Automatic retinopathy detection using deep learning and medical findings

Melissa de la Pava Rodriguez

Universidad Nacional de Colombia
Systems and Industrial Engineering Department
Engineering Department
Bogotá, Colombia
2021

Automatic retinopathy detection using deep learning and medical findings

Melissa de la Pava Rodriguez

In partial fulfilment of the requirements for the degree of:
Master of Systems and Computing Engineering

Advisor:
Fabio Augusto González Osorio, Ph.D.
Systems and Industrial Engineering Department

Coadvisor:
Oscar Julián Perdomo Charry, Ph.D.

Research Field:
Applied computing
Research Group:
MindLab Research Group

Universidad Nacional de Colombia
Systems and Industrial Engineering Department
Engineering Department
Bogotá, Colombia
2021

Dedication

To my mother

Acknowledgements

I want to express my gratitude to my mother Estela, for always being there for me. This would not be possible without her unconditional support. I want to thank Hugo and Maria Helena for their caring. I am thank full to my advisor Fabio González for his guidance and trust in me to develop this project and to my coadvisor Oscar J. Perdomo for his continuous encouragement and advice. I am grateful for the opportunity to be part of the MindLab group, learning and interchanging ideas with the members. I want to thank to the Universidad Nacional de Colombia for all the knowledge and opportunities it has provided to me.

This work was financially supported by COLCIENCIAS with the research project "Detección Temprana de dano ocular en Diabéticos usando un sistema de Inteligencia Artificial en Imágenes de Fondo de Ojo", under the code quipu No. 202010013045. The support is gratefully acknowledge.

Abstract

Diabetic retinopathy (DR) is the result of a complication of diabetes affecting the retina. It can cause blindness if left undiagnosed and untreated. The ophthalmologist performs the diagnosis by screening each patient and detecting in ocular imaging the lesions caused by DR, namely, microaneurisms, hemorrhages, cotton wool spots, venous beading and neovascularization. However, the analysis of ocular findings is cumbersome, time-consuming, and demanding. Due to the insufficient amount of trained specialists to diagnose the illness, and the actual growing population with DR, it is important to develop a method to assist the DR diagnosis. This thesis presents two approaches for the automatic classification of DR using eye fundus images. The first one utilizes convolutional neural networks, transfer learning and shallow machine learning classifiers to identify the main ocular lesions related to DR and then use them to diagnose the illness. The second one is a multitask model which predicts simultaneously ocular lesions and DR. These approaches follow a similar workflow to that of clinicians, providing information that can be interpreted clinically to support the prediction. To achieve this goal a subset of the kaggle EyePACS and the Messidor-2 datasets, are labeled with ocular lesions by a certified opthalmologist. The kaggle EyePACS subset is used as training set and the Messidor-2 dataset is used as test set for both, the lesions and DR classification models. The results indicate that both methods achieve results comparable with state-of-the-art performances. The best results are obtained using the first approach with a multi layer perceptron as classifier for the automatic detection of DR, however, the multitask approach lead to similar results and has a simpler architecture.

Keywords: ocular lesions, diabetic retinopathy, convolutional neural networks, transfer learning, multitask models, shallow machine learning classifiers.

Resumen

Título en español: Detección automática de retinopatía diabética usando aprendizaje profundo y hallazgos médicos.
La retinopatía diabética (RD) es el resultado de una complicacion de la diabetes que afecta la retina. Puede causar ceguera si no se diagnostica ni se trata. El diagnóstico de esta enfermedad se hace mediante el escaneo de cada paciente y el análisis de imágenes oculares para detectar lesiones causadas por la RD, como microaneurismas, hemorragias, manchas algodonosas, arrosamiento venoso y neovascularización. Sin embargo, el análisis de las lesiones oculares es engorroso, lento y exigente. Debido a la cantidad insuficiente de especialistas capacitados para diagnosticar la enfermedad y al crecimiento actual de la población con RD, es importante desarrollar un método para ayudar en el diagnóstico de esta enfermedad. Esta tesis presenta dos enfoques para la clasificación automática de la RD utilizando imágenes de fondo de ojo. El primero utiliza redes neuronales convolucionales, transferencia de aprendizaje y clasificadores clásicos de aprendizaje de máquina para identificar las principales lesiones oculares relacionadas con la RD y luego usarlas para diagnosticar la enfermedad. El segundo es un modelo multitarea que predice simultáneamente lesiones oculares y RD. Estos enfoques siguen un flujo de trabajo similar al de los médicos, proporcionando información que puede interpretarse clínicamente para respaldar la predicción. Para lograr este objetivo, un subconjunto de las bases de datos kaggle EyePACS y Messidor-2 fueron etiquetados con lesiones oculares por un oftalmólogo certificado. El subconjunto de kaggle EyePACS se utiliza como conjunto de entrenamiento y el de Messidor-2 se utiliza como conjunto de prueba tanto para los modelos de detección de lesiones, como para los de clasificación de RD. Los resultados indican que ambos enfoques logran desempeños comparables con los métodos del estado del arte. Los mejores resultados se obtienen utilizando el primer enfoque con un perceptrón multicapa como clasificador para la detección automática de RD, sin embargo, el enfoque multitarea conduce a resultados similares y tiene una arquitectura más simple.

Palabras clave: lesiones oculares, retinopatía diabética, redes convolucionales, transferecia de aprendizaje, modelo multitarea, clasificadores clásicos de aprendizaje de máquina.

Esta tesis de maestría se sustentó el 13 de enero a las 02:00 pm,
y fue evaluada por los siguientes jurados:
Angel Cruz (Ph.D)
Universidad de los Llanos
Jorge Camargo (Ph.D)
Universidad Nacional de Colombia

Contents

1. Introduction

Diabetic retinopathy (DR) is an ocular disease that affects patients with undiagnosed, untreated, or undertreated diabetes mellitus. DR causes damage to the vessels in the eye inducing leakage of fluid within the retina, exudates, and intraretinal hemorrhages. If the disease progresses it may lead to decreased vision and even blindness [54]. This illness appears in 40% to 45% of diabetic patients at any time of their lives, from which about 5% face vision-threatening complications [20, 7].

An ophthalmologist performs the DR diagnosis through a meticulously visual inspection of the retina according to the ICDR grading system [68], which determines the DR level according to the presence of retinal lesions such as microaneurysms (MA), hemorrhages (H), exudates (EX), cotton wool spots (CWS), intraretinal microvascular abnormalities (IRMA), venous beading (VB), and neovascularization (NV) [15, 54]. The initial clinical signs of DR are MA, that allow leakage of blood from the affected capillaries, they appear as small reddish dots on the superficial layers of the retina. MA tends to have weak walls that can break, leading to H of variable size and shape [45]. Damaged retinal vessels allow lipoproteins precipitates to leak, forming EX that have a yellowish color and an irregular shape. These EX exhibit a characteristic brightness in comparison to the contrast presented in MA and H. CWS are lesions caused by ischemia of the nerve fiber layer in the retina and emerge as yellowish-white spots with irregular edges. Signs of more advanced DR include irregular constriction and dilation of venous vessels in the retina, known as venous beading [14], as well as neovascularization, which appears as red fronds of abnormal vascular networks. These vessels may show up arranged in a radiating pattern or without a distinct pattern [31].

As much as 95% of the cases of vision loss and blindness can be prevented with regular screening and appropriate management [10]. Despite the importance of a timely diagnosis fewer than half of patients with DR are aware of their condition [54]. Due to the asymptomatic nature of the early stages of this disease [41], diabetic patients should be screened frequently with a fundus exam to look for early signs and make an on-time detection [9, 45]. However, access to specialized care by an ophthalmologist is limited for some populations, because most of the ophthalmologists in the world are concentrated in urban areas and big cities. The development of an automated detection system for DR could improve access to specialized care by reducing the time, cost, and effort of screening [51, 43]. In addition, the diabetic population is expected to increase 54% by 2030, while the projected increase of ophthalmologists is only 2% [27]. Thus, the need to integrate automatic detection methods in the screening process responds to these challenges in the present and future panorama of

diabetic retinopathy.

Dealing with the previous issues, we propose a model for the automatic detection of DR that relies on the identification of its ocular lesions. The proposed method is different from state-of-the-art models based on Convolutional Neural Networks (CNNs), because it follows the workflow of the clinicians in the process of DR diagnosis. This approach has the advantage of providing additional information regarding the lesions found in the input image what improves the interpretability of the model. The experimental results show that this strategy also improves the overall classification performance of the system.

1.1. Problem statement

We propose a novel deep learning method for the automatic classification of DR. This model incorporates information about the medical findings that are considered by ophthalmologists in the process of diagnosis. A Computer-assisted diagnosis (CAD) system could lower the workload of ophthalmologists, support during difficult to diagnose cases, and allow more people to access a diagnosis. We identified some challenges in the design and development of a CAD system for DR:

1. Few examples of ocular lesions related to DR in the public datasets with eye fundus images: the available datasets that include labels for lesions related to DR are small, this is due to labeling images at a lesion level is costly, tedious, and time-consuming [43]. There are larger datasets with binary or grade labels for DR; however, they are imbalanced with the normal class as the dominant.

2. The available automatic detection methods are not robust enough: many of the state-of-the-art methods for automatic detection of DR can not be used in real applications because they are unable to generalize.

3. Lack of interpretability: most of the work done on the automatic detection of DR is based on binary or grade classification, few efforts have been done in the improvement of the interpretability of the models. However, in medical applications the user needs to know why the models predict what they predict, to gain trust and confidence in the system [53].

The main drawback of the actual approaches, is that they are specialized in the individual classification of either DR or its ocular lesions. However, information related to the ocular lesions could be useful to ophthalmologists. Regarding this issue, it is proposed to design a deep learning method for the automatic identification of clinically meaningful findings for the automatic diagnosis of DR in retinal fundus images.

Considering the aforementioned, the main research question is:

- How to incorporate medical findings in a deep learning model for identifying diabetic retinopathy in retinal fundus photographs?

1.2. Objectives

1.2.1. General Objective

- To develop a classification model based on deep learning to support the diagnosis of DR using eye fundus images incorporating medical findings.

1.2.2. Specific Objectives

- To develop a deep learning model to identify aneurysms, hemorrhages, cotton wool spots, neovascularization, and venous beating in eye fundus images.
- To develop a strategy to incorporate medical findings into a deep learning model for diabetic retinopathy prediction.
- To systematically evaluate the model on test data sets.

1.3. Main contributions

This work presents a methodology to include ocular lesions information in the DR automatic detection. The main contributions are as follow:

- A new dataset with fine-grained lesions annotations for microaneurysms, hemorrhages, cotton wool spots, venous beading, neovascularization, exudates and global labels for diabetic macula edema, and referable condition for 3209 images from the kaggle EyePACS dataset and 1689 images from the Messidor-2 dataset.
- A strategy that follows the workflow of the clinicians in the DR diagnosis, which increases the interpretability of the model decision. The method for the detection of DR relies on the identification of its related retinal abnormalities and provides both, DR detection along with the classification of 5 lesions used by ophthalmologists to diagnose this illness.
- Exploration of two different strategies to incorporate lesion related to DR in the illness automatic detection.

The following papers have been published or accepted in conferences:

- Toledo-Cortés S., **delaPava M.**, Perdomo O., González F.A. (2020) Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification. In: Fu H., Garvin M.K., MacGillivray T., Xu Y., Zheng Y. (eds) Ophthalmic Medical Image Analysis. OMIA 2020. Lecture Notes in Computer Science, vol 12069. Springer, Cham. https://doi.org/10.1007/978-3-030-63419-3_21

- Perdomo-Charry, O., Pérez-Pérez, A., **delaPava-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

- Accepted in SIPAIM2021 conference: **delaPava M.**, Ríos-Calixto, H., Rodríguez-Alvira, F., Perdomo O., González F.A. (2020). A deep learning model for classification of diabetic retinopathy in eye fundus images based on retinal lesion detection.

2. Related works

In this Chapter we present an overview of the most representative methods associated with the automatic detection of DR using eye fundus images. We classify them into three categories, the classification of DR-related ocular lesions, the detection of DR, and the classification of DR based on the identification of ocular lesions.

The first methods proposed for the detection of DR or its related lesions use classical image processing techniques involving a stage to build a handcrafted feature set. However, the most recent methods are based on CNNs which extract a representation of the images that allow to identify DR patients [71]. In contrast to classical image processing systems, which use predefined features as an intermediate stage for classification, CNNs models can independently extract the appropriate representation and directly classify an image.

2.1. Diabetic retinopathy-related lesions detection

Most of the methods reported regarding the automatic detection of DR-related lesions are focused on the detection of red lesions, namely, MA, H, and EX.

The methods proposed usually extract retinal structures as a preprocessing step. Followed by a two-stage approach, in which a set of potential candidates are identified and later refined using classical image processing methods and hand-crafted features [43].

Alaguselvi and Murugan present an approach that uses the contrast limited adaptive histogram equalization method and the matched filter to remove retinal blood vessels, the optic disc, and noise from eye fundus images. Then a morphology method is applied to identify H, MA, and EX [5]. Similarly, Chen et al., preprocess the images by suppressing the retinal blood vessels and use a multi-scale sparse coding-based learning algorithm to learn the individualized retinal background and identify salient lesions [13]. Alternatively, Amin et al., present the combination of Gabor filter mathematical morphology, statistical and geometric features to detect EX and to grade DR using different ensembles of classifiers [6]. Qiao, Ying, and Zhou preprocessed the images using an enhancement for dark lesions on the edge of curvelet and morphological closures. Then they perform a candidate lesions detection for MA using a matched Filter with a 2-dimensional Gaussian kernel and a candidate extraction maximizing mutual information [48].

