, particularly in the vascular system. Glaucoma, on the other hand, usually only concerns the optic nerve head (in the zone of the optic disc).

2.2. Related Datasets

In the field of ophthalmic and histopathology research, the availability and effective use of data sets plays a crucial role in the advancement of diagnostic methods. These datasets, often consisting of retinal images, tomographies, whole slide images and associated medical records, provide invaluable resources for training and validating machine learning models. We will focus on three types of images: eye fundus images, adaptive optics scanning laser ophthalmoscope images, and whole slide images.

Eye Fundus Images (EFI)

An eye fundus image (see Figure **2-1**) is a photograph of the inner surface of the eye, including the retina, optic disc, macula and posterior pole (the central part of the retina). This image is taken using a special low-power microscope with an attached camera called a fundus camera. The resulting image can provide a wealth of information about the health of the eye and may show signs of various eye diseases such as glaucoma, macular degeneration, retinal detachment or diabetic retinopathy. Fundus photography is a common part of comprehensive eye examinations, particularly for patients with known eye disease or at risk of certain systemic diseases that can affect the eye, such as diabetes or hypertension [91].

Adaptive Optics Scanning Laser Ophthalmoscopy (AOSLO) Images

An Adaptive Optics Scanning Laser Ophthalmoscope is an imaging tool that allows clinicians and researchers to visualise the retina, the light-sensitive tissue at the back of the eye, at a cellular level (see Figure **2-2**). The adaptive optics component of the device corrects for distortions in the image, often caused by the outer layers of the eye, resulting in a much clearer and more detailed view of the retina than traditional imaging techniques. This technology allows individual photoreceptor cells (the cells responsible for converting light into electrical signals), blood vessels and other retinal cells to be seen in vivo, which can help diagnose and monitor various eye diseases. It can also be used in research to study the structure and function of the retina in great detail [18]. AOSLO is not yet widely used in clinical practice, but it has significant potential to improve our understanding and treatment of many eye diseases [36].

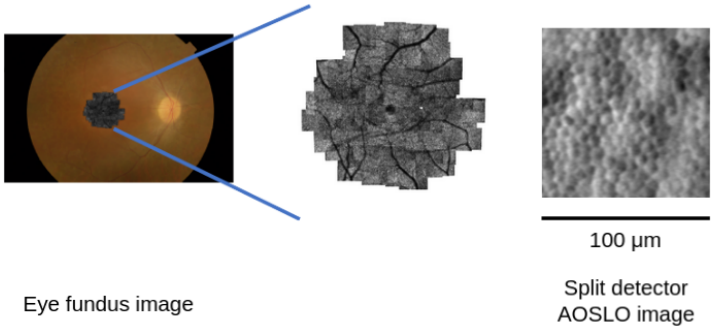


Figure 2-2.: Sample of a split detector AOSLO image. This image shows the detailed structure of the retina. The distinctive patterns of light and dark areas correspond to individual photoreceptor cells.

Whole Slide Images (WSI)

Whole Slide Images (WSIs) (see Figure **2-3**), also known as digital slides, are high-resolution digital representations of entire histological glass slides used in medical image analysis for histopathology [67]. WSIs are created by scanning whole glass slides with specialised slide scanners, capturing the entire tissue section at high magnification. These digital images preserve the spatial and morphological details of tissue samples, providing a comprehensive view that can be examined and analysed by pathologists and researchers. WSIs have revolutionised histopathology by enabling remote access, facilitating collaboration and supporting quantitative analysis using machine learning algorithms. By digitising histological slides, WSIs offer numerous benefits, including easier sharing and the potential for automated image analysis to aid in the diagnosis, prognosis and research of various diseases [81].

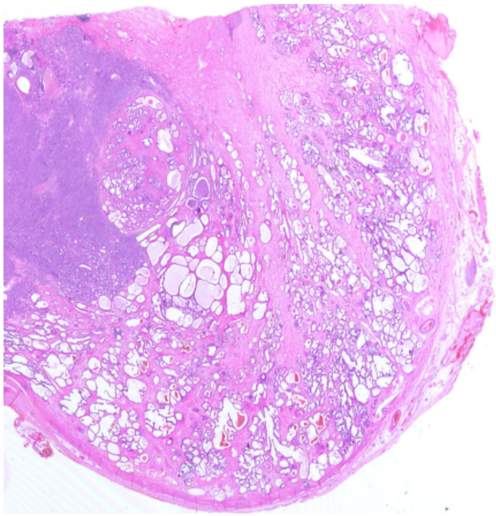


Figure 2-3.: An example of a Whole Slide Image from the TCGA-PRAD dataset [67], illustrating the detailed histopathological features associated with prostate cancer at a cellular level. This image aids researchers in studying the morphological variations across different stages of prostate cancer.

The global drive to improve and innovate diagnostic techniques has led to the creation of numerous public and private datasets. These compilations contain ocular fundus images or whole slide images of patients affected by a variety of diseases. A comprehensive summary of these accessible public datasets is presented in table **6-4**.

Table 2-1.: A summary of some of the available public datasets on ophthalmology and histopathology.

Dataset	**Disease and description**
DRIVE	DR. 40 eye fundus images.
E-OPHTHA	DR. Segmentation for exudates and microaneurysms. 463 eye fundus images
EYE PACS	DR. 88702 eye fundus images.
APTOS	DR. 13000 eye fundus images.
ONHSD	DR and Glaucoma. 49 eye fundus images.
HRF	DR and Glaucoma. 90 eye fundus images, with segmentation of clinical signs.
MESSIDOR	DR and DME. 1200 eye fundus images.
iDRID	DR and DME. 516 eye fundus images. Labeled with the degree of DR and DME.

Dataset	**Disease and description**
STARE	DR and AMD. 400 eye fundus images with masks of the blood vessels in black and white.
ARIA	DR and AMD. 143 eye fundus images.
DRIONS-DB	Glaucoma. 110 eye fundus images. Segmentation of the optic nerve head.
ORIGA-650	Glaucoma. 650 eye fundus images.
INSPIRE-AVR	Glaucoma. 40 eye fundus images.
RIM-ONE	Glaucoma. 783 eye fundus images. Segmentation of optic disc and optic cup.
ACHIKO-K	Glaucoma. 258 eye fundus images.
DRISHTI-GS	Glaucoma. 101 eye fundus images. Segmentation of optic disc and optic cup.
RIGA	Glaucoma. 760 retinal fundus images.
REFUGE	Glaucoma. 1200 eye fundus images. Segmentation of optic disc and optic cup.
PAPILA	Glaucoma. 488 eye fundus images. Segmentation of optic disc and optic cup, and structured clinical data.
CUNEFARE	Stargardt's Disease. 260 AOSLO images. Coordinates of the location of photoreceptor cells.
PANDA	PCa. 21135 images. Segmentation according to Gleason Score.
TCGA-PRAD	PCa. More than 20000 samples with WSI and clinical data records.
PESO	PCa. 102 WSI samples with Gleason score.
TCIA-Ovarian	Ovarian Cancer. 288 WSI from 78 patients, and clinical information regarding tratment with bevacizumab.
ERCPMPv1	Colorectal cancer. Histologic Images from 192 patients with colorectal polyps.
BreCaHAD	Breast Cancer. 162 breast cancer histopathology images.

2.3. Performance Evaluation of Machine Learning Models

Evaluating the performance of machine learning models, especially regression and multimodal learning models in the field of ophthalmology and histopathology, is crucial to ensure their diagnostic accuracy, reliability and clinical applicability. Several metrics and methodologies can be applied, including but not limited to the following:

- Accuracy: The most basic metric is accuracy, defined as the number of correct predictions made by the model divided by the total number of predictions. Although easy to understand, accuracy can be misleading if classes are unbalanced. Mathematically, it is defined as:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

- Confusion Matrix: The confusion matrix can give us a more comprehensive view of the model's performance, detailing true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). From this we can derive other useful metrics such as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).
- Receiver operating characteristic (ROC) curve and area under the ROC curve (AUC-ROC): The ROC curve is a plot of the true positive rate (TPR, or sensitivity) versus the false positive rate (FPR, or 1-specificity) for the different possible cutpoints of a diagnostic test. The AUC-ROC can provide a single measure of a model's performance, where 1.0 indicates perfect classification and 0.5 indicates a model that is no better than chance.

$$\text{Sensitivity (TPR)} = \frac{TP}{TP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP}$$

- Precision (Positive Predictive Value): Precision is defined as the number of true positives divided by the sum of true positives and false positives. It is a measure of a classifier's exactness.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall (Sensitivity or True Positive Rate): Recall is defined as the number of true positives divided by the sum of true positives and false negatives. It quantifies a classifier's completeness.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- F1-Score: The F1-Score is the harmonic mean of Precision and Recall, and ranges from 0 to 1. It is a way of combining the precision and recall of the model, and it is defined as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Mean Squared Error (MSE) and Mean Absolute Error (MAE): For regression and ordinal regression models, MSE and MAE are commonly used metrics that provide information on how far the model predictions are, on average, from the actual values.

  $$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2, \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

  where Y_i represents the actual values, $\hat{Y}_i$ represents the predicted values, and n is the number of observations.

- Clinical Validation: Finally, model performance needs to be evaluated in a clinical context. This often involves prospective validation with new patients and, ultimately, randomized controlled trials to demonstrate that use of the model improves patient

When interpreting these metrics, it's important to consider not only their absolute values, but also how they compare with other models and with clinicians' needs. For example, in a disease where false negatives can have serious consequences, a model with higher sensitivity may be preferred, even if it has lower overall accuracy.

2.4. Machine Learning for Medical Applications

Machine learning has been successfully applied to the analysis of a wide range of medical data. Analysis of histopathological images, detection of heart disease, study of neuronal pathologies, treatment of bone fractures, drug development, genome research, etc. [77]. The recent increase in the understanding and automatic analysis of disease-related data is largely due to the increase in information and systems that can extract appropriate symptomatic patterns. Machine learning has served to facilitate what is generally manual work, reducing time and facilitating access to medical resources.

In the specific case of ophthalmological and histopathological diseases, significant progress has come from the development of computer vision models, which have been successfully applied to the task of identifying objects in natural images for more than a decade. Thanks to model training techniques such as *transfer learning* and *fine tuning*, it has been possible to harness the computational power of deep learning and successfully learn representations to handle visual information from medical images [73].

2.4.1. Deep Learning

Deep learning is a branch of machine learning that includes a whole family of algorithms that share a common feature: an architecture organised by hierarchical layers. The history of deep learning begins in the 40s, with simple algorithms that were essentially variations of linear regression methods [130]. Then, in the early 60s, the perceptron appeared, giving birth to the first artificial neural network (ANN). In 1965, the first deep learning architecture appeared, using feed-forward propagation in a multi-layer structure.

In short, an ANN consists of a set of computational units (neurons) organised in layers and interconnected in such a way that information moves through the layers from the input to the output layer. The input data is transformed depending on the parameters of the connections and the activation function of each neuron. The practical problem is to find the parameters that better describe the data patterns, and to achieve this, a cost function should be minimised. The cost function could measure the error rate or the precision and accuracy of the predictions made by the network, and it should be carefully chosen depending on the model and the task. In any case, the cost function depends on the parameters of the network links. Since 1970, backpropagation was introduced as a method to minimise the cost function. Backpropagation tracks the error in the opposite direction of the information flow into the network and is now part of the state of the art in the training process of neural networks [130].

2.4.2. Ordinal Regression

Regression analysis aims to establish a relationship between a dependent variable (or response variable) and one or more independent variables (or features). The primary aim is to understand how the dependent variable changes in relation to the independent variables. There are numerous types of regression models, each of which has its place depending on the nature of the data and the problem at hand. They have proven to be incredibly robust and useful in many fields, including economics, psychology, biology, and medical imaging and diagnostics.

However, while traditional regression models have proven incredibly useful, there are many cases where the response variable is not on a continuous scale, but rather an ordinal one. This brings us into the realm of ordinal regression. Ordinal regression, also called ordinal classification, comes into play when the dependent variable is ordinal, i.e. when the possible output values have a clear order or ranking, but the intervals between these values are not necessarily known or constant. An general example in medical diagnostics might be predicting the stage of a disease, which might be classified as *none*, *mild*, *moderate* or *severe*. Categorical classification models are not suitable for this type of problem because they do not take into account the inherent order in the response categories [8]. Ordinal regression models, on the other hand, do take this order into account and therefore tend to provide more accurate and meaningful predictions [8].

Ordinal regression tasks are not exclusive to the medical field. Therefore the development of this approach has occurred alongside the rest of machine learning, as an intermediate field between regression and classification. According to Gutierrez et al. [60], the taxonomy of previously proposed methods for ordinal regression are: first, *naïve* approaches, which are standard machine learning models for nominal classification or metric regression, second, *ordinal binary decomposition* approaches, which break down the problem into several binary sub-problems [46], and third, in which our proposals are framed, *threshold* models, which are based on a predictor that yields a real value, which is then approximated to an integer value. Depending on the particular problem, different models might perform better than others, so there is not an optimal ordinal regression approach [60].

2.4.3. Ordinal Regression for Medical Data Analysis

The use of regression models in tasks related to medical diagnosis should be naturally understood, as a disease is usually a process that follows stages according to some order. The progression of a disease is also generally continuous rather than discrete, and the study of several diseases has led to the establishment of a discrete scale for their progression, which is then used in the diagnostic procedure to take certain actions to prevent the disease from reaching the next stage. For example, in ophthalmology, particularly in DR, the international standard identifies 5 different stages from complete absence of the disease to *proliferative retinopathy* [146]. For PCa, the Gleason system does the same: it determines the severity of the cancer on a 5-stage scale. Curiously, however, most research into these diseases approaches diagnosis as a binary problem [81]. The main reason for this is the ultimate purpose of the diagnosis.

Again, for DR we have 5 stages: *healthy*, *mild DR*, *moderate DR*, *severe DR*, and *proliferative retinopathy* (see Figure **2-4**) [135]. In this order, the stages indicate the severity of the disease. However, from this scale there is a binarization of the diagnosis: *referable DR* vs. *non-referable DR* [158]. *Referable* means that the patient is fit to be seen by an specialised ophthalmologist. A patient who is *not referable* is a patient who can be treated with standard procedures without the need for a specialist to examine the case. Therefore the next step in treatment is ultimately a binary decision, even if the diagnosis is not purely binary. This is one of the reasons why machine learning models have traditionally approached these problems as binary categorical classification. And they have been very successful [60]. Models based on deep neural networks have shown excellent performance on many of the data sets mentioned above. The question is, why bother with an ordinal scale at all? The answer is: interpretability [144] [6] [146]. When the ordinal information in the data is taken into account, representation learning suddenly has to take into account a more complex structure. This opens the door to a world of possibilities where it makes sense to re-understand the distribution of the data and how that distribution can explain the course of a disease. And it opens the door to a richer interpretation of the results, which are no longer simply a 0 or a 1, but a position on a broader scale. It begins to make sense to incorporate probabilistic methods, and new performance metrics are introduced.

Therefore, while it is true that there have been some applications of ordinal regression models in the medical field, there is not a clear and well-defined trend. Recently, ordinal

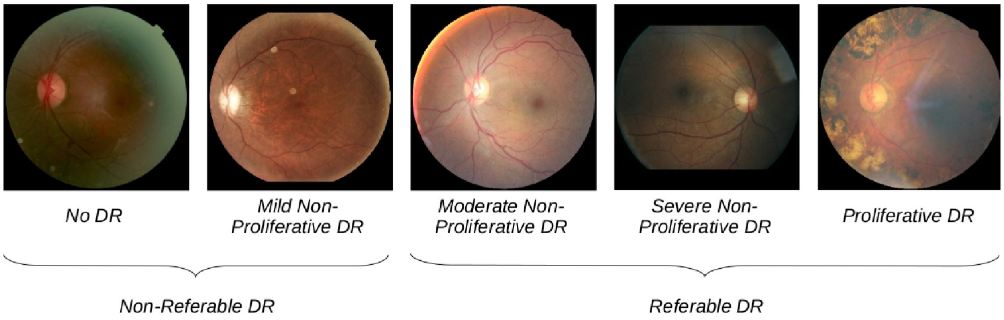


Figure 2-4.: Five possible grades for DR diagnosis [135]. Healthy cases correspond to grade 0. Grades 0 and 1 correspond to *non-referable* DR cases, while grades 2, 3 and 4 correspond to *referable* DR. Samples extracted from EyePACS dataset [40].

regression by binary classifiers has been applied to facial age estimation [109] [138], and diagnosis of Alzheimer's disease [89], taking advantage of the inherent ordinal severity of brain degeneration.

In addition, some challenges have been created from which useful datasets such as PANDA have emerged [17], and in general attempts have been made to adapt typical models so that their predictions span the spectrum of diagnostic possibilities [6], or to use this additional ordinal information tacitly during training [143].

Regarding our application of interest, most of the works for PCa have been focused on classifying Whole Slide Images (WSI) by *low* and *high* GS [81]. To train a model on 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 [67], the authors classify patches between *low*, and *high* GS, using various CNNs, and summarizing the patches at WSI level by a GS majority vote. Another approach by Tolkach et al. [149] uses a NASNetLarge CNN, and summarizes the GS of the patches by counting the probabilities per class. In Karimi et al. [69], they proposed training three CNNs for patches of different sizes and summarizing the probabilities by logistic regression.

Recently, however, there has been a growing interest in GS grading. Proof of this is the *Prostate cANcer graDe Assessment* (PANDA) Challenge [17], and the recently proposed CNN architectures, which include a combination of an atrous spatial pyramid pooling and a regular CNN [90], an Inception-v3 CNN with a support vector machine (SVM) [94], and a DeepLabV3+ with a MobileNet as the backbone in [72]. In [106], the authors use an InceptionV3 with a k-nearest-neighbor classifier to summarize the patch-level predictions in a heatmap. In [65], the authors implemented in parallel a categorical and an ordinal classification for Gleason patterns, training similar models with different loss functions, from the same data features. However, they used a softmax to return probabilities that, as mentioned before, cannot be interpreted directly as a probability distribution [154]. Other techniques for GS grading include Support Vector Machine Feature-Recursive Feature Elimination [126], and learning representations from *bag-of-words* features [159].

On the DR side, most works have been focused on a binary diagnosis based on deep neural networks [111] [28]. Tian et al. [143] used a deep CNN as the backbone for a model trained to optimize a combination of a metric loss and a focal loss function for soft labels, aiming to use the ordinal information of the DR stages. Additionally, Teresa Araujo et al. [6] proposed the DR|GRADUATE, a deep learning-based model whose last layer had the same number of neurons as classes, and with a Gaussian filter applied to the output. The model was trained with a loss function that controlled both the classification entropy and the standard deviation of the distribution. This strategy allowed them to infer, in addition to the DR grade, the uncertainty of the prediction.

Lastly, uncertainty quantification in ordinal regression has been analyzed in many studies to obtain more interpretable models for cases where reliability is important to the end user [133]. Studies have been carried out on the uncertainty of machine learning algorithms for organ classification [102] and on the estimation of tissue parameters in the operating room [3]. Furthermore, estimating the uncertainty of a model's prediction has been of great interest because it reduces the consequences of the blind use of the model's inference [71]. This is particularly relevant in medical settings, where misdiagnoses can have serious consequences for patients. Correspondingly, Leibig et al. [85] analyzed uncertainty information from deep neural networks for DR detection. The authors tested dropout-based Bayesian uncertainty estimation against alternative techniques, such as direct analysis of the softmax output of the network. They claimed that Bayesian approaches perform better for uncertainty estimation and showed that uncertainty-aware decisions can improve the overall grading process. Another approach is presented in [92], where stochastic batch normalization is used to calculate the uncertainty of the prediction of a model for DR level intervals estimation.

2.4.4. Multimodal Learning

Data may take a variety of forms and carry a variety of information about a single objective. For example, in speech recognition, humans combine auditory and visual information to distinguish between similar sounds of certain consonants [98]. In this case, visual information is in the movements of facial muscles [137]. Or, in many cases, a graphic is much better at describing the non-obvious information in a text. The aim of multimodal learning is to find the best way of combining information from different sources so that they complement each other and produce better results than if the sources were analyzed separately.

This multimodality problem leads to a representation problem. How do we learn the right representation for each modality? Data from different modalities may have different statistical properties, so a simple concatenation of representation features is not necessarily a good strategy [156]. You then have to look for latent spaces, and different strategies have been developed to find these possible spaces. However, representation is not the only challenge in multimodal learning. According to Baltrusaitis et al. [9], the challenges of multimodal machine learning include translation, alignment, fusion and co-learning.

Among these challenges, data fusion emerges as a core phase in the multimodal design process. Fusion is essentially looking for the best way to combine information from different sources in a way that complements and extracts information from individual sources. Various classifications can be made. In terms of the stage at which fusion takes place, there are three types of fusion: early, late and hybrid. In early fusion, a selection or combination of representations from each modality is made before the problem is solved. Several examples of this can be found in the literature [23], where fusion is performed at the feature level and then fed as input to a neural network. In some cases, early fusion gave better results than late fusion at the decision level. Late fusion combines the results of independent models to draw conclusions. An example of this can be found in [119], where fusion is performed with the high-level features of visual and audio modalities for (again) an emotion recognition task. In any case, the fusion mechanism and the stages of fusion are part of the settings which should be under examination in a model construction.

2.4.5. Multimodal Learning for Medical Data Analysis

In contrast to ordinal regression, multimodal applications are more common in the medical domain [105]. Many studies have shown that combining different sources of information generates models with greater predictive power and robustness [5]. Furthermore, multimodal analysis can be used to find clinical correlations between data [76]. Diagnosis of brain pathologies emerges as one of the most popular applications with multimodal information [99]. These multimodal approaches include combinations of many data sources: MRIs combined with encephalograms [116], demographic data with cerebrospinal fluid examinations [116], etc. Good results have also been reported for multimodal approaches in glioblastoma classification [76], skin lesions [169], and thyroid cytopathology [74], combining not only visual data but also textual and categorical information such as clinical outcomes of chemotherapy and radiotherapy or patient metadata.

And it has been used not only for diagnostic tasks, but also for segmentation and retrieval. For some diseases, such as glaucoma, segmentation is one of the first steps in the analysis process [113], as the geometric characterization of the optic disc (the area where the optic nerve enters the retina of the eye) and the optic cup (the depression in the optic disc where the optic nerve fibres meet to exit the eye to the brain) is one of the main indicators used by specialists diagnose this disease [58] [59]. Again, brain-related problems are among the most popular applications. Segmentation of brain tumours [152], delineation of brain tissue [75] on MRI images, segmentation of brain gliomas from MRI images [33] and segmentation of white, grey and cerebroespinal fluid [108] [68], are just a few examples where multimodal approaches have shown that using more than one source of information leads to better results than using the sources separately.

In the context of histopathology data analysis, kernel alignment [20] has been used in medical image collections, allowing the introduction of expert knowledge into the domain to improve the semantic representation of images, thus improving the performance of algorithms in image retrieval tasks. Non-negative matrix factorization (NMF) has also been used for summarisation tasks [21]. The NMF strategy is based on latent topic analysis, where summary elements are selected from the latent factors obtained in the matrix factorization process. In [66], the first multimodal method is presented that combines Whole Side Images and text from histopathology reports for retrieval tasks. The model is based on the extraction of visual features using deep learning and very simple fusion techniques, and the dataset they experimented with is a subset of The Cancer Genome Atlas (TCGA) [61] collection. The same dataset was subsequently used for retrieval tasks [31] using matrix factorisation techniques, and for the diagnostic task using a multimodal semantic alignment model using kernel functions [81].Finally, more recently, there have also been multimodal efforts to find correlations between genomic and histological data [26], with the notable fact that this uses the tensor product as a mechanism for fusing representations.

Regarding the application of multimodal strategies in the diagnosis of eye-related diseases, there are not too many published papers, beyond the fact that the results are promising. The first work exploring this idea is presented in [114] where, from scanning laser ophtalmoscope images, anatomical features such as blood vessels and the optical disc are extracted and then superimposed on the same image. The goal was to make an automatic mapping of the retina. Following this came works that combined angiographic images and images taken under natural light with a green filter, in order to diagnose DR [173]. Information fusion from stereo and temporal images of the retina was also explored in [122].

More recently, most work has focused on the process of segmentation and classification of glaucoma and AMD. Golabbakhsh et al. [51] proposed the registration of the retina vessels combining information from eye-fundus images and OCT. Segmentation of retina vessels is useful in the diagnosis of AMD and DR. This combination of fundus images and OCT has also been explored in [101], [160] and [171], not only for the AMD diagnosis but for the segmentation of the OC and OD, useful in glaucoma diagnosis. This segmentation task has also been used directly in combination with eye fundus images in [25], using the Texture of Projection features and Bag of Visual Words to capture shape and textual changes, as well as color information from eye fundus images. Although their source of information is only images, Chakravarty et al. [25] showed that the extraction of several features, obtained separately from the same source, can be combined within the same model to provide complementary information. They experimented with merging information at different stages of the process and found the best results by intermediate fusion, it means, within the classifier. A similar concept was developed in [127] where structural and non structural features are extracted from the eye fundus images, and then correlated in a late fusion module. For their part, Perdomo et al. [113] at the MindLab proposed the combination of deep neural networks with morphological features in the detection and classification of glaucoma. Image analysis and extraction of morphological features is done separately and then mixed within the network in a layer, then go through a fully connected layer and the output layer.

For their part, Ahmed El Tanboly et al. [43] combined information from OCT and OCTA (Optic Coherence Tomography Angiography) to detect early signs of DR. Their method include an independent feature extraction process followed by a feature fusion (a merging). The OCT provided information about 12 layers of the retinal tissue, and the OCTA provided information about the vessels. The final diagnosis process is performed by a Support Vector Machine with a linear kernel, outperforming the results obtained separately with each source. OCT and OCTA have also been combined in [153] for the diagnosis of dry-AMD.

Regarding the combination of images and text, the first work on this subject was developed by Schlegl et al. [129], where they combine a CNN to analyse OCT images, with semantic information extracted from medical reports. The fusion consist in a concatenation of parameters into a fully connected layer, and they report an improvement in the performance of retinal tissue classification tasks.

Other works combine fundus images with angiograms, performing pattern matching to fuse images and segment retinal blood vessels [97], and to make retinal image registration [83]. And then there is the combination of autoflorescense images, or images in Otsu binary modality [155] [14], to perform segmentation of reticular pseudodrussen and glaucoma diagnosis. A summary of the related literature can be found in Table **6-2**.

2.5. Quantum Machine Learning

At the intersection of quantum physics and machine learning, quantum machine learning (QML) seeks to exploit the unique properties of quantum computers and the mathematical foundations of quantum mechanics to develop and improve the functionality of machine learning algorithms [53]. The surge in attention to this field is largely due to the potential of quantum computers to manipulate massive amounts of data and perform computations at speeds beyond the reach of classical computers. Quantum computers, based on the principles of quantum mechanics, promise exponential acceleration for certain types of computation, potentially solving the bottlenecks caused by the computational demands of large, complex machine learning enterprises [15].

Progress in quantum machine learning has been rapid in recent years. Researchers have developed quantum versions of widely used machine learning algorithms such as support vector machines, decision trees, and neural networks [24]. Quantum computers, including both gate-based and quantum annealers, have been built by a wide range of companies and research groups, driving the development of QML. As a result, QML has been widely adopted in applications such as pattern recognition, natural language processing, and optimization.

The growing use of QML can be seen across a wide range of industries. In drug discovery, for example, QML has the potential to speed up conventional processes, which tend to be slow, costly, and inconsistent [15]. Quantum Generative Adversarial Networks (Q-GANs) are being investigated for the generation of smaller molecules, with the potential to significantly accelerate drug discovery and the development of innovative drugs [24]. In financial markets, QML has demonstrated the potential to perform tasks such as asset pricing, volatility prediction, exotic option outcome prediction, fraud detection, stock selection, hedge fund selection, algorithmic trading, market making, financial forecasting, accounting and auditing, and risk assessment faster and more accurately than classical algorithms [140].

However, QML still faces challenges. Despite encouraging signs that machine learning models can be trained to higher accuracy with less data than classical techniques, there are still limitations. Even the most powerful quantum computers in existence today lag behind today's most powerful servers. Nevertheless, quantum computing technology has considerable potential for development and expansion [24].

Therefore, a closer line of research is related to the development of quantum-inspired models. These models aim to develop learning methods and techniques using the mathematical formalism associated with quantum mechanics. There is no need to implement them on quantum computers (although is possible), and they can be directly integrated into the most popular machine learning frameworks as Tensorflow and Pytorch [53].

If you consider a set of data samples as an ensamble of quantum particles that lie in a feature space, quantum mechanics has natural mechanisms to capture the quantum state of the entire set of samples. In the end, this is nothing more than a probabilistic representation of the data. Density matrices are the objects that allow us to do this. Understanding the elements that make up the density matrix generated by the quantum representation of the data set as learnable parameters of a representation learning model is what González et al. proposed in [53]. This can be used to generate models for classification, regression, clustering, density estimation, etc. Furthermore, this can be used to develop multimodal methods, exploiting the underlying flexibility in handling representations through kernel functions. The exploration of these models in the field of medical applications is a large part of the development of this work.

Table 2-2.: A summary of recent publications on multimodal approaches to ophthalmological and histopathology data analysis. DR stands for Diabetic Retinopathy, AMD for Age related Macular Edema, OCT for Optical Coherence Tomography, OCTA for Optical Coherence Tomography Angiography and PCa for prostate cancer

Reference	**Data**	**Remarks**
[43] [153]	OCT and OCTA	Diabetic Retinopathy diagnosis. Fusion of probabilities from independent SVM models.
[113] [127] [128]	Eye-fundus image and morphological, textural, intensity features.	Glaucoma diagnosis based on image segmentation and morphological features, and also based on OC and OD segmentation combined with texture and intensity features. Fusion is acheved by layer concatenation.
[160] [171]	Eye-fundus images and OCT	Segmentation of retina vessels, OC and OD. Diagnosis of AMD. Fusion is achieved by the concatenation of multimodal features.
[83]	Eye fundus images, angiograms and OCT	Retinal image registration for multimodal alignment.
[168]	WSI and clinical data	Breast cancer recurrence prediction. A concatenation of features is input into a multilayer perceptron.
[26]	WSI and genomics data	Diagnosis of several types of cancer using tensor fusion on low dimensional feature spaces.
[31]	WSI and textual reports	Retrieval system based on multimodal kernel semantic embedding.
[81]	WSI and textual reports	PCa Diagnosis based on a multimodal latent semantic alignment.

Part I.

Deep Learning-based Models

3. Deep Regression for Cone Density Estimation and Genetic Ophthalmic Diseases Diagnosis

This chapter presents two models based on deep learning for the automatic analysis of microscopic images of the retina, namely AOSLO. These images allow the specialist to observe in detail the cellular mosaic of the retina, which is affected in different ways when different diseases are present. A manual analysis of these images then involves identifying each cell, and then proceeding to statistical analysis of the alterations present. A single image of this type can account for more than 200 cells, making this manual marking process time-consuming. The first is a regression model that makes use of a modified U-Net to estimate the cell density in the images. The model is trained to automatically return information about the location and count of the cells. From this information, the second model applies a CNN to classify cases between healthy patients and patients with a degenerative disease. The results show that indeed, explicit cell location and cell count information improves the performance of the diagnostic models.

