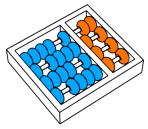
Universidade Estadual de Campinas
Instituto de Computação



José Ramon Trindade Pires

Image Analytics Techniques for Diabetic Retinopathy Detection

Técnicas de Análise de Imagens para Detecção de Retinopatia Diabética

CAMPINAS
2018

José Ramon Trindade Pires

Image Analytics Techniques for Diabetic Retinopathy Detection

Técnicas de Análise de Imagens para Detecção de Retinopatia Diabética

Tese apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Supervisor/Orientador: Prof. Dr. Anderson de Rezende Rocha
Co-supervisor/Coorientador: Prof. Dr. Jacques Wainer

Este exemplar corresponde à versão final da Tese defendida por José Ramon Trindade Pires e orientada pelo Prof. Dr. Anderson de Rezende Rocha.

CAMPINAS
2018

Agência(s) de fomento e nº(s) de processo(s): CAPES
ORCID: https://orcid.org/0000-0002-0023-1971

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

P665i

Pires, Ramon, 1989-
Image analytics techniques for diabetic retinopathy detection / José Ramon Trindade Pires. – Campinas, SP : [s.n.], 2018.

Orientador: Anderson de Rezende Rocha.
Coorientador: Jacques Wainer.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Retinopatia diabética. 2. Diagnóstico auxiliado por computador. 3. Cegueira. 4. Aprendizado de máquina. 5. Inteligência artificial. I. Rocha, Anderson de Rezende, 1980-. II. Wainer, Jacques, 1958-. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Técnicas de análise de imagens para detecção de retinopatia diabética
Palavras-chave em inglês:
Diabetic retinopathy
Computer-aided diagnosis
Blindness
Machine learning
Artificial intelligence
Área de concentração: Ciência da Computação
Titulação: Doutor em Ciência da Computação
Banca examinadora:
Anderson de Rezende Rocha [Orientador]
Agma Juci Machado Traina
Roberto Marcondes Cesar Junior
Ricardo da Silva Torres
Levy Boccato
Data de defesa: 20-12-2018
Programa de Pós-Graduação: Ciência da Computação

Universidade Estadual de Campinas
Instituto de Computação

José Ramon Trindade Pires

Image Analytics Techniques for Diabetic Retinopathy Detection

Técnicas de Análise de Imagens para Detecção de Retinopatia Diabética

Banca Examinadora:

- Prof. Dr. Anderson de Rezende Rocha
  IC/Unicamp
- Profa. Dra. Agma Juci Machado Traina
  ICMC/USP
- Prof. Dr. Roberto Marcondes Cesar Junior
  IME/USP
- Prof. Dr. Ricardo da Silva Torres
  IC/Unicamp
- Prof. Dr. Levy Boccato
  FEEC/Unicamp

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 20 de dezembro de 2018

Acknowledgements

Faith, courage and determination. It was a long and arduous journey. Despite all the obstacles, I am here, performing a dream! Giving up never crossed my mind. That's because I always had family, friends, and professors who encouraged me and motivated me to reach my goals. And here are my sincere acknowledgements.

I would first like to thank God for giving me the opportunity to embark on this research project, give me strength to persevere and complete it satisfactory. "Have I not commanded you? Be strong and courageous. Do not be afraid; do not be discouraged, for the Lord your God will be with you wherever you go." Joshua 1:9.

I am profoundly grateful to my parents: José Correia (in memorian) and Maria Lisboa. My mom, every day contacting me to hear my voice, to know how my day was. Thanks, mom, for all your prayers, guidance and advice.

I am grateful to all my siblings — brothers (Ney, Nildo, Cani, Marcelo, and Valério) and sisters (Gracinha, Têca, Mônica, Livinha, Ana, and Débora) — who have provided me emotional support in my life. Thanks also to my adorable grandmother Aurelina, and my nieces and nephews. I love you all!

A very special gratitude goes to my advisor Anderson Rocha. I will always remember how much he motivated me to come to Campinas — the Brazilian Silicon Valley — before we even met ourselves. One of the most important and successful decisions in my life. Anderson has supported me along the way, and his office was always open whenever I ran into a trouble spot in my research. He consistently steered me in the right the direction.

With a special mention to Professors Eduardo Valle, Sandra Avila, Jacques Wainer and Herbert F. Jelinek. I am gratefully indebted to them for their valuable comments and collaborations for constructing the manuscripts. Thanks also to Alexandre for the cooperation.

My sincere thanks also goes to José Augusto Stuchi, Flavio Pascoal Vieira and Diego Lencione, from Phelcom Technologies, for providing to our research the opportunity of moving towards real life application, in the direction of democratizing the recommended annual examination. I'm really hoping for the success and growth of the company.

And finally, last but by no means least, my thank goes also to everyone in the Recod Lab. It was really great sharing the laboratory with all of you during last years. Great moments I'll never forget. Recod rocks!

I can not forget to mention the gratitude to Unicamp. This is a university that supports the student at all times. I would also like to thank Samsung, for financial support; and to Google Research for the award and for believing in the potential of this work. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo

Retinopatia Diabética (RD) é uma complicação a longo prazo do diabetes e a principal causa de cegueira da população ativa. Consultas regulares são necessárias para diagnosticar a retinopatia em um estágio inicial, permitindo um tratamento com o melhor prognóstico capaz de retardar ou até mesmo impedir a cegueira. Alavancados pela evolução da prevalência do diabetes e pelo maior risco que os diabéticos têm de desenvolver doenças nos olhos, diversos trabalhos com abordagens bem estabelecidas e promissoras vêm sendo desenvolvidos para triagem automática de retinopatia. Entretanto, a maior parte dos trabalhos está focada na detecção de lesões utilizando características visuais particulares de cada tipo de lesão. Além do mais, soluções artesanais para avaliação de necessidade de consulta e de identificação de estágios da retinopatia ainda dependem bastante das lesões, cujo repetitivo procedimento de detecção é complexo e inconveniente, mesmo se um esquema unificado for adotado. O estado da arte para avaliação automatizada de necessidade de consulta é composto por abordagens que propõem uma representação altamente abstrata obtida inteiramente por meio dos dados. Usualmente, estas abordagens recebem uma imagem e produzem uma resposta — que pode ser resultante de um único modelo ou de uma combinação — e não são facilmente explicáveis. Este trabalho objetivou melhorar a detecção de lesões e reforçar decisões relacionadas à necessidade de consulta, fazendo uso de avançadas representações de imagens em duas etapas. Nós também almejamos compor um modelo sofisticado e direcionado pelos dados para triagem de retinopatia, bem como incorporar aprendizado supervisionado de características com representação orientada por mapa de calor, resultando em uma abordagem robusta e ainda responsável para triagem automatizada. Finalmente, tivemos como objetivo a integração das soluções em dispositivos portáteis de captura de imagens de retina. Para detecção de lesões, propusemos abordagens de caracterização de imagens que possibilitem uma detecção eficaz de diferentes tipos de lesões. Nossos principais avanços estão centrados na modelagem de uma nova técnica de codificação para imagens de retina, bem como na preservação de informações no processo de *pooling* ou agregação das características obtidas. Decidir automaticamente pela necessidade de encaminhamento do paciente a um especialista é uma investigação ainda mais difícil e muito debatida. Nós criamos um método mais simples e robusto para decisões de necessidade de consulta, e que não depende da detecção de lesões. Também propusemos um modelo direcionado pelos dados que melhora significativamente o desempenho na tarefa de triagem da RD. O modelo produz uma resposta confiável com base em respostas (locais e globais), bem como um mapa de ativação que permite uma compreensão de importância de cada pixel para a decisão. Exploramos a metodologia de explicabilidade para criar um descritor local codificado em uma rica representação em nível médio. Os modelos direcionados pelos dados são o estado da arte para triagem de retinopatia diabética. Entretanto, mapas de ativação são essenciais para interpretar o aprendizado em termos de importância de cada pixel e para reforçar pequenas características discriminativas que têm potencial de melhorar o diagnóstico.

Abstract

Diabetic Retinopathy (DR) is a long-term complication of diabetes and the leading cause of blindness among working-age adults. A regular eye examination is necessary to diagnose DR at an early stage, when it can be treated with the best prognosis and the visual loss delayed or deferred. Leveraged by the continuous expansion of diabetics and by the increased risk that those people have to develop eye diseases, several works with well-established and promising approaches have been proposed for automatic screening. Therefore, most existing art focuses on lesion detection using visual characteristics specific to each type of lesion. Additionally, handcrafted solutions for referable diabetic retinopathy detection and DR stages identification still depend too much on the lesions, whose repetitive detection is complex and cumbersome to implement, even when adopting a unified detection scheme. Current art for automated referral assessment resides on highly abstract data-driven approaches. Usually, those approaches receive an image and spit the response out — that might be resulting from only one model or ensembles — and are not easily explainable. Hence, this work aims at enhancing the lesion detection and reinforcing referral decisions with advanced handcrafted two-tiered image representations. We also intended to compose sophisticated data-driven models for referable DR detection and incorporate supervised learning of features with saliency-oriented mid-level image representations to come up with a robust yet accountable automated screening approach. Ultimately, we aimed at integrating our software solutions with simple retinal imaging devices. In the lesion detection task, we proposed advanced handcrafted image characterization approaches to detecting effectively different lesions. Our leading advances are centered on designing a novel coding technique for retinal images and preserving information in the pooling process. Automatically deciding on whether or not the patient should be referred to the ophthalmic specialist is a more difficult, and still hotly debated research aim. We designed a simple and robust method for referral decisions that does not rely upon lesion detection stages. We also proposed a novel and effective data-driven model that significantly improves the performance for DR screening. Our accountable data-driven model produces a reliable (local- and global-) response along with a heatmap/saliency map that enables pixel-based importance comprehension. We explored this methodology to create a local descriptor that is encoded into a rich mid-level representation. Data-driven methods are the state of the art for diabetic retinopathy screening. However, saliency maps are essential not only to interpret the learning in terms of pixel importance but also to reinforce small discriminative characteristics that have the potential to enhance the diagnostic.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Every eleventh person in the world suffers from diabetes mellitus, a disorder of sugar metabolism, whose prevalence is expected to reach every tenth person by 2040 [43]. Diabetes sufferers are 25 times more likely to suffer from sight loss resulting from diabetic retinopathy (DR), a major long-term microvascular complication, and the leading cause of blindness in high-income countries [43]. In the U.S. alone, 7.7 million people aged 40+ years have diabetic retinopathy [113], and the prevalence is even larger in developing countries with a shortage of ophthalmologists and optometrists [38].

Diabetes affects 415 million people worldwide, and this number is set to rise beyond 642 million in 22 years [43]. Yet, with 46.5% of cases currently undiagnosed, a vast amount of people with diabetes are progressing towards complications unawares [43]. In addition, three quarters of diabetics live in low and middle income countries that lack the luxury of making the recommended annual examination. In particular, more than 13 million people in Brazil have diabetes, and this number may increase to 19.2 million in 2035.

For progressive diseases, early diagnosis has a huge impact on prognosis, allowing corrective or palliative measures before irreversible organ damage takes place. In the case of DR, early detection is crucial to prevent vision loss. Therefore, screening patients for early signs of DR pathology is important to prevent the disease or limit its progression. The World Health Organization and professional organizations such as the American Academy of Ophthalmology recommend eye examinations at least once a year for diabetic patients [69]. However, in disfavored, rural or isolated communities, the access to healthcare professionals — particularly to ophthalmology specialists — is difficult or not possible, therefore reducing opportunities for early detection and timely treatment of DR. In practice, communities often lack screening opportunities to adopt frequent consultations and a continuous follow-up program [42].

Without early diagnosis and treatment, DR progresses from mild nonproliferative DR to moderate and severe nonproliferative DR before the occurrence of proliferative DR, in which there is growth of abnormal new retinal blood vessels [118]. In its distinct stages, DR is characterized by the presence of red (microaneurysms and hemorrhages) and bright (hard exudates, cotton wool spots) lesions as well as neovascularization. Figure 1.1 shows some examples of retinas with signs of DR in the mild and moderate nonproliferative stages, proliferative stage as well as with diabetic macular edema.