Few methods have been published regarding the automatic detection of NV. Hassan et al., describe a method that combines image normalization, compactness classifier, morphology-based operator, Gaussian filtering, and thresholding techniques to highlighted NV in fundus

images [25].

Other works use CNNs for the detection of lesions, which avoids building a set of features, as required by the classical ML classification techniques. Chudzik et al., use transfer learning along with a patch-based CNN with a Dice loss function for the detection of MA, this method requires three stages, namely, preprocessing, patch extraction, and classification [15]. Lam et al., implement a multitask approach where a GoogleNet architecture is used to classify 5 classes, specifically, normal, MA, dot-blot H, EX, or CWS, and high-risk lesions such as NV, VB, and scarring [35]. A dataset of macula-centered retinal fundus images is constructed in [57], it has labels for 12 abnormal findings. This dataset is used to train a network for the classification of each finding, the method also generates a heatmap that highlights their location. Alternatively, Orlando et al., use an ensemble of CNN and classical ML classifiers. They identified MA and H using CNN and hand-crafted features, where lesion candidates are discriminated by a random forest classifier using final augmented features [44].

2.2. Diabetic retinopathy classification

In 2008 were published the earliest methods for the automatic detection of DR using eye color fundus images. These first works usually have three steps: preprocessing, hand-crafted feature extraction and classification. For example in Li et al., perform a preprocessing step based on histogram equalization, morphological operators, and binarization followed by a feature extraction step, in which morphological operations are used to obtain six features namely, red, green, and blue layer of perimeter and layer of area, these features are finally used to train a neural network and classify the images [38].

However, most of the recent methods for the automatic detection of DR involve CNNs. The method proposed by Gayathri et al., uses a CNN to extract features of retinal fundus images and use them to grade DR using classical ML classifiers as support vector machine, AdaBoost, naive Bayes, random forest, and J48 [21]. Wan et al., implement transfer learning and hyper-parameter tuning, using AlexNet, VggNet, GoogleNet, and ResNet architectures for the classification of DR [64]. A similar approach is presented by Ashikur et al., using the AlexNet and GoogleNet architectures [8], and by Gulshan et al., implementing the Inception-v3 architecture for DR grading [23]. Li et al., use CNN models and transfer learning for the binary classification of DR, they explore three different strategies: fine-tuning just some layers of pretrained networks, use the pretrained models as features extractors to train a support vector machine ,and fine-tune all the layers of pretrained networks, leading this last one to the best results [39].

Leibig et al., state that difficult to diagnose cases of DR are unavoidable, so the prediction of the automatic diagnostic methods should be associate with a measure of the uncertainty of the decision. To do so, they evaluate drop-out based Bayesian uncertainty measures in deep learning methods for the binary classification of DR. The authors show that the inclusion of uncertainty in the predictions can improve diagnostic performance [36].

Other approaches include ensembles of CNNs. Qummar et al., present an ensemble of five CNNs, namely, Resnet50, Inceptionv3, Xception, Dense121, and Dense169 to classify the different stages of DR [49]. Gulshan et al., use an ensemble of 10 Inception-V3 networks pretrained on the ImageNet dataset, and fine-tuned them for the binary classification of DR, the authors use the Messidor-2 dataset as the test set, however, they re-labeled privately the images [24]. Voets et al., re-implemented the main method in Gulshan et al., and evaluate it using the available labels of the Messidor-2 dataset, however, they were not able to get the same reported performances, showing the challenges faced when trying to reproduce the results of deep learning methods [63].

Other implementations are inspired in the process of diagnosis of the clinicians. Wang et al., propose a weakly supervised learning framework, called Zoom-in-Net. It is based on CNNs that mimics the zoom-in process of the clinicians to identify lesions in the images, the proposed method classifies DR and highlights suspicious regions [67]. Zeng et al., implement a siamese-like CNN that analyzed images from both eyes of a patient, as a clinician does to diagnose DR [73].

Some methods have been used under real clinical scenarios leading to promising results to integrate them in the DR diagnosis system. Rajalakshmi et al., take retinal images of 301 patients using a smartphone-based device, called Remidio "Fundus on phone" (FOP). The images taken are graded for DR and diabetic macular edema (DME) by a reference standard, the ophthalmologists, and an artificial (AI) DR screening software named, EyeArtTM. After comparing both diagnoses, they found that the AI-based grading algorithm along with smartphone-based imaging can be used to reliably identify patients with DR [50]. Similarly, Abramoff et al., demonstrate that AI methods can bring specialty-level diagnostics to primary care settings, and present the first autonomous AI diagnostic system authorized by the FDA in any field of medicine, which detects more than mild DR and DME [4].

2.3. Diabetic Retinopathy automatic detection using ocular lesions information

In this section, we present an overview of the most representative methods that detect or extract features of ocular lesions for the classification of DR.

The most common approach contains three stages: a first segmentation step, a second step that extracts manually a set of features from DR-related lesions, and a third step that uses these features to classify or grade DR. Sharif et al., present a model that combines independent component analysis with a curve fitting technique to remove retinal blood vessels and the optic disc from eye fundus images. A feature set of DR-related lesions is built, which includes the count number of EX, H, and MA, they also incorporate the mean and standard deviations of candidate regions on the images. Then, a multi-class Gaussian Bayes classifier and a multi-SVM are trained with the feature set to grade DR [56]. Similarly, Abdelmaksoud

et al., use the U-Net network to segment EX, MA, H, and blood vessels. Then a set of features as co-occurrence matrix, areas, and bifurcation points count are calculated to grade DR using a SVM [2]. Paing et al., propose a method that uses histogram matching, morphological opening, and canny method to segment blood vessels, EX and MA. Some features as area and counts are estimated to classify the stages of DR using a customized artificial neural network [45]. Akram et al., implement a method that initially removes blood vessels and optic disc. Then, a set of morphological descriptors such as shape, color, and statistical features are used to train a weighted combination of multivariate m-Mediods and a Gaussian mixture model to identify potential candidates for MA, H, and EX. Finally, the fundus image is graded on a DR scale according to medical conditions that include the type and count of lesions [61].

Other approaches perform the feature extraction skipping the initial segmentation step. Seoud et at. initially detect MA and H using a strategy to extract shape features with morphological image flooding, that discriminate between lesions and other structures. They calculate a set of shape features that do not require segmentation of the candidates and those features are later used for the identification of the candidate regions [55].

All these methods use only red lesions for the detection of DR and most of them implement classical ML methods. However, recently the role of bright lesions for DR grading has also been investigated along with deep learning techniques.

Recent works propose models to automatically perform feature extraction from DR-related lesions to classify the illness. Yang et al., present a two-stages CNN-based algorithm trained with image patches and a weighted lesion map on the input, which can detect DR severity and its related lesions [70]. Wang et al., design a hierarchical multi-task deep learning framework for the classification of the severity of DR and its related lesions such as MA, H, CWS, VB, NV, and others [65]. Similarly, Zago et al., investigate a lesion localization model using two CNNs patch-based approaches. Additional information such as a lesion probability map and the maximum value of a probability map are used to classify DR [71]. Recently, Zhou et al., partially released pixel-level and image-level fine-grained annotations of lesions related to DR named the FGADR dataset. They reported individual modules for the segmentation of lesions to extract features that are integrated into a grading model to improve DR classification results [76].

Our model is different from the previous works because:

1. It does not require the segmentation of retinal structures or lesions as a preprocessing step.

2. Our method follows the same opthalmologist workflow in which initially A, H, CWS, VB, and NV are identified to later diagnose DR. Our model provides both lesions and DR classification.

3. In many previous works, the test set in unclear, private, or the DR labels from public datasets are re-labelled privately. Our DR detection model is evaluated in the Messidor-

2 dataset with the already publicly available labels and all the lesions labels used to train the models will be released.

3. Datasets

To perform the experiments in this work, we build a dataset that provides labels for DR-related lesions. To be able to compare our performance with state-of-the-art results we selected frequently used datasets for DR detection and annotated DR-related lesions on them. This Chapter presents an overview of the existing publicly available datasets with DR and DR-related lesions annotations, the detailed description of our dataset, and its comparison with the currently available.

3.1. Diabetic retinopathy grading datasets

We selected two datasets to be partially labeled with ocular lesions related to DR. The selection is made between the currently publicly available datasets with eye fundus images and DR annotations used in the state-of-the-art to train and evaluate models for the DR automatic detection. An overview of these datasets is presented briefly as follows:

Kaggle-EyePACS [32]: it is the largest set of high-resolution retina images with 35,126 training images and 53,576 testing images taken under a variety of imaging conditions, some images contain artifacts, are out of focus, underexposed, or overexposed. The images are labeled with left or right labels and noisy grades on a scale of 0, 1, 2, 3, 4, which stand for no DR, mild, moderate, severe, and proliferative DR respectively.

Kaggle-APTOS2019 [59]: this dataset has 3,662 training images with public labels for DR on a scale of 0 to 4 and 1,928 testing images with private labels. The images were taken under a variety of imaging conditions, some images contain artifacts, are out of focus, underexposed, or overexposed light.

DIR-5K [42]: this dataset was build analyzing 5,000 patients. The images present a range of varied image resolutions. Trained human readers under quality control management annotated eight labels including normal, diabetes, glaucoma, cataract, AMD, hypertension, myopia, and other diseases/abnormalities.

Messidor [29]: it contains 1200 eye fundus color images of the posterior pole taken with the same camera with a 45-degree field of view, the sizes of the images vary between 1440×960, 2240×1488 or 2304×1536 pixels. It provides DR grading on a scale of 0 to 3, and grades for risk of macula edema on a scale of 0 to 2.

Messidor-2 [30]: it is an extension of the Messidor dataset. It contains 529 examinations (1058 images) from the Messidor dataset and 345 examinations (690 images) never-before-published. Overall, the Messidor-2 dataset contains 1748 macula-centered eye fundus images

that came in pairs and were acquired with a 45-degree field of view with sizes between 1440×960 and 2304×1536 pixels. The dataset provides a spreadsheet with the image pairing, but, it does not contain annotations for DR. However, Abramoff et al., released on its webpage the images with a binary classification for DR, which is used in the experiments here performed [3].

DDR [37]: this dataset has 13673 fundus images, from which 6266 are healthy and 6256 have DR. The images in this dataset have DR grading (no DR, mild DR, moderate DR, severe DR, proliferative DR), pixel-level annotations for 4 different lesions and gradable labels. It has 6835 training images (6320 are gradable), 2733 validation images (2503 are gradable) and 4105 test images (3759 are gradable).

3.2. Diabetic retinopathy lesions datasets

We present an overview of the datasets that include ocular lesions related to DR and compare them with the proposed dataset to clarify its contributions.

DRIVE [58]: the Digital Retinal Images for Vessel Extraction dataset was made publicly available in 2004. It has 40 color fundus photographs acquired with 45-degree field of view with 768×584 pixels, from which, 7 contain pathologies, namely exudates, hemorrhages, and pigment epithelium changes. The dataset includes the segmentation of the vessels in all the images, which are JPEG compressed.

ImageRet [34, 33]: this dataset is subdivided into two sub-databases, DIARETDB0 with 130 images and DIARETDB1 with 89 images, which were made publicly available in 2008. The images were acquired with a 50 field of view at a size of 1500¯1152 pixels in PNG format. The images were annotated for 4 DR-related lesions, namely, microaneurysms, hemorrhages, and hard and soft exudates.

STARE [28]: this dataset consists of 400 images with labels for 39 possible manifestations. It also provides, blood vessel segmentation for 40 images, two artery/vein labelings of 10 images, and optic nerve segmentation for 80 images. The data can be found in [22].

IDRiD [40]: this dataset provides data to perform three different tasks: segmentation, classification, and localization:

- Segmentation: it has 81 images with segmentation for microaneurysms, hemorrhages, hard exudates, and soft exudates. It also includes segmentation of the optic disk.

- Disease grading: it has 516 images divided in the train (413 images), and the test set (103 images). It provides severity grades for DR, and DME.

- Localization: it has 516 images divided into the train set (413 images) and the test set, with optic disc center location and fovea center location.

FGADR [75]: is a large-scale fine-grained annotated DR dataset with 2,842 images, which is partially available to the date. It is divided into two sets with pixel-level DR-related lesion annotations and image-level labels:

- Seg-set: contains 1,842 images with both pixel-level lesion annotations and image-level grading labels for DR in a range 0 to 4, annotated by three ophthalmologists. The lesions annotated are microaneurysms (MA), hemorrhages (H), hard exudates (EX), cotton wool spots (CWS), intra-retinal microvascular abnormalities (IRMA), and neovascularization (NV). It also provides image-level labels for laser mark (LM) and proliferate membrane (PM).

- Grade-set: contains 1,000 images with grading labels on range 0 to 4 for DR, annotated by six ophthalmologists, which gives it high annotation confidence.

e-ophtha [18]: it is made of two sub-databases with pixel-level annotations:

- e-ophtha-MA: it has 148 binary masks for microaneurisms or small hemorrhages and 233 images with no lesion.

- e-ophtha-EX: it contains 47 images with binary masks for exudates and 35 images with no lesion.

Each folder in the dataset corresponds to a patient, which may have more than one color fundus image in jpeg format.

DRiDB [47]: this dataset has 50 images with annotations for both, pathologies and normal eye fundus structures. The experts segmented areas related to microaneurysms, hemorrhages, neovascularizations, hard and soft exudates. They also marked the blood vessels, optic disc, and macula. The dataset also provides grading for DR for each image in the dataset.

DDR [37]: The dataset include segmentation and labels for 4 different lesions. This dataset has 13673 fundus images from which 757 have at least one lesion. There is a total of 601 images with hemorrhages, 570 with microaneurisms , 239 have soft exudates and 486 present hard exudates.