The work in this chapter as been published as:

Toledo-Cortes, S., *Dubis, A. M., González, F.A., Müller, H. Deep Density Estimation for Cone Counting and Diagnosis of Genetic Eye Diseases From Adaptive Optics Scanning Light Ophthalmoscope Images. In: Translational Vision Science Technology 12 (2023), 11, Nr. 11, S. 25–25. – ISSN 2164–2591* [147]

3.1. Introduction

The world of ophthalmology has been transformed by our ability to image the inner back section of the eye. An inspection of the eye fundus allows specialists to find signs of degenerative diseases that may even go beyond the visual system such as, for instance, retinopathy caused by diabetes mellitus. In general, the density and regularity in the pattern of retinal photoreceptor cells are affected in different ways by various diseases [165]. However, even with the best clinical cameras, by the time that the macro-indicative changes in a disease are detected, the loss of hundreds of thousands of retinal cells cannot be quantified [107].

New technology now exists to visualize individual cells in the living eye. Split detector Adaptive Optics Scanning Light Ophthalmoscopy (AOSLO) [124] is a non-invasive retinal imaging technique that corrects for optical distortion and allows high-resolution images of the photoreceptor layer to be acquired in living patients [38]. AOSLO has allowed one to look at retinal tissue with microscopic detail (and make measurements), in a way that was previously only possible in dead patients. This has allowed the quantitative study of alterations at the cellular level caused by different diseases [19]. Indeed, it has been used to describe the degeneration of the photoreceptor structure in the retina by various quantitative measurements in several inherited retinal diseases such as Stargardt disease (STGD1) and retinitis pigmentosa (RP) [38, 36, 104].

Retinitis pigmentosa refers to a group of eye diseases of genetic origin. It affects the photoreceptor cells of the retina (first the rods, then the cones [22]), first causing loss of peripheral and night vision which is, in fact, the earliest associated symptom [107, 32]. However, the RP diagnosis is hard as there is no specific test, and the actual diagnosis is reached after various exams such as OCT, angiography, electroretinography, and genetic testing. Therefore, by the time that an accurate diagnosis is reached, the patients are in an advanced stage of the disease [32]. Stargardt's, on the other hand, is also a genetic disease that affects vision in the macula and has the potential to lead to complete loss of central vision [150]. It does not cause complete vision loss because patients retain peripheral vision. STGD1 is usually diagnosed between the ages of 10 and 20 years, but, as with RP, the diagnosis process is not straightforward and requires a series of different tests [115].

Furthermore, unfortunately, there is no cure for RP or STGD1 [32, 64]. However, several genetic, cell, and drug therapies are being studied with promising results. The success of these studies depends on the progress of the clinical development pipeline, especially the early and fast detection and diagnosis of positive cases [32].

Is in this sense that AOSLO imaging arises then as a promising diagnosis tool [18]. However, the manual analysis (which is the gold standard for this task) that allows the identification and labelling of photoreceptors in the images is a time-consuming procedure [104], and this prevents the implementation of this type of images in diagnostic and research processes [36]. Therefore, the development of automatic methods to support specialists and speed up the analysis of AOSLO images is of great importance. This chapter proposes and develops models based on deep learning to:

- Estimate cone density in split detector AOSLO images, allowing quantitative analysis at the cellular level.
- Support the diagnosis of retinitis pigmentosa and Stargardt disease from the analysis of split detector AOSLO images.

To achieve these goals we introduce two new models: CoDE (**Co**ne **D**ensity **E**stimation) and CoDED (CoDE-**D**iagnosis). CoDE is a deep density estimation model that learns to generate cone density maps from the original split detector AOSLO images. CoDED is an extension of CoDE intended to support the diagnosis of specific diseases. With these methods and the proposed experimental setup, we show that:

1. High precision cone density estimation can be achieved with no need of patch-based analysis and complex post-processing stages, as those reported in the state of the art.
2. Computer-assisted diagnosis of genetic retinal pathologies such as RP and STGD1 is feasible using split detector AOSLO image analysis.

Usually, disease diagnosis performed from AOSLO images has been based on statistical characterizations. Stargardt disease effects has been studied by means of cone spacing analysis and statistical analysis [27], and something similar happens for retinitis pigmentosa [107]. Some machine learning applications have been developed for the identification of cones in patients with choroideremia [104], and for patients with Stargardt disease [38]. Previous studies have shown promising results classifying Stargardt disease patients and healthy patients using optical coherence tomography images [131], by training deep learning models with relatively small datasets. However, while this study shows the feasibility of using a deep learning model for these tasks, to our knowledge, to date, diagnosing Stargardt disease and retinitis pigmentosa from AOSLO images has not yet been reported in the state of the art using machine learning techniques.

On the other hand, the cone density estimation task has already been studied using machine learning methods. Cunefare et al. [35] proposed a model for cone counting, which requires an analysis based on 32 × 32 pixel patches extracted from the original AOSLO image. This means a high time-consuming pre-processing is needed as each AOSLO image has to be partitioned to generate a set of patches which in turn need to be labeled according to the presence or absence of a cone. The training process is then carried out at the patch level. The goal with this process is to obtain a model capable of identifying the presence of a cone at a patch level. Therefore, the inference process requires the image to be analyzed from all possible 32 × 32 pixel patches centered on each pixel of the original image, generating at the end a global heat map indicating the probability of the presence of a cone at pixel level. This heat map must be processed to determine a final estimate of the location and the number of cones in the image. While the model achieves promising results, the necessary pre-processing and post-processing for both training and final cone counting estimation makes the implementation a cumbersome procedure, difficult to operate in end-to-end architectures, and therefore hard to update if new images are available. However, the dataset they use is freely available [35] and is used as a baseline reference for the evaluation of our models.

Davidson et al. [38], proposed a multidimensional recurrent neural network to segment the cones. For training, this approach requires the segmentation masks of the images. The model combines convolutional layers with multidimensional long-short-term memory blocks, to capture near- and far dependencies between pixels, and involve this information in the segmentation task. Again, the implementation of such a model requires extra work to generate masks for segmentation which, as in the case of Cunefare et al. [35], is a major disadvantage in terms of practicality.

Beyond cone counting in AOSLO images, the general task of counting objects in pictures is recurrent in the state of the art [47, 87]. It is not only at the microscopic level that it is necessary to count objects. It can also be necessary to count animals or people, and many methods have been developed in the last decades to automatically perform this task [63]. A recurrent idea is to use filters on the images to represent map densities [47]. In this sense, an interesting approach that does not require as much pre-processing and mask generation for the cell counting task is presented by Xie et al. [166]. They proposed a method for counting objects in images based on the estimation of the density map. Knowing the location (coordinates) of the objects to be counted, a density map can be generated, whose integral matches the number of elements. Then, the model (a U-Net based architecture [123]) is trained to generate the corresponding density map from the original image. This technique is used to count bacteria in images of microscopic samples and is shown to be robust and easy to implement, as it requires no pre-processing as patch extraction and no post-processing of the results.

With this background in mind, we propose a method that builds on the idea presented by Xie et al. [166], but enhances and adapts the segmentation model backbone, and adds a linear correction at the top to fine-tune the estimation of the number of cones in the image. Unlike Davidson et al. [38] and Cunefare et al.[35], there is no need for additional annotations apart from the location of the cones, nor patch extraction, nor masks for segmentation. The method is end-to-end trainable and can serve as a basis for other architectures.

This chapter is organized as follows: Section 3.2 presents the framework for CoDE and the subsequent diagnosis model: CoDE-D. Section 3.3 presents the experimental results. Finally, section 3.4 presents the conclusions of this work.

3.2. Methods

Cone counting is made based on the estimate of a density map created from the annotations on each AOSLO image. This density map image can later be used as input to a deep convolutional neural network (CNN) model to perform a diagnosis task. The details of each procedure are explained below.

3.2.1. CoDE: Cone Density Estimation

The complete architecture of the proposed CoDE model used for the cone density estimation task is shown in Figure **6-3**. Inspired by the method presented by Xie et al. [166], the model is trained to generate a density map from the original AOSLO image, using an U-Net architecture with an Xception backbone. The integral of this density map is a first approximation of the cone count in the image, which is then fitted by means of linear correction.

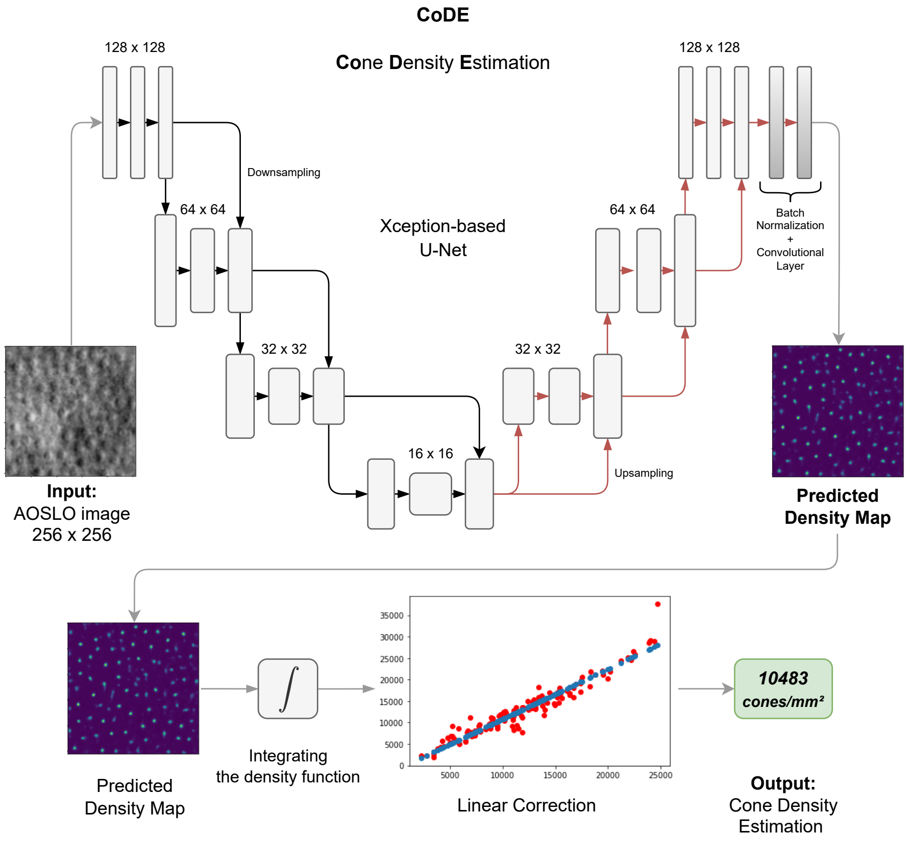


Figure 3-1.: CoDE model architecture for cone density estimation on split detector AOSLO images. The original image is input into a modified Xception-based U-Net [166, 29] to generate a density map of the cones in the original image. The integral over this density map is linearly corrected to provide an accurate estimation of the number of cones in the image.

Density Map Estimation

We use an Xception-based U-Net model [29] as the backbone for the density map phase. This model consists of four blocks. The entry block comprises a 2D convolution layer with 32 filters, a batch-normalization, and a rectifier activation function. The subsequent blocks consist of 2D separable convolution layers, each followed by batch-normalization and max-pooling. At the lowest point of the U-Net, the feature map has a size of $16 \times 16 \times 256$. Four up-sampling blocks composed of transposed convolution layers and batch-normalization are used to restore the corresponding density map. The kernel size in all the convolutional layers is 3×3. To adapt the model to generate density maps, inspired by the implementation of the cell-counting U-Net presented by Xie et al. [166], a batch-normalization layer and a single-filter convolutional layer are added on top of the model. As a result, the output of the model is a single-channel image. This output is intended to be a density map of the cones present in the original AOSLO image.

To create the ground truth of the density maps, we generated the corresponding density map for each AOSLO image using the available information of the coordinates of all the cones (see Figure **3-2**). A Gaussian filter is applied to each point marked as the center of a cone. The resulting product is a single-channel image whose integral matches the total number of cones. We trained the model as a regressor using a Mean Squared Error (MSE) loss function. All the implementation details are available in the Github repository [145].

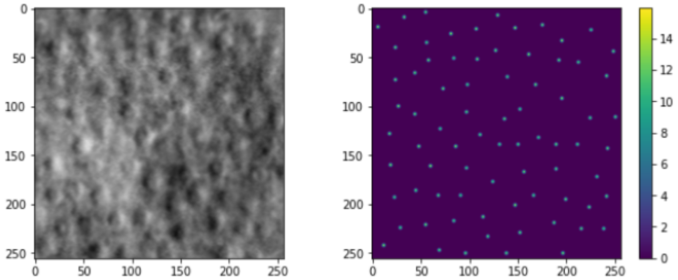


Figure 3-2.: On the left: original split detector AOSLO sample. On the right: ground truth density map. For the training process of the CoDE model, the ground truth density map is generated by means of the known coordinates of the center of the cones. A Gaussian filter is applied over each point, which generates a density map whose integral matches the number of cones in the original AOSLO image.

Linear Correction

The U-Net stage of the CoDE model is able to infer the cone density map and a proper number on the count of the cones. The density map shows the approximate location of each cone, and the count of the cones is the result of the sum all over the density map. While this could be enough for the task, this estimate of the number of cones in the image can be fine-tuned by a linear regression set on top of the model. This regression allows a correction on the cone density estimation, especially for high-density samples.

3.2.2. CoDED: CoDE Diagnosis

We approached Stargardt disease and retinitis pigmentosa diagnosis as a multi-class classification problem, having also a control (healthy) group. Taking advantage of the information learned by CoDE about the cone density, the inferred density map is used as input into a deep image classifier.

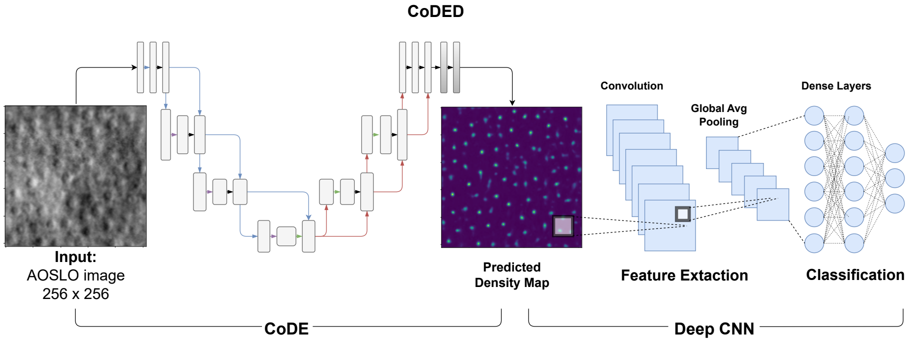


Figure 3-3.: CoDED model architecture for disease diagnosis on AOSLO images. The density map predicted by CoDE is the input of a deep CNN model. The convolutional block of the CNN works as feature extractor and the final classification is performed by a three-layer fully-connected perceptron.

The convolutional block of the deep CNN model is used as the feature extractor. The learned representation is provided by the last average pooling layer. These features then serve as the input to the classification block, which consists of a multilayer perceptron whose output is a layer with three neurons.

3.2.3. Experimental Set Up

The training process of the models follows the standard design of machine learning experiments. Details on the data sets used and the hyperparameter settings for training and evaluation metrics are presented below.

Datasets

We made experiments with two different AOSLO image datasets. The first, henceforth referred to as *Cunefare*, is a publicly available baseline dataset presented by Cunefare et al. [35] and uploaded to a Github repository [34]. This set consists of 264 split detector AOSLO samples with annotations of the center coordinates of the cones. This set is partitioned into 184 images for training and 80 for testing, and is utilized in the experimental setup for the cone density estimation task.

The second dataset, henceforth referred to as *Dubis*, also consists of 264 split detector AOSLO images. Of these, 60 samples correspond to control cases, 65 to patients with Stargardt disease, and 139 to patients with retinitis pigmentosa. In addition to disease labeling, each image also includes the location, as pixel coordinates, of the center of each cone present in the image. This annotation was made by an expert in the field. For this dataset, 184 images were used for training and validation, and 80 for testing. During the training of all models, a validation rate of 20% is defined. This setup was used identically for both the diagnosis task and the cone density estimation task.

Table 3-1.: *Cunefare* dataset partition for train and test [35]. This dataset is available in a Github repository [34], and has 264 split detector AOSLO images and is used in this work for the cone density estimation task.

Partition	No. of samples
Train	184
Test	80

Table 3-2.: *Dubis* dataset partition for train and test. This dataset has 264 split detector AOSLO images, labeled for three classes: normal, Stargardt disease and retinitis pigmentosa. This dataset was used in this work for the cone density estimation and for the disease diagnosis task.

Diagnosis	Train	Test
Normal	42	18
Stargardt disease	44	21
Retinitis pigmentosa	98	41
Total	184	80

The size of the images in both datasets is standardized to 256×256 pixels. Normalization of all images is also performed for the cone density estimation task. All details of the datasets are summarized in Table **3-1** and Table **3-2**.

Cone Density Estimation

The proposed model for cone counting in AOSLO images is trained from scratch, optimizing an MSE loss function using RMSprop. The learning rate was explored and ultimately set at 10^{-3}. Parameters for the per-pixel classification layer [29] were also explored, with the number of filters set at 3 and the kernel size at 5. The final output convolutional layer for density prediction comprises a single filter, linear activation, and orthogonal initialization, as implemented by Xie et al. [166]. For data augmentation, vertical and horizontal flips, random rotations, and random width and height shifts were set up as described by Voets et al. [158].

The model presented by in [35] constitutes our baseline for cone density estimation. Results are reproduced using the available code [34]. The method, called Adaptive Filtering and Local Detection (AFLD), is based on a CNN trained over small patches that may or not contain a single cone. Therefore, the inference process implies to classify all of the possible patches, one for each pixel, so that each pixel is assigned a score related to how likely it is to be a cone. With this, a heat map is predicted for the whole image, that need to be processed to locate the final predicted cone locations.

Stargardt Disease and Retinitis Pigmentosa Diagnosis

The approach to disease diagnosis is treated as a three-class classification task. For the CoDE stage, we initialize the model parameters with those learned from the cone density estimation task training. This implies that the classification model can leverage the combined *Cunefare* and *Dubis* train set partitions, despite the *Cunefare* dataset not possessing any disease labeling.

In the deep CNN stage, the convolutional block is taken, and the output of the average pooling layer is set as input for a multilayer perceptron. This perceptron consists of two dense layers with dropout regularization prior to the output layer. The first dense layer matches the size of the average pooling of the backbone CNN, and the second layer comprises 1024 neurons. The output consists of a dense layer with three neurons. A transfer learning procedure was undertaken for the convolutional block of the deep CNN model, utilizing the weights learned from the training with ImageNet [125].

We explored four well-known deep CNN's on top of CoDE for this task: Xception, ResNet, Inception-V3 and MobileNet-V2. The reason for using these models is the good performance they have previously shown in medical-related image analysis [82, 158]. In each case, the training was carried out with the train partition of the *Dubis* dataset (see Table **3-2**), optimizing a categorical cross-entropy loss by means of RMSProp. Learning rate was explored and set at 10^{-4}. The data augmentation configuration was the same as that used for eye fundus image analysis by Toledo et al. [144] and fully described by Krause et al. [79]. Single CNN models (Xception, ResNet, Inception-V3 and MobileNet-V2), fine-tuned from ImageNet with *Dubis* dataset in a similar experimental set up to the composed CoDED model, constitute our baseline for this task.

3.3. Experimental Evaluation

All methods were implemented in Python using TensorFlow and Keras [30]. Code and results are publicly available in a Github repository [145].

3.3.1. Cone Density Estimation

CoDE method is capable of accurately estimate the location of the cones thanks to the predicted density map given by the U-Net stage (see Figure **3-4**). As explained in Section 3.2.1, the integral on this predicted density map is a first approximation of the final count of the number of cones, which is fine-tuned by means of a final linear correction. As each sample image constitutes a square of $100\mu m$ on each side, the calculation of cone density is easy and straightforward after counting.

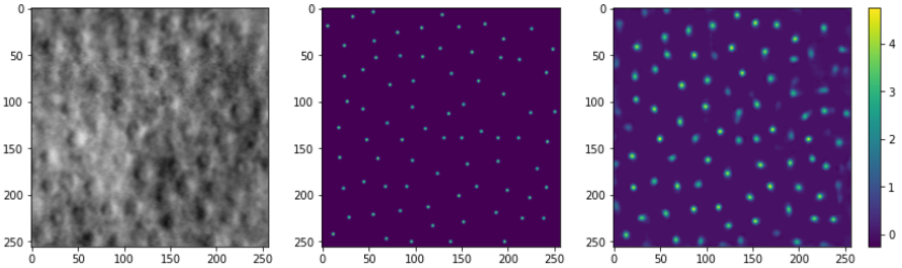


Figure 3-4.: On the left: original split detector AOSLO image sample (from the *Dubis* test partition). Center: ground truth density map indicating the location of each cone according to the manual annotation. On the right: predicted density map given by CoDE.

Several experimental procedures were performed using the different train partitions of each dataset (see Table **3-1** and Table **3-2**) and a joint dataset. To compare directly with our baseline, we report the performance of the models using Bland-Altman plots. A Bland-Altman plot allows one to analyze the level of agreement of two methods for the measurement of a quantity and, unlike the mean square error or mean absolute error, it gives an idea of whether the model is over- or underestimating the expected outcome. By setting the manual measurement of the cones as ground truth, the plot gives information on the difference between the measurement given by the method and the ground truth, with reference to the magnitude of these measurements. Training with *Cunefare* train partition only, allows us to directly compare the performance of the proposed CoDE method. The results on the *Cunefare* test set are shown in Figure **3-5**. Compared with our baseline AFLD [35], CoDE has a mean difference closer to zero, indicating a higher concordance level compared to the gold standard.

By training the CoDE model with joint training partitions of both *Cunefare* and *Dubis* datasets, and evaluating on the *Cunefare* test set, our method reports a mean difference in density estimation of -159 cells per square millimeter, and a 95% confidence interval between -1883.27 and 1564.4 cells per square millimeter. This outperforms any previous result reached for this task (see Figure **3-6**). For the same model evaluated on *Dubis* test set there is a mean difference of 306.82 cells per square millimeter, with a 95% confidence interval between -2561.73 and 3175.39 cells per square millimeter (see Figure **3-7**).

3.3.2. Stargardt Disease and Retinitis Pigmentosa Diagnosis

Classification results of CoDED using Xception, ResNet, Inception-V3 and MobileNet-V2 on *Dubis* test set for Stargardt disease, retinitis pigmentosa, and control subjects are reported in Table **3-3**. For each model, 20 independent trials were performed. The mean and standard deviation of the classification metrics calculated in a weighted average manner are reported. Confusion matrix on the test set for the best model, reached with the CoDED-Inception-V3,

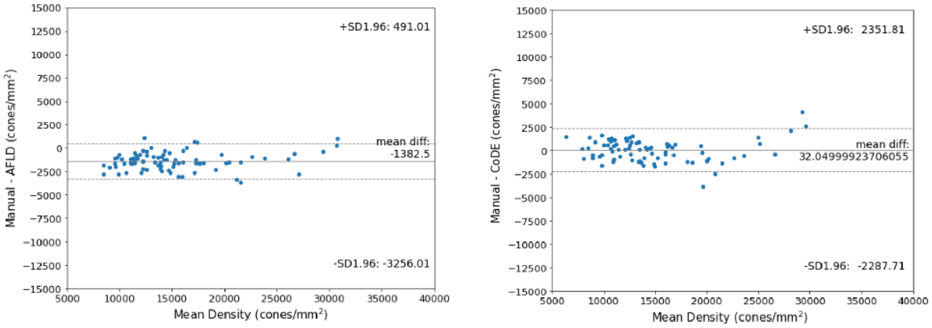


Figure 3-5.: Bland-Altman plots comparing the performance of cone density estimation at *Cunefare* test set. This is for a model trained with *Cunefare* train partition set only. The figure on the left shows the results for the baseline AFLD [35]. The figure on the right shows the results for the proposed CoDE. Note that although CoDE measurements presents a slightly higher standard deviation compared with AFLD, the mean difference of CoDE is much closer to zero.

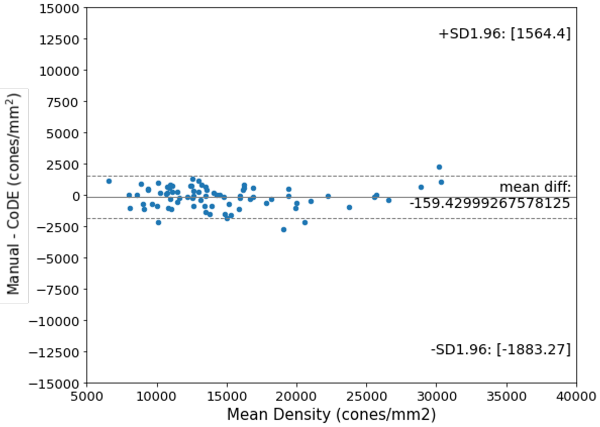


Figure 3-6.: Bland-Altman plot on *Cunefare* test set for the proposed CoDE method trained with a joint *Cunefare* and *Dubis* train partition set. Standard deviation is lower and the mean difference closer to zero, when compared with AFLD results.

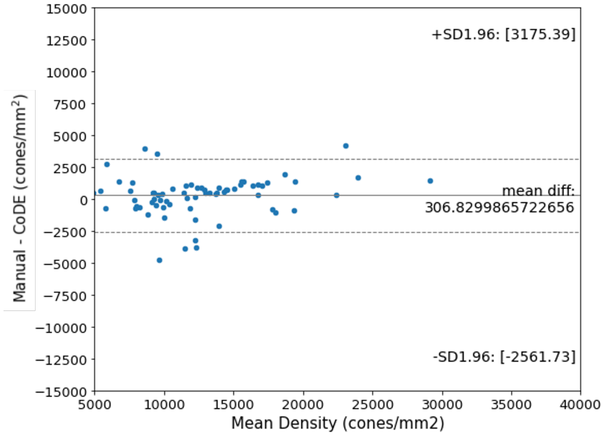


Figure 3-7.: Bland-Altman plot on *Dubis* test set for the proposed CoDE method trained with a joint *Cunefare* and *Dubis* train partition set.

is shown in Figure **3-8**.

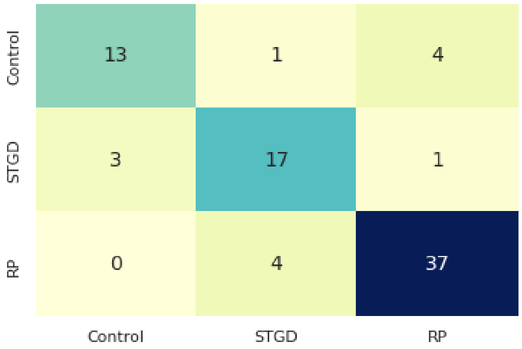


Figure 3-8.: Confusion matrix for the best model of CoDED-Inception-V3 on the *Dubis* test set. *Control* refers to healthy samples, *STGD* stands for Stargardt disease and *RP* for retinitis pigmentosa.

3.3.3. Discussion

From the results in Table **3-3**, we can see that CoDED-Inception-V3 is the model with the best performance. Generally, for all the deep CNN models that were tested, the mean performance improved when the deep CNN was applied on top of the CoDE model. The best overall model was found to be CoDED-Inception-V3, which achieved performance in the *Dubis* test set of 84% for accuracy, 84% for weighted average F1-score, 84% for weighted precision, and 84% for weighted recall (see Figure **3-8**).

Overall, the most remarkable aspect of these results is the fact that it is indeed possible to make a classification of a disease from a small sample of the cellular pattern (such as the

Table 3-3.: Classification performance on *Dubis* test set of the combined CoDE+DeepCNN models explored for the Stargardt disease and retinitis pigmentosa diagnosis. The mean and standard deviation over 20 trials are reported. Multi-class precision, recall, and F1 score are computed in a manner that involves a weighted average.

Model	Accuracy	Precision	Recall	F1 Score
Xception	0.681±0.04	0.706±0.03	0.681±0.04	0.652±0.06
ResNet	0.510±0.01	0.292±0.04	0.510±0.01	0.3615±0.02
Inception-V3	0.726±0.04	0.762±0.03	0.726±0.04	0.730±0.03
MobileNet-V2	0.706±0.04	0.729±0.04	0.706±0.04	0.708±0.03
CoDED-Xception	0.737±0.03	0.766±0.03	0.737±0.03	0.738±0.03
CoDED-ResNet	0.711±0.04	0.767±0.03	0.711±0.04	0.712±0.04
CoDED-Inception-V3	**0.768**±0.05	**0.794**±0.04	**0.768**±0.05	**0.770**±0.04
CoDED-MobileNet-V2	0.695±0.04	0.754±0.04	0.695±0.04	0.702±0.04

given by an AOSLO image). Deep learning models can learn to distinguish between different lesion patterns to discriminate between one disease and another, and the performance of the model is boosted when the cellular pattern is easily distinguishable as in the density maps generated by CoDE.

3.4. Conclusions

In this chapter we presented CoDE, a method for automatic cone density estimation on split detector AOSLO images. While machine learning techniques have previously been applied to cone density estimation, this proposal demonstrates the possibility of implementing this task with a model that does not necessitate exhaustive patch-based analysis or mask-based segmentation. This simplification of the method's implementation eliminates the need for substantial pre-processing effort and the generation of additional annotations beyond the cone coordinates. Moreover, this method is capable of generating estimates of cone locations, and, with the assistance of a regressor, it can automatically count the cones present in an image to provide an accurate density estimate. Evaluations on two different datasets lead to the conclusion that this method demonstrates robustness and good generalization capacity, making it directly competitive with state-of-the-art models. Additionally, the lack of a need for post-processing, as required in previous approaches, enables end-to-end training and produces a model that can be effortlessly updated if additional samples become available.

Based on CoDe, we also presented CoDED, a deep CNN-based approach for the automatic classification of Stargardt disease and retinitis pigmentosa cases from split detector AOSLO retinal images. The use of transfer learning techniques on different deep CNNs, coupled with the advantage of the pre-trained CoDE (which uses a larger dataset than the disease-labeled intended for this task), made it possible to demonstrate the feasibility of classifying Stargardt disease and retinitis pigmentosa in relation to control patients within a single model. The single deep CNN models performed well, however, it was demonstrated that the classification performance was enhanced when executed via the density maps inferred by CoDE. Thus, it is implied that the CoDED model leverages the explicit cone pattern, which is more distinguishable in the density maps than in the original AOSLO images.

Apart from the fact that the classification results are good and that CoDED-Inception-V3 was shown to have the best performance, the most important conclusion is that a small sample of the cellular pattern of the macular region of the retina is sufficient for these computational models to be considered as a reliable tool to assist medical staff in the diagnostic process of these diseases. This can help to streamline the usual diagnostic process, which requires numerous tests and the time and knowledge of specialised ophthalmological staff, which in turn can lead to greater coverage of the population at lower cost. Therefore, these results open the door to further research and development of methods to improve these diagnostic support tools. In the long term, the scientific and social benefits are potentially great.

Overall, this methods show the feasibility of deep machine learning models to speed up the analysis of split-detector AOSLO images, thus promoting and facilitating the development and use of this type of image in the research and treatment of genetic retinal pathologies.

4. Deep Probabilistic Regression for Diabetic Retinopathy Grading

This chapter begins to establish how to use ordinal regression to analyse eye fundus images. As mentioned earlier, the diagnosis of diabetic retinopathy is associated with a 5-level gradation, which can then be simplified into two categories for treatment purposes. Here we propose to use a hybrid model based on deep learning and Gaussian processes. A Gaussian process is a probabilistic regression model. This hybrid model allows, among other things, to combine the representational power of deep learning with the ability to generalise from small data sets of Gaussian process models, and also to quantify the uncertainty of the predictions. The results show that quantifying the uncertainty in the predictions improves the interpretability of the method as a diagnostic support tool.