In this century, the research and development of a large and varied set of automated

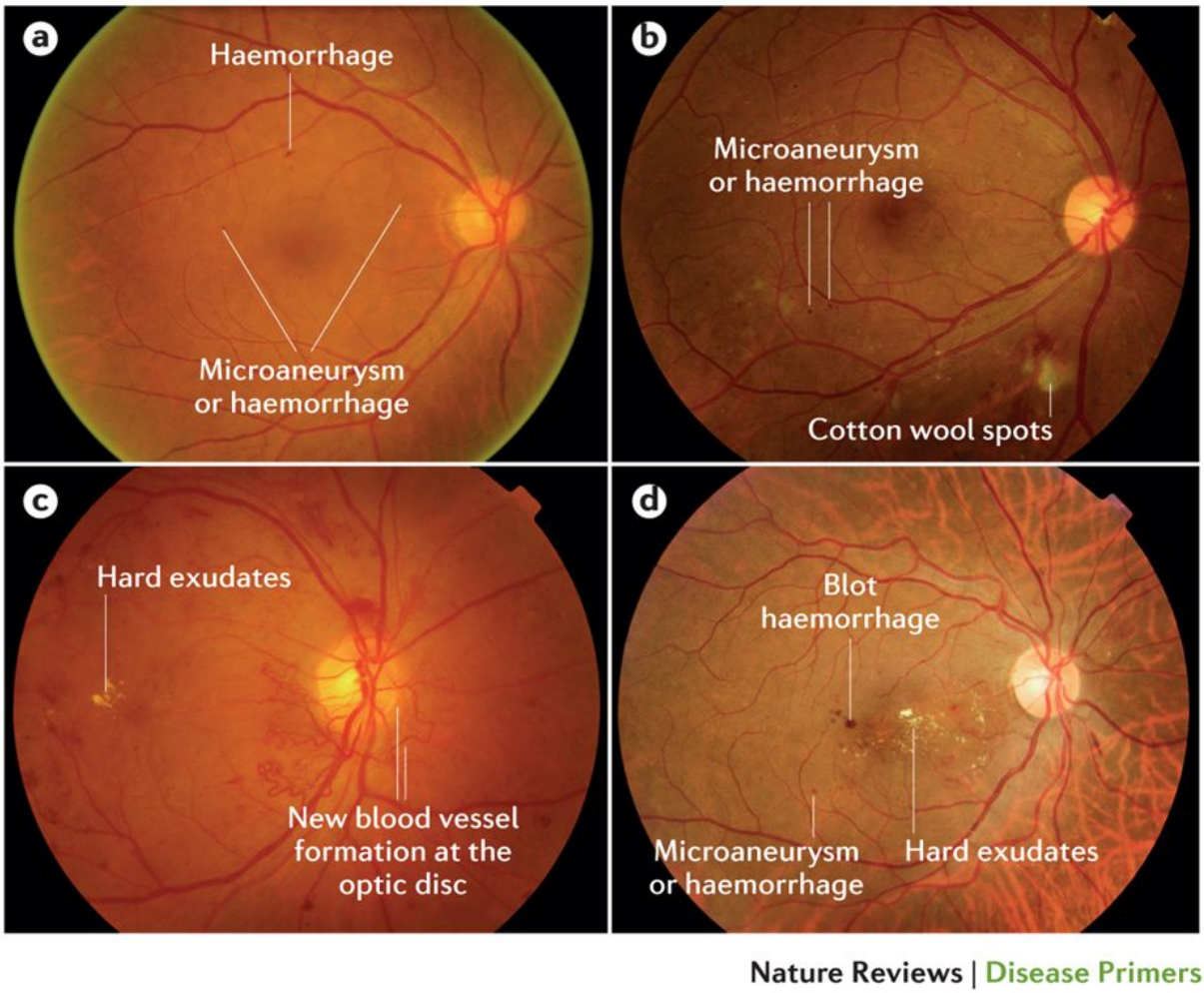


Figure 1.1: Diabetic retinopathy lesions and stages. a) Mild nonproliferative DR with microaneurysms and hemorrhages. b) Moderate nonproliferative DR with microaneurysms, hemorrhages and cotton wool spots. c) Proliferative DR showing neovascularization (new blood vessels) at the optic disc. d) Diabetic macular edema showing hard exudates at the fovea centre. Extracted from [118] – license CC BY (see Appendix A).

systems for diabetic retinopathy detection is trying to improve that scenario [4, 28, 29, 32, 35, 66, 67, 100]. Those researches are based on handcrafted features, whose manual extraction pipeline must ensure the extracted features are robust to the variances in the objects. The existing handcrafted proposals focus on the detection of DR signs using specific visual characteristics that vary among the different lesions [28, 29, 32, 35, 66, 67]. More recently, a few unified DR lesion detectors have been proposed, using novel handcrafted approaches that can easily be adapted to many kinds of retinal lesions [46, 49, 80, 87]. Unified approaches have evolved into advanced data-driven methods, that have pushed ahead the research advances in lesion detection and screening of diabetic retinopathy [84]. In data-driven approaches, the learning procedure is compelled by data distribution, rather than by intuition or prior knowledge regarding the data.

Computer-aided diagnosis may solve the dilemma of lacking screening opportunities by automatically deciding who should be referred to an ophthalmologist for further investigation. However, in general, the automated system that adopts the handcrafted manner

must identify a specific type of lesion that occurs both in isolation and in combination with other types of lesions, and make accurate decisions on the need to refer the patient to a specialist for further assessment.

Automatically deciding on whether or not the patient should be referred to the ophthalmic specialist is difficult, and a hotly debated research aim [1, 2, 12, 18, 25, 34, 41, 80, 84, 86, 101]. Automated or semi-automated detection of referable retinopathy is an important tool for managing the dispensation of care, reducing the specialist's workload and identifying patients in more critical need of ophthalmic review and treatment. Referability must consider not only the presence/absence of individual lesions, but directly or indirectly aggregate all signs and provide a succinct prediction for the need of further assessment and referral for DR and follow-up by a retinal specialist. The indirect discriminative pattern aggregation involves data-driven approaches, whose methodology aims at gathering enough information directly from the available data in high-level abstractions in order to design effective models capable of approximating very complex functions.

At the time we have started this work, most handcrafted methods for referral decisions were based on detecting distinct DR lesions (in most cases, employing different *ad hoc* models for each type of lesion) and then adding an additional decision layer for combining the scores of the lesion detectors and making a final decision [1, 12, 25, 80, 82, 86]. Those models are complex and cumbersome to implement. In addition, referral prediction accuracies were often limited.

Those hierarchical approaches ("lesion-first, referral-later") assume that preliminary lesion detection is necessary and sufficient for the later decision on referability. Those assumptions are questionable: although conventional machine-learning techniques often demand those staged decisions, current art based on data-driven methodology (deep learning) — so called image-based decisions — infer directly from the pixels [55]. Moreover, a preliminary stage of lesion detection discards information that may prove useful for the later stage of referability. Although such image-based systems raise legitimate concerns about spurious associations with confounders in training data, and about their sensitivity to adversarial images [2, 40, 65, 106], they have become the *de facto* gold standard for visual recognition tasks.