3.3. Diabetic Retinopathy Ocular Lesions Labels (DROLL) dataset:

Most of the available datasets have either grading/binary labels for DR or labels for some ocular lesions. The FGADR dataset is the only one that provides both, however,it is only partially available to the date.

Since our objective is to classify DR using lesions information, we decide to select two datasets employed in the literature for the detection of DR, to create and release a large-scale dataset with fine-grained labels of six ocular lesions and two diagnostic labels. The kaggle-EyePACS is selected since it is the largest available and therefore it could contain more images with lesions examples. The Messidor-2 is also selected since it includes most of the images from the Messidor dataset and both are widely used to evaluate models in the literature. However, we use the available DR labels to evaluate our models so our performances could be compared with state-of-the-art results. The construction details of the DROLL dataset are explained as follows:

1. *Dataset construction:* an ophthalmologist with supra-specialty in the retina from the Fundación Oftalmológica Nacional from Colombia selected 3209 retinal images from the kaggle-EyePACS dataset. The ophthalmologist performed the manual choice of these images based on the quality of fundus images and the presence of DR-related lesions. The Messidor-2 dataset was also analyzed by the specialist who determined that 1689 images are suitable for lesion level annotation. The expert labeled the images with six ocular findings: aneurysms, hemorrhages, cotton wool spots, venous beading, neovascularization, and exudates. In addition, these images were manually annotated with binary labels for referable and non-referable patients and the grade of DME on a scale of 0 to 3 for no DME, mild, moderate, and severe.

2. *Annotation criteria:* the whole images were evaluated in their entire area as follows: first, a meticulous analysis looking for any lesions on the optic nerve and the macular region is performed. Then, the extra-macular retina and outside the vascular arches areas are examined in detail to find possible DR-related lesions. Finally, each image was classified in a fine-grained way as positive for that specific finding if at least one ocular lesion was identified. Regardless of whether it was a single injury or many, or whether it affected the center of the macula or its periphery. In addition, the images were labeled as DME in an image-level way according to the ETDRS scale. Also, images were classified as a referable image if at least an ocular lesion was present in all image areas, regardless of which one.

The DROLL dataset is subdivided into two subsets:

- DROLL kaggle-EyePACS set: it has fine-grained labels for 3209 images from the kaggle-EyePACS dataset.

- DROLL Messidor-2 set: it has fine-grained labels for 1689 images from the Messidor-2 dataset.

The models for the detection of findings in this work are trained using the DROLL kaggle-EyePACS set and tested using the DROLL Messidor-2 set.

A detailed description of the labels of ocular lesions in the sets of the DROLL dataset is presented in Table **3-1**. In this Table is shown that the DR-related findings MA, H, and CWS are the most common lesions in the DROLL dataset, due to these ocular findings appear in the initial stages of subjects with DR. On the other hand, VB and NV have few examples in the dataset, these lesions are scarce to find because they only appear in the advanced grades of DR. Finally, since EX are used to grade the risk of DME both conditions have the same number of examples. The kaggle-EyePACS set was selected trying to include as many ocular findings as possible, this is reflected in the proportion of referable examples, which implies that almost 80% of the images have at least one ocular lesion.

The Table **3-2**, presents the description of the labels in the DROLL dataset for DME, DME grades (DME1, DME2, DME3), referable and DR, the DR labels analysed in this table, are the publicly available corresponding to the images labeled by the FON experts. The kaggle EyePACS DROLL set has a bigger proportion of examples of this conditions than the Messidor-2 DROLL set. The referable condition is the most commonly found in the sets and most of DME examples belong to the grade 2.

Table 3-1.: Description of DR-related ocular lesions labels from the customized Kaggle EyePACS (Training test) and Messidor-2 (Test set) data sets.

	Ocular lesions labels					
DROLL set	MA	H	CWS	VB	NV	EX
Kaggle EyePACS	1728 53.83%	1427 44.45%	656 20.44%	722 22.49%	142 4.42%	1821 56.44%
Messidor-2	876 51.86%	443 26.22%	133 7.87%	47 2.78%	15 0.89%	196 11.60%

Table 3-2.: Description of the labels for DME, DR and referible conditions from the customized Kaggle EyePACS (Training test) and Messidor-2 (Test set) data sets.

	Image-level labels					
DROLL set	REF	DME	DME1	DME2	DME3	DR
Kaggle EyePACS	2559 79.72%	1821 56.7%	419 13.05%	1122 34.95%	280 8.72%	2016 62.80%
Messidor-2	893 52.87%	196 11.61%	44 2.61%	102 6.04%	50 2.96%	377 22.32%

We compare in Table **3-3** the DROLL dataset with others publicly available with fine-grained labels of DR-related ocular lesions. It can be noticed that the DROLL dataset is the largest currently available with lesion-level labels and it is the only one that includes all the DR-related lesions used by the ophthalmologist to diagnose DR. We hope that this dataset

opens opportunities to develop new approaches for the classification of retina lesions and to improve interpretability in the DR detection models, leading to more reliable methods for the clinicians.

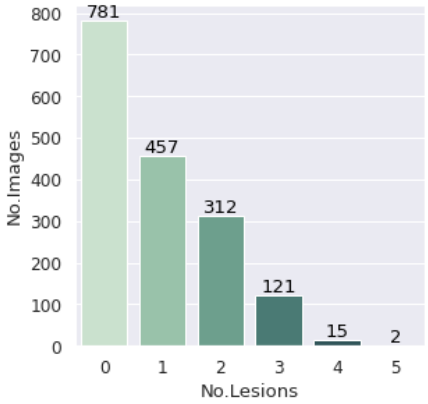
Table 3-3.: Comparison of ocular lesion labels between the DROLL dataset and public dataset. pigment epithelium changes (PEC), intraretinal microvascular abnormalities (IRMA).

Dataset	MA	H	CWS	VB	NV	EX	PEC	IRMA	Images	Imgs with lesions
DRIVE		X				X	X		40	7
DIARETDB0	X	X	X		X	X			130	110
DIARETDB1	X	X	X			X			89	84
STARE	X				X				400	361
IDRiD	X	X	X			X			81	81
FGADR	X	X	X		X	X		X	1842	1842
e-ophtha	X					X			463	195
DRiDB	X	X	X		X	X			50	
DDR	X	X	X			X			13673	757
Our dataset	X	X	X	X	X	X			4899	3075

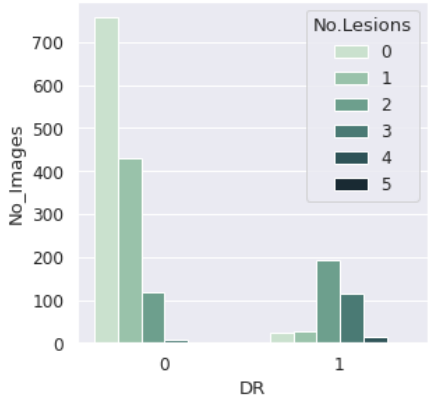
Fig. **3-1** a) presents the number of images in the Messidor-2 DROLL set that are annotated as healthy (represented with the number 0) or have one or multiple of the following ocular lesions: A, H ,CWS, VB and NV. This figure shows that most of the images with ocular lesions in the Messidor-2 DROLL set have one or two of them at the same time, examples with up to 3 different lesions are more scare. This is due to it is difficult to find patients with such an advance level of DR. In Fig. **3-1** b) the public labels for DR of the Messidor-2 dataset are analyzed along with the DROLL lesions labels, most of the images annotated as healthy in the public labels, have either, 0 or 1 lession and almost none have more than 3 different lesions. Consistently, most of the images annotated as DR patients in the public labels, present up to 2 different lesions according to the DROLL annotations. Fig. **3-1** c) presents the correlation matrix between the different lesions. It shows that the patients that have the latter DR findings, like VB or NV also present the early signs of the illness such as A and H.

Fig **3-2** a) shows the number of images in the kaggle EyePACS DROLL set that are annotated as healthy or have one or multiple of the following lesions: A, H ,CWS, VB and NV. In contrast with Fig. **3-1** a), this set has a more even number of patients with 1, 2 and, 3 different ocular lesions. Fig. **3-2** b), shows a similiar behavior to Fig. **3-1** c) in which patients with letter DR signs also present the early lesions of the illness. In Fig. **3-2** c) the public DR labels of the kaggle EyePACS dataset are analyzed along with the DROLL lesions labels, it is shown that most of the healthy patients have maximum one ocular DR-related lesions and

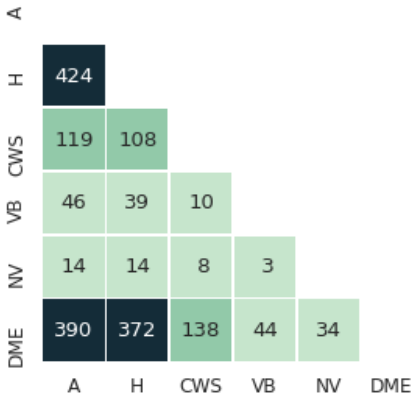
all the images with up to 3 different DR-related lesions are labeled as DR patients. Fig. **3-2** d) shows the number of different lesions in patients with different grades of DR, it can be noticed that the higher the degree of DR, the greater the predominant number of ocular DR-related lesions in the images.



(a) Number of images with 0 up to 5 different lesions.

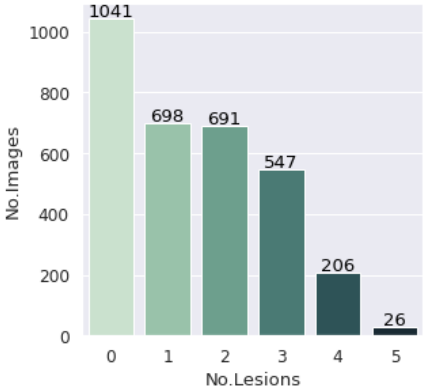


(b) Number of healthy and diabetic retinopathy images with 0 up to 5 different lesions.

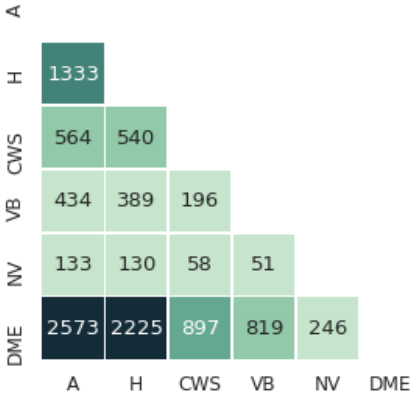


(c) Count number of pairwise cases for each lession in the DROLL dataset.

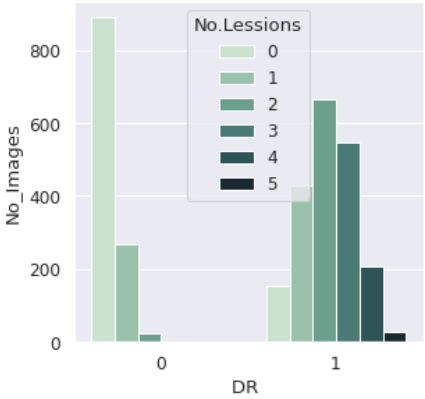
Figure 3-1.: Description of the number of images with ocular lesions in the Messidor-2-DROLL set.



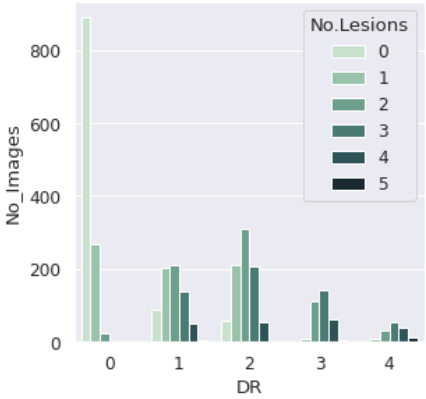
(a) Number of images with 0 up to 5 different lesions.



(b) Count number of pairwise cases for each lession in the DROLL dataset.



(c) Number of healthy and diabetic retinopathy images with 0 up to 5 different lesions.



(d) Number of images with 0 up to 5 different lesions by diabetic retinopathy grade.

Figure 3-2.: Description of the number of images with ocular lesions in the kaggle EyePACS-DROLL set dataset.

4. Detection of diabetic retinopathy lesions in eye fundus images.

Aiming to do the automatic detection of DR more relatable to the actual process clinicians do, we initially attempt to identify ocular lesions related to DR. This Chapter presents a transfer learning strategy that uses pre-trained CNNs as feature extractors and classical ML methods as classifiers. Alternatively, we train a multitask method that predicts simultaneously ocular lesions, and DR. We also apply t-SNE to visualize the features used to detect the lesions. All the methods are trained using the kaggle EyePACS DROLL set and evaluated using the Messidor-2 DROLL set.

4.1. Lesions detection using convolutional neural networks and shallow machine learning models

Conventional ML techniques are limited in image recognition applications because the data on its raw form needs to be mapped into a set of features insensitive to some variations on the input that are irrelevant in the discrimination process, but sensitive to other characteristics that some times are minor but relevant to discern between classes. Traditionally, domain experts have had to build handcrafted features to train ML methods, however, this can be avoided using a general-purpose learning procedure like deep learning, which allows learning automatically a representation from raw data [69, 43].

Deep learning is a representation learning method, with multiple levels of representations that can learn complex functions by applying multiple non-linear transformations consecutively. The deeper layers of a deep learning model learn general descriptors, while the higher layers are capable to identify elements important in the discrimination of different classes in high-dimensional data [17]. The first implementations include networks with full connectivity between adjacent layers, however, it was found that CNNs are easier to train and generalize better. CNNs have been used in computer vision applications in the past, but, they have gain more popularity since the ImageNet competition in 2012, in which they were used to classify 1000 different classes using a dataset of about a million images, along with efficient use of graphics processing units (GPUs), dropout regularization, and effective data augmentation [60].