The work in this chapter as been published as:

***Toledo-Cortés S.**, de la Pava M., Perdomo O., González F.A. Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification. In: Ophthalmic Medical Image Analysis. OMIA 2020. Lecture Notes in Computer Science, vol 12069. Springer, Cham., 2020, S. 206–215* [144].

4.1. Introduction

Diabetic Retinopathy (DR) is a consequence of *Diabetes Mellitus* that manifests itself in the alteration of vascular tissue. When an alteration in the correct blood supply occurs, lesions such as microaneurysms, hemorrhages and exudates appear [162]. These lesions can be identified in eye fundus images, one of the fastest and least invasive methods for DR diagnosing. Although early detection and monitoring are crucial to prevent progression and loss of vision [161], in developing countries approximately 40% of patients are not diagnosed due to lack of access to the medical equipment and specialist, which puts patients of productive age at risk of visual impairment [162, 170]. Therefore, to facilitate access to rapid diagnosis and speed up the work of professionals, many efforts have been made in the development of machine learning models focused on the analysis of eye fundus images for automatic DR detection.

For medical image analysis, deep Convolutional Neural Networks represent the state of the art. These methods work by means of filters that go through the image and exploit the natural structure of the data, being able to detect increasingly complex patterns. However, the success of these deep learning models depends on the availability of very large volumes of data, and this is not always the case for medical image datasets. For instance, one of the largest public-available image dataset for DR detection is EyePACS [40], which has 35126 samples for training. For this reason, training a deep learning model for this problem from scratch is not always feasible [70]. Instead, fine-tuning of pretrained models is preferred, as it allows the models to refine a general knowledge for an specific tasks. However, the number of specific sample images is not always enough to make a tuning that produces good final performances [70].

Classical machine learning methods such as Gaussian Processes (GP), on the other hand, were originally designed to work well with small data sets [44]. They have different advantages over deep neural network models, as lower number of parameters to train, convex optimization, modularity in model design, the possibility to involve domain knowledge, and in the case of Bayesian approaches, they allow the calculation of prediction uncertainty [163]. The latter would be useful in medical applications, as it gives to the final user an indication of the quality of the prediction [86].

This work presents and evaluates a hybrid deep learning-Gaussian process model for the diagnosis of DR, and prediction uncertainty quantification. Taking advantage of the representational power of deep learning, features were extracted using an Inception-V3 model, fine-tuned with EyePACS dataset. With these features we proceed to train a GP regression for DR grading. Our framework shows that:

1. The performance of the proposed hybrid model trained as a regressor for the DR grade, allows it to improve binary classification results when compared with the single deep learning approach.

2. Gaussian processes can improve the performance of deep learning methods by leveraging their ability to learn good image representations, when applied for small datasets analysis.

3. The integration of GP endows the method with the ability to quantify the uncertainty in the predictions. This improves the usability of the method as a diagnostic support tool. Furthermore the experimental results show that the predictions uncertainty is higher for false negatives and false positives than for true positives and true negatives respectively. This is a high valued skill in computational medical applications.

Many approaches have been proposed for the DR binary detection, most of them based in deep neural networks [111]. Some of them combine deep models with metric learning techniques, as in [174], where an Inception-V3 is trained and embedded into siamese-like blocks. The final DR binary or grade prediction is given by a fully-connected layer. In [49], a customized deep convolutional neural network to extract features is presented. The features and multiple metadata related to the original fundus image are used to trained a gradient boosting classifier to perform the DR prediction. In [88] an Inception-V3 model is once again fine-tuned using a private set of eye fundus images, but not with binary labels, but with five DR grade labels. The results are reported using a subset of the Messidor-2 dataset [2, 39]. This makes performance comparison impossible with many other results, including those presented in this chapter. Better results were reported by Gulshan et al. in [56], where an ensemble of ten Inception-V3 models, pretrained on ImageNet, are fine-tuned on a non-public eye fundus image dataset. The final classification is calculated as the linear average over the predictions of the ensemble. Results on Messidor-2 were reported, with a remarkable 99% AUC score. In [157], Voets et al. attempted to reproduce the results presented in [56], but it was not possible since the original study used non-public datasets for training. However, Voets et al. published the source code and models, and details on training procedure and hyperparameters are published in [157] and [79].

In relation to convolutional neural networks uncertainty estimation using GP, some work has been done specially outside the DR automatic detection context, as in [167], where a framework is developed to estimate uncertainty in any pretrained standard neural network, by modelling the prediction residuals with a GP. This framework was applied to the IMDB dataset, for age estimation based in face images. Also in [16], a GP on the top of a neural networks is end-to-end trained, which makes the model much more robust to adversarial examples. This model was used for classification in the MNIST and CIFAR-10 datasets. To our knowledge, this is the first proposal that implements a GP to quantify the uncertainty of a model predictions of DR diagnosis.

The rest of the chapter is organized as follows: section 4.2 introduces the theoretical framework for the method. Section 4.3 presents the experimental set up and results. Finally, in Section 4.4 the discussion of the results and conclusions are presented.

4.2. Method: Deep Learning Gaussian Process For Diabetic Retinopathy Diagnosis (DLGP-DR)

The overall strategy of the proposed Deep Learning Gaussian Process For Diabetic Retinopathy grade estimation (DLGP-DR) method comprises three phases, and is shown in Figure **4-1**. The first phase is a pre-processing stage, described in [157], which is applied to all eye fundus image datasets used in this work. This pre-processing eliminates the very dark images where the circular region of interest is not identified, eliminates the excess of black margin, and resizes the images to 299×299 pixels. The second phase is a feature extraction. An Inception-V3 model, trained with ImageNet and fine-tuned with EyePACS dataset is used as feature extractor. Each sample is then represented by a 2048-dimensional vector. The third and final task is the DR diagnosis, which is performed by a GP regressor.

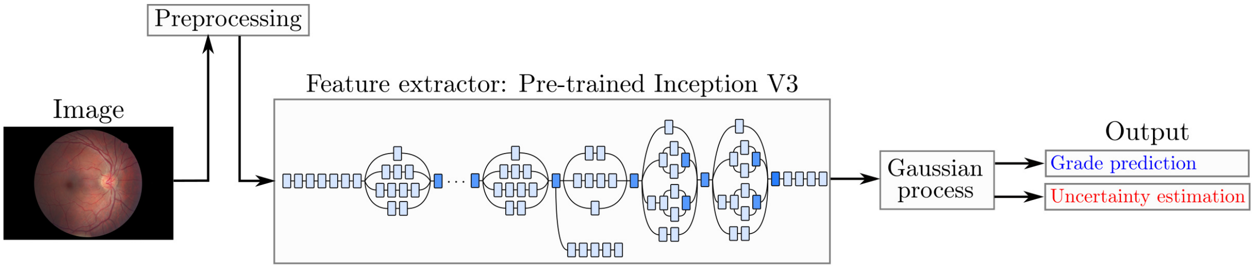


Figure 4-1.: Proposed DLGP-DR model. Fine-tuned Inception-V3 is used as feature extractor. The extracted features are then used to train a Gaussian process.

4.2.1. Feature Extraction

Many previous works have used deep learning models for the diagnostic of DR. Recently, Voets et al. [157] attempted to replicate the results published in [56], by fine-tuning an assembly of ten pretrained Inception-V3 networks. While Voets et al. were not able to achieve the same results reported in [56], most of the implementation details, as well as the specific partitioning for the training and test sets are publicly accessible, and were used in this study in the fine tuning of an Inception-V3 model. Once trained, the feature extraction is achieved by defining the global average pooling layer of the network as the output of the model, and use it to predict all the images in the datasets. Thus, each image will be represented by 2048 features which are used to train and evaluate the GP model.

4.2.2. Gaussian Processes

Gaussian processes are a Bayesian machine learning regression approach that are able to produce, in addition to the predictions, a value of uncertainty about them [44]. The method requires as input a covariance given by a kernel matrix. The kernel matrix would be the gram matrix computed over the training set with a Radial Basis Function (RBF). This RBF kernel depends on two parameters which will be learned during the training process. We performed a Gaussian process regression, where the labels are the five grades of retinopathy present in the EyePACS dataset. From the prior, the GP calculates the probability distribution of all the functions that fit the data, adjusting the prior from the evidence, and optimizing the kernel parameters. Predictions are obtained by marginalizing the final learned Gaussian distribution, which in turn yields another normal distribution, whose mean is the value of the prediction, and its standard deviation gives a measure of the uncertainty of the prediction. Thus, an optimized metric (attached to a RBF similarity measure) is learned from the data, used to estimate the DR grade.

This GP can be adapted to do binary classification. One simple way to do this is defining a linear threshold in the prediction regression results. The standard way however, consist in training a GP with binary labels and filtering the output of the regression by a sigmoid function. This results in a Gaussian Process Classifier (GPC). In any case, the predictions

Table 4-1.: Details of Messidor-2 dataset used for testing. Class 0 correspond to non-referable cases.

Class	Test Samples
0	1368
1	380

Table 4-2.: Details of the subset and final partition of the EyePACS dataset used for training and testing. This is the same partition used in [157]. Grades 0 and 1 correspond to non-referable patients, while grades 2, 3, and 4 correspond to referable cases.

Grade	Train Samples	Test Samples
0	37209	7407
1	3479	689
2	12873	0
3	2046	0
4	1220	694

of a GPC are not longer subject to a normal distribution, and the uncertainty can not be measured. Therefore, the GPC will not take part in this study.

4.3. Experimental Evaluation

4.3.1. Datasets

Experiments were performed with two eye fundus image datasets: EyePACS and Messidor-2. EyePACS comes with labels for five grades of DR: grade 0 means no DR, 1, 2, and 3 means non-proliferative mild, moderate and severe DR, while grade 4 means proliferative DR. For the binary classification task, according to the International Clinical Diabetic Retinopathy Scale [4], grades 0 and 1 corresponds to non-referable DR, while grades 2, 3, and 4 correspond to referable DR. In order to achieve comparable results with [157], we took the same EyePACS partition used for training and testing (see Table **6-4**). This partition was constructed only to ensure that the proportion of healthy and sick examples in training and testing was the same as that reported in [56]. EyePACS train set is used for training and validation of the Inception-V3 model. Then, the feature extraction described in Section 3 is applied. The extracted features are used for training the DLGP-DR model. The evaluation is performed on the EyePACS test set and on the Messidor-2, which is a standard dataset used to compare performance results in DR diagnosis task. Datasets details are described in Table **6-4** and in Table **6-2**.

4.3.2. Experimental Setup

Fine-tuning was made to an Inception-V3 network, pretrained on ImageNet and available in Keras [139]. The model was trained for binary DR classification task. The data augmentation configuration for horizontal reflection, brightness, saturation, hue, and contrast changes, is described in [79], and it is the same used in [157] and in [56]. The top layer of the Inception-V3 model is removed and replaced by two dense layers of 2048 and 1 neurons. BinaryCrossentropy was used as loss function and RMSprop as optimizer, with a learning rate of 10^{-6} and a decay of 4×10^{-5}. The performance of the model is validated by measuring the AUC in a validation set consisting of 20% of the training set.

Once the model is trained, the average pooling layer from the Inception-V3 model is used as output for feature extraction. The extracted features from the Inception-V3 are normalized and used to train a GP regressor over the five DR grade labels, it means, to perform the DR grading task. RBF kernel plus a white kernel were used as prior for the Gaussian process, using the Scikit-Learn implementation in Python. RBF length scale parameter was initiated at 1.0, with bounds at $(10^{-2}, 10^{3})$. White noise level parameter was started at 10^{-5}, with bounds at $(10^{-10}, 10)$. The output of the DLGP-DR is a continuous number indicating the DR grade.

Two baselines were defined to compare the DLGP-DR performance. Results reported by Voets es al. [157] constitute the first baseline of this study. The second baseline is an extension of the Inception-V3 model with two dense layers trained on the same feature test as the Gaussian process, which is called as *NN-model* hereafter.

4.3.3. EyePACS results

DLGP-DR is evaluated in the EyePACS test partition. The results are binarized using a threshold of 1.5 (which is coherent with referable DR detection), and compared with baselines in Table **4-3**. In addition, although uncertainty estimation is not used to define or modify the prediction, DLGP-DR uncertainty is analysed for false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN). As mentioned before, referable diabetic retinopathy is defined as the presence of moderate, severe and proliferative DR. So, the false negatives are calculated as the patients that belong to grade 4 but are classified as grades 0 and 1. The false positives are calculated as the patients belonging to grades 0 and 1 but classified in grade 4. The results are shown in Figure. **4-2** and Figure. **4-3**.

4.3.4. Messidor-2 results

For Messidor-2 dataset, the predictions given by DLGP-DR are binarized using athe same threshold of 1.5 used for EyePACS. Based on the results of the uncertainty measured in the EyePACS test dataset, those samples predicted negative for which the standard deviation

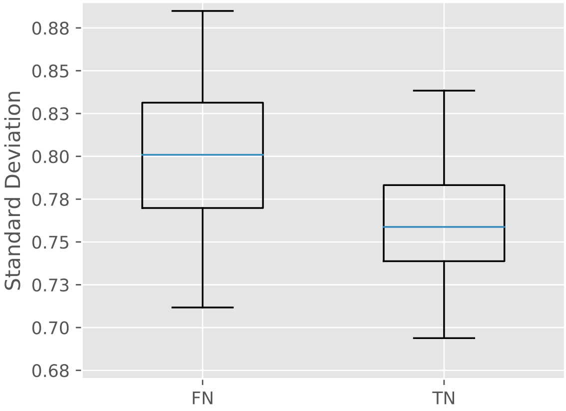


Figure 4-2.: Standard deviation for samples predicted as negative (non-referable) instances by DLGP-DR. FN: false negatives, TN: true negatives.

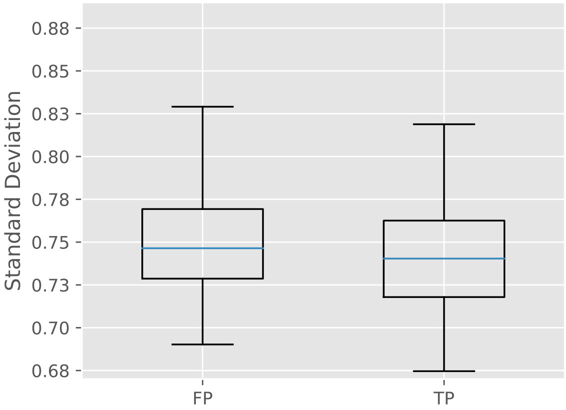


Figure 4-3.: Standard deviation for samples predicted as positive (referable) instances by DLGP-DR. FP: false positives, TP: true positives.

Table 4-3.: Comparison performance of DLGP-DR for binary classification in EyePACS test partition used in [157]. As it is not the standard EyePACS test set, comparison is not feasible with other similar studies.

Description	Sensitivity	Specificity	AUC
Voets 2019 [157]	0.906	0.847	0.951
NN-model	0.9207	0.85	0.9551
DLGP-DR	**0.9323**	**0.9173**	**0.9769**

Table 4-4.: Comparison performance of DLGP-DR for binary classification in Messidor-2. Referenced results from [157] were directly extracted from the respective documents.

Description	Sensitivity	Specificity	AUC
Voets 2019 [157]	**0.818**	0.712	0.853
NN-model	0.7368	0.8581	0.8753
DLGP-DR	0.7237	**0.8625**	**0.8787**

was higher than 0.84, were changed to positive. The results are reported and compared with the baselines in the Table **4-4**.

4.3.5. Discussion

Results reported in Table **4-3** shows that DLGP-DR outperforms specificity and AUC score of the *NN-model* and outperforms all the metrics reported by Voets et al. [157]. As observed in Table **4-4**, DLGP-DR outperforms both baselines for specificity and AUC scores. Although Gulshan et al. have reported 0.99 for AUC score in Messidor-2 [56], as Voets et al. comments in [157], the gap in the results may be due to the fact that the training in that study was made with other publicly available images and with a different gradation made by ophthalmologists. Further Overall, this shows that the global performance of the DLGP-DR exceeds that of a neural network-based classifier. In addition, in Figure **4-2** and Figure **4-3** the box-plot shows that the standard deviation is higher for false positives and false negatives. This means, that the DLGP-DR model has bigger uncertainties for wrong classified patients than for well classified. which provides the user a tool to identify wrong predictions. This behavior is especially visible for false negatives, which is the most dangerous mistake in medical applications, because a ill patient can leave out without a needed treatment.

4.4. Conclusions

In this chapter we took a deep learning model fine-tuned on the EyePACS dataset as feature extractor. The final task of DR classification and grading was carried out by means of a Gaussian process. For DR binary classification, the proposed DLGP-DR model reached better results than the original deep learning model. We also showed that a fine DR grade classification improve the binary classification performance of the original model.

Also, the DLGP-DR enables an uncertainty analysis. This analysis showed that the model could allow the identification of both, false negatives and false positives. The former are important due to the high cost of classifying a patient as healthy when it is not, and the later because they increase the costs of health care. The comparison between the Gaussian process and a neural network classifier for DR grades, showed once again that Gaussian processes are better tools for the analysis of medical images, for which datasets are usually far small to be analyzed entirely with deep learning techniques.

Overall, we demonstrate that the integration of deep learning and classical machine learning techniques is highly feasible in applications with small datasets, taking advantage of the representational power of deep learning and the theoretical robustness of classical methods.

Part II.

Quantum Measurement-based Models

5. Quantum Measurements

From this point on, the methods presented in this thesis are based on quantum-inspired models. The quantum domain has increasingly intersected with the field of machine learning, giving rise to a new area of research that seeks to exploit the principles of quantum mechanics to develop new methods for learning from data. In this chapter, we will explore this convergence, focusing on the concept of quantum measurements and their applications in machine learning.

Some parts of this chapter have been published as:

Gonzalez, F.A., Gallego, A., ***Toledo-Cortés S.****, Vargas-Calderon, V. Learning with density matrices and random features. In: Quantum Machine Intelligence 4 (2022), S. 23* [54].

5.1. Introduction

Quantum machine learning (QML) refers to a wide research area that combines the principles of quantum mechanics and machine learning, resulting in several approaches [176]. The various related developments can be categorised according to the classical or quantum nature of the data and the classical or quantum nature of the computing systems involved. Much work has been done on analysing classical data on quantum computers, trying to show the processing speed of these devices. The area that analyses classical data on classical devices, borrowing mathematical structures from quantum theory to develop machine learning methods, is the area that interests us.

The model presented in *Classification with Quantum Measurements* [54], namely Quantum Measurement Classification (QMC) is precisely a quantum-inspired approach with classical implementation, and offers a novel method for supervised machine learning based on the mathematical formalism that underpins quantum mechanics. QMC uses projective quantum measurement to construct a prediction function, where the relationship between input and output variables is represented as the state of a bipartite quantum system. The state is estimated from training samples by means of a density matrix, providing an innovative framework for performing classification tasks.

A density matrix, also known as a density operator, is a matrix used in quantum mechanics to describe the statistical state of a quantum system [53]. In other words, it provides a way to describe the state of a quantum system that is in a statistical ensemble, a collection of quantum systems prepared in a specific way, which may not be perfectly known or perfectly controlled. This concept is particularly useful when dealing with mixed states as opposed to pure states in quantum mechanics.

A pure state is a state with complete knowledge, described by a vector in a Hilbert space. For instance, the state of a quantum bit (or qubit, the fundamental unit of quantum information) can be a state $|0\rangle$ or $|1\rangle$ or a superposition of both. However, in many cases, we don't have complete knowledge of a quantum state or we have a system that is in a statistical mixture of states, also known as a mixed state. The density matrix is a mathematical tool that allows us to handle these mixed states. While the formalism will be dealt with in the next sections, we can proceed as follows: the density matrix ρ for a mixed state is defined as

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \tag{5-1}$$

where $|\psi_i\rangle$ represents the states of the system and p_i are the probabilities for each state. The diagonal elements of the density matrix give the probability of the outcome of a measurement, and the off-diagonal elements give information about the coherence between different states. One of the key features of a density matrix is that it allows us to represent both the quantum and classical uncertainty of quantum systems and to express different statistical operations such as measurement, system combination and expectations as linear algebra operations. This makes it a powerful tool in quantum mechanics and quantum computing, because it allows these models to be implemented in existing machine learning frameworks that can take advantage of common computing resources, such as GPU's.

Building a density matrix that aggregates the information from the data is then the basis of QMC. This involves several issues. First, each sample in the database must be represented in a quantum form. In practical terms, this means that the norm of the representation vector is equal to 1. With this representation, the mixed state of the system is constructed as the average of the individual states, and this in itself constitutes the training of the QMC model [54]. This means that no parameter optimisation is required. The inference task is done directly by projecting the operator defining the density matrix onto the space associated with the labels, and this works very well for simple tasks.

However, using this approach as a building block within other machine learning models may involve optimization processes. In *Learning with Density Matrices and Random Features* [54], the QMC framework is extrapolated and the idea of using density matrices to describe the statistical state of a quantum system is generalised, making the density matrix a trainable object as well. An optimization is then performed by means of gradient descent. While working with a matrix can lead to quadratic memory consumption, the Schmidt decomposition [80] allows to optimise the calculation process by allowing the density matrix to be implicitly manipulated by a number of fundamental components. This approach effectively uses the powerful formalism of density matrices to represent both quantum and classical uncertainties of quantum systems, and to express statistical operations such as measurement, system combination and expectations as linear algebraic operations. This allows not only classification models, but also density estimation and regression models, highlighting the versatility and power of density matrices in dealing with quantum systems, and their potential application to machine learning tasks.

The methods proposed in [54] use Random Fourier Features (RFF) [117] as the quantum map, given the aforementioned need to encode the data as a quantum state. While there are several options for this quantum mapping, the RFF proved to be an efficient mapping mechanism. However, this encoding can be implicit in the inference process, and ultimately everything can be posed in terms of inner products of the representations, so kernel functions can be applied. This then introduces *Quantum Kernel Mixtures* [52]. This mechanism provides an effective representation of joint probability distributions of both continuous and discrete random variables, and facilitates the construction of differentiable models for density estimation, inference, and sampling. The method can be integrated into end-to-end deep neural models and provides a differentiable, compositional and reversible inference procedure. Taken together, these developments highlight the exciting potential of quantum measurements in machine learning. This chapter covers the entire description of methods based on quantum measurements, step by step. First, we give a definition of what density matrices are, how they are constructed, and how they are used for inference processes. We will then see how the Schmidt decomposition allows us to simplify the computations to ultimately express the entire information flow in terms of kernel functions and how this allows us to build a highly versatile, modular processing layer that can be integrated into any gradient descent framework.

5.2. Density Matrix

The forecoming models assume that each sample in a data set corresponds to the state of quantum systems in an ensamble. This implies several things. First, that each sample of the data set must be represented by a vector $\psi \in H$ such that $||\psi|| = 1$ and H is a Hilbert space. Using Dirac's notation [50], we say that $|\psi\rangle$ belongs to the space H and that $\langle\psi|$ is the corresponding element in the dual space H^*. It follows that $\langle\phi|\psi\rangle$ corresponds to the inner product of the vectors ϕ and ψ, while $|\phi\rangle\langle\psi|$ corresponds to the outer product. Now, let's define what a density matrix is.

Definition 1 (Density Matrix) *A density matrix associated with a system is a positive semi-definite Hermitian operator that has a trace value of one and operates on the Hilbert space of the system.*

Given the above, and given an ensamble of quantum systems represented by $\{|\psi^i\rangle\}_{i=1}^n$, the quantum state of that ensamble can be represented by a density matrix defined by

$$\rho = \sum_{i=1}^{n} p_i |\psi^i\rangle \ \langle\psi^i|, \tag{5-2}$$

in what is known as a mixed state. Here, p_i is the probability associated to $|\psi^i\rangle$, therefore we must have $\sum_{i=1}^{n} p_i = 1$. The matrix ρ is an object that describes everything that can be physically measured about the set $\{|\psi^i\rangle\}_{i=1}^{n}$.

Now, given a dataset $\{(x_i, y_i); x_i \in X, y_i \in Y\}_{i=1}^{n}$ and the corresponding quantum feature maps

$$\psi_{\mathcal{X}} : X \to H_{\mathcal{X}} \quad x \to |\psi_{\mathcal{X}}(x)\rangle, \qquad\qquad \psi_{\mathcal{Y}} : Y \to H_{\mathcal{Y}} \quad y \to |\psi_{\mathcal{Y}}(y)\rangle, \tag{5-3}$$

each sample (x, y) of the dataset must be represented as a composition of two parts: a representation $|\psi_{\mathcal{X}}\rangle \in H_{\mathcal{X}}$ and a representation $|\psi_{\mathcal{Y}}\rangle \in H_{\mathcal{Y}}$. Then a bipartite system can be constructed, where the representation space is $H_{\mathcal{X}\mathcal{Y}} := H_{\mathcal{X}} \otimes H_{\mathcal{Y}}$, and each sample of the dataset is represented by means of $|\psi^i\rangle := |\psi^i_{\mathcal{X}}\rangle \otimes |\psi^i_{\mathcal{Y}}\rangle$, which leads to the next density matrix:

$$\rho_{\mathcal{X}\mathcal{Y}} = \sum_{i=1}^{n} p_i |\psi^i_{\mathcal{X}}\rangle \otimes |\psi^i_{\mathcal{Y}}\rangle\langle\psi^i_{\mathcal{X}}| \otimes \langle\psi^i_{\mathcal{Y}}|. \tag{5-4}$$

We will say that $\rho_{\mathcal{X}\mathcal{Y}}$ describes the quantum state of the system as defined by the data set.

5.3. Quantum Measurements

Given the Eq. 5-4, the question arises on how to use $\rho_{\mathcal{X}\mathcal{Y}}$ to make inference or, in other words, how to make measurements on the system $\rho_{\mathcal{X}\mathcal{Y}}$. A quantum measurement is the operation that allow us to make inference over new data samples. For this we require first a measurement operator to act over the system $\rho_{\mathcal{X}\mathcal{Y}}$. To understand how we define the measurement operator and how we manipulate it, let's recall the Born rule.

Born rule *The probability of $|\phi\rangle$ in a system represented by the state $|\psi\rangle$ is given by $p(\phi|\psi) = |\langle\phi|\psi\rangle|^2$.*

Born rule can be extended to a mixed state ρ:

$$p(\phi|\rho) = \mathrm{Tr}(|\phi\rangle\langle\phi|\rho) = \langle\phi|\rho|\phi\rangle. \tag{5-5}$$

Then, we can apply this to the system $\rho_{\mathcal{X}\mathcal{Y}}$. Let's start by explaining what a measuring operator is.

Definition 2 (Measurement Operator) *A measurement operator is a matrix $M \in \mathbb{R}^{n\times n}$ such that $M^\dagger M = I$, where $M^\dagger$ is the conjugate transpose of M.*

Then, the probability of measuring M in the system $\rho_{\mathcal{X}\mathcal{Y}}$ is given by

$$p(M|\rho_{\mathcal{X}\mathcal{Y}}) = \mathrm{Tr}(MM^\dagger\rho_{\mathcal{X}\mathcal{Y}}), \tag{5-6}$$

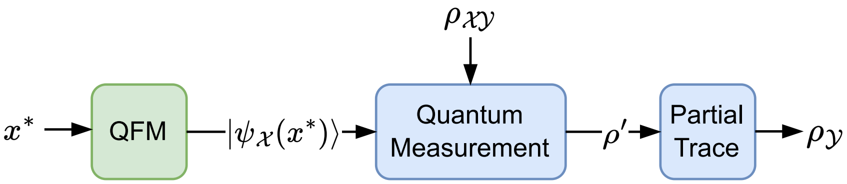


Figure 5-1.: QMC model. The inference process for a new sample x^* consists of first transforming it with a quantum feature map (QMF) $\psi_{\mathcal{X}}$ and then performing the measurement and partial trace to obtain $\rho_{\mathcal{Y}}$.

and the resulting status after the measurement is

$$\rho'_{\mathcal{XY}} = \frac{M^{\dagger}\rho_{\mathcal{XY}}M}{\mathrm{Tr}(M^{\dagger}\rho_{\mathcal{XY}}M)}. \tag{5-7}$$

The partial trace of $\rho'_{\mathcal{XY}}$ with respect to the space $H_{\mathcal{X}}$ corresponds to the reduced state of ρ' in $H_{\mathcal{Y}}$:

$$\rho_{\mathcal{Y}} = \mathrm{Tr}_{\mathcal{X}}[\rho'_{\mathcal{XY}}]. \tag{5-8}$$

This procedure is known as the collapse of the wave function and is a projection of the system defined by $\rho_{\mathcal{XY}}$ onto a direction of space $H_{\mathcal{X}}$ defined by the operator M. More precisely, given a new sample $|\psi^*_{\mathcal{X}}\rangle \in H_{\mathcal{X}}$ on which one wants to calculate its counterpart in $H_{\mathcal{Y}}$, the correspondent measurement operator M is constructed as

$$M = |\psi^*_{\mathcal{X}}\rangle\langle\psi^*_{\mathcal{X}}| \otimes \mathrm{Id}_{H_{\mathcal{Y}}}, \tag{5-9}$$

where $\mathrm{Id}_{H_{\mathcal{Y}}}$ is de identity operation in $H_{\mathcal{Y}}$. That is, the operator M is constructed in such a way that it contains information about the state of the sample in the space $H_{\mathcal{X}}$, and summarises the uncertainty about the state in $H_{\mathcal{Y}}$ through $\mathrm{Id}_{H_{\mathcal{Y}}}$.

5.3.1. Quantum Measurement Classification (QMC)

In short, given a dataset $\{(x_i, y_i); x_i \in X, y_i \in Y\}_{i=1}^{n}$ and the quantum feature maps $\psi_{\mathcal{X}}$ and $\psi_{\mathcal{Y}}$, Quantum Measurement Classification or QMC (see Figure **5-1**) is a model for categorical classification that works by constructing $\rho_{\mathcal{XY}}$ as defined in Eq. 5-4. This simple computation is the training. Then, according to Eq. 5-7 to Eq. 5-9, the inference process for a new sample x^* results in a density matrix $\rho_{\mathcal{Y}}$, such that $\rho_{\mathcal{Y}rr}$ corresponds to $p(Y = r|x^*, \rho_{\mathcal{XY}})$.

5.3.2. Density Matrix Decomposition

A density matrix ρ representing a mixed state is susceptible to efficient optimization. A gradient-based optimization is allowed by the Schmidt decomposition of the density matrix.