However, as the capacity of decision methods progresses and becomes practically autonomous, it also becomes hard to understand and interpret them. These concerns are notably prevalent in medical imaging — it is important not only to provide robust and accurate decision-making, but also intuitive explanations about *how* and *why* a particular decision was taken. These concerns were heightened recently with efforts for ensuring machine-learning accountability expressed as strategic notes by the European Political Strategy Centre [16] and standards by the scientific community [22].

1.1 Questions and Goals

Our work is divided into three meaningful paths: *Diabetic retinopathy lesion detection*, *Referable diabetic retinopathy detection* and *Validation with images from portable devices*. For the two first paths, we proposed novel handcrafted mid-level (two-tier) image repre-

sentations. Regarding automated examination of need of referral we go deeper and also investigate and propose novel data-driven models. We finally showcase an accountable and effective screening approach by using rich and comprehensive local- and global-based representations. For the third path, we optimize a data-driven model to evaluate the quality of images captured with low-cost devices.

For providing effective computer-aided diagnosis, effective image representations are mandatory. We use a two-tiered image representation method that rests upon the extraction of low-level local features and their aggregation into mid-level Bag-of-Visual-Words (BoVW) representation. The mid-level BoVW consists of two operations: the *coding* of the low-level feature vectors using the codebook, and the *pooling* of the codes, which are combined into a single aggregated feature vector [13].

Diabetic Retinopathy Lesion Detection

Given our interest in detecting individual DR lesions by using novel representation methods, we investigate the following questions keeping in mind our main purpose of unified screening solutions:

Q1.1: Can we combine the advantages of both *hard* and *soft* codings and provide a good balance for designing efficient and effective DR lesion detectors?

Q1.2: Can we preserve information in the pooling process and still obtain satisfactory results?

Q1.3: Can we improve the results employing global codebooks?

Q1.4: Can we diagnose patients in an ethnic group (e.g., isolated indigenous communities), even using models trained with data from other different ethnic groups?

Referable Diabetic Retinopathy Detection

Given our interest in preserving relevant information and providing non-lesion-dependent methods for referral decisions, we elaborate the following questions to guide and justify a research to such direction:

Q2.1: Can we forgo the detection of individual DR lesions and still have an effective referral decision?

Q2.2: Can sophisticated handcrafted mid-level features improve direct referral decision?

Q2.3: Can we confirm the suitability of handcrafted direct referral in a second, independent, dataset?

Q2.4: Can we learn in the data-driven manner highly abstract features that leverage referable DR detection without the need of manual feature engineering?

Q2.5: Can we diagnose retinal images collected under different acquisition conditions with data-driven models?

Q2.6: Can we transfer knowledge acquired with a different but similar task in the context of diabetic retinopathy screening?

Q2.7: Can we provide an effective and accountable data-driven solution for referable diabetic retinopathy detection?

Q2.8: Can we enhance the performance of the referral model by applying two-tiered data-driven image representation?

Q2.9: Can we combine global data-driven and local saliency-oriented two-tier representations for more accurate decision-making?

Validation with images from portable devices

Given our interest in integrating our software solutions with simple retinal imaging devices, we elaborate the following question to guide our research toward validation over a new dataset captured with portable devices. This work involves only the validation in terms of quality assessment since the dataset for presence/absence of referable diabetic retinopathy is still under preparation at the time of writing — collection in hospitals and annotation by experts.

Q3.1: Can we assess the quality of images from portable devices with data-driven models trained with images from high-cost instruments?

The investigations we perform to answer the questions from **Q1.1** to **Q1.4** and from **Q2.1** to **Q2.3** involve handcrafted methodologies, while from **Q2.4** to **Q2.8** and the question **Q3.1** we explore data-driven representation techniques. Finally, in the **Q2.9** question we combine both approaches.

Concerning our goals expressed by the questions above, some of them devised in the beginning of the work and others reformulated along the advances of the research, our main purpose under the scientific viewpoint was providing experts with higher-level retinal image representations that preserve valuable visual characteristics. We aimed at extracting local features by using supervised learning with cutting-edge convolutional neural networks, and incorporating them in advanced mid-level representations. By doing so, we come up with an entire framework able to extract this abstract learned information from retinal images, use it as a local description in a mid-level representation, and explore different classification methods such as Neural Networks and Random Forest for producing reliable computer-aided diagnoses.

The predominant aim was enhancing the results reported in the literature by designing and reporting methods for handcrafted two-tiered image representation, composing and delineating sophisticated data-driven models employing novel and also consolidated techniques, and ultimately incorporating supervised learning of features with advanced and saliency-oriented mid-level image representations.

Additionally, our secondary goals included, but were not limited to:

- Providing online and accessible retinal image classification software packages. Our expectation was producing a system that empowers the user to upload a retinal image and receive, in real time, the diagnostic about DR lesions and DR referral.

- Investigating simpler solutions suitable for the context of an embedded system, in which the image acquisition process and the computational resources tend to be limited.

- Integrating our software solutions with simple retinal imaging devices. The Brazilian startup Phelcom Technologies[1], for instance, is producing the Eyer, that aim at transforming the smartphone in a portable and connected retinal camera. This acquisition method, combined with our approaches, might result in a powerful and low-cost apparatus for eye examination.

1.2 Awards and Partnerships

In the course of this work, we established relevant partnerships and gathered significant awards. Our project received the Google Latin American Research Awards twice. With that compelling award, we became a Google's partner. Our proposal to participate in these prestigious awards concerns gathering enough information directly from the available data in order to design a more effective, unified, and less human-centered classification system. We intended to design methods for image representation, incorporating supervised learning of features with advanced mid-level image representations.

The project was awarded in 2016 and again in 2017. In 2017, there were 281 projects submitted but only 27 awarded[2].

During the last year of the doctoral program, we have also partnered with the aforementioned startup Phelcom Technologies. The main purpose was combining the developed image analytics techniques for DR detection with the Phelcom's hardware solutions for portable and low-cost image acquisition. We set forth a partnership, in which we mentored the startup by implementing solutions suitable for embedded systems, and validating with images from the devices. The scope of the partnership program with Phelcom encompasses:

1. Investigate the assessment of images captured with Phelcom devices using the techniques and methods developed in scope of the research on diabetic retinopathy in the IC/Unicamp under the coordination of Prof. Anderson Rocha.