Nevertheless, training a CNN from scratch has some challenges. Since a large number of

weights in a network are randomly initialized, a limited number of training examples can lead to a local minimum of the cost function causing an unwanted performance. That is why CNNs require a large training dataset, what in turn leads to the need for extensive computational and memory resources. However, large datasets are not always available, for example in the medical domain, where expert annotations are costly, tedious, and time-consuming. Other scenarios that make it difficult to train a model from scratch are the unbalance of available datasets, and the overfitting problems which make it necessary to iteratively adjust the network parameters, which is tedious, time-consuming, and demands expertise [60, 43]. To tackle this issue some common practices include:

- Fine-tune: it is a training strategy in which the weights of a network trained for a specific task with a large amount of labeled data are used to initialized other network with the same architecture. This network is taken as starting point of a new training process for a different task. The last fully connected layer of the pre-trained network is usually replaced with another that matches the number of classes for the new task. However, a larger number of randomly initialized layers and regularizations can be explored. Initially, only the layers randomly initialized are trained and then deeper layers are also tuned. A common practice that leads to good results is to only fine-tune the last layers of the network due to the early layers of a CNN learn general image features and subsequent layers learn more specific image features. Nevertheless, it can be necessary to fine-tune the entire network, if the applications are very different [60].

- Transfer learning: it is a transmission process that modifies the knowledge learned from a source domain to improve the prediction performance of a target domain that only has a small training dataset [46]. In this method, the last fully connected layer of a CNN trained for a general task is deleted and the remaining network is viewed as a feature extractor. This new configuration is used to extract features from a new dataset to train a classifier for a new task [64].

The pipeline of our proposed method for the detection of DR-related lesions is depicted in Figure **4-1**. The lesion detection is based on a transfer learning strategy in which a single network is used as a feature extractor to train independent classifiers for the detection of each finding.

Initially, the eye fundus images are preprocessed to resize them to 512×512 pixels and remove the black edges. Then we implement transfer learning using:

- An Inception-V3 pretrained on ImageNet and fine-tuned on the DR detection. The training details and the results are reported in Chapter 5.

- The pre-trained backbone models for DR detection reported by He et al. [26]. The authors implement an embedding of two attention blocks into different backbones networks, namely, DenseNet121, Xception, ResNet50, and MobileNet. He et al. train

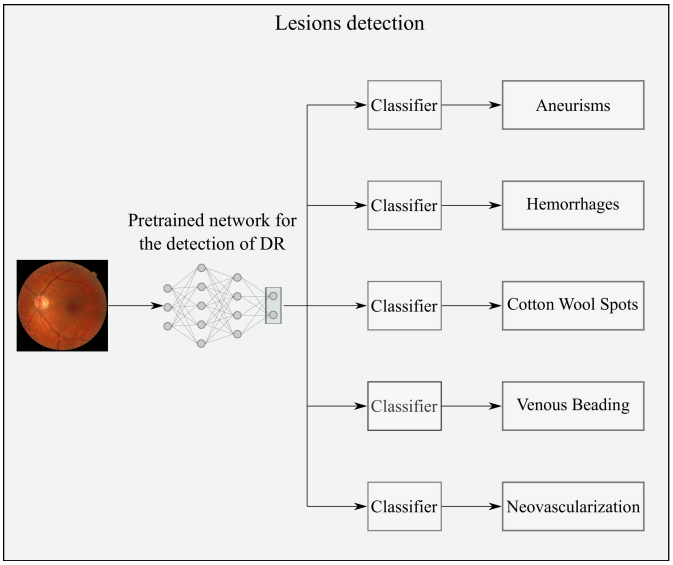


Figure 4-1.: An overview of the proposed method for ocular lesions detection.

their networks using two different datasets, specifically, the training set of the Kaggle EyePACS dataset and the DDR dataset. The files of the models are named accordingly to the backbone and the dataset used as training set. The ones used as feature extractors in the experiments in this work are: xception_CAB_EyePACS.h5, xception_CAB_DDR.h5, resnet50_CAB_EyePACS.h5, densenet121_CAB_EyePACS.h5 and densenet121_CAB_DDR.h5.

Finally, multiple shallow classifiers are trained for the detection of DR-related lesions using the features extracted from each of the previous CNNs and the fine-grained labels of the Kaggle-EyePACS DROLL set. The best overall performing method is selected as the classifier for the detection of lesions.

The classifiers and the parameters explored for the automatic detection of DR-related lesions are:

1. Support vector machine (SVM): the parameters explored are the regularization parameter (C), the kernel, and the gamma (γ) kernel coefficient.

2. Gaussian Process (GP): the kernels "RBF" and "Mattern" are evaluated. The kernel parameters explored are lower length scale bound, upper length scale bound, lower noise level, and upper noise level.

3. Multilayer Perceptron (MLP): the number of layers, number of neurons, activations and learning rate are explored.

4.2. Lesions detection using a multi-task model

Multi-task is a training method that uses a shared representation to train machine learning models to perform multiple related tasks simultaneously [16]. It is a kind of inductive transfer in which a model can be improved by an inductive bias introduced by auxiliary tasks that help the model to generalize better, which can help to avoid overfitting on one single task [52]. However, the training process of this kind of method can face negative transfer or destructive interference, which means that the learning tasks have conflicting needs, so improve the performance of one of them leads to worsening the performance in the other [16].

Multi-task learning methods have been classified into two main groups:

- Hard parameter sharing: it is the most conventional strategy used along with neural networks. In this approach, the same hidden layers of a model are shared and used as a feature extractor. Then individual output layers for each task are build to perform multiple tasks. This outline is called *shared trunk* [52, 16].

- Soft parameter sharing: in this approach, each task has its task-specific model, however, the parameters are kept similar using a joint objective function. It can be seen as a kind of multitask optimization [16].

There are multiple optimization methods in literature to train a multi-task model. However the most commonly used is the *loss weighting*. In this strategy the individual loss functions for different tasks are combined into an aggregated single one, usually as a weighted sum of the task-specific loss functions [16].

We implement a multi-task approach using a shared trunk structure and a loss weighting loss function for the integration of lesions information on the DR automatic detection. The overall structure of the method is shown in Figure **4-2**.

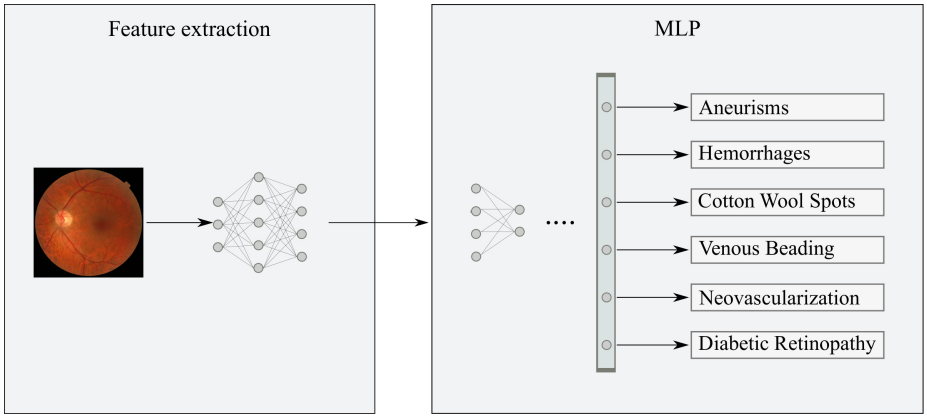


Figure 4-2.: Multi-task model for the detection of DR and its related ocular lesions.

We implemented as feature extractor the same architectures mentioned in Section 4.1. We use as classifier a MLP using fully connected layers and dropout layers in between. The last

fully connected layer has six neurons, one to classify each lesion and an additional one for the prediction of DR. We include class weights and explored the following hyperparameters during the network design process: number of layers, number of neurons, learning rate, and dropout. A weighted binary cross-entropy is used as the loss function. In specific, the weights were calculated based on the number of examples of each ocular lesion.

4.3. Results

We perform a systematic exploration and evaluation of the SVM, GP, MLP and multi-task methods for the detection of DR-related lesions using the features obtained from each of the following pre-trained models: xception_CAB_EyePACS.h5, xception_CAB_DDR.h5, resnet50_CAB_EyePACS.h5, densenet121_CAB_EyePACS.h5, densenet121_CAB_DDR.h5, vgg16_CAB_EyePACS.h5, vgg16_CAB_DDR.h5 and the fine-tuned Inception-V3. After an initial exploration, it is found that the best overall performing methods published by [26], are the ones trained using the EyePACS dataset. A more exhaustive exploration is made using only the CNNs trained using this dataset, all the comprehensive results are presented in Anexo A and the AUC for the best performing classifier for each CNN explored is presented in Table **4-1**.

Table 4-1.: The best classifier and AUC for each backbone explored in the automatic detection of ocular lesions.

Finding		xception	resnet50	densenet121	vgg16
MA	AUC	0.7837	0.7820	**0.8055**	0.8003
	Classifier	MLP	MLP	MLP	MLP
H	AUC	0.9303	0.9129	**0.9370**	0.9284
	Classifier	MLP	MLP	GP classifier	Multi-task
CWS	AUC	0.8904	0.8622	**0.9021**	0.8633
	Classifier	MLP	MLP	MLP	MLP
VB	AUC	0.8496	0.7006	**0.8538**	0.8172
	Classifier	MLP	Multi-task	Multi-task	Multi-task
NV	AUC	0.9651	0.9693	**0.9693**	0.9464
	Classifier	MLP	GP regressor	GP regressor	Multi-task

The best results are obtained using the backbone in the densenet121_CAB_EyePACS model, which is selected as feature extractor for the lesion detection task. The detailed results of the detection of MA, H, CWS, VB, and NV using this model are reported in Table **4-2**, where the best overall performing classifier is the MLP. The findings MA and VB presented greater difficulty to be detected than H and NV. This may be because, sometimes H are big and NV are usually accompanied by other lesions since it appears in the latter

stages of the disease, which can help the model to detect them. The predictions obtained using the backbone in the densenet121_CAB_EyePACS model as feature extractor and the MLP as classifier are used in the design of the DR detection model in Chapter 5.

Table 4-2.: Results of the detection of DR-related ocular lesions using the densenet121_CAB_EyePACS backbone model as a feature extractor. The parameters reported are sensitivity (Se), specificity (Sp) and threshold (tr).

Finding	Metric	GP classifier	GP regressor	SVM	MLP	Multi-task
MA	AUC	0.8011	0.7865	0.7516	**0.8034**	0.7973
	Sp	0.7109	0.7023	0.8733	0.6974	0.7011
	Se	0.7043	0.7141	0.5833	0.7146	0.7226
	tr	0.38	0.25	0.05	0.19	0.25
H	AUC	**0.9370**	0.9332	0.8503	0.9340	0.9273
	Sp	0.8451	0.8472	0.9775	0.8459	0.8603
	Se	0.8690	0.8600	0.7088	0.8804	0.8419
	tr	0.38	0.23	0.05	0.26	0.37
CWS	AUC	0.8492	0.8825	0.6884	**0.8950**	0.8833
	Sp	0.7667	0.7622	0.9485	0.8219	0.8271
	Se	0.7594	0.8571	0.4135	0.8496	0.8045
	tr	0.32	0.13	0.05	0.3	0.4
VB	AUC	0.7525	0.7299	0.6432	0.8084	**0.8538**
	Sp	0.6461	0.6565	0.8526	0.7618	0.7728
	Se	0.6808	0.6808	0.455	0.6808	0.8085
	tr	0.34	0.23	0.2	0.64	0.48
NV	AUC	0.9401	**0.9693**	0.6205	0.9650	0.9516
	Sp	0.8393	0.9074	0.9164	0.8900	0.9068
	Se	0.8666	0.9333	0.3333	0.9333	0.9333
	tr	0.25	0.1	0.5	0.45	0.57

In Figure **4-3** are presented more details of the results on the detection of ocular lesions using the densenet121_CAB_EyePACS model as feature extractor and the MLP as classifier. The images on the left present the distribution of the predictions of the models, it can be seen that the model achieves to predict the images with the lesion closer to 1 and the images without the lesions closer to 0. In the images in the middle, is shown how multiple metrics of the methods for the detection of each finding change with the threshold. The intersection between the curves of the sensitivity and specificity is used as the threshold of the model and corresponds to the one reported in Table **4-2**. In the same Figure, in the last column is also shown the confusion matrix for each method for the detection of ocular lesions.