Proposition 1 (Schmidt Decomposition) *Suppose $H_{\mathcal{X}}$ and $H_{\mathcal{Y}}$ are Hilbert spaces of dimensions n and m respectively, with $n \geq m$. Given any $|\psi\rangle$ belonging to $H_{\mathcal{X}\mathcal{Y}}$, we can find orthonormal sets $\{u_1, \ldots, u_m\}$ from $H_{\mathcal{X}}$ and $\{v_1, \ldots, v_m\}$ from $H_{\mathcal{Y}}$ such that we can represent $|\psi\rangle$ as $\sum_{i=1}^{m} \alpha_i u_i \otimes v_i$, where each α_i is a non-negative real number or zero, and these α_i's are unique modulo permutations.*

In particular for our purpose, this decomposition allows us to approximate $\rho_{\mathcal{X}\mathcal{Y}}$ by means of a set of scalars β_i and vectors $|v_i\rangle$:

$$\rho_{\mathcal{X}\mathcal{Y}} \approx \sum_{i=1}^{m} \beta_i |v_i\rangle \ \langle v_i|, \tag{5-10}$$

in which the number of components of the decomposition, m, becomes an hyper-parameter of the model and $\{\beta_i, |v_i\rangle\}_{i=1}^{m}$ are learnable parameters such that $\beta_i \in \mathbb{R}$ and $v^i \in H_{\mathcal{X}\mathcal{Y}}$. Analogously, $\rho_{\mathcal{Y}}$ can be expressed in a decomposed form as

$$\rho_{\mathcal{Y}} \approx \sum_{i=1}^{m} \gamma_i |\hat{v}^i\rangle \ \langle \hat{v}^i|, \tag{5-11}$$

where $\gamma_i \in \mathbb{R}$ and $\hat{v}^i \in H_{\mathcal{Y}}$. The important thing about this is that it is possible to make some sort of change of basis for the representation of the density matrices, and then it is possible to optimize this basis, rather than optimizing the density matrix directly. According to [52], the next proposition states the way to perform the quantum measurement when using decomposed matrices:

Proposition 2 *Given a dataset $\{(x_i, y_i); x_i \in X, y_i \in Y\}_{i=1}^{n}$, some quantum feature maps $\psi_{\mathcal{X}} : X \to H_{\mathcal{X}}$ and $\psi_{\mathcal{Y}} : Y \to H_{\mathcal{Y}}$, and given the density matrix*

$$\rho_{\mathcal{X}\mathcal{Y}} = \sum_{i=1}^{m} \beta_i |\psi_{\mathcal{X}}(x_i)\rangle \otimes |\psi_{\mathcal{Y}}(y_i)\rangle \langle \psi_{\mathcal{X}}(x_i)| \otimes \langle \psi_{\mathcal{Y}}(y_i)|, \tag{5-12}$$

and a new sample $x^* \in X$, then $p(y|x^*, \rho_{\mathcal{X}\mathcal{Y}})$ can be calculated as

$$p(y|x^*, \rho_{\mathcal{X}\mathcal{Y}}) = \rho_{\mathcal{Y}} = \sum_{i=1}^{m} \gamma_i |\psi_{\mathcal{Y}}(y_i)\rangle \langle \psi_{\mathcal{Y}}(y_i)|, \tag{5-13}$$

where

$$\gamma_i = \frac{\beta_i \langle \psi_{\mathcal{X}}(x^*), \psi_{\mathcal{X}}(x_i)\rangle^2}{\sum_{j=1}^{m} \beta_j \langle \psi_{\mathcal{X}}(x^*), \psi_{\mathcal{X}}(x_j)\rangle^2}. \tag{5-14}$$

This opens the way for using quantum measurements in a modular way within deep learning-based models.

5.4. Quantum Maps

Given the need for quantum maps, there are a variety of options for features and labels. For example, one-hot encoding is a natural representation for the labels. A softmax encoding is also proposed in [53]. On the other hand, for the features, given that the feature space induced by a kernel function is a Hilbert space, any quantum feature map should implicitly define a kernel function. In this scenario, the Random Fourier Features (RFF) [118] emerges as a way to approximate a quantum feature map from a kernel function.

5.4.1. Kernel Functions

Kernel functions are maps that allow an implicit representation of information in high-dimensional spaces. Kernel-based methods rely on the transformation of data from its original space X to a new space $\mathcal{H}$ (called Reproducing Kernel Hilbert Space or RKHS) where the task (either classification, regression, clustering, etc) can be done linearly. Lets call ϕ the function that is responsible for the transformation, $\phi : X \to \mathcal{H}$. Depending on the specific kernel, $\mathcal{H}$ might be infinite-dimensional, so the computations can not be done explicitly in $\mathcal{H}$; they must be done implicitly. This is when the kernel takes importance. A kernel function associated with ϕ and $\mathcal{H}$ is is a finitely positive semi-definite function such that for any $x_1, x_2 \in X$

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle_{\mathcal{H}}. \tag{5-15}$$

The kernel function is a shortcut to work in the feature space $\mathcal{H}$ without knowing explicitly the data representation in it. This is what is called the *Kernel Trick* [44]. The kernel gives information of the metric, and as the metric characterizes the space, just with the pairwise interaction of the elements you can have a complete characterization of what happens in $\mathcal{H}$. In conclusion, kernel functions are attractive and useful tools for transforming data from its original representation space to more convenient spaces. Any kernel function k such that $k(x, x) = 1$ implicitly defines a quantum feature mapping. However, as this mapping ϕ is implicit, we cannot use it directly in Eq. 5-14. This is where random Fourier features become important.

5.4.2. Random Fourier Features

The Random Fourier Features (RFF) is a method to create a feature map of the data $\phi_{\text{rff}}(x) : X \subset \mathbb{R}^n \to \mathbb{R}^D$ in which the dot product of the samples in the $\mathbb{R}^D$ space approximates a shift invariant kernel k. A shift invariant kernel is such that $k(x_1 - x_2) = k(x_2 - x_1)$. 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 $k(x_1 - x_2)$, and sampling i.i.d. $b_1, \cdots, b_D \in \mathbb{R}$ from a uniform distribution in $[0, 2\pi]$.
In our context, the shift invariant kernel is the Radial Basis Function (RBF) or Gaussian kernel given by

$$k_{\text{rbf}}(x_1, x_2) = e^{-\frac{1}{2\sigma^2}\|x_1 - x_2\|^2}, \tag{5-16}$$

where σ and the number D of RFF components would become hyper-parameters of the final models.

5.5. Quantum Measurement Ordinal Regressor (QMOR)

In the light of the above, Quantum Measurement Ordinal Regressor (QMOR) [54] is a differentiable probabilistic ordinal regressor model that uses a trainable density matrix, $\rho_{\mathcal{X}\mathcal{Y}}$, to represent the joint probability distribution of inputs and labels. A full diagram of the model is shown in Figure **5-2**. A QMR layer receives a RFF encoded input sample $|\phi_{rff}(x)\rangle$, and then builds a prediction operator $M = |\phi_{rff}(x)\rangle \langle\phi_{rff}(x)| \otimes \text{Id}_{\mathcal{H}_\mathcal{Y}}$. Inference is made by performing a quantum measurement on the training density matrix $\rho_{\mathcal{X}\mathcal{Y}}$ (see Eq. 5-12). Then $\rho_\mathcal{Y}$ (see Eq. 5-13) encodes in $\rho_{\mathcal{Y}rr}$, with $r \in \{0, \ldots, N-1\}$ (supposing we have N different grades), the posterior probability over the labels. The expected value represents the final prediction:

$$\hat{y} = \sum_{r=0}^{N-1} r\rho_{\mathcal{Y}rr}. \tag{5-17}$$

We use *one-hot* encoding for the labels, which are categorical for ordinal regression tasks. The model is trained using stochastic gradient descent to minimize a *Mean Squared Error* loss function with a variance term whose relative importance is controlled by hyper-parameter α:

$$L = \sum (y - \hat{y})^2 + \alpha \sum_r \rho_{\mathcal{Y}rr}(\hat{y} - r)^2. \tag{5-18}$$

Then, the learnable parameters of the model are the following:

- $\{w_1, \cdots, w_D\} \subset \mathbb{R}^n$, the components of the random Fourier features. Those are randomly initialized by sampling from the Fourier transform of the Gaussian kernel (Eq. 5-16).

- $\{\beta_i, |v_i\rangle\}_{i=1}^m$ where $\beta_i \in \mathbb{R}$ and $v^i \in H_{\mathcal{X}\mathcal{Y}}$. These are the weights and vectors of the decomposed density matrix $\rho_{\mathcal{X}\mathcal{Y}}$ (see Eq. 5-10). The vectors can be initialized by randomly chosen m samples from the training set and encoding them with ϕ_{rff} for the part $H_\mathcal{X}$ and with *one-hot* for the part $H_\mathcal{Y}$. β_i are initialized in $\frac{1}{m}$ for all i.

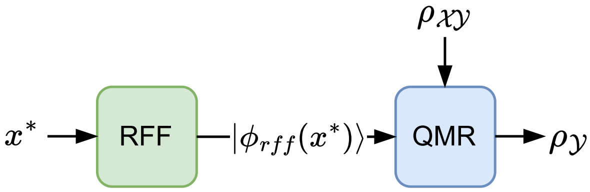


Figure 5-2.: QMR model for regression. A new sample x^* is first passed through a Random Fourier Features (RFF) layer and then through the quantum measurement.

5.6. Kernel Quantum Measurement Unit (KQMU)

Given Eq. 5-14 and the kernel trick in Eq. 5-15, it follows that the inference process described in Proposition 2 can be completely stated in terms of kernel functions, since the entire computation depends only on inner products in the feature space $H_{\mathcal{X}}$. Then, we would no longer need quantum feature maps, and we would use kernel functions directly. Given a kernel $k_{\mathcal{X}}$, the Eq. 5-14 would look like this:

$$\gamma_i = \frac{\beta_i k_{\mathcal{X}}(x^*, x_i)^2}{\sum_{j=1}^{m} \beta_j k_{\mathcal{X}}(x^*, x_j)^2}. \tag{5-19}$$

Furthermore, the inference process can be generalised to handle mixed input states. In other words, as we have said, until now the proposed models receives a pure state representing a new sample x^*. It turns out that quantum measurement can be extended to measure mixed states over mixed states. The following proposition summarises the whole idea above and generalises the process described in Proposition 2.

Proposition 3 *Given a dataset* $\{(x_i, y_i); x_i \in X, y_i \in Y\}_{i=1}^{n}$, *some kernel functions* $k_{\mathcal{X}}$ *and* $k_{\mathcal{Y}}$ *and their corresponding feature maps* $\psi_{k_{\mathcal{X}}}$, $\psi_{k_{\mathcal{Y}}}$, *and given the density matrix*

$$\rho_{\mathcal{X}\mathcal{Y}} = \sum_{i=1}^{m} \beta_i |\psi_{k_{\mathcal{X}}}(x_i)\rangle \otimes |\psi_{k_{\mathcal{Y}}}(y_i)\rangle\langle\psi_{k_{\mathcal{X}}}(x_i)| \otimes \langle\psi_{k_{\mathcal{Y}}}(y_i)|, \tag{5-20}$$

and a mixed state

$$\rho_{\mathcal{X}} = \sum_{i=1}^{m'} \alpha_i |\psi_{k_{\mathcal{X}}}(x'_i)\rangle\langle\psi_{k_{\mathcal{X}}}(x'_i)|, \tag{5-21}$$

then $p(y|\rho_{\mathcal{X}}, \rho_{\mathcal{X}\mathcal{Y}})$ *can be calculated as*

$$p(y|\rho_{\mathcal{X}}, \rho_{\mathcal{X}\mathcal{Y}}) = \rho_{\mathcal{Y}} = \sum_{i=1}^{m} \gamma_i |\psi_{k_{\mathcal{Y}}}(y_i)\rangle\langle\psi_{k_{\mathcal{Y}}}(y_i)|, \tag{5-22}$$

where

$$\gamma_i = \sum_{l=1}^{m'} \frac{\alpha_l \beta_i k_{\mathcal{X}}(x'_l, x_i)^2}{\sum_{j=1}^{m} \beta_j k_{\mathcal{X}}(x'_l, x_j)^2}. \tag{5-23}$$

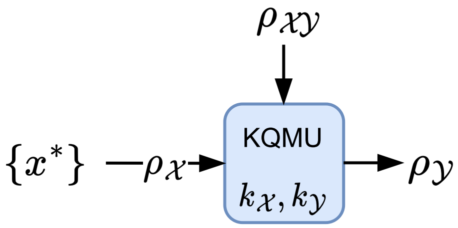


Figure 5-3.: KQMU schema. This layer relies on kernel functions $k_{\mathcal{X}}$, $k_{\mathcal{Y}}$ that implicitly define quantum maps and is a processing unit based on quantum measurements for density matrices between different feature spaces.

Proposition 3 allows us to build a unit for processing density matrices between different feature spaces. We will call this unit KQMU: Kernel Quantum Measurement Unit (see Figure **5-3**). The input and output of this layer are the matrices $\rho_{\mathcal{X}}$, $\rho_{\mathcal{Y}}$ which represent in a probabilistic way the information about data lying in $H_{\mathcal{X}}$ and $H_{\mathcal{Y}}$ respectively. Quantum measurement allows us to go from space $H_{\mathcal{X}}$ to the space $H_{\mathcal{Y}}$, and it also allows us to pass through intermediate representation spaces. This allows the unit not only to be coupled to the output of a model, but also to be used as an intermediate processing layer, and again to be used freely hand in hand with deep learning based models.

In addition, we most note that, regarding the learnable parameters of the layer, namely $\{\beta_i, |v_i\rangle\}_{i=1}^m$, these may be initialized by $\beta_i = \frac{1}{m}$ and randomly chosen m samples $\{x_i, y_i\}_{i=1}^m$ from the training data to define $|v_i\rangle = |\psi_{k_{\mathcal{X}}}(x_i)\rangle \otimes |\psi_{k_{\mathcal{Y}}}(y_i)\rangle$. This is just the same initialization used for QMOR. However, in QMOR the optimization occurs explicitly in an approximation of the feature space $H_{\mathcal{X}}$. In KQMU, this optimization occurs implicitly in $H_{\mathcal{X}}$ but explicitly in X. Therefore after the training, we finished with an optimal set of weights and vectors $\{\hat{\beta}_i, \hat{x}_i\}_{i=1}^m$. According to Eq. 5-23, the inference process for a new sample relies almost entirely on the similarity (given by the kernel) between the new sample and the learned $\{\hat{x}_i\}_{i=1}^m$. This is, $\{\hat{x}_i\}_{i=1}^m$ is a set of representative optimal vectors of the dataset (not necessarily present in the training set). We will therefore call these vectors the *prototypes* of the model.

5.6.1. Multimodal Kernel Fusion

A multimodal extension of KQMU follows directly from the properties of the kernels. Basically, fusion is achieved by using a kernel that merges the representation of each modality from their respective kernels. How this is done is explained below.

Without loss of generality, let's suppose we have only two different modalities: V and T. Let be k_V and k_T the kernel functions used for those data modalities respectively. Let $x_v^1, x_v^2 \in X_v$ be two samples of the first modality and $x_t^1, x_t^2 \in X_t$ their respective counterparts of the second modality. Then, by the properties of kernel functions [132], it is well known that the function $k_{V\oplus T}$ defined as

$$k_{V\oplus T}(x_v^1 \oplus x_t^1, x_v^2 \oplus x_t^2) = k_V(x_v^1, x_v^2) + k_T(x_t^1, x_t^2), \tag{5-24}$$

is a kernel function itself. This means the Cartesian product of the feature spaces induced by k_V and k_T matches the feature space induced by kernel defined as the sum of k_V and k_T. Analogously, the multiplication of k_V and k_T results in the kernel that induces the tensor product of the feature spaces induced by k_V and k_T:

$$k_{V\otimes T}(x_v^1 \otimes x_t^1, x_v^2 \otimes x_t^2) = k_V(x_v^1, x_v^2) \times k_T(x_t^1, x_t^2). \tag{5-25}$$

Therefore we must highlight that by using kernel functions we can build and learn representations in a joint multimodal feature space created from the Cartesian or tensor product of each modality feature space. And, although still being a Cartesian or tensor product, the main difference from any explicit fusion approach (like those from neural network based models) is that with the kernel functions we never have to deal with the explicit representation of the feature space and therefore we do not have any computational restrictions related to a high dimensional representation.

5.7. Conclusions

In this chapter we set the foundation for the quantum-based models to come. The method presented in [53] for supervised machine learning based on the mathematical formalism supporting quantum mechanics, reveals an interesting framework where the relationship between input and output variables is embodied as the state of a bipartite quantum system. The use of projective quantum measurement to construct a prediction function marks a unique departure from traditional machine learning methods and opens up new directions for research and application.

The use of quantum density matrices in probabilistic deep learning provides a simpler yet effective mechanism for representing joint probability distributions of both continuous and discrete random variables. This approach enables the construction of differentiable models for density estimation, classification and regression, and facilitates their integration into end-to-end deep neural network models.

In addition, being able to use kernel functions directly as quantum feature maps allows great flexibility in the handling of the representations and, fundamentally, allows the adaptation of these models in multimodal applications.

6. Deep Quantum Measurement Regression for Diabetic Retinopathy and Prostate Cancer Grading

This chapter approaches disease diagnosis as a purely ordinal regression problem. Unlike the Gaussian process, which is a continuous regression model that fits a discrete regression, the quantum measurement approach allows the problem to be posed directly as a discrete regression, while retaining the ability to measure uncertainty given the probabilistic nature of the predictions.

The work in this chapter as been published as:

Toledo-Cortés S., *Useche D.H., Müller H., González F.A. Grading diabetic retinopathy and prostate cancer diagnostic images with deep quantum ordinal regression. In: Computers in Biology and Medicine 145 (2022), S. 105472. – ISSN 0010–4825* [148].

6.1. Introduction

The stages of a disease are not categorical. The degenerative process of a disease is not a discrete jump from one class to another but a progressive continuum [93]. These stages are therefore an attempt of the specialists to discretize a continuous behavior. While not completely accurate, this information is useful in the generation of automatic systems if a model with an appropriate descriptive capability is used. However, the information of the relative distance between the different grades of a disease is disregarded when a categorical classification model is used. The way to exploit this grading is therefore through a regression model. In addition, if a probabilistic regression model is implemented, the predictions can be interpreted as probability distributions over the range of the labels. Hence, one can infer the stage of the disease in a non-categorical way, providing more statistical information, for instance, the uncertainty of the predictions.

In the medical field, deep CNNs have been demonstrated to be effective at analyzing images and visual content of all kinds, from X-rays to diagnose osteoporosis [84], to MRIs to diagnose brain conditions [175]. Although many diseases present different stages on a progressive scale and in many cases this information is available, binary labels are usually favored [93]. However, two drawbacks arise from addressing the task of classifying grade-based medical images as a categorical problem with conventional neural networks: first, the ordinal information of the grades is not taken into account for the training process, and second, the predictions of the models, usually subject to a softmax activation function, cannot be interpreted as probability distributions [154].

In this chapter, we present the Deep Quantum Ordinal Regressor (DQOR), a deep probabilistic model capable to combine a CNN with the Quantum Measurement Regression (QMR) [55, 54]. DQOR is intended as a diagnostic support tool for the medical specialist which allows to:

1. Predict posterior probability distributions over the grades range. Unlike other probabilistic methods such as Gaussian processes, these are explicit discrete distributions.
2. In the case of patch-based image analysis, integrate patch posterior distributions into a single whole-slide image distribution using a simple, yet powerful probability-based strategy.
3. Quantify the uncertainty of the predictions. This enriches the model as a diagnostic support tool, which in safety-critical applications, provides the method with a first-level of interpretability.
4. Improve the posterior prognosis-oriented binary diagnosis, based on an ordinal grade-label end-to-end training.

To show the effectiveness of our DQOR proposal, we test it on two grade-based diagnostic tasks: prostate cancer (PCa) diagnosis, and diabetic retinopathy (DR) diagnosis.

Early PCa detection allows for greater treatment options and a higher 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 [45]. 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 label the tissue regions with Gleason patterns on a scale from 1 to 5. The sum of the two most dominant Gleason patterns gives the final Gleason score. Hence, the Gleason score ranges from 2 to 10. However, in practice, the specialists only consider the highest five grades, from 6 to 10, since biopsies with a grade below 3 are not taken into account [90]. The higher the grade, the more advanced cancer. Although the automatic classification of PCa with CNNs has been widely studied, the usual approach has been as multi-class or binary classification of low risk (6-7 GS) vs high risk (8-10 GS) tasks [81, 136].

Something similar happens with DR. Early DR diagnosis allows preventing most of the severe consequences of the disease, including complete blindness [161]. As we mentioned in previous chapters, an early and effective diagnosis of DM consists of the inspection of the retinal tissue by means of an eye fundus image. By counting the number of visible lesions, an ophthalmologist specializing in the retina can give a diagnosis of the disease, on a five-level severity scale from 0 to 4, being 0 a negative case of DR, and 4 a case of proliferative DR [4]. Many approaches have treated the problem from a multi-class classification perspective, or as a binary task, where a diagnosis of 0 or 1 corresponds to a case of *non-referable* DR and a case of 2, 3, or 4 corresponds to a case of *referable* DR [135].

In the medical field, while it is true that there have been some applications of ordinal regression models, there is not a clear and well-defined trend. Recently, ordinal regression by binary classifiers has been applied to facial age estimation [109] [138], and diagnosis of Alzheimer's disease [89], taking advantage of the inherent ordinal severity of brain degeneration.

In addition to the predictive value, for medical applications, it is desirable to obtain a probability distribution over the possible output stages. Furthermore, the distribution describing the probability of belonging to a disease stage should be unimodal, so it is expected to use unimodal distributions for ordinal classification tasks [13]. Various models have been proposed where predictions are forced to follow Poisson or binomial distributions over the possible outputs [12, 13], showing that, when needed, the ordinal approach improves the results compared to the conventional cross-entropy approach.

Our method manages to capture the ordinal structure of the data, based on a strong theoretical framework, and with a greater versatility of integration. DQOR allows the predictions to be actual probability distributions over the range of degrees, without the need to force them with the softmax or other activation functions. This in turn allows local predictions to be integrated into global predictions, in the case of PCa, and also allows the variance to be interpreted naturally as a measure of uncertainty. Finally, unlike classical probabilistic methods, DQOR can be trained with gradient descent, enabling its integration with conventional deep learning architectures.

This chapter is organized as follows: Section 6.2 presents the theoretical framework of the method, and Section 6.3 presents the experimental setup. To validate our approach, we compare the performance of our model with state-of-the-art deep learning-based models, and with various closely related classification and regression methods. In Section 6 we present the conclusions of this work.

6.2. Method: Deep Quantum Ordinal Regressor (DQOR)

The overall architecture of the proposed Deep Quantum Ordinal Regressor (DQOR) is described in Figure **6-1**. We use a deep CNN as a feature extractor. The extracted features are then used as inputs for the QMR method [54]. QMR uses density matrices for regression problems and works as a density estimator. It requires an additional feature mapping from the inputs to get a quantum state-like representation. This is achieved using a random Fourier features approach [118]. The regressor yields a discrete posterior probability

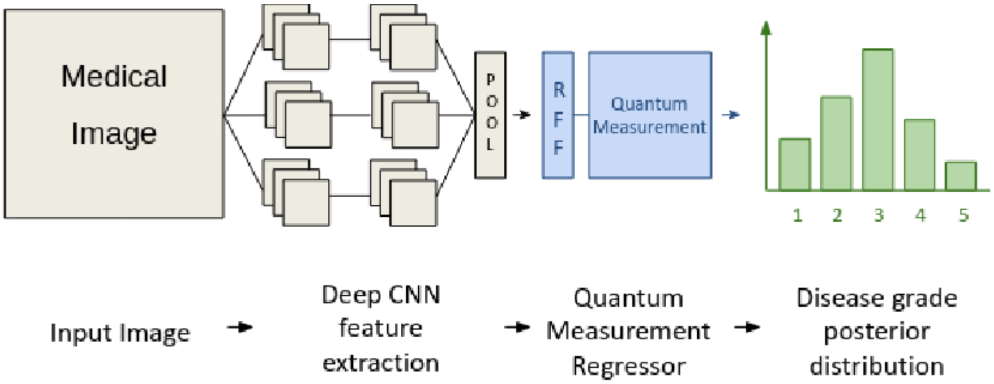


Figure 6-1.: Overview of the proposed DQOR method for medical image analysis. A deep CNN is used as feature extractor for the input image. Those features are the input for the QMR regressor model, which yields a posterior probability distribution over the possible grades of the disease.

distribution from which we get the final grade prediction and a measure of the uncertainty.

6.2.1. Feature Extraction

Medical and non-medical automatic image analysis relies on deep convolutional neural networks. The representational power of these models has shown remarkable results on computer vision and therefore we use them for feature extraction [57]. Regardless of the CNN architecture, these models conserve a basic structure: an input layer for the image, followed by a series of convolutional blocks and a pooling layer that summarizes all the information extracted from the convolutions. Usually, this layer is connected to a series of dense layers that perform the final classification of the image. Instead, we use the output of the pooling layer to feed the Quantum Measurement Regressor.

6.2.2. Patch-based Analysis Summarization

Patch-based image analysis is preferred or rather needed in some applications. This is the case of prostatic cancer diagnoses with WSI. We require the additional step of summarizing the predictions of the patches to reach a prediction of a whole slide image. The most straightforward procedure is the majority vote (MV), as reported in most previous works [81, 110]. In the majority vote, the image's prediction is decided according to the grade with the highest number of predictions among the patches of the image. However, as in [149], DQOR admits a probability vote procedure (PV); 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,

$$P(r|I) = \frac{P(r,I)}{P(I)} = \frac{\sum_{i=1}^{n} P(r|p_i,I)P(p_i|I)P(I)}{P(I)} = \frac{1}{n}\sum_{i=1}^{n} P(r|p_i). \tag{6-1}$$

The final prediction may correspond to the grade with the highest probability or with the expected value of the distribution.

6.3. Experimental Set Up

The specific details of the experimental procedures are described below. The implementation was developed in Python, and the code is available at `https://github.com/stoledoc/DQOR`.

6.3.1. QMR Hyperparameter Optimization

As described in Chapter 5.4 and 5.5, the RFF and QMR layer of the DQOR requires five hyperparameters to be set before training. In addition to the usual learning rate, we have σ and D as hyperparameters of the RFF embedding, σ controls the spread of the objective Gaussian kernel that we try to approximate and D corresponds to the number of random Fourier features, which determines the dimension of the embedding space. Also, the training density matrix $\rho_{\mathcal{X}\mathcal{Y}}$ (see Eq. 5-10) depends on the number of eigencomponents, and the loss function (see Eq. 5-18) on the parameter α to control the variance of the predictions.

Due to the high number of possible hyperparameter combinations, we made a random search to look for the optimal setup. However, there are some important aspects to take into account to refine this process, which are presented below.

From a subset of the original dataset, we generated more than 3000 combinations of hyperparameters to train the model. In each case, we established an early stopping callback to halt the training process after 10 epochs with no improvement in the validation loss. We recorded the final MAE scores of the validation data set of all the resulting models. From this information, we made a statistical analysis of the hyperparameter sensibility of the model, by measuring the relative change of the MAE in comparison to the relative change of each hyperparameter. We looked at the density distribution of these combinations of hyperparameters as a function of the relative change of the MAE (see Figure **6-2**). We conclude that the more sensitive parameters of the QMR are learning rate and σ.

For the random search of the whole datasets, we set a range from 10^{-8} to 10^{-2} for the learning rate. Since σ measures the dispersion of the data, we took the mean of the pairwise distances between the data samples and the mid-point of the data for the range of search. For the number of random features, D, we explored 2^7, 2^8, 2^9, 2^{10} and 2^{12}, taking into account that the output of the Xception and the Inception-V3 is a 2048-dimension vector. For the number of eigenvectors, we explored five different fractions of the chosen D:

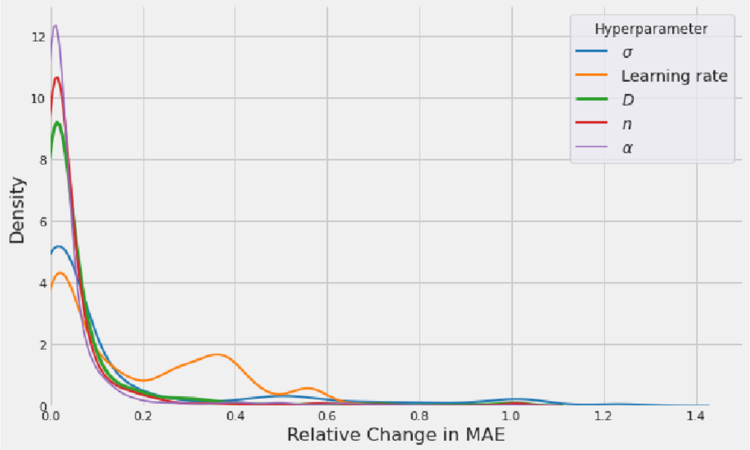


Figure 6-2.: Density plot of the ratio between the relative change of MAE and relative change of each hyperparameter. Although the mode of all distributions is close to zero, it can be noted that the variances of the learning rate and σ distributions are higher in comparison with the other three hyperparameters. This implies that the sensitivity of the model, measured against the variance of the MAE in a validation set, is higher in these two parameters.

D, $D/2$, $D/4$, $D/8$, $D/16$. Finally, for the α parameter, we set the searched range from 0 to 1.

6.3.2. Prostate Cancer

The setup of the DQOR applied to the PCa image analysis is described in Figure **6-3**.

Dataset

We used images from the TCGA-PRAD dataset, which contains samples of prostate tissue with GS from 6 to 10. This data set is publicly available via The Cancer Genome Atlas (TCGA) [67]. To directly compare our results with the baseline [81], we used the same subset and partition for train and test, the details of the partition are presented in Table **6-1**. The process to extract the patches from WSI is described in [67].

We used the images' patches to train the model. To obtain predictions at the level of WSIs, a process of summarization was carried out. Each patch was labeled with the same GS of the WSI from which it belongs. Although it is not clear that a GS can be assigned to a single patch, our methodology focused on showing the effectiveness of the regression approach by comparing it with previous works which use the labels of patches in the same manner, however, in theory, a GS can be set to each patch of a WSI.

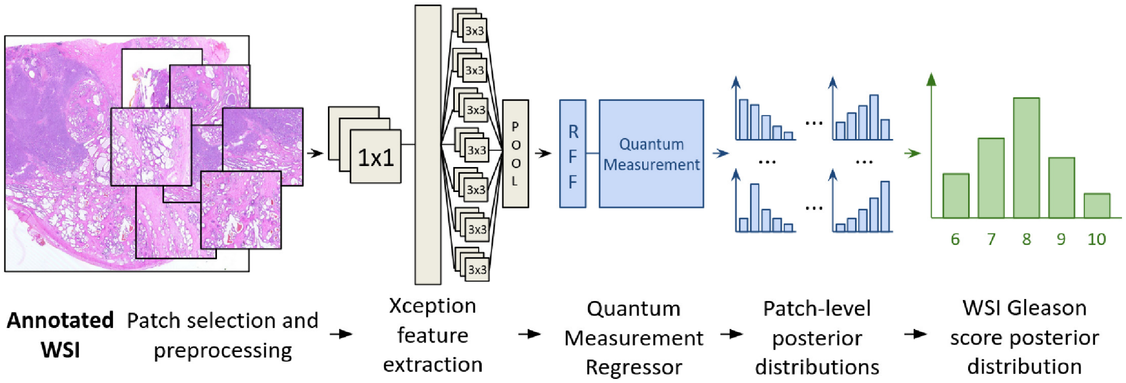


Figure 6-3.: Overview of the proposed DQOR method for prostate tissue grading. The Xception network was used as a feature extractor of the images patches. Those features were the input for the QMR regressor model which yielded a posterior probability distribution by patch over the Gleason scores. Finally, those distributions were summarized into a single discrete probability distribution of the WSI.

Table 6-1.: Details of the subset and final partition of the TCGA dataset used for training and testing. This is the same partition used in [81].

Risk	Gleason Score	Train	Validation	Test
Low	6	11	4	4
Low	7	53	17	17
High	8	23	8	8
High	9	50	17	16
High	10	4	2	1

Feature Extraction

The model presented in [81] was used as a feature extractor. It 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. The augmentation procedure and training details are described in [81]. From the output of the last average pooling layer of the model, we got a 2048-dimensional vector representing each image patch.