2. Investigate the adaptation of existing solutions from this thesis, and also ones recently published in the literature, in the context of diabetic retinopathy for lesion detection and need of referral for the images acquired by Phelcom.

[1] `https://www.phelcom.com.br/en/home_page` (accessed February 13, 2019)

[2] `https://brasil.googleblog.com/2017/08/conheca-os-vencedores-do-programa-de.html` (accessed February 13, 2019)

3. Investigate the feasibility of implementing techniques developed in this thesis and published in the literature on the image acquisition equipment produced by the company.

1.3 Scientific Contributions

We envision this research contributing to the areas of Computer Vision, Machine Learning, Medicine, and Biomedical Engineering. Our expectations towards scientific contributions embraced:

- Advance the state of the art by proposing new effective and robust image representation approaches, preserving relevant visual and structural information for improving retina image analysis.
- Advance the approaches for retina image analysis by proposing efficient and effective computer-aided methods for higher-level decisions (i.e., referable diabetic retinopathy detection), discarding the heretofore indispensable lesion detection task.
- Advance the state of the art by designing a two-tiered image representation that incorporates supervised learning of features with advanced mid-level image representations for improving retina image analysis.
- Integrate our solutions regarding DR lesion detection and referral decision with simple and portable retinal imaging solutions.

With these contributions, we expected to push the current retina research boundaries forward, providing experts and practitioners cutting-edge technology for daily activities in their work.

1.4 Text Organization

This thesis is organized as follows. Chapter 2 surveys the state of the art of the main topics related to handcrafted and data-driven image characterization/classification (Sections 2.1 and 2.2) and retinal image analysis: DR Lesion Detection (Section 2.3) and Referable DR Detection (Section 2.4), Retinal Image Quality Assessment (Section 2.5) and Accountable Retinal Image Analysis (Section 2.6). Chapter 3 presents the proposed methodologies of the work. Chapter 4, presents the adopted experimental protocol (datasets, validation protocol and metrics). Chapters 5 and 6 present the results which respond the questions regarding lesion DR detection and referable DR detection, respectively. Chapter 7 shows initial advances into integrating the solution with a low-cost camera, with results for quality assessment. Finally, Chapter 8 summarizes some advances and discoveries, and concludes the work.

Chapter 2

State of the Art and Related Concepts

We divide the state of the art and related concepts into six sections. Section 2.1 describes the Bag of Visual Words and its extensions for mid-level image representation, while Section 2.2 introduces the concept of Convolutional Neural Networks. The next three sections survey the state of the art for retinal image analysis: DR Lesion Detection, Referable DR Detection, and Retinal Image Quality Assessment. Section 2.3 presents some long-established methods as well as the new characterization concept used in recent works regarding lesion detection. Section 2.4 discusses the traditional approach for referral assessment based upon the presence, location and number of lesions. Section 2.5 shows some handcrafted and data-driven proposals for evaluating the quality of fundus images. Finally, Section 2.6 introduces the concepts related to accountable machine learning and presents some recent works concerning comprehensive methods towards diabetic retinopathy detection.

2.1 Bag of Visual Words and Extensions

A two-tiered feature extraction scheme, based upon the creation of an aggregation of encoded local features became a staple of the image classification literature. The technique was popularized by the work of Sivic and Zisserman [97], who made explicit an analogy with the traditional bag-of-words representation used in information retrieval [9]. That formalism from information retrieval is reformulated for local image descriptors as "visual words" by associating the low-level local features with the elements of a codebook, which is aptly named a "visual dictionary". The number of visual words for a given image is represented as a histogram named bag of visual words (BoVW), and used as a mid-level representation.

Learning the codebook is a challenge for BoVW representations. The traditional way involves unsupervised learning over a set of low-level features from a training set of images. K-means clustering [60], for example, can be used on a sample of these features and the k centroids be employed as codewords. There is also considerable variation throughout the literature on the *size* of the codebook, ranging from a few hundred codewords up to hundreds of thousands.

The metaphor of "visual word" should not be taken too literally. While textual words

are intrinsically semantic, visual words are usually appearance-based only. Moreover, the BoVW model was considerably extended since the seminal work of Sivic and Zisserman. New ways of encoding the local descriptors using the codebook were proposed, as well as new ways of aggregating the obtained codes. This stretched the metaphor of "visual word" too much, and a more formal model was proposed by Boureau et al. [13], making explicit the operations of *coding* and *pooling*. Therefore, the BoVW formalism evolved into a meta-model for which myriads of variations are possible, based upon the combinations of low-level descriptors, codebook learning, coding and pooling.

The coding and pooling operations can be conveniently understood in the matrix form proposed by Precioso and Cord (see Figure 2.1, adapted from [8, 83]). Their formalism starts with the choice of the codebook (e.g., by sampling or learning on the low-level feature space) as an indexed set of vectors, $\mathcal{C} = \{\mathbf{c}_i\}$, $i \in \{1, \ldots, M\}$, where $\mathbf{c}_i \in \mathbb{R}^d$. Then, the low-level local features for each image, which are represented by the index set $\mathcal{X} = \{\mathbf{x}_j\}$, $j \in \{1, \ldots, N\}$, where $\mathbf{x}_j \in \mathbb{R}^d$ is a local feature and N is the number of salient regions, points of interest, or points in a dense sampling grid on the image are extracted. The final BoVW vector representation encodes a relationship between X and C [8, 13].

$$
\mathbf{H} = \begin{array}{c} \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_m \\ \vdots \\ \mathbf{c}_M \end{array}
\begin{array}{c}
\begin{array}{ccccccc} \mathbf{x}_1 & \ldots & \mathbf{x}_j & \ldots & \mathbf{x}_N \end{array} \\
\left[\begin{array}{ccccc}
\alpha_{1,1} & \ldots & \alpha_{1,j} & \ldots & \alpha_{1,N} \\
\vdots & & \vdots & & \vdots \\
\alpha_{m,1} & \ldots & \alpha_{m,j} & \ldots & \alpha_{m,N} \\
\vdots & & \vdots & & \vdots \\
\alpha_{M,1} & \ldots & \alpha_{M,j} & \ldots & \alpha_{M,N}
\end{array} \right] \\
\Downarrow \\
f : coding
\end{array}
\Rightarrow g : pooling
$$

Figure 2.1: The BoVW model illustrated in a convenient matrix form, highlighting the relationships between the low-level features $\mathbf{x}_j$, the codewords $\mathbf{c}_m$ of the visual dictionary, the encoded features α_m, the coding function f and the pooling function g.