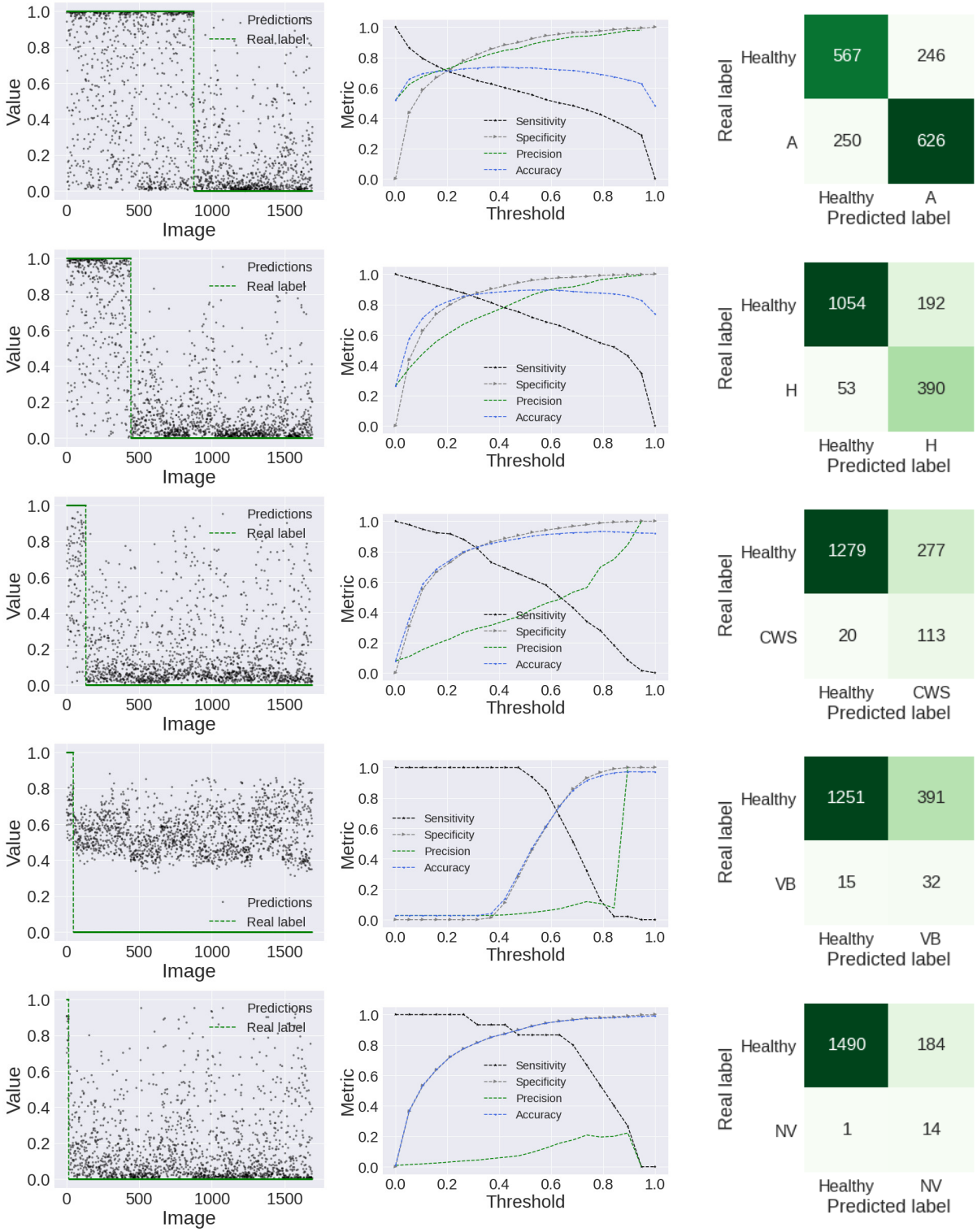


Figure 4-3.: Predictions visualization, threshold analysis and confusion matrix using a MLP as classifier and densenet121_CAB_EyePACS backbone as feature extractor.

Figure 4-4.: Visual representation of the images with and without microaneurysms (MA), hemorrhages (H), cotton wool spots (CWS), venous beading (VB), neovascularization (NV) in the Messidor-2 DROLL set using the t-SNE technique.

4.3.1. Graphical representation

The T-distributed Stochastic Neighbor Embedding (t-SNE) strategy is applied to each ocular finding using the extracted features obtained using the backbone in the densenet121 _CAB_EyePACS model. This strategy helps to visualize the models in a 2D space, getting a better representation to improve the interpretability of the model.

The principal component analysis (PCA) is applied to perform a dimensionality reduction to 15 of the features obtained from the Messidor-2 dataset, which is used to implement the t-SNE method and get a two-dimensional map of the features, as shown in Figure **4-4**. MA and H present a better visual cluster separation than CWS and VB. The CWS and VB representations show an overlap of the negative over the positive classes. The few NV examples limited a proper qualitative analysis of the ability to separate the data using the t-SNE representation.

5. Diabetic Retinopathy detection using lesions information.

Ophthalmologists diagnose DR based on the presence or absence of ocular lesions, namely, microaneurisms, hemorrhages, cotton wool spots, venous beading, and neovascularization. In this Chapter we present a strategy that follows their workflow, by integrating lesion information into two DR automatic detection methods:

1. In Section 5.1, we use the lesions classification model from Chapter 4 to build a feature set by concatenating the prediction of each lesion, and train a DR detection method.

2. In Section 5.2, we develop a multitask approach in which DR is predicted along with its related ocular lesions.

We also propose two base-lines to compare the performance of our approaches for the detection of DR:

1. We fine-tune the Inception-V3 architecture pre-trained on ImageNet using the kaggle EyePACS training set. The last fully connected layer is replaced by a randomly initialized MLP with dropout. Multiple configurations of the MLP are evaluated using the features obtained from the Inception-V3 network pretrained on ImageNet. Specifically, the number of layers, number of neurons, activation, and learning rate are analyzed. The best-performing architecture is used to replace the original last fully connected layer in the pretrained model. In the final training process, we froze the architecture of the Inception-V3 network, in such a way that only the new elements are trained for 10 epochs, then the latter half of the architecture is tuned for the new task, preserving the early half of the weights of the model. The results are presented in Table **5-2**, under the Method name 'Fine-tuned Inception-V3'.

2. We evaluate the complete architectures reported by [26] and used as feature extractor in Chapter 4 on the Messidor-2 dataset. The performance of each model is reported in Table **5-2**.

5.1. Lesions predictions for DR detection

The pipeline of the proposed method for the automatic detection of DR using lesions information is depicted in Figure **5-1**. The method has two main stages, the first one is lesion

detection, which is described in Section 4.1 and the second one is the DR classification, presented here. In the second stage, training and test feature sets are build by concatenating the predictions of the lesion detection model over the kaggle EyePACS DROLL set and the Messidor-2 dataset respectively. In this way each image in the datasets is represented by a 5-dimensional vector. The training feature set is used to train different classifiers for the binary classification of DR. The same classifiers (SVM, GP, and MLP) and parameters described in Section 4.1 for the detection of DR-related lesions, are also explored for the DR automatic detection. The best results are reported and compared with state-of-the-art methods using the Messidor-2 dataset as the test set.

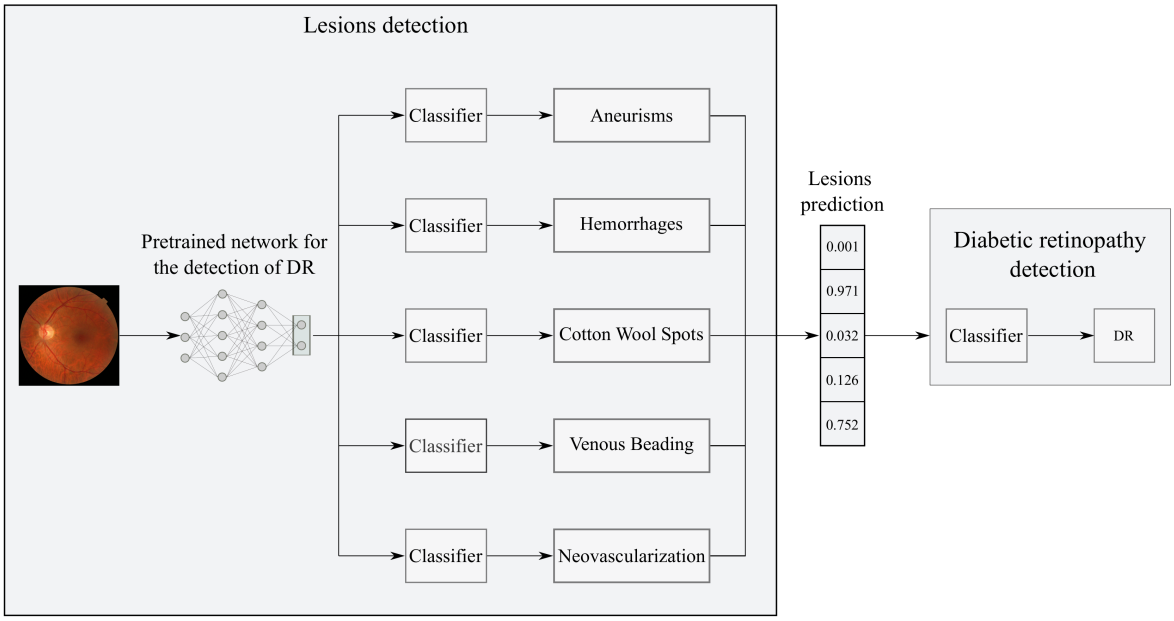


Figure 5-1.: An overview of the proposed method for ocular lesions detection and DR classification. The model is organized in three consecutive stages as follows: lesion detection (first block), lesion prediction (second block) and DR detection (third block).

The lesions detection model is made up of the backbone in the densenet121_CAB_EyePACS model as feature extractor and the MLP as the classifier. Our model is intended to improve the interpretability of the automatic DR detection model, however, it is important to determine if our approach affects the performance of the model by adding the intermediate step for lesion detection. So, for a fair comparison, we tested the models from He et al. in the same test set used to evaluate our approach, the Messidor-2 dataset. It is important to note that we did not reannotate the labels for DR of the Messsidor-2 dataset, all the DR detection models are evaluated with the already publicly available annotations so they can be compared with state-of-the-art methods.

5.2. Multitask model

The performance of the detection of DR of the multi-task method described in the Chapter 4 in Section 4.2, is reported in this Chapter.

5.3. Results

Once the training and test features sets are built, the models SVM, GP, MLP, and multi-task are explored and evaluated for the binary prediction of DR. The results are presented in Table **5-1**. The best performing classifier, namely, MLP, is compared with the baseline methods and state-of-the-art approaches in Table **5-2**.

Classifier	AUC	Specificity	Sensitivity	Threshold
SVM	0.9393	0.9794	0.6737	0.05
GP classifier	0.9392	0.8544	0.8674	0.44
GP regressor	0.9392	0.8582	0.8621	0.45
MLP	0.9482	0.8750	0.8859	0.38
Multitask	0.9479	0.8783	0.8859	0.36

Table 5-1.: DR detection using support vector machine (SVM), Gaussian process classifier (GP classifier), Gaussian process regressor (GP regressor) and multi layer perceptron (MLP)

The proposed method trained with the kaggle EyePACS DROLL set and evaluated in Messidor-2 shows competitive performance in AUC and consistent results in sensitivity and specificity when compared to the baseline models trained for the DR detection task as shown in Table **5-2**. Despite some state-of-the-art methods have better performances, they only classify DR, unlike our approach that also provides information about the DR-related lesions, which provides interpretability and makes the DR detection process more familiar for clinicians, since it follows a workflow similar to their own.

Table 5-2.: Comparison of proposed method with state-of-the-art DR detection methods tested on Messidor-2 dataset. The top 3 results for each metric are marked in bold$^{(1)}$, italic$^{(2)}$, and underline$^{(3)}$ respectively.

Training set (Set of images)	Method	AUC	Sensitivity	Specificity
EyePACS train (35126)	densenet121_CAB_DDR [26]	0.765	0.359	**0.990**$^{(1)}$
EyePACS train (35126)	densenet121_CAB_EyePACS [26]	0.782	0.395	*0.954*$^{(2)}$
EyePACS train (35126)	resnet50_CAB_EyePACS [26]	0.884	0.753	0.801
EyePACS train (35126)	xception_CAB_DDR [26]	0.908	0.877	0.861
EyePACS train (35126)	xception_CAB_EyePACS [26]	0.908	<u>0.880</u>$^{(3)}$	0.855
EyePACS train (35126)	Fine-tuned Inception-V3	0.8447	0.8011	0.7023
EyePACS custom (57146)	Voets et al. [62]	0.800	0.737	0.697
EyePACS custom (28102)	Zhou et al. [74]	**0.960**$^{(1)}$	-	-
EyePACS custom (75137)	Gargeya and Leng [20]	<u>0.940</u>$^{(3)}$	**0.930**$^{(1)}$	0.870
IDRiD	Zago et al. [71]	0.944	0.900	0.87
EyePACS custom (3209)	densenet121_CAB_EyePACS backbone and MLP	*0.9482*$^{(2)}$	*0.8859*$^{(2)}$	<u>0.875</u>$^{(3)}$
EyePACS custom (3209)	Multitask	0.9479	0.8783	0.8859

5.3.1. Graphical representation

Initially, using the principal component analysis technique we perform a dimensionality reduction to 3, which is used to implement the t-SNE method and get a two-dimensional map of the features. The result is shown in Figure **5-2**, where the DR and no DR examples have few overlapping.

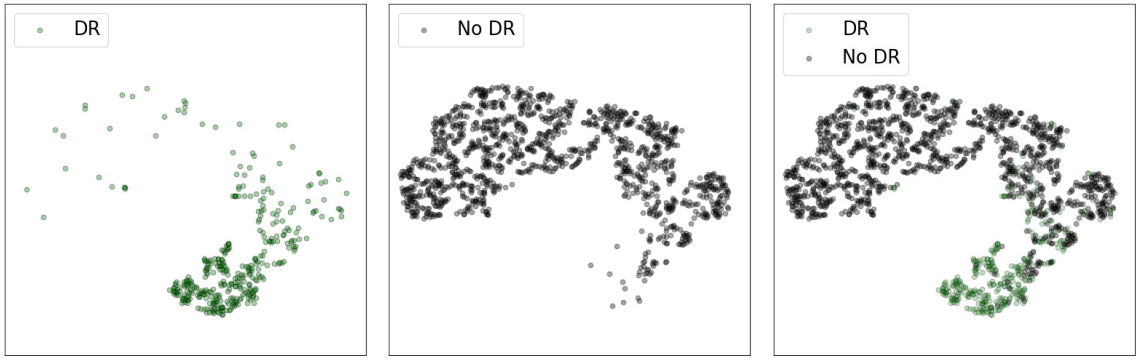


Figure 5-2.: Visual representation of the images with and without diabetic retinopathy (DR) in the Messidor-2 DROLL set using the t-SNE technique.