Quantum Measurement Regression

For the QMR, hyper-parameter tuning of the model was performed by generating 25 different random configurations choosing the best combination. Henceforth, we created an embedding of 1024 RFF components with σ equals to 2^6. Also, the density matrix was trained with 32 eigenvalues. For the loss function (See eq. 5-18) the value of 0.4 was selected for α, and the learning rate was set to 6×10^{-5}.

Baseline

We extended the feature extractor with a conventional feed-forward neural network as a baseline in this work. Called *DLC-PCa* hereafter, it consists of 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. The learning rate was set to 10^{-7}, as in the baseline [81]. We also explored two closely related methods to QMR: the Density Matrix Kernel Density Classification (DMKDC) [54] and the Gaussian process. DMKDC is a differentiable classification method, which applies the 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. A Gaussian process (GP) [120] is a powerful Bayesian approach to regression problems. Through a kernel covariance matrix, the GP calculates and iteratively updates 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 Gaussian kernel. The prediction process consists of marginalizing the learned Gaussian distribution from which the mean would be the actual predicted value and its standard deviation an indicator of the uncertainty. We also explored deep Gaussian processes (DGP) [37], which also use RFF to approximate the covariance function. For those experiments, another hyper-parameter random search was made, finally setting the number of RFF to 1024 and the learning rate to 2×10^{-7} in a single layer schema.

6.3.3. Diabetic Retinopathy

The setup of the DQOR applied to eye fundus images' analysis is presented in Figure **6-4**.

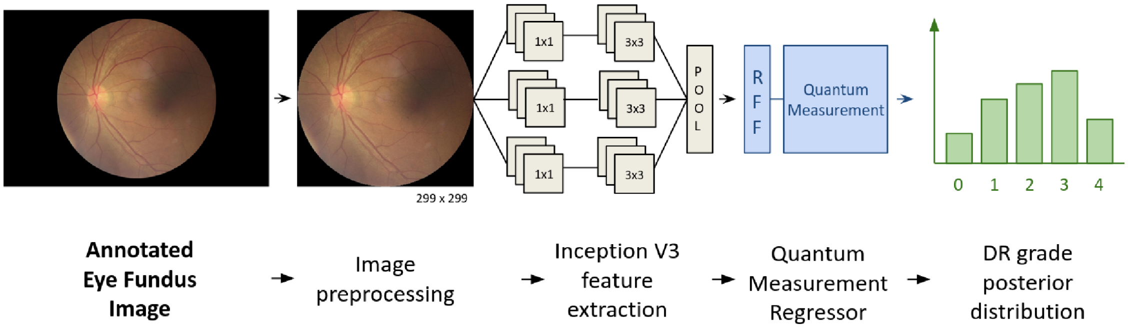


Figure 6-4.: Overview of the DQOR for diabetic retinopathy grading. An Inception-V3 network was used as feature extractor for the eye fundus image. These features were the input for the QMR regressor model, which yielded a posterior probability distribution over the DR grades.

Datasets

To directly compare the DQOR performance with the baselines of the state-of-the-art, we worked with EyePACS [40] and Messidor-2 [2] datasets. EyePACS is one of the largest publicly available datasets of eye fundus images. Each sample is labeled as one grade of the five grade scale from 0 to 4, where 0 stands for a healthy case, 1 for mild non-proliferative DR, 2 for moderate non-proliferative DR, 3 for severe non-proliferative DR, and 4 for proliferative DR. To compare our method with our baselines [157] [144], we kept the same partition for training and testing, which are described in Table **6-2**. This data configuration of EyePACS has, however, a drawback: in the test data set there are no samples of DR grades 2 and 3. This is justifiable in the binary context for which it was originally designed, but it is not a fair evaluation for the ordinal classification case. Therefore, we set up a second partition for EyePACS, called EyePACS-b, based on the former one, but moving some samples from the train set to the test set. The details of the EyePACS-b partition are described in Table **6-3** Regarding Messidor 2, it is a standard dataset in the field for testing. It consists of 1748 eye fundus images. While DR grades are not provided, we used them to show the effectiveness of our proposal for a *referable/non-referable* diagnosis. Details of Messidor-2 are described in Table **6-4**.

Feature Extraction

We used the model presented in [144] as a feature extractor, which is also publicly available and consists of an Inception-V3 network trained on ImageNet and fine-tuned on eye fundus images with the EyePACS train partition. This network had already been used as a feature extractor in a model for automatic DR grading. In such a model, the training was made in two independent stages, one for the Inception-V3 and another for the Gaussian process.

Table 6-2.: Details of the subset and final partition of the EyePACS dataset used for training and testing. This partition is the same used in [144] and in [157].

Referable	Grade	Train	Test
No	0	37209	7407
No	1	3479	689
Yes	2	12873	0
Yes	3	2046	0
Yes	4	1220	694

Table 6-3.: Details of the subset and final partition of the EyePACS-b partition used for a fair ordinal regression evaluation with samples in all the grades, in contrast to the partition used in [144] and in [157].

Referable	Grade	Train	Test
No	0	37209	7407
No	1	3479	689
Yes	2	10298	2575
Yes	3	1637	409
Yes	4	1220	694

Table 6-4.: Details of Messidor-2 dataset used for testing. Messidor-2 is used to compare the performance of the model in a purely binary task (*referable/non-referable*).

Referable	Total
No	1368
Yes	380

However, the Inception-V3 was trained for the binary *referable / non-referable* DR diagnosis. The same training setup was used to train an Inception-V3 on EyePACS-b. From the output of the last average pooling layer of the Inception-V3, we got a 2048-dimensional vector representing each eye fundus image.

Quantum Measurement Regression

We performed a random search for the QMR hyper-parameters fixing the Inception-V3 stage and generating 25 different random configurations. As result, we chose an embedding of 128 RFF components, 8 eigencomponents, and σ was set to 2^5. For the loss function (Eq. 5-18), α was set at 0.6, optimized with a learning rate of 7×10^{-5}.

Baseline

Similar to the baseline used for the PCa grading case, an extension of the feature extractor model with two dense layers was set up as a baseline for this task (called *DLC-DR* hereafter). Correspondingly, we report the results from the deep Gaussian process [37] and the DMKDC model [54]. We also present the results of the Gaussian process approach proposed in [144].

6.4. Experimental Results and Discussion

To measure the performance of an ordinal regression method requires taking into account the severity of misclassified samples. Usual categorical classification metrics, such as accuracy or F1-score, are not appropriate to estimate the actual performance of an ordinal classification. For example, given a sample whose actual label is grade 3, it is more severe if a model classifies the sample as grade 1 than grade 2. The separation between the predictions and the actual labels of the models is especially relevant in the medical field. Therefore, it is required a metric that quantifies the magnitude of the misclassification error. Between all the possible metrics, *Mean Absolute Error* (MAE) is currently one widely used measure in ordinal regression, both for evaluation and the loss function of the models [48]. MAE is a metric that penalizes misclassifications according to their distance to the true labels. Therefore, in addition to the categorical classification metrics, we also measured and reported MAE on the test data sets.

Finally, since it is desirable to measure the performance of the models for the binary classification task, we binarized the regressor predictions and compared this strategy with state-of-the-art models which were built for this specific purpose. Accuracy was used to measure the performance in the PCa case, and sensitivity, specificity, and AUC for the DR case.

Table 6-5.: Patch-level multiclass results of the dense layers classifier model DCL-PCa, Gaussian process GP, DGP, and density matrix-based models DMKDC, DQOR.

Method	Accuracy	Macro F1	MAE
DLC-PCa [81]	**0.593**	**0.359**	0.698
GP [120]	0.399	0.255	0.777
DGP [37]	0.265	0.169	1.013
DMKDC [54]	0.584	0.377	0.717
DQOR	0.515	0.317	**0.6807**

6.4.1. Prostate Cancer

WSI scores were summarized utilizing MV and PV. The prediction methods at the WSI level were also applied to the baseline models. In the dense layer classifiers, the summarization was made from the softmax output, as in [149]. In the DMKDC, the summarization 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 at patch-level and at WSI-level are reported in Table **6-5** and Table **6-6** respectively. In terms of multi-class accuracy at the patch level, the DLC-PCa model obtained the highest results. This was expected since this model is trained to optimize the categorical cross-entropy loss function. The difference with the regression approach is noticeable in the MAE, for which DQOR reached the best performance.

At the WSI level, the best multi-class accuracy was also reached with the DLC-PCa model and with the DMKDC with probability vote. Regarding the regression performance, the DQOR obtained the lowest MAE at the WSI level.

In general, it should be highlighted that higher performances in the categorical classification do not imply higher performances for ordinal classification. The difference between the model with the highest accuracy and the model with the lowest MAE is shown in Figure **6-5**. The DQOR confusion matrix indeed presents a higher concentration of samples around the diagonal, showing that the model takes advantage of the probability distributions and the inherent ordering of the GS grades.

By considering the predicted GS of 6 or 7 as *low* GS, and the predicted GS of 8, 9, or 10 as *high* GS, we binarized the results and computed the accuracy to make a direct comparison with previous works using the same dataset. The results are reported in Table **6-7**.

The results reported in [67] and [121] were obtained by training the binary labels of the WSIs with CNNs. The model presented in [81] is a multimodal approach, which used text reports as additional information to enrich the predictions of the WSIs, this model makes inferences from visual information alone, and also used binary labels. We can see that DQOR reached the highest binary accuracy and hence, it performed better on the gradation task. This approach was beneficial for the posterior binarization of the model.

Table 6-6.: WSI-level results. For each model, two summarization procedures were applied, majority vote (MV) and probability vote (PV).

Method	Accuracy	Macro F1	MAE
DLC-PCa MV [81]	**0.608**	0.354	0.7173
GP MV [120]	0.391	0.233	0.739
DGP MV [37]	0.174	0.059	0.935
DMKDC MV [54]	0.608	0.354	0.717
DQOR MV	0.587	**0.361**	0.695
DLC-PCa PV [81]	**0.608**	0.354	0.717
DMKDC PV [54]	**0.608**	0.354	0.717
DQOR PV	0.587	0.356	**0.652**

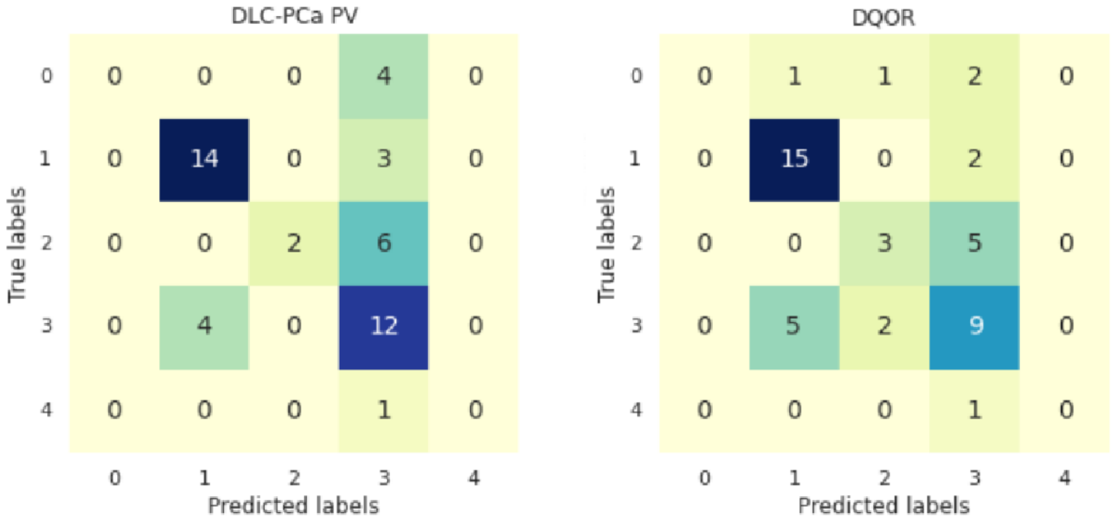


Figure 6-5.: Confusion matrices of the WSI grade predictions for DLC-PCa (left) and for DQOR (right) in the TCGA test partition. WSI prediction is obtained using the probability vote.

Table 6-7.: Results at WSI-level of *low* risk vs *high* risk.

Method	Accuracy
Google LeNet [67]	0.7352
Modified AlexNet[121]	0.769
M-LSA [81]	0.770
DQOR	**0.782**

Table 6-8.: Comparison on EyePACS test partition results. Sensitivity, specificity and AUC for binary classification and MAE for grading.

Description	Sensitivity	Specificity	AUC	MAE
DLC-DR [144]	0.7867	0.9643	0.9471	0.3166
Voets et al. (2019) [158]	0.906	0.847	0.951	–
GP [144]	**0.9323**	0.9173	0.9769	0.7750
DGP [37]	0.3703	0.6196	0.4947	1.3566
DMKDC [54]	0.6473	0.8787	0.9135	0.4051
DQOR	0.8660	**0.9809**	**0.9805**	**0.2871**

6.4.2. Diabetic Retinopathy

The results for DR grading on the EyePACS test set are reported in Table **6-8**. As previously mentioned, the methods tested in this work generated a prediction for a five-grade range, and a binarization of the results was performed by defining a threshold on the predicted value. Henceforth, we report the ROC-AUC score to directly compare it with the state of the art (see Figure **6-6** and Figure **6-7**). For Messidor-2 we only report the binary classification performance in Table **6-9**.

For the ordinal regression task, DQOR was the best performing model according to MAE. Furthermore, for both partitioned datasets, it reported the highest AUC for the binary *referable / non-referable* task.

Also, the results of DQOR using the EyePACS-b partition are reported in Table **6-10** and the confusion matrix is shown in Figure **6-8**, next to the confusion matrix for DLC-DR. DQOR had the lowest MAE and the highest AUC. Note that DLC-DR was the model with the highest sensitivity, and the second-highest AUC, just below DQOR. However, it was not the model with the second-lowest MAE. The confusion matrices show again that the results of the DQOR had a lower dispersion around the diagonal, which improved the results for the subsequent binary task.

In general, it is noticeable that for both the ordinal and the binary classification tasks, our proposed DQOR improves the performance of the previous models, which justifies once again the importance of using the different stages of the disease for an automatic diagnosis.

6.4.3. Uncertainty Quantification

In addition to the performances of the method on the classification and regression tasks, DQOR offers an uncertainty quantification based on the variance of the predicted distribution for each sample. For the PCa diagnosis, we analyzed the statistical behavior of the predicted variance on the test data set at the WSI level, grouping the samples according to whether or not they were correctly classified on the binary task. Figure **6-9** shows box-

Table 6-9.: Comparison on Messidor-2 results. Sensitivity, specificity and AUC for binary classification.

Description	Sensitivity	Specificity	AUC
DLC-DR [144]	0.6105	**0.9715**	0.8624
Voets 2019 [158]	**0.8181**	0.712	0.853
GP [144]	0.7237	0.8625	0.8787
DGP [37]	0.4026	0.5782	0.4960
DMKDC [54]	0.5906	0.5316	0.5864
DQOR	0.7974	0.9291	**0.9239**

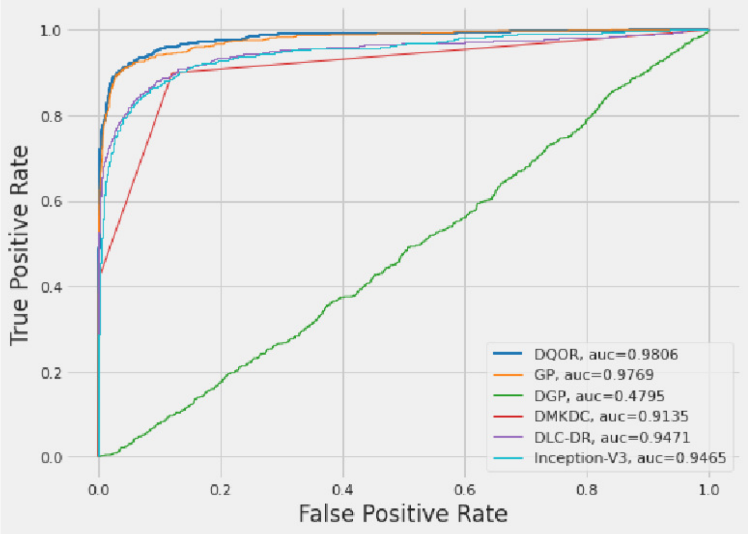
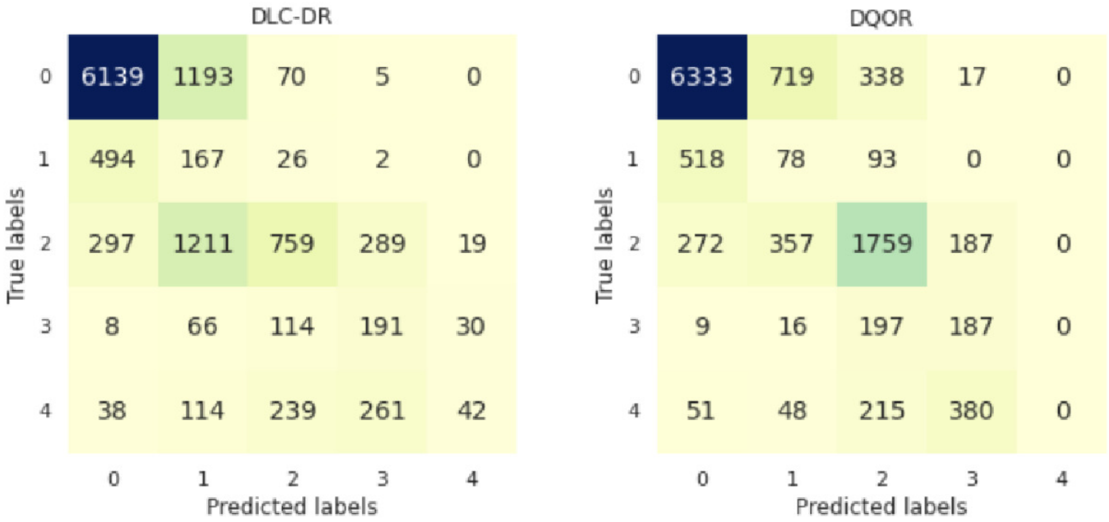


Figure 6-6.: ROC curves plot for EyePACS test set.

Table 6-10.: Comparison on EyePACS-b test partition results. Sensitivity, specificity and AUC for binary classification and MAE for grading.

Description	Sensitivity	Specificity	AUC	MAE
DLC-DR [144]	**0.9517**	0.7308	0.9363	0.4702
GP [144]	0.9454	0.5903	0.8385	0.5939
DGP [37]	0.2504	0.7510	0.5018	1.4954
DMKDC [54]	0.9166	0.8121	0.8736	0.4164
DQOR	0.9152	**0.8461**	**0.9438**	**0.3872**

Figure 6-7.: ROC curves plot for Messidor 2.



DLC-DR

True labels \ Predicted labels	0	1	2	3	4
0	6139	1193	70	5	0
1	494	167	26	2	0
2	297	1211	759	289	19
3	8	66	114	191	30
4	38	114	239	261	42

DQOR

True labels \ Predicted labels	0	1	2	3	4
0	6333	719	338	17	0
1	518	78	93	0	0
2	272	357	1759	187	0
3	9	16	197	187	0
4	51	48	215	380	0

Figure 6-8.: Confusion matrices of the predictions of DLC-DR (left) and DQOR (right) in the EyePACS-b test partition.

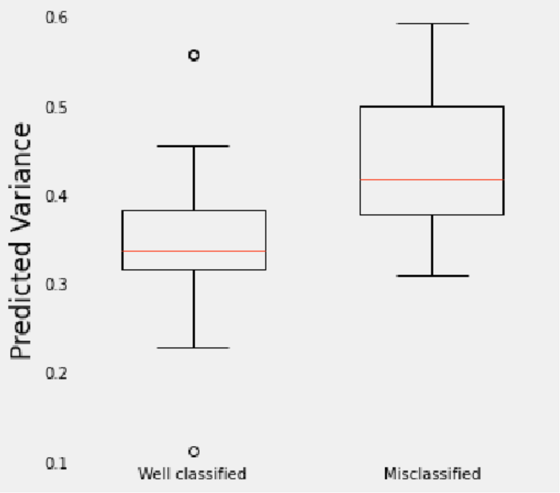


Figure 6-9.: Box plot of the predicted variance on TCGA test samples at WSI-level, grouped by classification status on the low risk vs. high risk GS diagnosis task.

plots of the predicted variance for each group. A similar procedure was performed for the DR diagnosis (see Figure **6-10**). As expected, DQOR predicts low uncertainties on well-classified samples in comparison with miss-classified samples. For the case of DR diagnosis, it is remarkable that for the EyePACS and Messidor-2 datasets the range of the variances are directly comparable, and they have similar statistical behavior. The uncertainty quantification provides the specialist with a more interpretable result, from which he may decide whether to trust or not on the model's prediction.

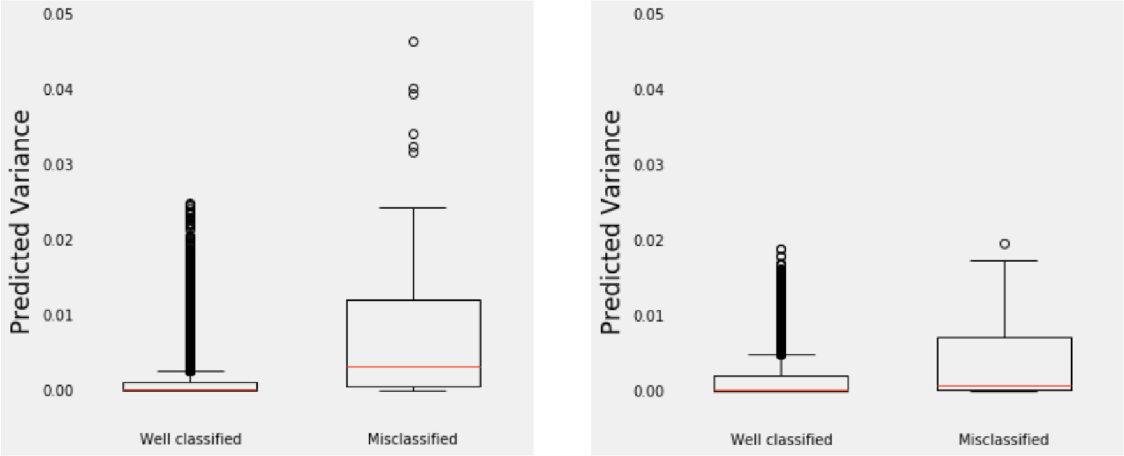


Figure 6-10.: Box plot of the predicted variance on EyePACS test samples (left) and Messidor-2 (right), grouped by their classification status on the *referable / non-referable* diagnosis task.

6.5. Conclusions

In this work, we presented a novel method for grade-based medical image analysis. Intended as a diagnostic support tool for the practitioner, the method combines the representational power of deep learning with the Quantum Measurement Regression method [54], which uses density matrices and random features to build a density estimator.

We tested our approach in two different tasks: the diagnosis of prostate cancer and diabetic retinopathy. In both cases, the diagnosis was based on a gradation by progressive levels. The training of the models was performed using the five available grades, and we reported the results for both ordinal regression and binary classification tasks. The latter were obtained by direct re-categorization over the regressor's predictions.

Compared with similar regression and classification methods, the results show that the while DQOR does not guarantee a better multi-class accuracy, it consistently allows obtaining better results in terms of MAE, which can be advantageous in medical applications, given the sensitivity to the magnitude of classification errors that a purely categorical metric does not have. Furthermore, by directly binarizing the results, we showed that training the models with the information of the grades improves the performance of the binary classification.

It should be highlighted that in these two medical applications the labels are presented on a progressive scale. Namely, the method takes advantage of the ordinal relationship of the labels, which is absent in the purely categorical tasks. In conventional multiclass classification problems the method is not expected to show improvements, on the contrary imposing an artificial order on the labels may negatively impact the performance.

Furthermore, unlike methods based solely on neural networks and other probabilistic models, DQOR predicts for each sample a discrete probability distribution over the range of labels.

This enables a robust integration of the results of patch-level images to a prediction on a whole slide image and offers the uncertainty of the prediction. In test cases, we showed that this uncertainty is significantly lower on well-classified samples in comparison to misdiagnosed samples and that the statistical behavior of this measurement is consistent across different datasets. This implies that the method is able to provide the level of confidence of its inference which can support the identification of misclassified samples. While this may require further research and statistical analysis, it is a highly valued feature for medical applications, where the goal is to prevent false positives and especially false negatives in a diagnostic process.

Overall, we demonstrated that unlike deep learning architectures and standard classification models, the combination of deep CNNs and Quantum Measurement Regression allows us to take advantage of the ordinal information of the stages of a disease in a probabilistic manner. This provides a better theoretical framework to deal with patch-based analysis, improves the performance in the binary prognosis-oriented diagnosis, and provides tools to quantify the uncertainty of the model for safety-critical applications.

7. Multimodal Deep Kernel Quantum Measurement for Prostate Cancer Grading and Glaucoma Diagnosis

In this chapter, two medical multimodal tasks are approached by means of a kernel-based fusion using a quantum measurement scheme. As we mentioned early in this thesis, medical data is inherently multimodal, since the diagnostic process for many diseases may involve several tests and procedures. Our approach allows us to build robust probabilistic kernel-based models that are able to integrate information from different sources, with the benefit of being extremely flexible to different training and inference conditions.

The work in this chapter is in preparation to be published as:

Toledo-Cortés S., *González F.A. (2023). Multimodal Quantum Kernel Fusion for Glaucoma Diagnosis.*

7.1. Introduction

Multimodal learning encompasses a whole family of machine learning algorithms that fuses different sources of information from the same object finding joint representations for the different modalities. Information fusion mechanisms are varied and depend on the target task and the data itself. These models have shown success in several applications such as image retrieval tasks, auto-labeling, image collection summarization, segmentation, classification, prediction, etc [11].

At the beginning of this thesis we mentioned that there are some key factors when designing a multimodal model. First, as in almost all machine learning applications, we have to deal with the problem of data representation [9]. So far, all the models presented in the previous chapters have used deep CNNs to solve this problem, since we have only been working with images. Next, we have the fusion. Three types of fusion schemes are always described in the literature [9]: early fusion, late fusion, and intermediate or hybrid fusion. This classification clearly depends on the point in the algorithmic flow at which the fusion takes place. For instance, the models presented in this chapter use intermediate fusion. However, in practice, there is an important aspect of multimodal development that is not always mentioned: interactions between the different modalities can occur in many ways, and there are different reasons why the data may not be complete for all modalities in the same sample.

We may have complementary data modalities when no single modality can be used on its own for the final task. The best examples of this can be found in the medical field, in the diagnostic process for many diseases. For instance Glaucoma, according to the Glaucoma Research Foundation [1], requires a diagnosis process that includes multiple tests. These may include a tonometry to measure the inner presure of the eye, a perimetry to measure the field of vision, a pachymetry to measure the thickness of the cornea, etc. Determining whether someone has glaucoma can often be challenging, and ongoing detailed assessment of the optic nerve is critical to both diagnosis and treatment. Ophthalmologists consider a variety of factors before deciding on the most appropriate treatment.

On the other hand, modalities may overlap if one modality contains more information than the other. In practice, there are several reasons why data may not be complete for all modalities in the same sample. There may be security or privacy constraints of many kinds, or, for example, in applications involving sociodemographic data, there may not be completeness of all possible data for each individual [95]. Medical applications are a special case. Sometimes diagnostic images and their corresponding textual reports (written by a specialist) are available. Medical images are more abundant than medical reports because the latter require far more specialised human resources [10]. The images and text can then be used to train a model for a diagnostic support task. In practice, however, the report of a new image is obviously not available. The main goal in these medical situations is to improve the performance of a single-modal model through multimodal training. A really useful model should only work with the images and make a diagnosis from them. That is, the model should be able to be trained multimodally, but it should be able to make inferences with only one modality. And the goal is clearly that multimodal training should lead to improvements in single-modal test performance.

Having such flexibility in a multimodal model is not an elementary task. Dealing with the absence of a modality implies looking for solutions beyond simply removing a slice of the input from the model. The model must be able to be independent of a modality. Various attempts have been made in the past to achieve this goal [96]. Frequently, this involves complex reconstruction processes, or looking for joint representation spaces that can be learned in a multi-modal fashion but which can be accessed through a single modality.

In this chapter, we propose a quantum-inspired deep kernel learning model which takes advantage of the probabilistic mechanisms of quantum mechanics, the fusion capabilities of kernel functions, and combines them with the representational power of deep learning models. This approach allows us to:

- Fuse the representation of modalities into a joint high or infinite dimensional space, in a way that would be impossible for a model based on deep learning alone.

- Obtain interpretable results by means of the inspection of the prototypes learned by the model.
- Tolerate the uncertainty of a modality, on the basis of the identity element, or neutral element, of the operation that defines the fusion.
- Enhance a single-modal model by training it with multi-modal data.
- Incorporate the previous advantages into a end-to-end gradient descent learning schema.

This chapter is organised as follows: in the next section we will briefly review the related work to put our contribution into perspective. Then, we will present our proposed method and the experimental set up. After that, we present and discuss the experimental results and finally we address the conclusions.

7.2. Method: Deep Multimodal Kernel Quantum Measurement Unit (MM-KQMU)

The basic architecture of the proposed Deep Multimodal Kernel Measurement Unit (MM-KQMU) is described in Figure **7-1**. As in previous chapters, we use a deep CNN as a feature extractor for the visual modality. For textual modalities, we will use TF-IDF for feature extraction, and in the case of structured clinical data, only a short pre-processing is needed. Each modality representation can be input into independent multiple layer encoders before being fed into a KQMU. Recall from Chapter 5 that a KQMU is a processing unit that receives a density matrix representing a pure or mixed state and operates in a feature space defined by modality-specific kernels. This means that we need to define kernel functions for each modality. Two kernel functions were explored for all upcoming experiments: the cosine kernel, defined by

$$k_{\cos}(x, y) = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}, \tag{7-1}$$

and Gaussian kernel or Radial Basis Function (RBF):

$$k_{\mathrm{rbf}}(x, y) = \exp\left(-\frac{1}{2\sigma^2}\|x - y\|^2\right). \tag{7-2}$$

These kernels define a quantum feature mapping, since for both cases we have $k(x, x) = ||x||^2 = 1$. Note that unlike the cosine kernel, the use of a Gaussian kernel implies the optimization of an additional parameter σ. This value determines how far the influence of a single training point extends. If necessary, σ is learned during training along with the rest of the parameters of the model described in Section 5.6.