The coding step transforms the low-level descriptors into a representation based upon the codewords, which is better adapted to the specific task and preserves relevant information, while discarding noise. Coding can be modeled by a function f: $\mathbb{R}^d \rightarrow \mathbb{R}^M$, $f(\mathbf{x}_j) = \alpha_j$ that takes the individual local descriptors $\mathbf{x}_j$ and maps them onto individual codes α_j. The classical BoVW model employs the "hard assignment" of a low-level descriptor to the closest codeword, and can be modeled by:

$$\alpha_{m,j} = 1 \text{ if } m = \arg\min_k \|\mathbf{c}_k - \mathbf{x}_j\|_2^2 \quad \text{else} \quad 0 \tag{2.1}$$

where $\alpha_{m,j}$ is the m^{th} component of the encoded descriptor.

Recent publications [13, 111], however, suggests that "soft" assignment schemes, which allow degrees of association between the low-level descriptors and the elements of the codebook, work better, avoiding both the boundary effects and the imprecision of hard

assignment [111]. Soft assignment scheme can be modeled by:

$$\alpha_{m,j} = \frac{K_\sigma(\|\mathbf{c}_m - \mathbf{x}_j\|_2)}{\sum_{\mathbf{c}\in\mathcal{C}} K_\sigma(\|\mathbf{c} - \mathbf{x}_j\|_2)}, \tag{2.2}$$

where $\mathbf{c}_m$ represents the codewords of the visual dictionary, $\mathbf{x}_j$ is the low-level local features for each image, and K_σ is the Gaussian kernel.

The pooling step takes place after the coding, and can also be represented by a function g: $\{\alpha_j\}_{j\in 1,\dots,N} \to \mathbb{R}^M$, $g(\{\alpha_j\}) = \mathbf{z}$. The classical BoVW corresponds to a "counting of words" (called sum-pooling) and can be modeled as:

$$g(\{\alpha_j\}) = \mathbf{z} : \ \forall m, \ z_m = \sum_{j=1}^{N} \alpha_{m,j} \tag{2.3}$$

This simplistic pooling has been criticized, and taking the maximum activation of each codeword (in a scheme aptly named max-pooling) is often much more effective [14]:

$$g(\{\alpha_j\}) = \mathbf{z} : \ \forall m, \ z_m = \max_{j\in\{1,\dots,N\}} \alpha_{m,j} \tag{2.4}$$

The vector $\mathbf{z} \in \mathbb{R}^M$ obtained from pooling is the BoVW representation, which is used for classification.

There are a number of choices to normalize the BoVW vector. For example, in the classical BoVW scheme, ℓ_1-normalization is often employed to turn a vector of occurrences into a vector of relative frequencies.

In the following, we overview two recent mid-level representation methods, alternatives to the BoVW: BossaNova [8] and Fisher Vector [74].

2.1.1 BossaNova

In order to keep more information than the BoVW during the pooling step, BossaNova [8] introduces a density-based pooling strategy, which computes the histogram of distances between the local descriptors and the codewords. More formally, BossaNova pooling function g estimates the probability density function of α_m: $g(\alpha_m) = \mathrm{pdf}(\alpha_\mathbf{m})$, by computing the following histogram of distances $z_{m,b}$:

$$\begin{aligned} g \ &: \ \mathbb{R}^N \longrightarrow \mathbb{R}^B, \\ &\quad \alpha_\mathbf{m} \longrightarrow g(\alpha_m) = z_m, \\ z_{m,b} \ &= \ \mathrm{card}\left(\mathbf{x}_j \mid \alpha_{m,j} \in \left[\frac{b}{B}; \frac{b+1}{B}\right]\right), \\ &\quad \frac{b}{B} \geq \alpha_m^{min} \ \text{ and } \ \frac{b+1}{B} \leq \alpha_m^{max}, \end{aligned} \tag{2.5}$$

where N denotes number of local descriptors in the image, B indicates the number of bins of each histogram z_m, $\alpha_{m,j}$ represents a dissimilarity (i.e., a distance) between codeword $\mathbf{c}_m$ and descriptor $\mathbf{x}_j$, and $[\alpha_m^{min}; \alpha_m^{max}]$ limits the range of distances for the descriptors

considered in the histogram computation.

In addition to that pooling strategy, Avila et al. [8] also proposed a localized soft-assignment coding that considers only the k-nearest codewords for coding a local descriptor, and keeps the representation compact.

In comparison to the standard BoVW representation, BossaNova significantly outperforms BoVW on many challenging image classification benchmarks [7, 8].

2.1.2 Fisher Vector

Fisher Vector [74] is the mid-level image representation with consistently best results in computer vision literature [17, 91]. Based upon the idea of Fisher information vectors [44] in the parametric space of Gaussian Mixture Models (GMM) estimated over the whole set of images, it extends the BoVW paradigm by encoding first- and second-order average differences between the descriptors and codewords.

Furthermore, Fisher Vector is a compact representation, since much smaller codebooks are required in order to achieve a good classification performance in general vision problems.

Formally, given a GMM with N Gaussians, let us denote its parameters by $\lambda = \{w_i, \mu_i, \sigma_i, i = 1 \dots N\}$, where w_i, μ_i and σ_i are respectively the mixture weight, mean vector and diagonal covariance matrix of Gaussian i. In the Fisher Vector framework, the D-dimensional descriptor $\mathbf{x}_j$ is encoding with a function $\Phi(\mathbf{x}_j) = [\varphi_1(\mathbf{x}_j), \dots, \varphi_N(\mathbf{x}_j)]$ into a $2ND$-dimensional space where each function $\varphi_i(\mathbf{x}_j)$ is defined by:

$$\begin{aligned} \varphi_i(\mathbf{x}_j) &: \mathbb{R}^D \longrightarrow \mathbb{R}^{2D}, \\ \varphi_i(\mathbf{x}_j) &= \left[\frac{\gamma_j(i)}{\sqrt{w_i}} \left(\frac{\mathbf{x}_j - \mu_i}{\sigma_i} \right), \frac{\gamma_j(i)}{\sqrt{2w_i}} \left(\frac{(\mathbf{x}_j - \mu_i)^2}{{\sigma_i}^2} - 1 \right) \right], \end{aligned} \tag{2.6}$$

where $\gamma_j(i)$ denotes the soft assignment of descriptor $\mathbf{x}_j$ to Gaussian i.