6. Diabetic macular edema and referible cases classification

Although it is out of the initial objectives of this thesis, in this chapter, we take advantage of the availability of labels for diabetic macular edema (DME) and referable cases (RF) in the DROLL dataset by developing two approaches to detect these conditions. In one of the approaches, we use the features from the backbone in the densenet121_CAB_EyePACS model and the same classifiers used in previous chapters. In the other one, we proposed a multitask approach in which DR, DME, RF, and ocular lesions are detected simultaneously.

6.1. Classification of diabetic macular edema

The vision-threatening retinopathy is classified as severe non-proliferative DR or diabetic macular edema (DME). Worldwide approximately 93 million people are affected by DR and 21 million are affected by DME. DME is a condition in which fluid is accumulated in the macula, and it is characterized by retinal thickening, exudates, hemorrhages with or without microaneurysms and blot hemorrhages in the macula region [12]. It is the most common cause of moderate visual loss in diabetic patients, although, it is not unique in them, its prevalence in this population is high, up to 42% in type 1 and 53% in type 2 diabetes mellitus patients [72]. It is important to detect DR and DME, because in addition to its visual complications consequences, they are signs of the presence of other diabetes complications in other organ systems [19].

The diagnostic of DME is made using optical coherence tomography, however, the cost and availability limit its use for screening. Rather, eye fundus images are used to identify EX, which is a good indicator of DME risk [12, 66]. EX are lipids and fluids that leak from damaged capillaries, they appear at different locations with variable shapes and sizes, with bright yellowish or white color [1].

We attempt to classify DME following the same methodology used for the detection of DR-related lesions in Chapter 4. We extract features using the backbone in the model densenet121_CAB_EyePACS and explore SVM, GP, and MLP as final classifiers. The method is depicted in Figure **6-1**.

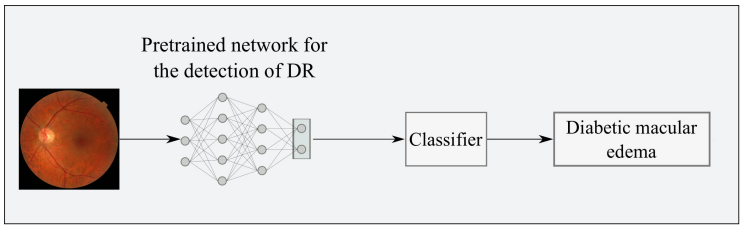


Figure 6-1.: Proposed model for the detection of DME cases.

6.2. Referable patients classification

Many diabetic patients receive a premature ophthalmologic referral since it is recommended to screen annually this population. A screening protocol in the primary care setting that automatically classifies patients with DR and healthy ones could help to reduce the referrals from 100% of diabetic patients, which is the actual recommendation to approximately 53% [11].

The DROLL dataset has labels for referable patients in which a patient is annotated as referable if DME or DR is present. We build a classification method for referable patients extracting features from the backbone in the model densenet121_CAB_EyePACS and exploring SVM, GP, and MLP as classifiers. Figure **6-2** shows the proposed method.

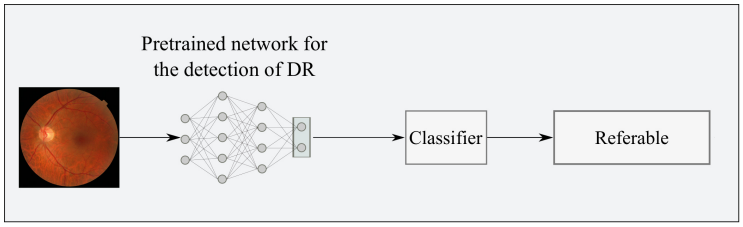


Figure 6-2.: Proposed model for the detection of RF cases.

6.3. Multi-task model for multi-conditions detection

We propose a multi-task method for the simultaneous detection of DR, DME, RF cases, and ocular lesions, namely, MA, H, CWS, VB, and NV. We implement the multi-task method using a shared trunk structure and a loss weighting loss function, similarly to the one presented in Chapter 4. The features from the densenet121_CAB_EyePACS backbone and a MLP as the classifier, are used to build the model. The overall structure of the method is presented in Figure **6-3**.

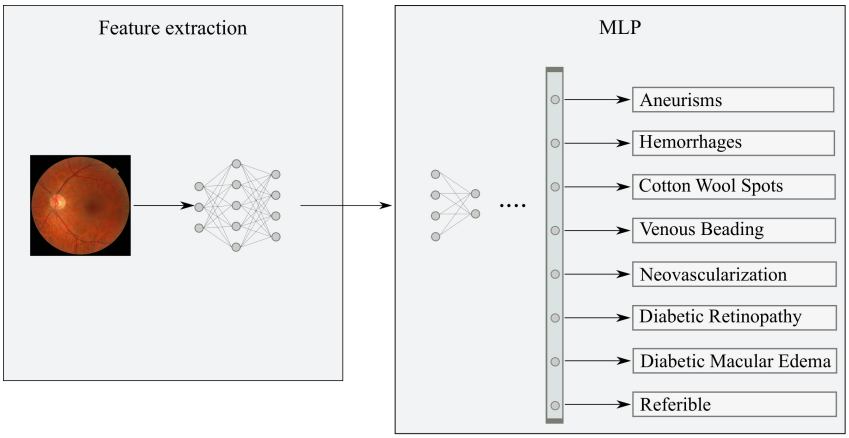


Figure 6-3.: Multitask model for the detection of ocular lesions, DR, DME and RF cases.

6.4. Results

In this section, we present the results of the methods for the detection of DME and RF cases using a single model and a multitask approach.

6.4.1. Diabetic macular edema detection

The results of the experiments for the detection of DME are shown in Table **6-1**, the multitask method leads to a better AUC metric than the GP and SVM. Although the MLP has a slightly lower AUC metric, it presents a higher specificity (Sp) and and sensitivity (Se) than the multi-task model.

Classifer	AUC	Specificity	Sensitivity	Threshold
GP regressor	0.8960	0.8064	0.8112	0.55
SVM	0.9393	0.7653	0.7676	0.2
MLP	0.9457	0.8761	0.8571	0.5
Multitask	**0.9475**	0.7349	0.7278	0.5

Table 6-1.: DME detection using multiple classifiers.

6.4.2. Referable cases detection

Table **6-2** shows the results of the detection for referable classes using multiple models. The multitask method presents a significantly higher performance than the other approaches, this can be due to the inductive bias introduced by the lesions, DR, and DME detection auxiliary tasks, which could help the model to generalize better.

Classifer	AUC	Specificity	Sensitivity	Threshold
GP regressor	0.7572	0.7010	0.6976	0.69
SVM	0.7204	0.6947	0.6998	0.9
MLP	0.7915	0.7522	0.7634	0.3
Multitask	**0.9512**	0.8822	0.8947	0.56

Table 6-2.: Referable using multiple classifiers.

6.4.3. Multitask model

Table **6-3** shows the results of the multi-tasks model implemented for the simultaneous detection of DR, DME, RF, and ocular lesions, namely, MA, H, CWS, VB, and NV. This results are compared in the same table with the best performances obtained in Chapters 4 and 5. The approach presented in this Chapter manages to improve the detection of MA, but especially of VB, getting an AUC of 0.9003, which is significantly higher than the obtained in Chapter 4 of 0.8538. This can be caused by the inductive bias introduced by the DME and RF task added. It is important to notice that the model has a close AUC to those in Chapter 5 in the tasks that did not manage to improve the performance.

Medical condition	AUC	Specificity	Sensitivity	Threshold
MA	**0.8059**	0.7146	0.7295	0.27
MA in Chapter 4	0.8034	0.6974	0.7146	0.19
H	0.9288	0.8596	0.8511	0.4
H in Chapter 4	**0.9370**	0.8451	0.8690	0.38
CWS	0.8811	0.8213	0.8120	0.29
CWS in Chapter 4	**0.8950**	0.8219	0.8496	0.3
VB	**0.9003**	0.8423	0.8298	0.53
VB in Chapter 4	0.8538	0.7728	0.8085	0.48
NV	0.9593	0.9301	0.8676	0.62
NV in Chapter 4	**0.9693**	0.9074	0.9333	0.1
DR	0.9475	0.8704	0.8833	0.32
DR in Chapter 5	**0.9479**	0.8783	0.8859	0.36

Table 6-3.: Comparison of the results for the detection of lesions and DR, between the multitask model proposed in this Chapter and the previous results.

7. Conclusions and future work

This thesis presents a novel strategy to improve the performance of known pretrained models for the detection of diabetic retinopathy. Our proposed model follows the current clinical workflow to perform the diagnosis of this decease. It uses a pretrained model as feature extractor to detect lesions related to diabetic retinopathy and combine the predictions to classify this illness. The most remarkable aspect of our proposed method is the combination of domain knowledge from ophthalmologist experts and the versatility of deep learning to support medical decision-making. We hypothesize that this model could be an outstanding option to gain greater acceptance and possible use in real-world applications by clinical staff.

We also propose a multi-task approach using a shared trunk structure to simultaneously detect aneurisms, hemorrhages, cotton wool spots, venous beading, neovascularization and diabetic retinopathy. This method has a end-to-end design in which a pretrained model for the detection of DR is used as feature extractor and a multi layer perceptron as classifier. The lesion detection results are similar to those obtained training a single model for each task and the diabetic retinopathy classification performance is slightly lower than the one achieved using ocular lesions for its detection. The multitask model is intuitive and simple to implement since only one model is used to predict all the tasks.

The results of this thesis in the diabetic retinopathy classification task are comparable with state-of-the-art performances, with the advantage that our proposed models also provide the prediction of ocular lesions.

Even though it is not within the scope of the thesis we perform initial experiments for the automatic detection of diabetic macular edema and referable patients. We implement individual models for the automatic classification of each of this conditions and compare the results with a multi-task model for the simultaneous detection of diabetic retinopathy, diabetic macular edema, referable cases and ocular lesions. The multi-task approach manage to either improve or to lead to similar results to those of the models trained specifically for each task. The effect of the inductive bias introduced by the auxiliary tasks is noticeable in the detection of venous beading and referable patients, since the multi-task method has an appreciable performance improvement when compared to the methods trained individually for this tasks. Therefore, a multi-task approach is able to lead to satisfactory performances in the simultaneous detection of multiple ocular conditions and lesions.

The future work comprises the systematic study of the proposed method with other datasets, the implementation of the fine-tuning strategy using the developed end-to-end models, and the design of a method that also provide the illness grading classification.

Finally, other illnesses like age-related macular degeneration using ocular lesions like exudates and drusen could be explored and evaluated.

Bibliography

[1] Automatic Detection of Hard Exudates in Color Retinal Images Using Dynamic Threshold and SVM Classification: Algorithm Development and Evaluation. In: *BioMed Research International* 2019 (2019), S. 13

[2] Abdelmaksoud, Eman ; El-Sappagh, Shaker ; Barakat, Sherif ; Abuhmed, Tamer ; Elmogy, Mohammed: Automatic Diabetic Retinopathy Grading System Based on Detecting Multiple Retinal Lesions. In: *IEEE Access* 9 (2021), Nr. January, S. 15939–15960. http://dx.doi.org/10.1109/ACCESS.2021.3052870. – DOI 10.1109/ACCESS.2021.3052870. – ISSN 21693536

[3] Abramoff, Michael D.: *Datasets and Algorithms.* https://medicine.uiowa.edu/eye/abramoff, 2015. – [Online; accessed 15-January-2020]

[4] Abràmoff, Michael D. ; Lavin, Philip T. ; Birch, Michele ; Shah, Nilay ; Folk, James C.: Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. In: *npj Digital Medicine* 1 (2018), Nr. 1. http://dx.doi.org/10.1038/s41746-018-0040-6. – DOI 10.1038/s41746–018–0040–6. – ISSN 2398–6352

[5] Alaguselvi, R. ; Murugan, Kalpana: Performance analysis of automated lesion detection of diabetic retinopathy using morphological operation. In: *Signal, Image and Video Processing* 15 (2021), Nr. 4, 797–805. http://dx.doi.org/10.1007/s11760-020-01798-x. – DOI 10.1007/s11760–020–01798–x. – ISSN 18631711

[6] Amin, Javeria ; Sharif, Muhammad ; Yasmin, Mussarat ; Ali, Hussam ; Fernandes, Steven L.: *A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions*

[7] Antal, BÃ¡lint ; Hajdu, AndrÃ¡s: An ensemble-based system for automatic screening of diabetic retinopathy. In: *Knowledge-Based Systems* 60 (2014), Nr. January, S. 20–27. http://dx.doi.org/10.1016/j.knosys.2013.12.023. – DOI 10.1016/j.knosys.2013.12.023. – ISSN 09507051

[8] Ashikur, Md ; Arifur, Md ; Ahmed, Juena: *Automated Detection of Diabetic Retinopathy using Deep Residual Learning*

[9] Beagley, Jessica ; Guariguata, Leonor ; Weil, Clara ; Motala, Ayesha A.: Global estimates of undiagnosed diabetes in adults. In: *Diabetes Research and Clinical Practice* 103 (2014), Nr. 2, 150–160. http://dx.doi.org/10.1016/j.diabres.2013.11.001. – DOI 10.1016/j.diabres.2013.11.001. – ISSN 18728227