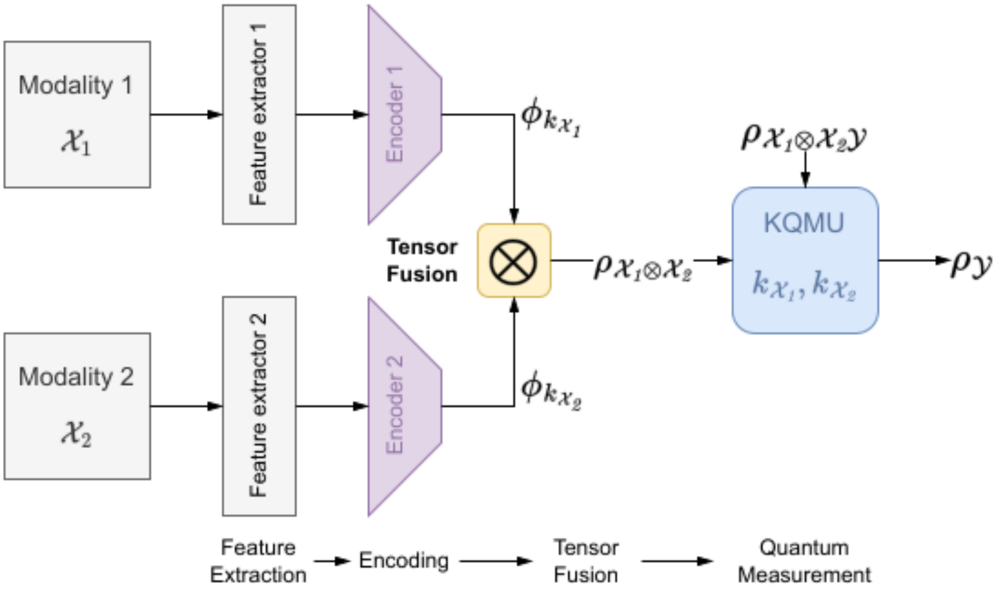


Figure 7-1.: Overview of the proposed MM-KQMU method for multimodal medical data analysis. Feature extraction and encoding can be performed independently for each modality. MM-KQMU performs a tensor kernel fusion and outputs a density matrix ρ_y suitable for classification or regression tasks.

7.2.1. Kernel Fusion

For multimodal fusion, MM-KQMU uses the tensor product between modalities. Although the Cartesian product is also an option, several explorations with different datasets consistently showed better performance when using the tensor product. This is understandable if we consider the following aspects:

- Interactions: A tensor product creates a new representation in which every single element of a vector of one modality interacts with every single element of the corresponding vector of the other modality. In the case of multimodal data, this means that a tensor product can capture interactions between different modalities (such as the relationship between a specific clinical feature and a region in an image) that a Cartesian product would miss.

- Dimensionality: Tensor products can result in higher dimensional tensors that can represent more complex relationships. For example, a tensor product of vectors results in a matrix that can represent a transformation from one vector space to another. This could be used, for example, to represent a transformation from image data to clinical data.

As we explained in Chapter 5, given two kernels k_V and k_T for visual and textual (or clinical) data, the tensor product kernel $k_{V\otimes T}$ is simply computed by the multiplication of the respective kernel functions:

$$k_{V\otimes T} = k_V \times k_T. \tag{7-3}$$

7.2.2. Missing Modality Flexibility

The implementation of KQMU in relation to PCa grading requires a missing modality flexibility. This is because for each sample we have a WSI and its corresponding report. We can use both sources of information to train the model, but it makes no sense to test with multimodal data: we have to test only with images. This is because this is intended to be a diagnostic support tool, so in practice the end user will not have a report and an image, but only the WSI to be analysed.

This means that the model should be able to make inferences in the absence of a modality. As an explicit Cartesian or tensor product of the feature spaces is usually managed as a concatenation or tensor fusion layer within the model [95] [172], this leads to limitations in terms of the length of the representations and the flexibility of the model when there is no information for a modality. However, in our model, when a modality is missing, Eq. 7-3 provide a natural solution to the problem: to multiply by 1 in the part of the kernel of the missing modality. In our case, for example when evaluating only with visual information, the kernel for inference would be

$$k_{V\otimes T} = k_V \times 1. \tag{7-4}$$

7.2.3. WSI Mixed State

Unlike QMR (see Chapter 6.2), KQMU allows the input to be a mixed state. This means, given a WSI and its corresponding set of patches $\{x_i\}_{i=1}^n$ in a feature space X, and given a suitable kernel function k_X, we can summarize the information of the patches directly from the input defining the next mixed state:

$$\rho_{\text{WSI}} = \sum_{i=1}^{n} \frac{1}{n} |\phi(x_i)\rangle\langle\phi(x_i)|, \tag{7-5}$$

where ϕ is the implicit feature mapping defined by k_X. Therefore, we need no longer a post-inference summarization process as majority vote or probability vote; KQMU inference produces predictions at the WSI level.

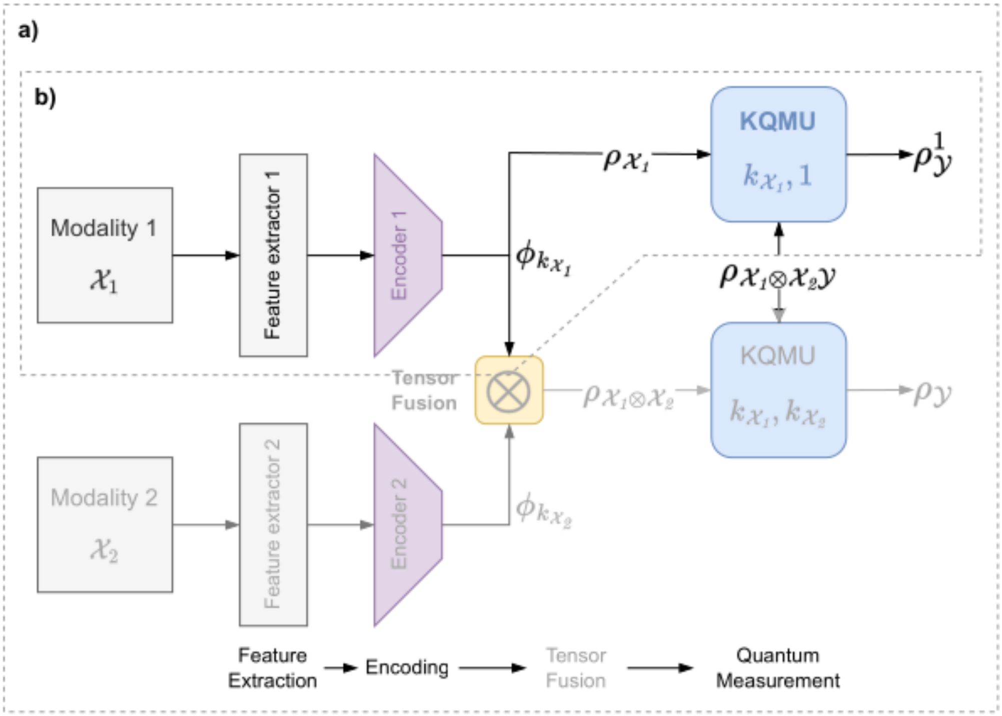


Figure 7-2.: MM-KQMU Missing Modality Flexibility. If needed, MM-KQMU can handle an absent modality for inference. Then, the model shown in a) can be trained with all available modalities, but only the model shown in b) is used for inference.

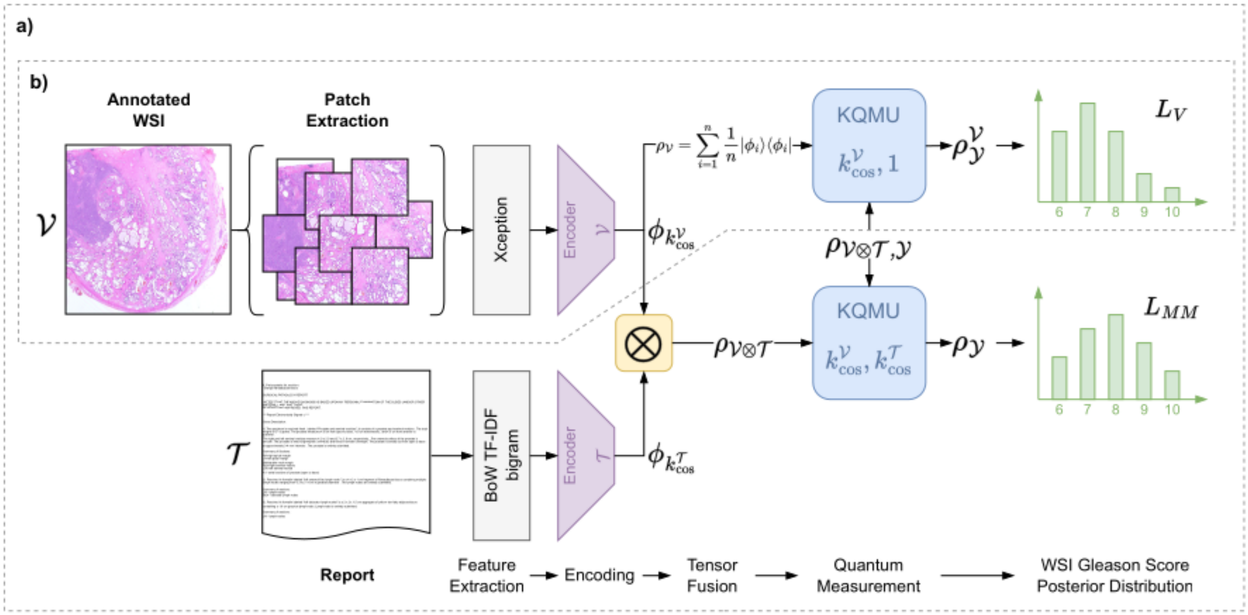


Figure 7-3.: Overview of the proposed MM-KQMU for PCa grading. The model shown in a) is trained using the two available modalities. The loss L_{MM} measures the multimodal performance. For inference we use the model shown in b), which only receives images. The loss L_V measures the visual performance. During training, a combination of L_V and L_{MM} is optimized.

7.3. Experimental Set Up

7.3.1. Prostate Cancer

The setup of the MM-KQMU applied to the PCa data analysis is described in Figure **7-3**.

Dataset

We used the TCGA PRAD dataset, keeping the same set up described in Chapter 6.3. Recall that this set contains prostate tissue samples with Gleason scores from 6 to 10. The details of the dataset partitions are shown in Table **7-1**.

The processing of the images is the same as that described in Chapter 6, i.e. we will use the same features obtained with a fine-tuned Xception. In this case, however, texts are used during model training. These texts include the report made by the specialist and therefore detailed information about the Gleason score at the WSI level.

To be consistent with our baseline, we use the same pre-processing as described in [81]. This pre-processing consist in the elimination of stop words when building the text vocabulary. The term frequency inverse document frequency (TF-IDF) method was used because

Table 7-1.: Details of the subset and final partition of the TCGA dataset used for training and testing [81].

Risk	Gleason Score	Train	Validation	Test
Low	6	11	4	4
Low	7	53	17	17
High	8	23	8	8
High	9	50	17	16
High	10	4	2	1

of its advantageous features. In particular, TF-IDF provides numerical stability, which is advantageous for our information fusion approach. Furthermore, it enhances the relevance of unique terms while downplaying the effect of more common terms. This well-considered balance contributes to a stronger and more meaningful representation of the textual data [81].

Multimodal KQMU

As we are interested in evaluating the model using images only, we use a multitask approach to train the model. This approach can be visualised in the Figure **7-3**. Essentially, the model is trained with a loss function that includes inference from multimodal data and, at the same time, inference with visual data only. This makes it possible to control performance on the final specific task to be assessed: inference from images only. As text contains more information, training the model with only a multimodal loss function runs the risk that the model learns to rely only on textual information during training and loses the ability to generalise with visual information.

Then, given a pair of WSI-Text and their respective representations ρ_{WSI}, ρ_{Text}, these pass through independent encoders. In both cases, the encoder has two dense layers of 1024 and 512 units. The fused representation (with the use of the kernel functions) is passed to the KQMU layer. The output of the KQMU is a density matrix $\rho_{\mathcal{Y}}$ such that

$$p(Y = r|\rho_{\text{WSI}}, \rho_{\text{Text}}, \rho_{\mathcal{XY}}) = \rho_{\mathcal{Y}rr}. \tag{7-6}$$

Given the flexibility of the model to make inference using only one modality (see Eq. 7-4), we can simultaneously calculate $\rho'_{\mathcal{Y}}$ such that

$$p(Y = r|\rho_{\text{WSI}}, \rho_{\mathcal{XY}}) = \rho'_{\mathcal{Y}rr}. \tag{7-7}$$

Therefore, it is straightforward to simultaneously compute a visual regression loss function L_V and a multimodal regression loss function L_{MM}. The total regression loss function will be a weighted sum:

$$L = \delta L_V + \epsilon L_{MM}. \tag{7-8}$$

The parameters δ and ϵ allow to control the importance given by the model to unimodal and multimodal performance and were set to 1/2. For both modalities we explored the kernels, the number of components of the density matrix decomposition and the α parameter (see Eq.5-18) of the regression loss functions. In both modalities, the best performing kernel was the cosine kernel with the number of components set to 128 and α set to 0.02. In addition, the components of the density matrix are randomly initialised with data from the training set. This was found to have a significant effect on the performance of the model. Furthermore, the initialisation of each component of the density matrix is crucial for the correct performance of the model. For this purpose, a random representative subset is selected to initialise the components of both features and labels, in both modalities. Training was performed with an Adam optimiser and a learning rate of 0.001.

Baseline

We compared the performance of our model with five related approaches: the extended version of the feature extractor trained as a categorical classifier (DLC-PCa) [81], the Density Matrix Kernel Density Classification (DMKDC) [54] (also a categorical classifier) and the Gaussian process (GP) [120], a regressor model. We also report results for deep Gaussian processes (DGP) [37] and, finally, the results of the Deep Quantum Ordinal Regressor (DQOR) [146] presented in Chapter 6.

For the *low* risk vs *high* risk binary classification task, we compare our models with the methods reported in [67] and [121]. Those are CNN models trained with the binary labels directly. We also report the results of model presented in [81], trained in a similar multimodal schema as ours, and tested only with images.

7.3.2. Glaucoma

The setup of the MM-KQMU applied to the Glaucoma diagnosis is described in Figure **7-4**.

Dataset

The Public Anonymous oPhtalmological Image and Label Analysis (PAPILA) dataset [78] is a robust and structured compilation of ophthalmological clinical data that was collected at the Department of Ophthalmology of the *Hospital General Universitario Reina Sofía*, HGURS, in Murcia, Spain, between 2018 and 2020. This dataset is comprised of records from 244 patients.

Each record in the database contains the age and sex of the patient and fundus images of the left and right eyes. These images were taken using a Topcon TRC-NW400 non-mydriatic retinal camera with a resolution of 2576 × 1934 pixels. The dataset also contains expert knowledge provided by the ophthalmologists. This includes the classification of the patient

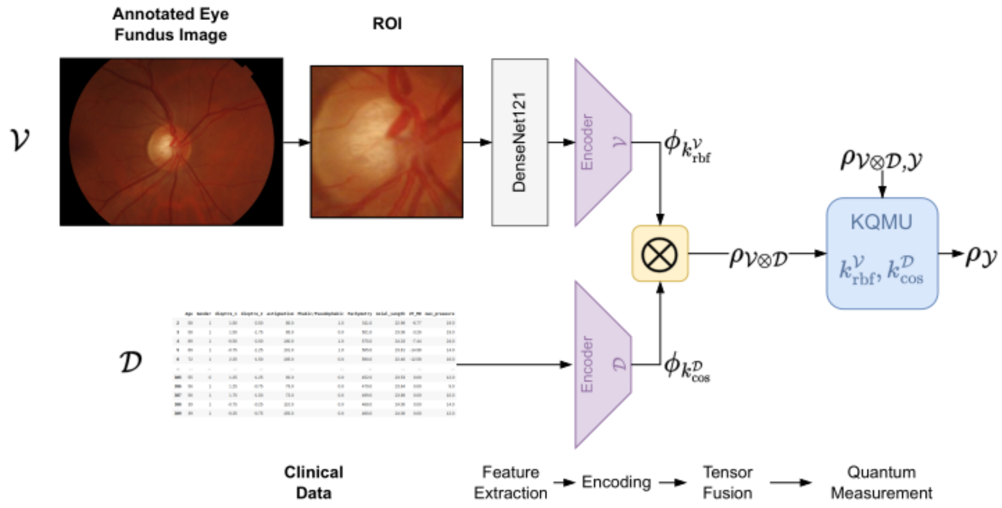


Figure 7-4.: Overview of the proposed MM-KQMU for glaucoma diagnosis. The eye fundus images are cropped according to the Region of Interest (ROI) containing the disc and optic cup. A fine-tuned DenseNet121 is then used as a feature extractor. Scaled clinical data is fed directly into the corresponding encoder. A Gaussian kernel is used for the visual part and a cosine similarity kernel for the clinical data.

into three categories: glaucomatous, non-glaucomatous and suspect. Diagnostic labels were assigned based on a comprehensive evaluation of the subject's clinical data and fundus imaging. In addition, segmentation of the optic disc (OD) and optic cup (OC) in the images of both eyes was performed by two experienced ophthalmologists. Clinical data collected from both eyes of all patients included:

- Refractive error, a vision problem that occurs when the shape of the eye does not bend the light correctly, resulting in a blurred image.
- Intraocular pressure (IOP), with normal values for healthy patients ranging from 10 mmHg to 21 mmHg.
- Central corneal thickness, which is relevant in patients with glaucoma, since it perturbs IOP measurements.
- Axial length, the distance between the anterior vertex and the posterior pole of the eye.
- Mean defect (MD) of both eyes, an overall value of the total amount of visual field loss compared to the normal sensitivity expected for the population group with the age of the patient.

- The condition of the crystalline lens, informing if the eye has the crystalline lens (phakic) or if it has been surgically removed (pseudophakic).

A data sample can be labelled as 0, 1 or 2. 0 if it is a *healthy* sample, 1 for *glaucoma* and 2 for *suspect*. This *suspect* categorization is based on the patient having a predisposing condition, such as hypertension, but no morphological deterioration of the optic nerve. In [78], the authors show that in the data set, the *suspect* category doesn't serve as a transition class from *healthy* to *glaucoma*, but seems to represent an indeterminate class between the two. This complexity further complicates the task of delineating a clear classification boundary. As a result, they choose to exclude the *suspect* samples from the dataset for training and evaluation of different classification methods. This exclusion does not imply any inherent problem with these samples, but is simply a methodological choice made for the sake of the experimental setup.

We defined a new partition of the dataset. As explained above, the dataset contains fundus images and clinical data (per eye) from 244 patients. That is, 488 image-data samples. The original work in which the dataset is presented [78] provides an set up with five train-test partitions in an 80-20 ratio. However, these partitions present a problem: there are always some samples from the same patient in both sets: for example, the data from the right eye in the train and the data from the left eye in the test. We have therefore defined a new experimental set up that corrects for this detail. The details of the new partition are described in the Table **7-2**.

Table 7-2.: Training and test distribution proposed for PAPILA dataset. Each sample in this table corresponds to a single eye. In this configuration, data from the same patient belongs to the same partition. Note that there are fewer samples for multimodal data. This is due to missing clinical data.

Data	Class	Train	Test
Eye Fundus Images	0	265	68
	1	71	16
	TOTAL	336	84
Images and Clinical Data	0	261	66
	1	62	15
	TOTAL	323	81

Multimodal KQMU

The architecture used for glaucoma diagnosis is described in the Figure **7-4**. Unlike the case of PCa, this task does not present an inference problem in the absence of one modality. Therefore, we trained the KQMU model for classification by training only the multimodal loss function. However, to prevent the model from suddenly learning to rely in only one modality, the multimodal training starts with prior unimodal trained encoders for each modality. This is explained below.

For the images with a resolution of 2576 × 1934, and following the setup presented in [78], we fine-tuned a DenseNet121. This deep CNN reports the best results when compared with VGG, MobileNet, InceptionV3, ResNet50 and Xception. The images are cropped according to the ROI defined by the manual annotation and reduced to a size of 224x224 (see Figure **7-4**). We use the same data enhancement settings as in chapters 3, 4 and 6, except for the horizontal and vertical translations, as we do not want to damage the radial symmetry of the ROI. On top of DenseNet121 we placed a dropout layer (with a factor of 0.2), followed by a layer of 1024 neurons and an output layer of 2 neurons. We did a warm-up with a learning rate of 10^{-4} for an Adam optimiser and used a batch size of 32. Fine tuning was done with a learning rate of 10^{-5} during 200 epochs, also using an early stop which gave the best performance on the validation set at epoch 125.

The KQMU visual input is the representation obtained with the DenseNet121, so that each image is represented by a vector of dimension 1024. This input is passed through an encoder consisting of a two-layer dense model. The kernel was explored between a cosine kernel and a Gaussian kernel, with the Gaussian kernel giving the best results. The number of components of the density matrix decomposition is explored and set to 64. First, the encoder and the KQMU are trained together. Then, starting from the trained encoder, the γ parameter of the kernel is initialised according to the range of the encoder outputs, and the components of the density matrix are also initialised. The model is then fine-tuned. This procedure proves to be more robust than simply training the encoder and KQMU simultaneously from scratch. As this is a binary classification task, we use a binary cross entropy as the loss function. The learning rate is explored and set to 0.001, and Adam is used as the optimiser.

Clinical data must be treated with caution. The measure of visual field loss is only taken from patients with a positive diagnosis, so they are not included in the training of the models. In addition, there are 11 patients with missing data for some eyes. After removing these missing data, we have 323 samples for training and 81 for testing. These data are simply standardised and used directly in the classification models. The best performing kernel was the cosine kernel. The number of components of the density matrix decomposition is explored and set to 64. The model uses a two-layer dense encoder, and the training procedure is the same. The learning rate is explored and set to 0.01 and Adam is used as the optimiser.

Finally, for multimodal training, we start with the previously trained encoders of each modality. As we are dealing with a more complex system, the number of components of the density matrix was raised to 256, and are initialized with random encoded samples. σ for visual kernel is initialized from the visual-only training results. The learning rate was set to 0.001 and Adam is used as optimizer.

Baseline

For each modality separately, and of course for the multimodal application, we compared the performance of our models with several related approaches. Regarding the clinical data, we compared our model with a KNN, which is, within the classical machine learning methods, the one that reported the best results. We also used an SVM, which according to hyperparameter exploration works better with a linear kernel. Furthermore we used an SVM on the features encoded by KQMU with the kernel used in KQMU. Analogously for images, our baseline is the DenseNet121 model used as feature extractor, and SVM models on the features given by DenseNet121 with a linear kernel, and on the features encoded by KQMU with the kernel learned by the same KQMU.

In the mutimodal application, we compare our model with a concatenation model, which has two encoders with the same configuration as the KQMU, and simply concatenates the outputs to pass a single vector to a final MLP. We also try a Gated Multimodal Unit (GMU) model [7], which is indeed similar to the concatenation model but, instead of concatenating, it performs a weighted sum of the representations, where the weighting is learned in time with all the model parameters.

7.4. Experimental Results and Discussion

7.4.1. Prostate Cancer

Experimental results of KQMU for PCa grading at WSI level are shown in Table **7-3**. We report image-only grading results for KQMU trained with images only (V-KQMU) and for KQMU trained with multimodal data (MM-KQMU). As in Chapter 6, we report classification and regression metrics for the grading task. We also report the *low/high risk* binary classification results in Table **7-4**.

As expected for the grading task, Table **7-3** shows that MM-KQMU significantly improves the regression metric compared to the state of the art, while sacrificing a little on the classification metrics. This is exactly the trend we saw with QMOR in Chapter 6, and it makes sense given that we are optimising an MSE-based loss function. Furthermore, it is surprising to see that the MAE already improves quite a bit with V-KQMU. This means that much of the performance improvement in MAE is due to the way the model input is constructed, or rather the way the patch summation process is done. KQMU performs an intrinsic summation process from the construction of the input as a mixed state of patches. This way of combining the information into a single object and processing all the patches of a WSI synchronously has a very positive effect on the final results.

Looking at the results in Table **7-4**, it is remarkable to see that the improvement in MAE for the PCa grading task translates into an improvement in accuracy for the binary PCa risk classification task. Again, the performance of V-KQMU shows that part of this improvement

Table 7-3.: WSI-level PCa grading results. For each model, two summarization procedures may be applied, majority vote (MV) and probability vote (PV). KQMU models do not require summarization procedure. V-KQMU stands for the model trained only with images and MM-KQMU stands for the model trained with multimodal data.

Method	Accuracy	Macro F1	MAE
DLC-PCa PV [81]	**0.608**	0.354	0.717
GP MV [120]	0.391	0.233	0.739
DGP MV [37]	0.174	0.059	0.935
DMKDC PV [54]	0.608	0.354	0.717
DQOR [146]	0.587	**0.356**	0.652
V-KQMU	0.565	0.340	0.587
MM-KQMU	0.522	0.326	**0.565**

Table 7-4.: Results at WSI-level of *low* risk vs *high* risk PCa classification. V-KQMU stands for the model trained only with images and MM-KQMU stands for the model trained with multimodal data.

Method	Accuracy
Google LeNet [67]	0.7352
Modified AlexNet[121]	0.769
M-LSA [81]	0.770
DQOR [146]	0.782
V-KQMU	0.804
MM-KQMU	**0.848**

is due to the ρ_{WSI} construction. However, here the effect of the MAE improvement in gradation is even more noticeable. That is, the improvement in MAE in grading means that a greater number of predictions are closer to the true value on the Gleason score scale, and this is sufficient to make the model robust to the binary task. Better accuracy in grading is not necessarily better if it is not accompanied by better MAE; it is of little use to have more samples correctly classified if the misclassified ones are far from the correct value.

7.4.2. Glaucoma

Results for glaucoma diagnosis on PAPILA dataset are reported in Tables **7-5** to **7-7**. For this task, we report binary classification metrics for single modality and multimodal models.

Table 7-5.: Results on the PAPILA dataset using only clinical data for training and test. In the case of SVM and KQMU, the kernel we used is given in parentheses. * indicates that the kernel is the same used by the KQMU model.

Model	Accuracy	Precision	Recall	F1 Macro	AUC
KNN	**0.80**	**0.45**	0.33	0.63	0.58
SVM (linear)	0.75	0.14	0.07	0.47	0.64
SVM (cos*)	0.75	0.37	**0.47**	0.63	0.58
KQMU (cos)	0.77	0.39	**0.47**	**0.64**	**0.77**

Table 7-6.: Results on the PAPILA dataset using only images for training and test. In the case of SVM and KQMU, the kernel we used is given in parentheses. * indicates that the kernel is the same used by the KQMU model.

Model	Accuracy	Precision	Recall	F1 Macro	AUC
DenseNet121	0.85	0.62	0.53	0.74	0.83
SVM (linear)	0.84	0.56	**0.67**	0.75	0.77
SVM (rbf*)	0.88	0.67	**0.67**	**0.80**	0.86
KQMU (rbf)	**0.89**	**0.75**	0.60	**0.80**	**0.91**

There are several things to highlight from the results. First, it is clear from Table **7-7** that our multimodal approach provides the best classification results. In each modality separately (Tables **7-5** and **7-6**), it is noticeable how the SVM improves significantly when using the encoded data and the kernels learned by the KQMU. Thus, while SVM and KQMU are closely related, the positive effect of probabilistic training and gradient descent on KQMU is evident. However, the best results are achieved by our respective models. In other words, KQMU dominates in all cases.

In multimodal applications, it is noteworthy that GMU does not perform better than concatenation. The poor performance of the multimodal SVM is also very interesting. For this model, we take the same representation learned by MM-KQMU and as kernel we use the product of the respective kernels of each modality. Surprisingly, in the unimodal case, this approach was a boost for the SVM, which was always the second best performer after KQMU. Thus, the same MM-KQMU fusion strategy used with the SVM simply achieves an average performance between the performance of each modality separately. In contrast, the MM-KQMU is able to improve the individual performances. That is, the success of MM-KQMU lies not only in fusing the information with a product of the kernel functions, but in adapting the whole model to the new combined metric.

Table 7-7.: Results on the PAPILA dataset using multimodal data for training and test. MM-KQMU stands for Multi-Modal KQMU. In the case of SVM and MM-KQMU, the kernel we used is given in parentheses. * indicates that the kernel is the same used by the KQMU model.

Model	Accuracy	Precision	Recall	F1 Macro	AUC
Concatenation	0.82	0.52	**0.80**	0.76	0.90
GMU	0.78	0.44	0.73	0.70	0.90
SVM (cos*×rbf*)	0.84	0.55	0.73	0.76	0.83
MM-KQMU (cos×rbf)	**0.89**	**0.69**	0.73	**0.82**	**0.92**

Interpretability

Apart from the numerical results, KQMU offers another advantage regarding potential implementation of these models. Recall from Section 5.6 that the inference process for a new sample depends entirely on the similarity given by the kernel between the new sample and the propotypes that the model learns and that make up the factorization of the training matrix $\rho_{\mathcal{X}\mathcal{Y}}$. In Eq. 5-23 it can be seen that the importance of each prototype is marked by the coefficients β_i. Thus, once the model is trained, we can check how many prototypes are relevant to the model, and look directly at samples that are close (according to the kernel) to those prototypes in the training set to get an idea of how the model compares when making a decision.

To illustrate the above idea, Figure **7-5** shows the decision region of MM-KQMU and the train samples using kernel PCA dimensionality reduction. This figure also shows the position of the learned prototypes: only 2 out of 256 prototypes remained such that $\beta_i > 0$. Not surprisingly, there is one prototype per class. The kernel consisting of the product of the cosine similarity and the Gaussian kernel corresponds to the similarity measure in this space, and we can use it to see which samples in the training set are closest to these prototypes. Figures **7-6** and **7-7** show the results of this exercise.

Thus, given a new sample consisting of a fundus image and clinical data, the MM-KQMU model is able not only to make a prediction, but also to justify its decision based on a comparison criterion with a specimen from the training data set. In practical terms, this takes the model to the next level of interpretability, providing medical staff with even greater decision-making tools.

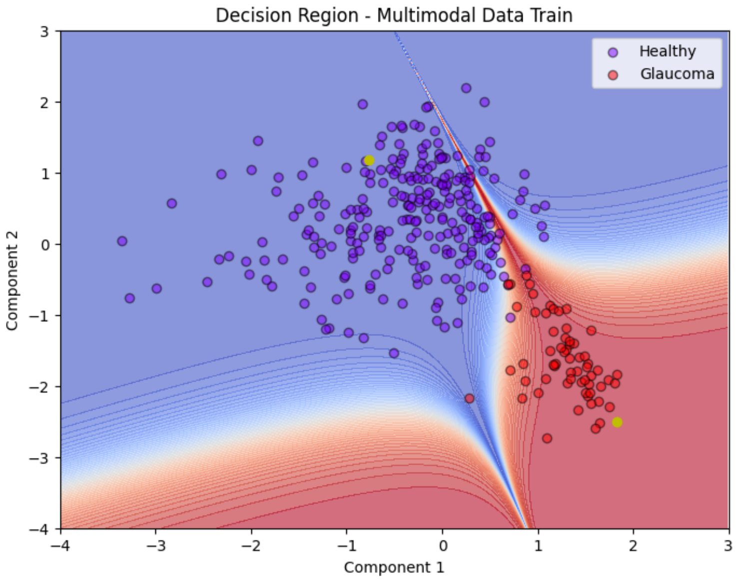
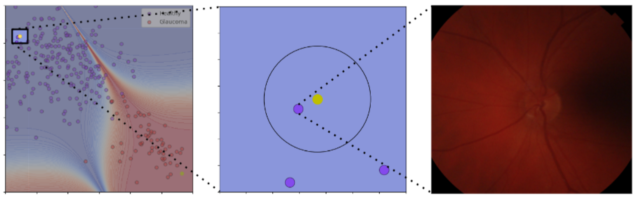


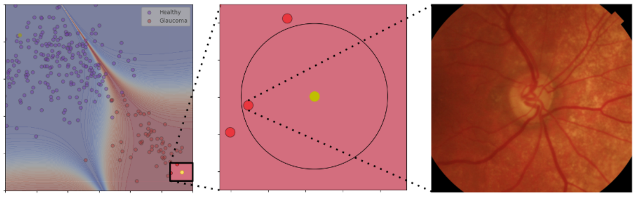
Figure 7-5.: Decision region of the MM-KQMU together with the training samples. The yellow dots correspond to the relevant prototypes trained by the model. Dimensionality reduction was performed using PCA.