We evaluate the Fisher Vector mid-level features given that it offers a more complete representation of the sample set, which we believe would be important for DR analysis. To the best of our knowledge, it has never been applied to this problem. More details on Fisher Vector representation can be found in [74, 91].

2.2 Convolutional Neural Networks

For decades, constructing a pattern recognition system with conventional approaches required a considerable domain expertise in order to convert raw data (such as size, shape, color and location of retina lesions) into a proper representation; and a posterior learning stage with algorithms capable of recognizing categorical patterns (e.g., whether or not a patient needs a follow-up).

Deep learning composes non-linear modules that, starting with raw data, transform the representation in one level into a more abstract representation [55]. With enough data, the multiple-level composition enables to learn very complex functions in an end-to-end manner: represent the data and categorize.

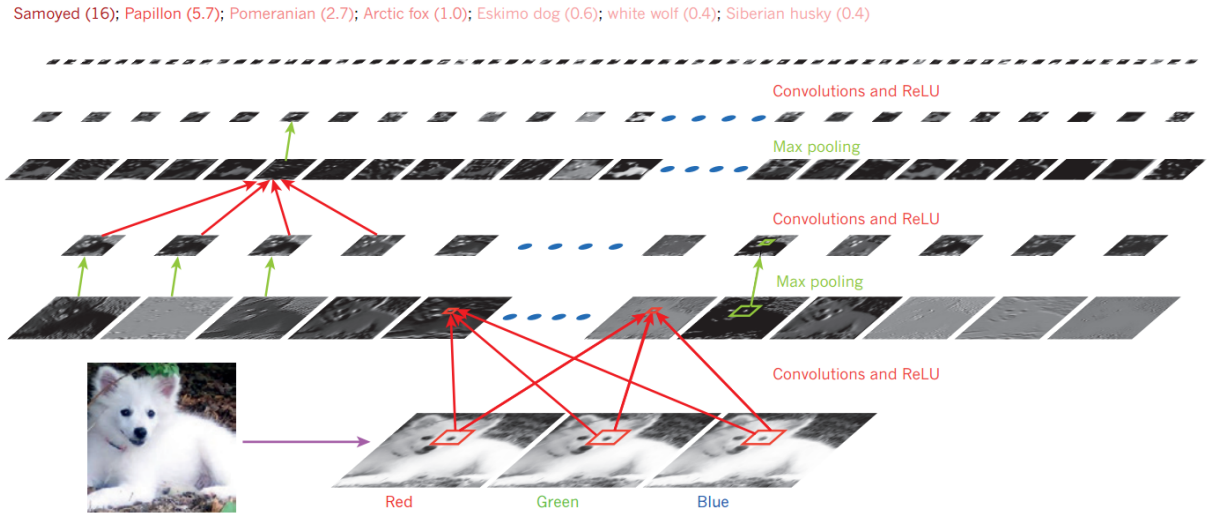


Figure 2.2: Representation of a Convolutional Neural Network. Example applied to a raw image with red, green, and blue (RGB) color channels. Each convolutional layer has a bank of filters (not represented in the image) and outputs a feature map. Information flows bottom up, with lower-level features, and a score is computed for each image class in output. Extracted from [55] – license CC BY (see Appendix A).

Convolutional Neural Networks (CNN) are deep-learning methods designed to deal with data in form of multiple arrays such as images. Inspired by biological processes and strongly founded on mathematical convolution operations, the first popular CNN was proposed in 1995 by Lecun and Bengio [54].

In CNNs, convolutional layers play the role of non-linear modules. Units within a convolutional layer are connected to local patches of its input (that might correspond to output of a previous layer) through a set of weights, and the weighted-summed local patches pass through a non-linearity. Those layers have the role of detecting local conjunction on features from previous layers [55]. Pooling layers (and its numerous variations) have the role of merging neighbouring features into one. It is expected that those neighbour features are semantically similar.

Commonly, convolutional, non-linear and pooling layers are stacked, followed by some dense layers (each unit is fully-connected to the input, followed by a non-linear unit). Weights of convolutional and dense layers are traditionally learned by using simple stochastic gradient descent – or more advanced optimization algorithms – and the parameter updating is performed by backpropagation procedure [55].

Figure 2.2 shows the architecture of a typical CNN, whose structure is composed of a series of stages [55]. As exposed, the first stages in the CNN are either convolutional layers or pooling layers. Convolutional units are organized as a feature map (represented horizontally) resulting from a convolution operation applied to the previous layer through a set of weights called a filter bank. The result of this convolution is passed through a non-linearity function (e.g., rectified linear unit). Pooling layers merges semantically similar features into one, typically computing the maximum of a local patch of units in one feature map.

2.2.1 Transfer Learning

Transfer learning aims to transfer knowledge learned in one or more source tasks to improve the learning process in a target task. Normally, the transfer is sought when there are not enough training samples in the target domain or when there is already a reasonable solution for a related (source) problem and it would be natural to leverage such knowledge while solving the target problem. Most of the existing work that exploit transfer of skills/knowledge implicitly assume that the source and target domains are related to each other [70].

Transfer learning normally appears in two distinct scenarios: feature extraction and fine-tuning. In the former case, we freeze network layers and correspondent parameters and use them just to extract features to the target task. In practice, this corresponds to remove the last dense layer and treat the CNN as feature extractor that will be followed by a classifier (e.g., Support Vector Machine) for the new dataset.

In turn, the latter case involves continuing the backpropagation, updating of the network internal parameters to better adjust them to the target domain. In this situation, it is possible and sometimes recommended to keep the earlier layers frozen, since they contain generic features that recognize edges and color blobs. Assuming the domains are similar, smaller learning rates might be used since the pre-learned weight can be distorted too quickly and too much.

2.3 Diabetic Retinopathy Lesion Detection

Although obtaining high sensitivities and specificities, in the beginning of the doctoral program the state of the art on aided diagnosis of DR used to be specialized for a specific type of lesion [6, 29, 32, 35, 36, 48, 53, 67, 90, 96, 100, 115, 121]. For bright lesion detection, sensitivities range from 70.5 to 100.0% and specificities from 84.6 to 99.7% [32, 36, 67, 90, 96, 100, 115]; for red lesion detection, sensitivities range from 77.5 to 97.0% and specificities from 83.1 to 88.7% [29, 48, 96]. A summary of results found in the literature is presented in Tables 2.1 and 2.2.