[10] Bhaskaranand, Malavika ; Ramachandra, Chaithanya ; Bhat, Sandeep ; Cuadros, Jorge ; Nittala, Muneeswar G. ; Sadda, Srinivas R. ; Solanki, Kaushal: The value of automated diabetic retinopathy screening with the EyeArt system: A study of more than 100,000 consecutive encounters from people with diabetes. In: *Diabetes Technology and Therapeutics* 21 (2019), Nr. 11, S. 635–643. http://dx.doi.org/10.1089/dia.2019.0164. – DOI 10.1089/dia.2019.0164. – ISSN 15578593

[11] Bresnick, George H. ; Mukamel, Dana B. ; Dickinson, John C. ; Cole, David R.: A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy. In: *Ophthalmology* 107 (2000), Nr. 1, S. 19–24. http://dx.doi.org/10.1016/S0161-6420(99)00010-X. – DOI 10.1016/S0161–6420(99)00010–X. – ISSN 01616420

[12] Carvalho, C. ; Pedrosa, João ; Maia, Carolina ; Penas, S. ; Carneiro, Â. ; MendonÃ§a, L. ; MendonÃ§a, A. M. ; Campilho, A.: A Multi-dataset Approach for DME Risk Detection in Eye Fundus Images. In: *ICIAR*, 2020

[13] Chen, Benzhi ; Wang, Lisheng ; Wang, Xiuying ; Sun, Jian ; Huang, Yijie ; Feng, Dagan ; Xu, Zongben: Abnormality detection in retinal image by individualized background learning. In: *Pattern Recognition* 102 (2020), 107209. http://dx.doi.org/10.1016/j.patcog.2020.107209. – DOI 10.1016/j.patcog.2020.107209. – ISSN 00313203

[14] Chen, Weijie ; Sahiner, Berkman ; Samuelson, Frank ; Pezeshk, Aria ; Petrick, Nicholas: Calibration of medical diagnostic classifier scores to the probability of disease. In: *Statistical Methods in Medical Research* 27 (2018), Nr. 5, S. 1394–1409. http://dx.doi.org/10.1177/0962280216661371. – DOI 10.1177/0962280216661371. – ISBN 0962280216

[15] Chudzik, Piotr ; Majumdar, Somshubra ; Calivá, Francesco ; Al-Diri, Bashir ; Hunter, Andrew: Microaneurysm detection using fully convolutional neural networks. In: *Computer Methods and Programs in Biomedicine* 158 (2018), S. 185–192. http://dx.doi.org/10.1016/j.cmpb.2018.02.016. – DOI 10.1016/j.cmpb.2018.02.016. – ISSN 18727565

[16] Crawshaw, Michael: Multi-Task Learning with Deep Neural Networks: A Survey. (2020). http://arxiv.org/abs/2009.09796

[17] Cun, Y. L. ; Boser, B. ; Denker, J. S. ; Henderson, D. ; Howard, R. E. ; Hubbard, W. ; Jackel, L. D.: Handwritten Digit Recognition with a Back-Propagation Network. In: *Advances in neural information processing systems* 2 (1989), S. 396–404

[18] Decencière, E. ; Cazuguel, G. ; Zhang, X. ; Thibault, G. ; Klein, J. C. ; Meyer, F. ; Marcotegui, B. ; Quellec, G. ; Lamard, M. ; Danno, R. ; Elie, D. ; Massin, P. ; Viktor, Z. ; Erginay, A. ; Laÿ, B. ; Chabouis, A.: *TeleOphta: Machine learning and image processing methods for teleophthalmology*

[19] Ding, Jie ; Wong, Tien Y.: Current epidemiology of diabetic retinopathy and diabetic macular edema. In: *Current Diabetes Reports* 12 (2012), Nr. 4, S. 346–354. `http://dx.doi.org/10.1007/s11892-012-0283-6`. – DOI 10.1007/s11892–012–0283–6. – ISSN 15344827

[20] Gargeya, Rishab ; Leng, Theodore: Automated Identification of Diabetic Retinopathy Using Deep Learning. In: *Ophthalmology* 124 (2017), Nr. 7, 962–969. `http://dx.doi.org/10.1016/j.ophtha.2017.02.008`. – DOI 10.1016/j.ophtha.2017.02.008. – ISSN 15494713

[21] Gayathri, S. ; Gopi, Varun P. ; Palanisamy, P.: A lightweight CNN for Diabetic Retinopathy classification from fundus images. In: *Biomedical Signal Processing and Control* 62 (2020), Nr. August, 102115. `http://dx.doi.org/10.1016/j.bspc.2020.102115`. – DOI 10.1016/j.bspc.2020.102115. – ISSN 17468108

[22] Goldbaum, M. H.: *STructured Analysis of the Retina Project*. `http://cecas.clemson.edu/~ahoover/stare`. Version: 1975

[23] 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. `http://dx.doi.org/10.1001/jama.2016.17216`. – DOI 10.1001/jama.2016.17216. – ISSN 15383598

[24] 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. http://dx.doi.org/10.1001/jama.2016.17216. – DOI 10.1001/jama.2016.17216. – ISSN 15383598

[25] Hassan, Siti Syafinah A. ; Bong, David B. ; Premsenthil, Mallika: Detection of Neovascularization in Diabetic Retinopathy. In: *Journal of Digital Imaging* 25 (2012), S. 437â€“444

[26] He, Along ; Li, Tao ; Li, Ning ; Wang, Kai ; Fu, Huazhu: CABNet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading. In: *IEEE Transactions on Medical Imaging* 40 (2021), Nr. 1, S. 143–153. http://dx.doi.org/10.1109/TMI.2020.3023463. – DOI 10.1109/TMI.2020.3023463. – ISSN 1558254X

[27] Heijden, Amber A. d. ; Abramoff, Michael D. ; Verbraak, Frank ; Hecke, Manon V. ; Liem, Albert ; Nijpels, Giel: *Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System*

[28] Hoover, A. D. ; Kouznetsova, V. ; Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. In: *IEEE Transactions on Medical Imaging* 19 (2000), March, Nr. 3, S. 203–210. http://dx.doi.org/10.1109/42.845178. – DOI 10.1109/42.845178. – ISSN 0278–0062

[29] Imaging Experts, ADCIS: A T.: *Messidor*. https://www.adcis.net/en/third-party/messidor/, 2008. – [Online; accessed 23-May-2019]

[30] Imaging Experts, ADCIS: A T.: *Messidor-2*. https://www.adcis.net/en/third-party/messidor2/, 2015. – [Online; accessed 25-January-2020]

[31] James Talks, Stephen ; Manjunath, Vina ; H W Steel, David ; Peto, Tunde ; Taylor, Roy: New vessels detected on wide-field imagingcompared to two-field and seven-field imaging:implications for diabetic retinopathy screeningimage analysis. In: *Br J Ophthalmol* 99 (2015), S. 1606–1609

[32] Kaggle: *Diabetic Retinopathy Detection*. https://www.kaggle.com/c/diabeticretinopathy-detection, 2015. – [Online; accessed 10-January-2020]

[33] Kauppi, T. ; Kalesnykiene, V. ; Kamarainen, J. K. ; Lensu, L. ; Sorri, I. ; Raninen, A. ; Voutilainen, R. ; Pietilä, J. ; Kälviäinen, H. ; Uusitalo, H.: The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: *BMVC 2007 - Proceedings of the British Machine Vision Conference 2007* (2007), S. 1–18. http://dx.doi.org/10.5244/C.21.15. – DOI 10.5244/C.21.15

[34] Kauppi, Tomi ; Kalesnykiene, Valentina ; Kamarainen, Joni-kristian ; Lensu, Lasse ; Sorri, Iiris ; Uusitalo, Hannu ; Kalviainen, Heikki

; Pietila, Juhani: DIARETDB0 : Evaluation Database and Methodology for Diabetic Retinopathy Algorithms. In: *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland.* (2006), 1–17. http://www.siue.edu/$\sim$sumbaug/RetinalProjectPapers/DiabeticRetinopathyImageDatabaseInformation.pdf

[35] Lam, Carson ; Yu, Caroline ; Huang, Laura ; Rubin, Daniel: Retinal Lesion Detection With Deep Learning Using Image Patches. (2018)

[36] 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. http://dx.doi.org/10.1038/s41598-017-17876-z. – DOI 10.1038/s41598–017–17876–z. – ISSN 20452322

[37] Li, Tao ; Gao, Yingqi ; Wang, Kai ; Guo, Song ; Liu, Hanruo ; Kang, Hong: Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening. In: *Information Sciences* 501 (2019), 511 - 522. http://dx.doi.org/https://doi.org/10.1016/j.ins.2019.06.011. – DOI https://doi.org/10.1016/j.ins.2019.06.011. – ISSN 0020–0255

[38] Li, Wong ; Acharya, U R. ; Venkatesh, Y V. ; Chee, Caroline ; Choo, Lim ; Ng, E Y K.: Identification of different stages of diabetic retinopathy using retinal optical images. 178 (2008), S. 106–121. http://dx.doi.org/10.1016/j.ins.2007.07.020. – DOI 10.1016/j.ins.2007.07.020

[39] Li, Xiaogang ; Pang, Tiantian ; Xiong, Biao ; Liu, Weixiang ; Liang, Ping ; Wang, Tianfu: Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In: *Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017* 2018-Janua (2018), Nr. 978, S. 1–11. http://dx.doi.org/10.1109/CISP-BMEI.2017.8301998. – DOI 10.1109/CISP–BMEI.2017.8301998. ISBN 9781538619377

[40] Meriaudeau, Prasanna Porwal; Samiksha Pachade; Ravi Kamble; Manesh Kokare; Girish Deshmukh; Vivek Sahasrabuddhe; F.: Indian Diabetic Retinopathy Image Dataset (IDRiD). (2018). http://dx.doi.org/10.21227/H25W98. – DOI 10.21227/H25W98

[41] Mookiah, Muthu Rama K. ; Acharya, U. R. ; Chua, Chua K. ; Lim, Choo M. ; Ng, E. Y. ; Laude, Augustinus: Computer-aided diagnosis of diabetic retinopathy: A review. In: *Computers in Biology and Medicine* 43 (2013), Nr. 12, S. 2136–2155. http://dx.doi.org/10.1016/j.compbiomed.2013.10.007. – DOI 10.1016/j.compbiomed.2013.10.007. – ISSN 00104825

[42] (ODIR), Ocular Disease Intelligent R.: *ODIR-5K*. https://academictorrents.com/details/cf3b8d5ecdd4284eb9b3a80fcfe9b1d621548f72, 2019. – [Online; accessed 13-May-2022]

[43] Orlando, JosAn ensemble deep learning based approach for red lesion detection in fundus imagesÃ© I. ; Prokofyeva, Elena ; Fresno, Mariana del ; Blaschko, Matthew B.: lista. An ensemble deep learning based approach for red lesion detection in fundus images. In: *Computer Methods and Programs in Biomedicine* 153 (2018), S. 115–127. http://dx.doi.org/10.1016/j.cmpb.2017.10.017. – DOI 10.1016/j.cmpb.2017.10.017. – ISSN 18727565

[44] Orlando, JosÃ© I. ; Prokofyeva, Elena ; Fresno, Mariana del ; Blaschko, Matthew B.: An ensemble deep learning based approach for red lesion detection in fundus images. In: *Computer Methods and Programs in Biomedicine* 153 (2018), S. 115–127. http://dx.doi.org/10.1016/j.cmpb.2017.10.017. – DOI 10.1016/j.cmpb.2017.10.017. – ISSN 18727565

[45] Paing, May P. ; Choomchuay, Somsak: Detection of Lesions and Classification of Diabetic Retinopathy Using Fundus Images. (2016). ISBN 9781509039401

[46] Pandeya, Y.R. ; Lee, J: Deep learning-based late fusion of multimodal information for emotion classification of music video. In: *Multimed Tools Appl* 80 (2021), S. 2887–â€“2905

[47] Prentasic, Pavle ; Loncaric, Sven ; Vatavuk, Zoran ; Bencic, Goran ; Subasic, Marko ; Petkovic, Tomislav ; Malenica-Ravlic, Maja ; Budimlija, Nikolina ; Tadic, Raseljka: Diabetic retinopathy image database(DRiDB): A new database for diabetic retinopathy screening programs research. In: *8th International Symposium on Image and Signal Processing and Analysis (ISPA)* (2013), S. 711–716

[48] Qiao, Lifeng ; Zhu, Ying ; Zhou, Hui: Diabetic Retinopathy Detection Using Prognosis of Microaneurysm and Early Diagnosis System for Non-Proliferative Diabetic Retinopathy Based on Deep Learning Algorithms. In: *IEEE Access* 8 (2020), S. 104292–104302. http://dx.doi.org/10.1109/ACCESS.2020.2993937. – DOI 10.1109/ACCESS.2020.2993937. – ISSN 21693536

[49] Qummar, Sehrish ; Khan, Fiaz G. ; Shah, Sajid ; Khan, Ahmad ; Shamshirband, Shahaboddin ; Rehman, Zia U. ; Khan, Iftikhar A. ; Jadoon, Waqas: A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection. In: *IEEE Access* 7 (2019), S. 150530–150539. http://dx.doi.org/10.1109/ACCESS.2019.2947484. – DOI 10.1109/ACCESS.2019.2947484. – ISSN 21693536

[50] Rajalakshmi, Ramachandran ; Subashini, Radhakrishnan ; Anjana, Ranjit M. ; Mohan, Viswanathan: *Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence*

[51] Resnikoff, Serge ; Lansingh, Van C. ; Washburn, Lindsey ; Felch, William ; Gauthier, Tina M. ; Taylor, Hugh R. ; Eckert, Kristen ; Parke, David ; Wiedemann, Peter: Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): Will we meet the needs? In: *British Journal of Ophthalmology* (2019), Nr. Md, S. 1–5. http://dx.doi.org/10.1136/bjophthalmol-2019-314336. – DOI 10.1136/bjophthalmol–2019–314336. – ISSN 14682079