Most similar sample to healthy prototype

Age	Gender	Astig.	Diop. 1	Diop. 2	Crystalline	Pachymetry (*µm*)	Axial Length (*mm*)	IOP (*mmHg*)
43	Male	10	-3.00	-1.00	Pseudo-phakic	600.0	25.22	16.0

Figure 7-6.: Most similar specimen from PAPILA train partition for the *healthy* class prototype (yellow dot). The similarity is given by the multimodal kernel learned by the MM-KMQU. In this case, the kernel value between the prototype and the shown sample is 0.9957.

Most similar sample to glaucoma prototype



Age	Gender	Astig.	Diop. 1	Diop. 2	Crystalline	Pachymetry (*μm*)	Axial Length (*mm*)	IOP (*mmHg*)
78	Female	90	-1.5	-2.25	Pseudo-phakic	543.0	22.16	16.0

Figure 7-7.: Most similar specimen from PAPILA train partition for the *glaucomatous* class prototype (yellow dot). The similarity is given by the multimodal kernel learned by the MM-KMQU. In this case, the kernel value between the prototype and the shown sample is 0.9908.

7.5. Conclusions

In this chapter, we introduced the MM-KQMU, a quantum kernel-inspired multimodal probabilistic approach that combines the power and fusion flexibility of kernel functions with the representation learning capabilities of deep learning models, all within a probabilistic framework. Our proposal has several advantages. First, thanks to kernel functions, our model allows us to learn a joint representation space for different data modalities via a tensor product, without having to deal with the high dimensionality constraints that result from these combinations. This allows us to take full advantage of the increased complexity of tensor product spaces, which is explained by their ability to capture component-by-component interactions within the representations of each modality. Second, kernel functions allow us to naturally handle the uncertainty associated with missing information when a modality is missing from the dataset. In this way, we achieve flexibility: the model execution pipeline is not broken, nor does it have to resort to a reconstruction approach when a modality is missing. Finally, the probabilistic framework of our model also provides better interpretability tools than the neural network based models that are usually the state of the art for this type of application.

We tested our models on two tasks: prostate cancer grading and glaucoma diagnosis. For prostate cancer, the multimodal data consists of whole slide images and medical reports of those images, with detailed diagnostic information about Gleason patterns and other clinical data. For glaucoma, we have both fundus images and structured clinical data. Although the two tasks are multimodal, they have significant differences in terms of the complementarity of the information and the way it is used in practice.

For prostate cancer diagnosis, the model should be tested with visual information only, as the reports will not be available in a real implementation. Our approach allows us to train with multimodal data and test with a single modality. Thanks to the multitask training scheme, we can simultaneously control multimodal and image-only performance. In this way, the model learns to balance the representation of both modalities in favour of image-only performance. The results show that MM-KQMU is able to improve single modal performance through multimodal training. For glaucoma, the two data modalities are complementary, as in medical practice neither images nor clinical data alone are conclusive, and therefore it makes sense to train and evaluate the model multimodally. The results show that MM-KQMU improves the performance of unimodal models and multimodal baseline models. Furthermore, we show that this improvement is not only based on the mechanism of fusing the representations of each modality, but also on the fact that the model can learn to adapt to this joint representation of the data.

Finally, MM-KQMU provides some clear interpretability tools. Since the model is based on density estimation mechanisms, it is possible to find prototypes for each class that explain the model's inference decisions. In this way, a prediction can be justified directly on the basis of comparison criteria with samples from the same dataset. This leads to the development of many other applications, such as retrieval models, and makes the model a real mechanism to help medical staff make sensible decisions.

Overall, what we have achieved in this chapter is to show that using kernel functions in the middle of a probabilistic framework using density matrices allows us to create flexible, versatile and interpretable multimodal models that are able to combine the adaptive power of deep learning and the representational power of kernel functions.

8. Conclusions and Future Work

In this thesis, we presented the development of machine learning models that support the medical diagnosis of various ophthalmological diseases and prostate cancer. The common denominator of the models here presented is their ability to take advantage of different data structures and different sources of information simultaneously. Thus, we proposed different models based on regression, ordinal regression and multimodal classification.

The challenges behind the design of such models are mainly related to the data representation learning, the possible fusion of information and the subsequent processing of the same, coupled with the ability to leverage all the structural information available in the data labels. In terms of representation learning, images or visual modalities are the most challenging for medical data. The other possible information modalities, such as medical reports or clinical data, are relatively easy to represent because they follow well-defined structures with a specific and not very diverse vocabulary. For text and structured data, term frequency-based representation and good cleaning and standardization processing were sufficient in our application cases. Images, on the other hand, were represented with features learned by deep convolutional neural network models. Although we are on the verge of a possible disruption due to transformer-based approaches, deep CNNs are still a computationally efficient tool for the initial analysis of medical images. The fact is, as we have seen throughout this thesis, that the representations learned by these models can have much better processing than that offered by a traditional multilayer perceptron.

We call the first part of this thesis *Deep Learning-based Models*, precisely because the main novelty of what we propose lies in the treatment of these deep learning-based representations. We show that by using information beyond simple categorical labels, in particular by using ordinal information, we can improve the performance of models for diagnostic tasks: Chapter 3 shows how cell count estimation helps in the subsequent diagnosis of genetic diseases, and in Chapter 4 we show that a regression on the grades of diabetic retinopathy helps in the binary categorical diagnosis of the severity of the same disease. This is the first major conclusion of this work: *The use of regression models that exploit the ordinal information of the target variable helps to improve the subsequent categorical performance associated with the diagnostic task in medical applications.*

The last sections of the first part of the work are a teaser of what is to come. In Chapter 4, we use not only a regression model, but a probabilistic regression model: Gaussian processes. There, we begin to see that using a probabilistic approach allows us to enrich the interpretation of the model results. Now we not only have predictions, but we can measure the uncertainty of the model about those predictions, i.e. we can measure how confident the model is about what it is predicting, and this is particularly important for these kinds of sensitive medical applications. When we are told that we are sick, we naturally wonder how confident the agent who made the decision is in that diagnosis. A probabilistic model can tell us. The next natural question should be: on what basis did it make that decision? But that question would be answered later.

The second part of the thesis is called *Quantum Measurement-based Models* because most of the novelty now lies in the information processing mechanism itself. Quantum computing has had an impact on many areas of knowledge, and machine learning has been extensively involved. Quantum mechanics provides a rich theoretical framework and mathematical objects to encode information and uncertainty, and also to make measurements on these objects. We have developed probabilistic machine learning models in which the entire flow of information can be handled by means of density matrices. This makes it possible to summarize, fuse and interpret all the available information for one task in many different ways.

Chapter 6 is a natural extension of the ideas presented in Chapter 4, using the quantum-inspired models presented in Chapter 5. We show again that ordinal information is of great help to representation learning, and that the uncertainty of the predictions correlates well with the effectiveness of the model. This brings us to the second major conclusion of the thesis: *the use of probabilistic methods in medical applications allows us to generate models that are interpretable from the quantification of uncertainty.*

The efficient implementation of models based on density matrices leads to factorized versions that allow a more intuitive interpretation of the measurement mechanism. It also allows the entire flow of information to be expressed in terms of the inner product of the representation spaces, allowing the direct use of kernel functions. This opens up a world of possibilities. Kernel functions bring with them a whole another robust theoretical underpinning, full of tools and possibilities. One of these is the easy fusion of information. This has allowed us to build versatile multimodal methods in Chapter 7, which show that integrating multiple sources of information through the tensor product of representation spaces allows us to achieve better performance. More importantly, we have obtained models that allow us to answer the question: on what basis does the model make a decision? Our models learn representative prototypes of the dataset that serve as a reference point for decision making. So when a model makes a decision, we can look at the comparison criteria that the model used to make that decision. This brings us to the third major conclusion of this thesis: *probabilistic models based on kernels and density matrices allow the generation of robust and flexible fusion mechanisms and highly interpretable models.*

There are many research directions that remain open in the development of this work. To start with, in relation to AOSLO image analysis, there are techniques that need to be explored to better fit the cell localization given by the CoDE model. Also, we need to approach the diagnosis of diabetic retinopathy using kernel functions instead of random Fourier features. This means using KQMU instead of DQOR. And glaucoma diagnosis can also be approached in an ordinal way. Going a little deeper, it should be noted that KQMU offers many application paths due to its symmetrical structure. The processing layer receives, transforms and returns density matrices that allow us to easily invert the flow of information in the graph of the model. This opens the door to the development of more complex but more versatile models to explore interrelationships between information sources, their intrinsic structure, and the labels that may be present. One can then think of exploring translations between modalities, or making more consistent use of the probabilistic distribution of labels, and also exploring new kernel functions.

Bibliography

[1] *Testing for Glaucoma.* `https://glaucoma.org/learn-about-glaucoma/testing-for-glaucoma/`. 2023

[2] Abràmoff, Michael D. ; Folk, James C. ; Han, Dennis P. ; Walker, Jonathan D. ; Williams, David F. ; Russell, Stephen R. ; Massin, Pascale ; Cochener, Beatrice ; Gain, Philippe ; Tang, Li ; Lamard, Mathieu ; Moga, Daniela C. ; Quellec, Gwénolé ; Niemeijer, Meindert: Automated analysis of retinal images for detection of referable diabetic retinopathy. In: *JAMA Ophthalmology* 131 (2013), Nr. 3, S. 351–357. – ISSN 21686165

[3] Adler, Tim J. ; Ardizzone, Lynton ; Vemuri, Anant ; Ayala, Leonardo ; Gröhl, Janek ; Kirchner, Thomas ; Wirkert, Sebastian ; Kruse, Jakob ; Rother, Carsten ; Köthe, Ullrich ; Maier-Hein, Lena: Uncertainty-aware performance assessment of optical imaging modalities with invertible neural networks. In: *International Journal of Computer Assisted Radiology and Surgery* 14 (2019), Nr. 6, S. 997–1007. – ISSN 18616429

[4] American Academy of Ophthalmology: International clinical diabetic retinopathy disease severity scale detailed table. In: *International Council of Ophthalmology* (2002)

[5] Andrearczyk, Vincent ; Müller, Henning: Deep multimodal classification of image types in biomedical journal figures. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11018 LNCS (2018), S. 3–14. – ISBN 9783319989310

[6] Araújo, Teresa ; Aresta, Guilherme ; Mendonça, Luís ; Penas, Susana ; Maia, Carolina ; Carneiro, Ângela ; Mendonça, Ana M. ; Campilho, Aurélio: DR—GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. In: *Medical Image Analysis* 63 (2020). – ISSN 13618423

[7] Arevalo, John ; Solorio, Thamar ; Montes-y-Gómez, Manuel ; González, Fabio A.: Gated multimodal units for information fusion. In: *5th International Conference on Learning Representations, ICLR 2017 - Workshop Track Proceedings* (2017)

[8] Baccianella, Stefano ; Esuli, Andrea ; Sebastiani, Fabrizio: Evaluation measures for ordinal regression. In: *2009 Ninth international conference on intelligent systems design and applications* IEEE, 2009, S. 283–287

[9] Baltrusaitis, Tadas ; Ahuja, Chaitanya ; Morency, Louis P.: Multimodal Machine Learning: A Survey and Taxonomy. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2019), Nr. 2, S. 423–443. – ISSN 19393539

[10] Bauer, Dominik F. ; Russ, Tom ; Waldkirch, Barbara I. ; Tönnes, Christian ; Segars, William P. ; Schad, Lothar R. ; Zöllner, Frank G. ; Golla, Alena K.: Generation of annotated multimodal ground truth datasets for abdominal medical image registration. In: *International Journal of Computer Assisted Radiology and Surgery* 16 (2021), Nr. 8, S. 1277–1285. – ISSN 18616429

[11] Bayoudh, K ; Knani, R ; Hamdaoui, F ; Mtibaa, A: A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. In: *Vis Comput* 38 (2022), Nr. 8, S. 2939–2970. – Epub 2021 Jun 10; PMID: 34131356; PMCID: PMC8192112

[12] Beckham, Christopher ; Pal, Christopher: A simple squared-error reformulation for ordinal classification. (2016), Nr. Nips

[13] Beckham, Christopher ; Pal, Christopher: Unimodal probability distributions for deep ordinal classification. In: *34th International Conference on Machine Learning, ICML 2017* 1 (2017), S. 647–655. ISBN 9781510855144

[14] Benzebouchi, Nacer E. ; Azizi, Nabiha ; Ashour, Amira S. ; Dey, Nilanjan ; Sherratt, R. S.: Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis. In: *Journal of Experimental and Theoretical Artificial Intelligence* 31 (2019), Nr. 6, S. 841–874. – ISSN 13623079

[15] Blunt, Nick S. ; Camps, Joan ; Crawford, Ophelia ; Izsák, Róbert ; Leontica, Sebastian ; Mirani, Arjun ; Moylett, Alexandra E. ; Scivier, Sam A. ; Sünderhauf, Christoph ; Schopf, Patrick ; Taylor, Jacob M. ; Holzmann, Nicole: Perspective on the Current State-of-the-Art of Quantum Computing for Drug Discovery Applications. In: *Journal of Chemical Theory and Computation* 18 (2022), Nr. 12, S. 7001–7023. – PMID: 36355616

[16] Bradshaw, John ; Matthews, Alexander G. de G. ; Ghahramani, Zoubin: Adversarial Examples, Uncertainty, and Transfer Testing Robustness in Gaussian Process Hybrid Deep Networks. In: *https://arxiv.org/pdf/1707.02476.pdf*`arXiv:1707.02476v1` *eprint* (2017), S. 1–33

[17] Bulten, Wouter ; Litjens, Geert ; Pinckaers, Hans ; Ström, Peter ; Eklund, Martin ; Kartasalo, Kimmo ; Demkin, Maggie ; Dane, Sohier. *The PANDA challenge: Prostate cANcer graDe Assessment using the Gleason grading system.* März 2020

[18] Burns, Stephen A. ; Elsner, Ann E. ; Sapoznik, Kaitlyn A. ; Warner, Raymond L. ; Gast, Thomas J.: Adaptive optics imaging of the human retina. In: *Progress in Retinal and Eye Research* 68 (2019), S. 1–30. – ISSN 1350–9462

[19] Burns, Stephen A. ; Elsner, Ann E. ; Sapoznik, Kaitlyn A. ; Warner, Raymond L. ; Gast, Thomas J.: Adaptive optics imaging of the human retina. In: *Progress in Retinal and Eye Research* 68 (2019), Nr. August 2018, S. 1–30. – ISSN 18731635

[20] Camargo, Jorge E. ; Caicedo, Juan C. ; Gonzalez, Fabio A.: A kernel-based framework for image collection exploration. In: *Journal of Visual Languages and Computing* 24 (2013), Nr. 1, S. 53–67. – ISSN 1045926X

[21] Camargo, Jorge E. ; González, Fabio A.: Multimodal latent topic analysis for image collection summarization. In: *Information Sciences* 328 (2016), S. 270–287. – ISSN 00200255

[22] Campochiaro, Peter A. ; Mir, Tahreem A.: The mechanism of cone cell death in Retinitis Pigmentosa. In: *Progress in Retinal and Eye Research* 62 (2018), S. 24–37. – ISSN 1350–9462

[23] In: Castellano, Ginevra ; Kessous, Loic ; Caridakis, George: *Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech.* Berlin, Heidelberg : Springer Berlin Heidelberg, 2008, S. 92–103. – ISBN 978–3–540–85099–1

[24] Cerezo, M ; Verdon, Guillaume ; Huang, Hsin-Yuan ; Cincio, Lukasz ; Coles, Patrick J.: Challenges and opportunities in quantum machine learning. In: *Nature Computational Science* 2 (2022), Nr. 9, S. 567–576

[25] Chakravarty, Arunava ; Sivaswamy, Jayanthi: Glaucoma classification with a fusion of segmentation and image-based features. In: *Proceedings - International Symposium on Biomedical Imaging* 2016-June (2016), Nr. i, S. 689–692. – ISBN 9781479923502

[26] Chen, Richard J. ; Lu, Ming Y. ; Williamson, Drew F. ; Chen, Tiffany Y. ; Lipkova, Jana ; Noor, Zahra ; Shaban, Muhammad ; Shady, Maha ; Williams, Mane ; Joo, Bumjin ; Mahmood, Faisal: Pan-cancer integrative histology-genomic analysis via multimodal deep learning. In: *Cancer Cell* 40 (2022), Nr. 8, S. 865–878.e6. – ISSN 18783686

[27] Chen, Yingming ; Ratnam, Kavitha ; Sundquist, Sanna M. ; Lujan, Brandon ; Ayyagari, Radha ; Gudiseva, V. H. ; Roorda, Austin ; Duncan, Jacque L.: Cone photoreceptor abnormalities correlate with vision loss in patients with Stargardt Disease. In: *Investigative Ophthalmology and Visual Science* 52 (2011), Nr. 6, S. 3281–3292. – ISSN 01460404

[28] Choi, Joon Y. ; Yoo, Tae K. ; Seo, Jeong G. ; Kwak, Jiyong ; Um, Terry T. ; Rim, Tyler H.: Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database. In: *PLOS ONE* 12 (2017), 11, Nr. 11, S. 1–16

[29] Chollet, Francois: *Image segmentation with a U-Net-like architecture*. `https://keras.io/examples/vision/oxford_pets_image_segmentation/`. 2020. – [Online; accessed 30-September-2021]

[30] Chollet, Francois [u. a.]: *Keras*. `https://github.com/fchollet/keras`. 2015. – [Online; accessed 01-Mar-2022]

[31] Contreras, Victor H. ; Lara, Juan S. ; Perdomo, Oscar J. ; González, Fabio A.: Supervised online matrix factorization for histopathological multimodal retrieval. In: Romero, Eduardo (Hrsg.) ; Lepore, Natasha (Hrsg.) ; Brieva, Jorge (Hrsg.): *14th International Symposium on Medical Information Processing and Analysis* Bd. 10975 International Society for Optics and Photonics, SPIE, 2018, S. 109750Y

[32] Cross, Nancy ; van Steen, Cécile ; Zegaoui, Yasmina ; Satherley, Andrew ; Angelillo, Luigi: Current and Future Treatment of Retinitis Pigmentosa. In: *Clinical Ophthalmology* 16 (2022), S. 2909–2921. – ISSN 11775483

[33] Cui, Shaoguo ; Mao, Lei ; Jiang, Jingfeng ; Liu, Chang ; Xiong, Shuyu: Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. In: *Journal of Healthcare Engineering* 2018 (2018). – ISSN 20402309

[34] Cunefare, David: *CNN-Cone-Detection*. `https://github.com/DavidCunefare/CNN-Cone-Detection`. 2017. – [Online; accessed 01-Mar-2022]

[35] Cunefare, David ; Fang, Leyuan ; Cooper, Robert F. ; Dubra, Alfredo ; Carroll, Joseph ; Farsiu, Sina: Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. In: *Scientific Reports* 7 (2017), Nr. 1, S. 1–11. – ISSN 20452322

[36] Cunefare, David ; Huckenpahler, Alison L. ; Patterson, Emily J. ; Dubra, Alfredo ; Carroll, Joseph ; Farsiu, Sina: RAC-CNN: multimodal deep learning based automatic detection and classification of rod and cone photoreceptors in adaptive optics scanning light ophthalmoscope images. In: *Biomedical Optics Express* 10 (2019), Nr. 8, S. 3815. – ISSN 2156–7085

[37] Cutajar, Kurt ; Bonilla, Edwin V. ; Michiardi, Pietro ; Filippone, Maurizio: Random feature expansions for Deep Gaussian Processes. In: *34th International Conference on Machine Learning, ICML 2017* 2 (2017), S. 1467–1482. ISBN 9781510855144

[38] Davidson, Benjamin ; Kalitzeos, Angelos ; Carroll, Joseph ; Dubra, Alfredo ; Ourselin, Sebastien ; Michaelides, Michel ; Bergeles, Christos: Automatic Cone Photoreceptor Localisation in Healthy and Stargardt Afflicted Retinas Using Deep Learning. In: *Scientific Reports* 8 (2018), Nr. 1, S. 1–13. – ISBN 4159801826350

[39] Decencière, Etienne ; Zhang, Xiwei ; Cazuguel, Guy ; Laÿ, Bruno ; Cochener, Béatrice ; Trone, Caroline ; Gain, Philippe ; Ordóñez-Varela, John R. ; Massin, Pascale ; Erginay, Ali ; Charton, Béatrice ; Klein, Jean C.: Feedback on a publicly distributed image database: The Messidor database. In: *Image Analysis and Stereology* 33 (2014), Nr. 3, S. 231–234. – ISSN 18545165

[40] Diabetic Retinopathy Detection of Kaggle: *EyePACS Challenge*. `www.kaggle.com/c/diabetic-retinopathy-detection/data`. – Accessed: 2019-10-15

[41] for Disease Control, Centers ; Prevention: *Prostate Cancer Statistics*. 2022. – Accessed: 2023-06-01

[42] for Disease Control, Centers ; Prevention. *Prostate Cancer Incidence by Age and Stage at Diagnosis, United States—20012019. USCS data brief, no 34*. 2023

[43] Eladawi, Nabila ; Eltanboly, Ahmed ; Elmogy, Mohammed ; Ghazal, Mohammed ; Fraiwan, Luay ; Aboelfetouh, Ahmed ; Riad, Alaa ; Keynton, Robert ; El-Azab, Magdi ; Schaal, Shlomit ; El-Baz, Ayman: Diabetic retinopathy early detection based on OCT and OCTA feature fusion. In: *Proceedings - International Symposium on Biomedical Imaging* 2019-April (2019), S. 587–591. – ISBN 9781538636411

[44] Ethem, Alpaydin: *Introduction to Machine Learning*. 3. The MIT Press, 2014

[45] Faraj, Sheila F. ; Bezerra, Stephania M. ; Yousefi, Kasra ; Fedor, Helen ; Glavaris, Stephanie ; Han, Misop ; Partin, Alan W. ; Humphreys, Elizabeth ; Tosoian, Jeffrey ; Johnson, Michael H. ; Davicioni, Elai ; Trock, Bruce J. ; Schaeffer, Edward M. ; Ross, Ashley E. ; Netto, George J.: Clinical validation of the 2005 isup gleason grading system in a cohort of intermediate and high risk men undergoing radical prostatectomy. In: *PLoS ONE* 11 (2016), Nr. 1, S. 1–13. – ISSN 19326203

[46] Frank, Eibe ; Hall, Mark: A Simple Approach to Ordinal Classification. In: De Raedt, Luc (Hrsg.) ; Flach, Peter (Hrsg.): *Machine Learning: ECML 2001.*

Berlin, Heidelberg : Springer Berlin Heidelberg, 2001. – ISBN 978–3–540–44795–5, S. 145–156

[47] Garcia Arnal Barbedo, Jayme: A Review on Methods for Automatic Counting of Objects in Digital Images. In: *IEEE Latin America Transactions* 10 (2012), Nr. 5, S. 2112–2124

[48] Garg, Bhanu ; Manwani, Naresh: Robust Deep Ordinal Regression under Label Noise. In: Pan, Sinno J. (Hrsg.) ; Sugiyama, Masashi (Hrsg.): *Proceedings of The 12th Asian Conference on Machine Learning* Bd. 129. Bangkok, Thailand : PMLR, 18–20 Nov 2020, S. 782–796

[49] Gargeya, Rishab ; Leng, Theodore: Automated Identification of Diabetic Retinopathy Using Deep Learning. In: *Ophthalmology* 124 (2017), Nr. 7, S. 962–969

[50] Gieres, François: Mathematical surprises and Dirac's formalism in quantum mechanics. In: *Reports on Progress in Physics* 63 (2000), dec, Nr. 12, S. 1893

[51] Golabbakhsh, M. ; Rabbani, H.: Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model. In: *IET Image Processing* 7 (2013), November, Nr. 8, S. 768–776. – ISSN 1751–9667

[52] González, Fabio A. ; Ramos-Pollán, Raúl ; Gallego-Mejia, Joseph A.: Quantum Kernel Mixtures for Probabilistic Deep Learning. (2023)

[53] González, Fabio A. ; Vargas-Calderón, Vladimir ; Vinck-Posada, Herbert: Classification with quantum measurements. In: *Journal of the Physical Society of Japan* 90 (2021), Nr. 4, S. 044002

[54] González, F.A. ; Gallego, A. ; Toledo-Cortés, S. [u. a.]: Learning with density matrices and random features. In: *Quantum Machine Intelligence* 4 (2022), S. 23

[55] González, Fabio A. ; Vargas-Calderón, Vladimir ; Vinck-Posada, Herbert: Classification with Quantum Measurements. In: *Journal of the Physical Society of Japan* 90 (2021), Nr. 4, S. 044002

[56] Gulshan, Varun ; Peng, Lily ; Coram, Marc ; Stumpe, Martin C. ; Wu, Derek ; Narayanaswamy, Arunachalam ; Venugopalan, Subhashini ; Widner, Kasumi ; Madams, Tom ; Cuadros, Jorge ; Kim, Ramasamy ; Raman, Rajiv ; Nelson, Philip C. ; Mega, Jessica L. ; Webster, Dale R.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. In: *JAMA - Journal of the American Medical Association* 316 (2016), Nr. 22, S. 2402–2410

[57] Gunawardhana, Piumi L. ; Jayathilake, Raviru ; Withanage, Yasiru ; Ganegoda, Gamage U.: Automatic Diagnosis of Diabetic Retinopathy using Machine Learning: A Review. In: *Proceedings of ICITR 2020 - 5th International Conference on Information Technology Research: Towards the New Digital Enlightenment* (2020). ISBN 9781665414753

[58] Guo, Z. ; Li, X. ; Huang, H. ; Guo, N. ; Li, Q.: Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018. – ISSN 1945–8452, S. 903–907

[59] Guo, Z. ; Li, X. ; Huang, H. ; Guo, N. ; Li, Q.: Deep Learning-Based Image Segmentation on Multimodal Medical Imaging. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 3 (2019), March, Nr. 2, S. 162–169. – ISSN 2469–7303

[60] Gutiérrez, Pedro A. ; Pérez-Ortiz, María ; Sánchez-Monedero, Javier ; Fernández-Navarro, Francisco ; Hervás-Martínez, César: Ordinal Regression Methods: Survey and Experimental Study. In: *IEEE Transactions on Knowledge and Data Engineering* 28 (2016), Nr. 1, S. 127–146. – ISSN 10414347

[61] Gutman, David A. ; Cobb, Jake ; Somanna, Dhananjaya ; Park, Yuna ; Wang, Fusheng ; Kurc, Tahsin ; Saltz, Joel H. ; Brat, Daniel J. ; Cooper, Lee A. ; Kong, Jun: Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. In: *Journal of the American Medical Informatics Association* 20 (2013), Nr. 6, S. 1091–1098

[62] Hashemi, Mehrdad ; Zandieh, Mohammad A. ; Talebi, Yasmin ; Rahmanian, Parham ; Shafiee, Sareh S. ; Nejad, Melina M. ; Babaei, Roghayeh ; Sadi, Farzaneh H. ; Rajabi, Romina ; Abkenar, Zahra O. ; Rezaei, Shamin ; Ren, Jun ; Nabavi, Noushin ; Khorrami, Ramin ; Rashidi, Mohsen ; Hushmandi, Kiavash ; Entezari, Maliheh ; Taheriazam, Afshin: Paclitaxel and docetaxel resistance in prostate cancer: Molecular mechanisms and possible therapeutic strategies. In: *Biomedicine Pharmacotherapy* 160 (2023), S. 114392. – ISSN 0753–3322

[63] He, Shenghua ; Minn, Kyaw T. ; Solnica-Krezel, Lilianna ; Anastasio, Mark A. ; Li, Hua: Deeply-supervised density regression for automatic cell counting in microscopy images. In: *Medical Image Analysis* 68 (2021), S. 101892. – ISSN 1361–8415

[64] Huang, Di ; Heath Jeffery, Rachael C. ; Aung-Htut, May T. ; McLenachan, Samuel ; Fletcher, Sue ; Wilton, Steve D. ; Chen, Fred K.: Stargardt disease and progress in therapeutic strategies. In: *Ophthalmic Genetics* 43 (2022), Nr. 1, S. 1–26. – ISSN 17445094

[65] Huang, Huikang ; Situ, Haozhen ; Zheng, Shenggen: Bidirectional Information Flow Quantum State Tomography. In: *Chinese Physics Letters* 38 (2021), Nr. 4, S. 1–6. – ISSN 17413540

[66] Jim, Oscar A. ; Cirujeda, Pol ; Henning, M: Combining Radiology Images and Meta – data for Multimodal Medical Case – based Retrieval. In: *VISCERAL book* (2017), S. 1–14

[67] Jiménez del Toro, Oscar ; Atzori, Manfredo ; Otálora, Sebastian ; Andersson, Mats ; Eurén, Kristian ; Hedlund, Martin ; Rönnquist, Peter ; Müller, Henning: Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score. In: *Medical Imaging 2017: Digital Pathology* 10140 (2017), S. 101400O. – ISBN 9781510607255

[68] Kamnitsas, K. ; Bai, W. ; Ferrante, E. ; McDonagh, S. ; Sinclair, M. ; Pawlowski, N. ; Rajchl, M. ; Lee, M. ; Kainz, B. ; Rueckert, D. ; Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10670 LNCS (2018), S. 450–462. – ISBN 9783319752372

[69] Karimi, Davood ; Nir, Guy ; Fazli, Ladan ; Black, Peter C. ; Goldenberg, Larry ; Salcudean, Septimiu E.: Deep Learning-Based Gleason Grading of Prostate Cancer from Histopathology Images - Role of Multiscale Decision Aggregation and Data Augmentation. In: *IEEE Journal of Biomedical and Health Informatics* 24 (2020), may, Nr. 5, S. 1413–1426. – ISSN 21682208

[70] Kaya, Mahmut ; Bilge, H.s: Deep Metric Learning: A Survey. In: *Symmetry* 11 (2019), 08, S. 1066

[71] Kendall, Alex ; Gal, Yarin: What uncertainties do we need in Bayesian deep learning for computer vision? In: *Advances in Neural Information Processing Systems* 2017-December (2017), Nr. Nips, S. 5575–5585. – ISSN 10495258

[72] Khani, Ali A. ; Fatemi Jahromi, Seyed A. ; Shahreza, Hatef O. ; Behroozi, Hamid ; Baghshah, Mahdieh S.: Towards Automatic Prostate Gleason Grading Via Deep Convolutional Neural Networks. In: *5th Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS 2019* (2019), Nr. December, S. 18–19. ISBN 9781728153506

[73] Kim, Hee E. ; Cosa-Linan, Alejandro ; Santhanam, Nandhini ; Jannesari, Mahboubeh ; Maros, Mate E. ; Ganslandt, Thomas: Transfer learning for medical image classification: a literature review. In: *BMC medical imaging* 22 (2022), Nr. 1, S. 69

[74] Kim, Tae H. ; Jeong, Dae J. ; Hahn, Soo Y. ; Shin, Jung H. ; Oh, Young L. ; Ki, Chang S. ; Kim, Jong W. ; Jang, Ju Y. ; Cho, Yoon Y. ; Chung, Jae H. ; Kim, Sun W.: Triage of patients with AUS/FLUS on thyroid cytopathology: Effectiveness of the multimodal diagnostic techniques. In: *Cancer Medicine* 5 (2016), Nr. 5, S. 769–777. – ISSN 20457634

[75] Kleesiek, Jens ; Urban, Gregor ; Hubert, Alexander ; Schwarz, Daniel ; Maier-Hein, Klaus ; Bendszus, Martin ; Biller, Armin: Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. In: *NeuroImage* 129 (2016), S. 460–469. – ISSN 10959572

[76] Kong, Jun ; Cooper, Lee A. ; Wang, Fusheng ; Gutman, David A. ; Gao, Jingjing ; Chisolm, Candace ; Sharma, Ashish ; Pan, Tony ; Van Meir, Erwin G. ; Kurc, Tahsin M. ; Moreno, Carlos S. ; Saltz, Joel H. ; Brat, Daniel J.: Integrative, multimodal analysis of glioblastoma using TCGA molecular data, pathology images, and clinical outcomes. In: *IEEE Transactions on Biomedical Engineering* 58 (2011), Nr. 12 PART 2, S. 3469–3474. – ISSN 00189294

[77] Kononenko, Igor: Machine learning for medical diagnosis: History, state of the art and perspective. In: *Artificial Intelligence in Medicine* 23 (2001), Nr. 1, S. 89–109. – ISSN 09333657

[78] Kovalyk, Oleksandr ; Morales-Sánchez, Juan ; Verdú-Monedero, Rafael ; Sellés-Navarro, Inmaculada ; Palazón-Cabanes, Ana ; Sancho-Gómez, José Luis: PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. In: *Scientific Data* 9 (2022), Nr. 1, S. 1–12. – ISBN 4159702201

[79] Krause, Jonathan ; Gulshan, Varun ; Rahimy, Ehsan ; Karth, Peter ; Widner, Kasumi ; Corrado, Greg S. ; Peng, Lily ; Webster, Dale R.: Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. In: *Ophthalmology* 125 (2018), Nr. 8, S. 1264–1272. – ISSN 15494713

[80] Krenn, Mario ; Malik, Mehul ; Fickler, Robert ; Lapkiewicz, Radek ; Zeilinger, Anton: Automated Search for new Quantum Experiments. In: *Phys. Rev. Lett.* 116 (2016), S. 090405

[81] Lara, Juan S. ; Contreras O., Victor H. ; Otálora, Sebastián ; Müller, Henning ; González, Fabio A.: Multimodal Latent Semantic Alignment for Automated Prostate Tissue Classification and Retrieval. In: Martel, Anne L. (Hrsg.) ; Abolmaesumi, Purang (Hrsg.) ; Stoyanov, Danail (Hrsg.) ; Mateus, Diana (Hrsg.) ; Zuluaga, Maria A. (Hrsg.) ; Zhou, S. K. (Hrsg.) ; Racoceanu, Daniel (Hrsg.) ;

JOSKOWICZ, Leo (Hrsg.): *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Cham : Springer International Publishing, 2020. – ISBN 978–3–030–59722–1, S. 572–581

[82] LARA, Juan S. ; CONTRERAS O, Victor H. ; OTÁLORA, Sebastián ; MÜLLER, Henning ; GONZÁLEZ, Fabio A.: 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)* Bd. 12265 LNCS, 2020. – ISBN 9783030597214, S. 572–581

[83] LEE, Jimmy A. ; LIU, Peng ; CHENG, Jun ; FU, Huazhu: A Deep Step Pattern Representation for Multimodal Retinal Image Registration. In: *The IEEE International Conference on Computer Vision (ICCV)* (2019), S. 5077–5086

[84] LEE, Ki S. ; JUNG, Seok K. ; RYU, Jae J. ; SHIN, Sang W. ; CHOI, Jinwook: Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. In: *Journal of Clinical Medicine* 9 (2020), Nr. 2. – ISSN 20770383

[85] LEIBIG, Christian ; ALLKEN, Vaneeda ; AYHAN, Murat S. ; BERENS, Philipp ; WAHL, Siegfried: Leveraging uncertainty information from deep neural networks for disease detection. In: *Scientific Reports* 7 (2017), Nr. 1, S. 1–14. – ISSN 20452322

[86] LEIBIG, Christian ; ALLKEN, Vaneeda ; AYHAN, Murat S. ; BERENS, Philipp ; WAHL, Siegfried: Leveraging uncertainty information from deep neural networks for disease detection. In: *Scientific Reports* 7 (2017), Nr. 1, S. 1–14

[87] LI, Daoliang ; MIAO, Zheng ; PENG, Fang ; WANG, Liang ; HAO, Yinfeng ; WANG, Zhenhu ; CHEN, Tao ; LI, Hui ; ZHENG, Yingying: Automatic counting methods in aquaculture: A review. In: *Journal of the World Aquaculture Society* 52 (2021), Nr. 2, S. 269–283

[88] LI, Feng ; LIU, Zheng ; CHEN, Hua ; JIANG, Minshan ; ZHANG, Xuedian ; WU, Zhizheng: Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. In: *Translational Vision Science and Technology* 8 (2019), Nr. 6. – ISSN 21642591

[89] LI, Hongming ; HABES, Mohamad ; FAN, Yong: Deep Ordinal Ranking for Multi-Category Diagnosis of Alzheimer's Disease using Hippocampal MRI data. In: *arXiv* (2017), sep

[90] LI, Yuchun ; HUANG, Mengxing ; ZHANG, Yu ; CHEN, Jing ; XU, Haixia ; WANG, Gang ; FENG, Wenlong: Automated Gleason Grading and Gleason Pattern Region

Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer. In: *IEEE Access* 8 (2020), S. 117714–117725. – ISSN 21693536

[91] Lim, G. ; Bellemo, V. ; Xie, Y. [u. a.]: Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review. In: *Eye and Vision* 7 (2020), S. 21

[92] Lim, Zhan W. ; Lee, Mong L. ; Hsu, Wynne ; Wong, Tien Y.: Building Trust in Deep Learning System towards Automated Disease Detection. In: *The Thirty-First AAAI Conference on Innovative Applications of Artificial Intelligence* (2018), S. 9516–9521

[93] Liu, Xiaofeng: Ordinal Regression with Neuron Stick-breaking for Medical Diagnosis. 2018. – Forschungsbericht. – 0–0 S

[94] Lucas, Marit ; Jansen, Ilaria ; Savci-Heijink, C. D. ; Meijer, Sybren L. ; de Boer, Onno J. ; van Leeuwen, Ton G. ; de Bruin, Daniel M. ; Marquering, Henk A.: Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies. In: *Virchows Archiv* 475 (2019), Nr. 1, S. 77–83. – ISSN 14322307

[95] Ma, Mengmeng ; Ren, Jian ; Zhao, Long ; Testuggine, Davide ; Peng, Xi: Are Multimodal Transformers Robust to Missing Modality? (2022), S. 18177–18186

[96] Ma, Mengmeng ; Ren, Jian ; Zhao, Long ; Tulyakov, Sergey ; Wu, Cathy ; Peng, Xi: SMIL: Multimodal Learning with Severely Missing Modality. In: *35th AAAI Conference on Artificial Intelligence, AAAI 2021* 3B (2021), S. 2302–2310. ISBN 9781713835974

[97] Manjula Sri, K.M.M. R.: Novel image pro-cessing techniquesto detect lesion us-ing lab view R. In: *India Conference (INDICON), 2011 Annual IEEE* (2011)

[98] Mcgurk, Harry ; Macdonald, John: Hearing lips and seeing voices. In: *Nature* 264 (1976), Nr. 5588, S. 746–748. – ISSN 00280836

[99] Menze, Bjoern ; Reyes, Mauricio ; Jakab, Andras ; Gerstner, Elisabeth ; Farahani, Keyvan ; Menze, Bjoern ; Reyes, Mauricio ; Jakab, Andras ; Gerstner, Elisabeth ; Kirby, Justin: Brain Tumor Image Segmentation (BRATS) 2013 To cite this version : NCI-MICCAI Challenge on Multimodal Brain Tumor Segmentation. (2013)

[100] MINSALUD: ANÁLISIS DE SITUACIÓN DE SALUD VISUAL EN COLOMBIA 2016. (2015). – ISBN 9780511993398

[101] Miri, Mohammad S. ; Abràmoff, Michael D. ; Lee, Kyungmoo ; Niemeijer, Meindert ; Wang, Jui K. ; Kwon, Young H. ; Garvin, Mona K.: Multimodal Segmentation of Optic Disc and Cup from SD-OCT and Color Fundus Photographs Using a Machine-Learning Graph-Based Approach. In: *IEEE Transactions on Medical Imaging* 34 (2015), Nr. 9, S. 1854–1866. – ISSN 1558254X

[102] Moccia, Sara ; Wirkert, Sebastian J. ; Kenngott, Hannes ; Vemuri, Anant S. ; Apitz, Martin ; Mayer, Benjamin ; De Momi, Elena ; Mattos, Leonardo S. ; Maier-Hein, Lena: Uncertainty-aware organ classification for surgical data science applications in laparoscopy. In: *IEEE Transactions on Biomedical Engineering* 65 (2018), Nr. 11, S. 2649–2659. – ISSN 15582531

[103] Mookiah, Muthu Rama K. ; Acharya, U. R. ; Chua, Chua K. ; Min, Lim C. ; Ng, E. Y. ; Mushrif, Milind M. ; Laude, Augustinus: Automated detection of optic disk in retinal fundus images using intuitionistic fuzzy histon segmentation. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 227 (2013), Nr. 1, S. 37–49. – ISSN 09544119

[104] Morgan, Jessica I. ; Chen, Min ; Huang, Andrew M. ; Jiang, Yu Y. ; Cooper, Robert F.: Cone identification in choroideremia: Repeatability, reliability, and automation through use of a convolutional neural network. In: *Translational Vision Science and Technology* 9 (2020), Nr. 2, S. 1–13. – ISSN 21642591

[105] Müller, Henning ; Ünay, Devrim: Medical Decision Support Using Increasingly Large Multimodal Data Sets. In: *Big Data Analytics for Large-Scale Multimedia Search* (2019), S. 317–336

[106] Nagpal, Kunal ; Foote, Davis ; Liu, Yun ; Chen, Po Hsuan C. ; Wulczyn, Ellery ; Tan, Fraser ; Olson, Niels ; Smith, Jenny L. ; Mohtashamian, Arash ; Wren, James H. ; Corrado, Greg S. ; MacDonald, Robert ; Peng, Lily H. ; Amin, Mahul B. ; Evans, Andrew J. ; Sangoi, Ankur R. ; Mermel, Craig H. ; Hipp, Jason D. ; Stumpe, Martin C.: Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. In: *npj Digital Medicine* 2 (2019), dec, Nr. 1, S. 1–10. – ISSN 23986352

[107] Nakatake, Shunji ; Murakami, Yusuke ; Funatsu, Jun ; Koyanagi, Yoshito ; Akiyama, Masato ; Momozawa, Yukihide ; Ishibashi, Tatsuro ; Sonoda, Koh H. ; Ikeda, Yasuhiro: Early detection of cone photoreceptor cell loss in retinitis pigmentosa using adaptive optics scanning laser ophthalmoscopy. In: *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 257 (2019), Nr. 6, S. 1169–1181. – ISSN 1435702X

[108] Nie, Dong ; Wang, Li ; Gao, Yaozong ; Shen, Dinggang: FULLY CONVOLUTIONAL NETWORKS FOR MULTI-MODALITY ISOINTENSE INFANT BRAIN IMAGE SEGMENTATION Dong. In: *Proc IEEE Int Symp Biomed Imaging* (2016)

[109] Niu, Zhenxing ; Zhou, Mo ; Wang, Le ; Gao, Xinbo ; Hua, Gang: Ordinal regression with multiple output CNN for age estimation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Bd. 2016-Decem, IEEE Computer Society, dec 2016. – ISBN 9781467388504, S. 4920–4928

[110] Otálora, Sebastian ; Perdomo, Oscar ; González, Fabio ; Müller, Henning: 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)* Bd. 10552 LNCS, 2017. – ISBN 9783319675336, S. 146–154

[111] Perdomo, Oscar. ; Gonzalez, Fabio.: A Systematic Review of Deep Learning Methods Applied to Ocular Images. In: *Ciencia e Ingenieria Neogranadina* 30 (2019), Nr. 1. ISBN 0000000314251

[112] Perdomo, Oscar ; Otalora, Sebastian ; Gonzalez, Fabio A. ; Meriaudeau, Fabrice ; Muller, Henning: OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes. In: *Proceedings - International Symposium on Biomedical Imaging* 2018-April (2018), Nr. Isbi, S. 1423–1426. – ISBN 9781538636367

[113] Perdomo Charry, Oscar J. ; Arevalo, John ; González, Fabio A.: Combining morphometric features and convolutional networks fusion for glaucoma diagnosis. (2017), Nr. November, S. 57. – ISBN 9781510616332

[114] Pinz, A. ; Bernogger, S. ; Datlinger, P. ; Kruger, A.: Mapping the human retina. In: *IEEE Transactions on Medical Imaging* 17 (1998), Aug, Nr. 4, S. 606–619. – ISSN 1558–254X

[115] Piotter, Elena ; McClements, Michelle E. ; Maclaren, Robert E.: Therapy approaches for stargardt disease. In: *Biomolecules* 11 (2021), Nr. 8, S. 1–28. – ISSN 2218273X

[116] Polikar, R. ; Tilley, C. ; Hillis, B. ; Clark, C. M.: Multimodal EEG, MRI and PET data fusion for Alzheimer's disease diagnosis. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010. – ISSN 1558–4615, S. 6058–6061

[117] Rahimi, Ali ; Recht, Ben: Random features for large-scale kernel machines. In: *Advances in neural information . . .* (2007), Nr. 1, S. 1–8. – ISBN 160560352X

[118] Rahimi, Ali ; Recht, Benjamin: Random features for large-scale kernel machines. In: *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, 2009. – ISBN 160560352X

[119] Ramirez, Geovany A. ; Baltrušaitis, Tadas ; Morency, Louis-Philippe: Modeling Latent Discriminative Dynamic of Multi-dimensional Affective Signals. In: D'Mello, Sidney (Hrsg.) ; Graesser, Arthur (Hrsg.) ; Schuller, Björn (Hrsg.) ; Martin, Jean-Claude (Hrsg.): *Affective Computing and Intelligent Interaction*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2011. – ISBN 978–3–642–24571–8, S. 396–406

[120] Rasmussen, Carl E. ; Williams, Christopher K. I.: *Gaussian processes for machine learning.* The MIT Press, 2006. – ISBN 026218253X

[121] Ren, Jian ; Hacihaliloglu, Ilker ; Singer, Eric A. ; Foran, David J. ; Qi, Xin: Unsupervised Domain Adaptation for Classification of Histopathology Whole-Slide Images. In: *Frontiers in Bioengineering and Biotechnology* 7 (2019), Nr. May, S. 1–12. – ISSN 2296–4185

[122] Ritter, N. ; Owens, R. ; Cooper, J. ; Eikelboom, R. H. ; Van Saarloos, P. P.: Registration of stereo and temporal images of the retina. In: *IEEE Transactions on Medical Imaging* 18 (1999), May, Nr. 5, S. 404–418. – ISSN 1558–254X

[123] Ronneberger, Olaf ; Fischer, Philipp ; Brox, Thomas: U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351 (2015), S. 234–241. – ISBN 9783319245737

[124] Roorda, Austin ; Romero-Borja, Fernando ; III, William J. D. ; Queener, Hope ; Hebert, Thomas J. ; Campbell, Melanie C.: Adaptive optics scanning laser ophthalmoscopy. In: *Opt. Express* 10 (2002), May, Nr. 9, S. 405–412

[125] Russakovsky, Olga ; Deng, Jia ; Su, Hao ; Krause, Jonathan ; Satheesh, Sanjeev ; Ma, Sean ; Huang, Zhiheng ; Karpathy, Andrej ; Khosla, Aditya ; Bernstein, Michael ; Berg, Alexander C. ; Fei-Fei, Li: ImageNet Large Scale Visual Recognition Challenge. In: *International Journal of Computer Vision (IJCV)* 115 (2015), Nr. 3, S. 211–252

[126] Sahran, Shahnorbanun ; Albashish, Dheeb ; Abdullah, Azizi ; Shukor, Nordashima A. ; Hayati Md Pauzi, Suria: Absolute cosine-based SVM-RFE feature selection method for prostate histopathological grading. In: *Artificial Intelligence in Medicine* 87 (2018), S. 78–90. – ISSN 18732860

[127] Salam, Anum A. ; Khalil, Tehmina ; Akram, M. U. ; Jameel, Amina ; Basit, Imran: Automated detection of glaucoma using structural and non structural features. In: *SpringerPlus* 5 (2016), Nr. 1. – ISSN 21931801

[128] Salam, Anum A. ; Akram, M. U. ; Wazir, Kamran ; Anwar, Syed M. ; Majid, Muhammad: Autonomous Glaucoma detection from fundus image using cup to disc ratio and hybrid features. In: *2015 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2015* (2016), Nr. c, S. 370–374. ISBN 9781509004805

[129] Schlegl, Thomas ; Waldstein, Sebastian ; Vogl, Wolf-Dieter ; Schmidt-Erfurth, Ursula ; Langs, Georg: Predicting Semantic Descriptions from Medical Images with Convolutional Neural Networks. In: *Information Processing in Medical Imaging* 9123 (2015), Nr. Chapter 58, S. 733–745. – ISBN 978–3–319–19991–7

[130] Schmidhuber, Jürgen: Deep Learning in neural networks: An overview. In: *Neural Networks* 61 (2015), S. 85–117. – ISBN 0893–6080

[131] Shah, Mital ; Roomans Ledo, Ana ; Rittscher, Jens: Automated classification of normal and Stargardt disease optical coherence tomography images using deep learning. In: *Acta Ophthalmologica* 98 (2020), Nr. 6, S. e715–e721

[132] Shawe-Taylor, John ; Cristianini, Nello: *Kernel Methods for Pattern Analysis*. New York, New York, USA : Cambridge University Press, 2004. – ISBN 9780521813976

[133] Singh, Amitojdeep ; Sengupta, Sourya ; Lakshminarayanan, Vasudevan: Explainable deep learning models in medical image analysis. In: *Journal of Imaging* 6 (2020), Nr. 6, S. 1–19. – ISSN 2313433X

[134] Society, American C. *Key Statistics for Prostate Cancer*. 2023

[135] Stolte, Skylar ; Fang, Ruogu: A survey on medical image analysis in diabetic retinopathy. In: *Medical Image Analysis* 64 (2020), S. 101742. – ISSN 13618423

[136] Ström, Peter ; Kartasalo, Kimmo ; Olsson, Henrik ; Solorzano, Leslie ; Delahunt, Brett ; Berney, Daniel M. ; Bostwick, David G. ; Evans, Andrew J. ; Grignon, David J. ; Humphrey, Peter A. ; Iczkowski, Kenneth A. ; Kench, James G. ; Kristiansen, Glen ; van der Kwast, Theodorus H. ; Leite, Katia R. ; McKenney, Jesse K. ; Oxley, Jon ; Pan, Chin C. ; Samaratunga, Hemamali ; Srigley, John R. ; Takahashi, Hiroyuki ; Tsuzuki, Toyonori ; Varma, Murali ; Zhou, Ming ; Lindberg, Johan ; Lindskog, Cecilia ; Ruusuvuori, Pekka ; Wählby, Carolina ; Grönberg, Henrik ; Rantalainen, Mattias ; Egevad, Lars ; Eklund, Martin: Artificial intelligence for diagnosis and grading of prostate cancer

in biopsies: a population-based, diagnostic study. In: *The Lancet Oncology* 21 (2020), Nr. 2, S. 222–232. – ISSN 14745488

[137] Summerfield, Quentin: L ipreading and audio-visual speech perception. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 335 (1992), Nr. 1273, S. 71–78

[138] Sun, Yunlian ; Tang, Jinhui ; Sun, Zhenan ; Tistarelli, Massimo: Facial Age and Expression Synthesis Using Ordinal Ranking Adversarial Networks. In: *IEEE Transactions on Information Forensics and Security* 15 (2020), S. 2960–2972. – ISSN 15566021

[139] Szegedy, Christian ; Vanhoucke, Vincent ; Ioffe, Sergey ; Shlens, Jon ; Wojna, Zbigniew: Rethinking the Inception Architecture for Computer Vision. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-December (2016), S. 2818–2826. – ISBN 9781467388504

[140] Tang, Yehui ; Yan, Junchi ; Hu, Guoqiang ; Zhang, Baohua ; Zhou, Jinzan: Recent progress and perspectives on quantum computing for finance. In: *Service Oriented Computing and Applications* 16 (2022), Nr. 4, S. 227–229

[141] Teng, Q. ; Liu, Z. ; Song, Y. ; Han, K. ; Lu, Y.: A survey on the interpretability of deep learning in medical diagnosis. In: *Multimed Syst* 28 (2022), Nr. 6, S. 2335–2355

[142] Threatt, Jennifer ; Williamson, Jennifer F. ; Huynh, Kyle ; Davis, Richard M.: Ocular disease, knowledge and technology applications in patients with diabetes. In: *American Journal of the Medical Sciences* 345 (2013), Nr. 4, S. 266–270. – ISSN 00029629

[143] Tian, Li ; Ma, Liyan ; Wen, Zhijie ; Xie, Shaorong ; Xu, Yupeng: Learning Discriminative Representations for Fine-Grained Diabetic Retinopathy Grading. (2020)

[144] Toledo-Cortés, Santiago ; De La Pava, Melissa ; Perdómo, Oscar ; González, Fabio A.: Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification. In: *Ophthalmic Medical Image Analysis. OMIA 2020. Lecture Notes in Computer Science, vol 12069. Springer, Cham.*, 2020, S. 206–215

[145] Toledo-Cortés, Santiago: *AOSLO-CNN Diagnosis and Counting.* `https://github.com/stoledoc/AOSLO-CNN_Diagnosis_Counting`. 2022. – [Online; accessed 01-Mar-2022]

[146] Toledo-Cortés, Santiago: *DQOR Code for Medical Image Grading with Deep Quantum Ordinal Regression.* `https://github.com/stoledoc/DQOR`. 2022. – [Online; accessed 01-Mar-2022]

[147] Toledo-Cortés, Santiago ; Dubis, Adam M. ; González, Fabio A. ; Müller, Henning: Deep Density Estimation for Cone Counting and Diagnosis of Genetic Eye Diseases From Adaptive Optics Scanning Light Ophthalmoscope Images. In: *Translational Vision Science Technology* 12 (2023), 11, Nr. 11, S. 25–25. – ISSN 2164–2591

[148] Toledo-Cortés, Santiago ; Useche, Diego H. ; Müller, Henning ; González, Fabio A.: Grading diabetic retinopathy and prostate cancer diagnostic images with deep quantum ordinal regression. In: *Computers in Biology and Medicine* 145 (2022), S. 105472. – ISSN 0010–4825

[149] Tolkach, Yuri ; Dohmgörgen, Tilmann ; Toma, Marieta ; Kristiansen, Glen: High-accuracy prostate cancer pathology using deep learning. In: *Nature Machine Intelligence* 2 (2020), jul, Nr. 7, S. 411–418. – ISSN 25225839

[150] In: Tsang, Stephen H. ; Sharma, Tarun: *Stargardt Disease*. Cham : Springer International Publishing, 2018, S. 139–151. – ISBN 978–3–319–95046–4

[151] Uppamma, P. ; Bhattacharya, S.: Deep Learning and Medical Image Processing Techniques for Diabetic Retinopathy: A Survey of Applications, Challenges, and Future Trends. In: *J Healthc Eng* 2023 (2023), S. 2728719

[152] Urban, G. ; Bendszus, M. ; Hamprecht, Fred A. ; Kleesiek, J.: Multi-modal Brain Tumor Segmentation using Deep Convolutional NeuralNetworks. In: *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winningcontribution*, 2014, S. 31–35

[153] Vaghefi, Ehsan ; Hill, Sophie ; Kersten, Hannah M. ; Squirrell, David: Multimodal Retinal Image Analysis via Deep Learning for the Diagnosis of Intermediate Dry Age-Related Macular Degeneration : A Feasibility Study. 2020 (2020)

[154] Vaicenavicius, Juozas ; Widmann, David ; Andersson, Carl ; Lindsten, Fredrik ; Roll, Jacob ; Schön, Thomas: Evaluating model calibration in classification. In: Chaudhuri, Kamalika (Hrsg.) ; Sugiyama, Masashi (Hrsg.): *Proceedings of Machine Learning Research* Bd. 89, PMLR, 16–18 Apr 2019, S. 3459–3467

[155] Van Grinsven, Mark J. ; Buitendijk, Gabriëlle H.S. ; Brussee, Corina ; Van Ginneken, Bram ; Hoyng, Carel B. ; Theelen, Thomas ; Klaver, Caroline C. ; Sánchez, Clara I.: Automatic identification of reticular pseudodrusen using multimodal retinal image analysis. In: *Investigative Ophthalmology and Visual Science* 56 (2015), Nr. 1, S. 633–639. – ISSN 15525783

[156] Vanegas, Jorge A.: *Large-scale Non-linear Multimodal Semantic Embedding Large-scale Non-linear Multimodal Semantic Embedding*, Dissertation, 2017

[157] Voets, Mike ; Møllersen, Kajsa ; Bongo, Lars A.: Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. In: *PLoS ONE* 14 (2019), Nr. 6, S. 1–11

[158] Voets, Mike ; Møllersen, Kajsa ; Bongo, Lars A.: Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. In: *PLoS ONE* 14 (2019), Nr. 6, S. 1–11. – ISBN 1111111111

[159] Wang, Daihou ; Foran, David J. ; Ren, Jian ; Zhong, Hua ; Kim, Isaac Y. ; Qi, Xin: Exploring automatic prostate histopathology image gleason grading via local structure modeling. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2015-Novem (2015), S. 2649–2652. – ISBN 9781424492718

[160] Wang, Weisen ; Xu, Zhiyan ; Yu, Weihong ; Zhao, Jianchun ; Yang, Jingyuan ; He, Feng ; Yang, Zhikun ; Chen, Di ; Ding, Dayong ; Chen, Youxin ; Li, Xirong: Two-Stream CNN with Loose Pair Training for Multi-modal AMD Categorization. (2019), S. 156–164. – ISBN 9783030322380

[161] Wells, John A. ; Glassman, Adam R. ; Ayala, Allison R. ; Jampol, Lee M. ; Bressler, Neil M. ; Bressler, Susan B. ; Brucker, Alexander J. ; Ferris, Frederick L. ; Hampton, G. R. ; Jhaveri, Chirag ; Melia, Michele ; Beck, Roy W.: Aflibercept, Bevacizumab, or Ranibizumab for Diabetic Macular Edema Two-Year Results from a Comparative Effectiveness Randomized Clinical Trial. In: *Ophthalmology* 123 (2016), Nr. 6, S. 1351–1359

[162] Wilkinson, C. P. ; Ferris, Frederick L. ; Klein, Ronald E. ; Lee, Paul P. ; Agardh, Carl D. ; Davis, Matthew ; Dills, Diana ; Kampik, Anselm ; Pararajasegaram, R. ; Verdaguer, Juan T. ; Lum, Flora: Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. In: *Ophthalmology* 110 (2003), Nr. 9, S. 1677–1682

[163] Wilson, Andrew ; Nickisch, Hannes: Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP). In: *Proceedings of the 32nd International Conference on Machine Learning*, JMLR: W&CP. Lille, France, 2015

[164] World Health Organisation: World report on vision. 2019 (14). – Forschungsbericht. – 180–235 S. – ISBN 9789241516570

[165] Wynne, Niamh ; Carroll, Joseph ; Duncan, Jacque L.: Promises and pitfalls of evaluating photoreceptor-based retinal disease with adaptive optics scanning light

ophthalmoscopy (AOSLO). In: *Progress in Retinal and Eye Research* (2021), Nr. October, S. 100920. – ISSN 18731635

[166] Xie, Weidi ; Noble, J. A. ; Zisserman, Andrew: Microscopy cell counting and detection with fully convolutional regression networks. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* 6 (2018), Nr. 3, S. 283–292. – ISSN 21681171

[167] Xin, Qiu ; Elliot, Meyerson ; Miikkulainen, Risto: Quantifying Point-Prediction Uncertainty in Neural Networks via Residual Estimation with an I/O Kernel. In: *ICLR 2020.*, 2019, S. 1–17. Addis Ababa, Ethiopia

[168] Yang, Jialiang ; Ju, Jie ; Guo, Lei ; Ji, Binbin ; Shi, Shufang ; Yang, Zixuan ; Gao, Songlin ; Yuan, Xu ; Tian, Geng ; Liang, Yuebin ; Yuan, Peng: Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. In: *Computational and Structural Biotechnology Journal* 20 (2022), S. 333–342. – ISSN 2001–0370

[169] Yap, Jordan ; Yolland, William ; Tschandl, Philipp: Multimodal skin lesion classification using deep learning. In: *Experimental Dermatology* 27 (2018), Nr. 11, S. 1261–1267. – ISSN 16000625

[170] Yau, Joanne W. ; Rogers, Sophie L. ; Kawasaki, Rho ; Lamoureux, Ecosse L. ; Kowalski, Jonathan W. ; Bek, Toke ; Chen, Shih J. ; Dekker, Jacqueline M. ; Fletcher, Astrid ; Grauslund, Jakob ; Haffner, Steven ; Hamman, Richard F. ; Ikram, M. K. ; Kayama, Takamasa ; Klein, Barbara E. ; Klein, Ronald ; Krishnaiah, Sannapaneni ; Mayurasakorn, Korapat ; O'Hare, Joseph P. ; Orchard, Trevor J. ; Porta, Massimo ; Rema, Mohan ; Roy, Monique S. ; Sharma, Tarun ; Shaw, Jonathan ; Taylor, Hugh ; Tielsch, James M. ; Varma, Rohit ; Wang, Jie J. ; Wang, Ningli ; West, Sheila ; Zu, Liang ; Yasuda, Miho ; Zhang, Xinzhi ; Mitchell, Paul ; Wong, Tien Y.: Global prevalence and major risk factors of diabetic retinopathy. In: *Diabetes Care* 35 (2012), Nr. 3, S. 556–564

[171] Yoo, Tae K. ; Choi, Joon Y. ; Seo, Jeong G. ; Ramasubramanian, Bhoopalan ; Selvaperumal, Sundaramoorthy ; Kim, Deok W.: The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment. In: *Medical and Biological Engineering and Computing* 57 (2019), Nr. 3, S. 677–687. – ISSN 17410444

[172] Zadeh, Amir ; Chen, Minghai ; Cambria, Erik ; Poria, Soujanya ; Morency, Louis P.: Tensor fusion network for multimodal sentiment analysis. In: *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings* (2017), S. 1103–1114. ISBN 9781945626838

[173] Zana, F. ; Klein, J. C.: A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform. In: *IEEE Transactions on Medical Imaging* 18 (1999), May, Nr. 5, S. 419–428. – ISSN 1558–254X

[174] Zeng, Xianglong ; Chen, Haiquan ; Luo, Yuan ; Ye, Wenbin: Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network. In: *IEEE Access* 7 (2019), Nr. c, S. 30744–30753

[175] Zhang, Daoqiang ; Wang, Yaping ; Zhou, Luping ; Yuan, Hong: Multimodal Classification of Alzheimer's Disease and Mild Cognitive Impairment. In: *Neuroimage* (2011). – ISBN 6176321972

[176] Zhang, Yao ; Ni, Qiang: Recent advances in quantum machine learning. In: *Quantum Engineering* 2 (2020), Nr. 1, S. e34