[52] Ruder, Sebastian: An Overview of Multi-Task Learning in Deep Neural Networks. (2017)

[53] Selvaraju, Ramprasaath R. ; Cogswell, Michael ; Das, Abhishek ; Vedantam, Ramakrishna ; Parikh, Devi ; Batra, Dhruv: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *Proceedings of the IEEE International Conference on Computer Vision* 2017-Octob (2017), S. 618–626. http://dx.doi.org/10.1109/ICCV.2017.74. – DOI 10.1109/ICCV.2017.74. – ISBN 9781538610329

[54] Sengar, Namita ; Dutta, Malay K.: lista.Automated method for hierarchal detection and grading of diabetic retinopathy. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 1163 (2017), Nr. July, 1–11. http://dx.doi.org/10.1080/21681163.2017.1335236. – DOI 10.1080/21681163.2017.1335236. – ISSN 2168–1163

[55] Seoud, Lama ; Hurtut, Thomas ; Chelbi, Jihed ; Cheriet, Farida ; Langlois, J. M.: Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening. In: *IEEE Transactions on Medical Imaging* 35 (2016), Nr. 4, S. 1116–1126. http://dx.doi.org/10.1109/TMI.2015.2509785. – DOI 10.1109/TMI.2015.2509785. – ISSN 1558254X

[56] Sharif, Muhammad ; Shah, Jamal H.: Automatic screening of retinal lesions for grading diabetic retinopathy. In: *International Arab Journal of Information Technology* 16 (2019), Nr. 4, S. 766–774. – ISSN 23094524

[57] Son, Jaemin ; Shin, Joo Y. ; Kim, Hoon D. ; Jung, Kyu H. ; Park, Kyu H. ; Park, Sang J.: Development and Validation of Deep Learning Models for Screening Multiple Abnormal Findings in Retinal Fundus Images. In: *Ophthalmology* 127 (2020), Nr. 1, 85–94. http://dx.doi.org/10.1016/j.ophtha.2019.05.029. – DOI 10.1016/j.ophtha.2019.05.029. – ISSN 15494713

[58] Staal, Joes ; Abràmoff, Michael D. ; Niemeijer, Meindert ; Viergever, Max A. ; Ginneken, Bram v.: Ridge-Based Vessel Segmentation in Color Images of the Retina. In: *IEEE Transactions on Medical Imaging* 23 (2004), Nr. 4, S. 501–509. http://dx.doi.org/10.1080/17455030500184511. – DOI 10.1080/17455030500184511. – ISSN 17455030

[59] Symposium, Asia Pacific Tele-Ophthalmology Society (.: *APTOS 2019 Blindness Detection*. https://www.kaggle.com/c/aptos2019-blindness-detection/data, 2019. – [Online; accessed 15-January-2020]

[60] Tajbakhsh, Nima ; Shin, Jae Y. ; Gurudu, Suryakanth R. ; Hurst, R. T. ; Kendall, Christopher B. ; Gotway, Michael B. ; Liang, Jianming: Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? In: *IEEE Transactions on Medical Imaging* 35 (2016), Nr. 5, S. 1299–1312. http://dx.doi.org/10.1109/TMI.2016.2535302. – DOI 10.1109/TMI.2016.2535302

[61] Usman Akram, M. ; Khalid, Shehzad ; Tariq, Anam ; Khan, Shoab A. ; Azam, Farooque: Detection and classification of retinal lesions for grading of diabetic retinopathy. In: *Computers in Biology and Medicine* 45 (2014), Nr. 1, 161–171. http://dx.doi.org/10.1016/j.compbiomed.2013.11.014. – DOI 10.1016/j.compbiomed.2013.11.014. – ISSN 00104825

[62] Voets, Mike ; Møllersen, Kajsa ; Bongo, Lars A.: Replication study: Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. (2018), 1–16. http://arxiv.org/abs/1803.04337

[63] 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. http://dx.doi.org/10.1371/journal.pone.0217541. – DOI 10.1371/journal.pone.0217541. – ISBN 1111111111

[64] Wan, Shaohua ; Liang, Yan ; Zhang, Yin: Deep convolutional neural networks for diabetic retinopathy detection by image classification. In: *Computers and Electrical Engineering* 72 (2018), 274–282. http://dx.doi.org/10.1016/j.compeleceng.2018.07.042. – DOI 10.1016/j.compeleceng.2018.07.042. – ISSN 00457906

[65] Wang, Juan ; Bai, Yujing ; Xia, Bin: Simultaneous Diagnosis of Severity and Features of Diabetic Retinopathy in Fundus Photography Using Deep Learning. In: *IEEE Journal of Biomedical and Health Informatics* 24 (2020), Nr. 12, S. 3397–3407. http://dx.doi.org/10.1109/JBHI.2020.3012547. – DOI 10.1109/JBHI.2020.3012547. – ISSN 21682208

[66] Wang, Yu T. ; Tadarati, Mongkol ; Wolfson, Yulia ; Bressler, Susan B. ; Bressler, Neil M.: Comparison of prevalence of diabetic macular edema based on monocular fundus photography vs optical coherence tomography. In: *JAMA Ophthalmology* 134 (2016), Nr. 2, S. 222–228. http://dx.doi.org/10.1001/jamaophthalmol.2015.5332. – DOI 10.1001/jamaophthalmol.2015.5332. – ISSN 21686165

[67] Wang, Zhe ; Yin, Yanxin ; Shi, Jianping ; Fang, Wei ; Li, Hongsheng ; Wang, Xiaogang: Zoom-in-Net: Deep mining lesions for diabetic retinopathy detection. In: *arXiv* 1 (2017), S. 267–275. http://dx.doi.org/10.1007/978-3-319-66179-7. – DOI 10.1007/978–3–319–66179–7. – ISBN 9783319661797

[68] 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. http://dx.doi.org/10.1016/S0161-6420(03)00475-5. – DOI 10.1016/S0161–6420(03)00475–5. – ISSN 01616420

[69] Y., LeCun ; Y., Bengio ; G, Hinton: Deep learning. In: *Nature* 521 (2015), S. 436–â€“444

[70] Yang, Y ; Li, T ; Li, W ; Wu, H ; Fan, W ; Zhang, W: Lesion Detection and Grading of Diabetic Retinopathy via Two-Stages Deep Convolutional Neural Networks. In: *Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer Assisted Intervention MICCAI 2017.* Bd. 10435, 2017. – ISBN 978–3–319–66184–1, 516–524

[71] Zago, Gabriel T. ; Andreão, Rodrigo V. ; Dorizzi, Bernadette ; Teatini Salles, Evandro O.: Diabetic retinopathy detection using red lesion localization and convolutional neural networks. In: *Computers in Biology and Medicine* 116 (2020), Nr. November 2019. http://dx.doi.org/10.1016/j.compbiomed.2019.103537. – DOI 10.1016/j.compbiomed.2019.103537. – ISSN 18790534

[72] Zander, Eckhard ; Herfurth, Sabine ; Bohl, Beate ; Heinke, Peter ; Kohnert, Klaus D. ; Kerner, Wolfgang ; Herrmann, Uwe: Maculopathy in patients with diabetes mellitus type 1 and type 2: Associations with risk factors. In: *British Journal of Ophthalmology* 84 (2000), Nr. 8, S. 871–876. http://dx.doi.org/10.1136/bjo.84.8.871. – DOI 10.1136/bjo.84.8.871. – ISSN 00071161

[73] Zeng, Xianglong ; Chen, Haiquan ; Luo, Yuan ; Ye, Wenbin: *Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network*

[74] Zhou, Lei ; Zhao, Yu ; Yang, Jie ; Yu, Qi ; Xu, Xun: Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images. In: *IET Image Processing* 12 (2018), Nr. 4, S. 563–571. http://dx.doi.org/10.1049/iet-ipr.2017.0636. – DOI 10.1049/iet–ipr.2017.0636. – ISSN 17519659

[75] Zhou, Y. ; Wang, B. ; Huang, L. ; Cui, S. ; Shao, L.: A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability. In: *IEEE Transactions on Medical Imaging* 40 (2021), Nr. 3, S. 818–828. http://dx.doi.org/10.1109/TMI.2020.3037771. – DOI 10.1109/TMI.2020.3037771

[76] Zhou, Yi ; Wang, Boyang ; Huang, Lei ; Cui, Shanshan ; Shao, Ling: A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability. In: *IEEE Transactions on Medical Imaging* 40 (2021), Nr. 3, S. 818–828. http://dx.doi.org/10.1109/TMI.2020.3037771. – DOI 10.1109/TMI.2020.3037771. – ISSN 1558254X

Appendices

A. Findings detection additional results

Table A-1.: Results of the detection of DR-related ocular lesions using the xception_CAB_EyePACS backbone model as a feature extractor.

Finding	Metric	GP classifier	GP regressor	SVM	MLP	Multi-task
MA	AUC	0.7714	0.7605	0.7464	**0.7837**	0.7822
	Sp	0.6654	0.6900	0.8696	0.7307	0.6913
	Se	0.7032	0.6929	0.5765	0.6872	0.7066
	tr	0.42	0.22	0.05	0.18	0.19
H	AUC	0.9116	0.9128	0.8594	**0.9303**	0.9267
	Sp	0.8411	0.8419	0.9719	0.8619	0.8676
	Se	0.8419	0.8307	0.7200	0.8623	0.8578
	tr	0.42	0.19	0.05	0.2	0.22
CWS	AUC	0.8252	0.8858	0.6870	**0.8904**	0.8765
	Sp	0.7519	0.8226	0.9762	0.8386	0.8079
	Se	0.7293	0.8120	0.3759	0.8195	0.7969
	tr	0.36	0.12	0.05	0.38	0.31
VB	AUC	0.5374	0.7299	0.5203	**0.8496**	0.8441
	Sp	0.5596	0.6565	0.9311	0.6626	0.7467
	Se	0.4681	0.6808	0.0638	0.8511	0.8511
	tr	0.35	0.18	0.1	0.47	0.45
NV	AUC	0.6847	0.8334	0.5328	**0.9651**	0.9564
	Sp	0.6804	0.9767	0.9988	0.9587	0.9229
	Se	0.7333	0.6	0.0667	0.8	0.8
	tr	0.29	0.1	0.1	0.57	0.52

Table A-2.: Results of the detection of DR-related ocular lesions using the vgg16_CAB_EyePACS backbone model as a feature extractor.

Finding	Metric	GP classifier	GP regressor	SVM	MLP	Multi-task
MA	AUC	0.7723	0.7637	0.7273	**0.8003**	0.7974
	Sp	0.6801	0.6937	0.9545	0.7269	0.7196
	Se	0.6952	0.6952	0.4817	0.7146	0.7215
	tr	0.27	0.17	0.05	0.13	0.15
H	AUC	0.9066	0.8944	0.8168	**0.9284**	**0.9284**
	Sp	0.8290	0.8202	0.9823	0.8571	0.8499
	Se	0.8375	0.8172	0.6320	0.8646	0.8736
	tr	0.35	0.3	0.05	0.2	0.16
CWS	AUC	0.8438	0.8220	0.5557	**0.8633**	0.8619
	Sp	0.7731	0.7828	0.9910	0.8065	0.7885
	Se	0.7895	0.7895	0.1203	0.8120	0.8045
	tr	0.19	0.12	0.05	0.38	0.26
VB	AUC	0.6631	0.6007	0.5	0.7489	**0.8172**
	Sp	0.6157	0.5944	1	0.8337	0.8008
	Se	0.6809	0.5532	0	0.6383	0.7446
	tr	0.32	0.28	0.05	0.56	0.51
NV	AUC	0.8935	0.8605	0.5206	0.9069	**0.9464**
	Sp	0.7993	0.8716	0.9749	0.8142	0.9217
	Se	0.8	0.8667	0.0667	0.8667	0.8666
	tr	0.14	0.07	0.1	0.37	0.68

Table A-3.: Results of the detection of DR-related ocular lesions using the resnet_CAB_EyePACS backbone model as a feature extractor.

Finding	Metric	GP classifier	GP regressor	SVM	MLP	Multi-task
MA	AUC	0.7674	0.7426	0.7524	**0.7820**	0.7636
	Sp	0.7072	0.6789	0.6961	0.7171	0.7060
	Se	0.6803	0.6632	0.6849	0.6952	0.6757
	tr	0.47	0.15	0.11	0.17	0.2
H	AUC	0.8762	0.8586	0.8508	**0.9129**	0.9047
	Sp	0.7945	0.7697	0.9221	0.8250	0.9318
	Se	0.7788	0.7855	0.7200	0.8262	0.7065
	tr	0.47	0.15	0.05	0.37	0.21
CWS	AUC	0.7284	0.8429	0.6515	**0.8622**	0.8500
	Sp	0.6696	0.8528	0.9524	0.7769	0.8657
	Se	0.6316	0.6842	0.3233	0.8045	0.6316
	tr	0.42	0.11	0.1	0.22	0.4
VB	AUC	0.4936	0.6553	0.5	0.6607	**0.7006**
	Sp	0.4762	0.6114	1	0.3922	0.5937
	Se	0.4893	0.6383	0	0.7447	0.6595
	tr	0.43	0.23	0.5	0.47	0.52
NV	AUC	0.7189	**0.9693**	0.5	0.9477	0.9446
	Sp	0.6368	0.9074	1	0.8817	0.9020
	Se	0.6	0.9333	0	0.8666	0.8
	tr	0.39	0.09	0.5	0.43	0.4