The development and implementation of single-lesion algorithms is a limitation for accurate referral as, in general, a method developed for one lesion cannot be directly applied to other lesions, preventing the development of a general framework for multi-lesion detection and referral. In order to overcome this, researches have proposed several multi-lesion schemes. Li et al. [57] implemented a real-time management tool for diabetic eye disease that focuses on the two main DR-related lesions: microaneurysms and hard exudates. However, their framework does not exploit a unique technique for the detection of both lesions simultaneously. Lesions are first detected using several image analysis criteria including texture measurements. This provides a content-based image retrieval framework once the microaneurysms and exudates have been detected in each image. The information is grouped together and a complete description of the retinal image is created as query, which is then compared to a database of past images with known diagnoses.

Another common limitation of using current algorithms for DR detection and classification is the need for complex and *ad hoc* pre- and post-processing of the retinal images,

Table 2.1: State of the art for the detection of bright lesions.

Work	Sens	Spec	AUC	Dataset	Approach
Sinthanayothin et al. [96]	88.5%	99.7%	–	30	Recursive Region-Growing Segmentation (RRGS) and thresholding
Niemeijer et al. [67]	95.0%	88.0%	95.0%	300	Each pixel is classified in a so-called lesion probability map
Sánchez et al. [90]	100%	90%	–	80	Mixture models and dynamic threshold for segmentation, followed by a postprocessing to distinguish the lesions
Giancardo et al. [36]	–	–	88.0%*	169 + 1200 + 89**	Features based on color, wavelet decomposition and exudate probability. Several classification algorithms
Fleming et al. [32]	95.0%	84.6%	–	13219	Multiscale morphological process followed by thresholding
Sopharak et al. [100]	80.0%	99.5%	–	60	Mathematical morphology methods followed by thresholding
Welfer et al. [115]	70.5%	98.8%	–	89	Mathematical morphology methods and thresholding

* AUC obtained for training on HEI-MED dataset and testing on Messidor dataset.
** HEI-MED, Messidor and ROC datasets, respectively

Table 2.2: State of the art for the detection of red lesions.

Work	Sens	Spec	AUC	Dataset	Approach
Sinthanayothin et al. [96]	77.5%	88.7%	–	23	RRGS in green channel
Jelinek et al. [48]	97.0%	88.0%	–	758	A microaneurysm (MA) detector notes the number of MAs and dot hemorrhages detected
Fleming et al. [29]	85.4%	83.1%	90.1%	1441	MA detection with emphasis on the role of the local contrast normalization
Giancardo et al. [37]	–	–	–	100*	Microaneurysm detection with Radon Cliff Operator
Antal and Hajdu [6]	–	–	90.0%	100*+ 120**	Combination of internal component of MA detectors
Lazar et al. [53]***	–	–	–	100*	Statistical measures of attributes on peaks are used in naïve Bayes classification. Scores are thresholded for a binary output
Zhang et al. [121]***	–	–	–	100*	Multiscale Correlation Filtering (MSCF) and dynamic thresholding for intensity-based detection and localization
Sánchez et al. [90]***	–	–	–	100*	Statistical approach based on mixture model-based clustering and logistic regression

* ROC dataset.
** Messidor dataset.
*** The authors used the performance measure applied in the Retinal Online Challenge.

depending on the lesion of interest. The pre- and post-processing address issues like image acquisition and field-of-view variations, or adaptations to take ethnicity of the patients into account [23, 35]. Pre-processing of retinal images may include standardizing the resolution of the image, normalizing color, segmenting and removing blood vessels [99], and detecting and removing the optic disk [5, 29]. For this task, morphological operators [39] are often employed as part of the pre-processing step [32, 100, 115].

Research in automated retinal lesion classification is becoming more general, bypassing the need for pre- and post-processing. Rocha et al. [87] proposed a unified framework for detection of both hard exudates and microaneurysms. The authors introduced the use of BoVW representations for DR lesion detection, creating a framework easily extendible to different types of retinal lesions. However, the BoVW model employed in that work was simple and chosen without any theoretical or experimental design analysis but rather from experimental results in other fields of image analysis [97]. This has opened up the opportunity for substantial improvements, which are explored in the doctoral program.

2.4 Referable Diabetic Retinopathy Detection

The simple presence of DR lesions is insufficient to warrant referral. For instance, a small number of microaneurysms in a safe region of the retina might be considered mild nonproliferative DR, without need of referral.

Following the International Clinical Diabetic Retinopathy (ICDR) severity scale [117], formulated by a consensus of international experts to standardize and simplify DR classification, a person is deemed to have referable DR if either or both eyes has moderate or severe nonproliferative DR, or proliferative DR, macular edema, or both. A person is deemed to have nonreferable DR if there is no nonproliferative DR or mild proliferative DR and no macular edema in both eyes.

Referable diabetic retinopathy detection is currently addressed in different ways [1, 2, 12, 25, 34, 76, 82, 84, 86]. Solutions vary from custom-tailored handcrafted lesion detectors (sometimes encoding expert-domain knowledge) to mid-level representations (combining a series of low-level descriptors) to data-driven lesion detectors (underpinned by neural network advances). In this section, we overview existing approaches to detect referable diabetic retinopathy.

We start with methods based upon the detection of individual lesions. Then, we discuss some approaches that exploit data mining to diagnose retinal images. Thus, we present some works which integrates telemedicine systems with automated screening systems. In sequence, we describe some recent methods that evaluate referral using deep learning. Afterwards, we show referral solutions that require the lesion detection but alleviate the complexity of traditional approaches employing unified detection techniques. Table 2.3 shows important prior work regarding referable diabetic retinopathy detection.

Many techniques for referral assessment rely upon examining the positioning of lesions on retinal landmarks, are complex and, frequently, tailored to each lesion. Some works rely on the number of microaneurysms [101], although their presence in one quadrant of the retina characterizes only mild nonproliferative diabetic retinopathy, without the

Table 2.3: State of the art for referable diabetic retinopathy detection.

Work	**AUC (%)**	**Dataset**	**Approach**
Decencière et al. [25], 2013	81.8*	12,500 examinations	diagnosis rules from decisions regarding exudates, microaneurysms and hemorrhages + contextual information
Quellec et al. [86], 2016	84.4	25,702 examinations	diagnosis rules from multigranular visual word histograms + contextual information
Bhaskaranand et al. [12], 2015	88.3	434,023 images/54,324 participants	image enhancement, followed by identification and description of regions that are classified into microaneurysms, exudates and hemorrhages, and posteriorly used for referral
Abràmoff et al. [1], 2013	93.7	1,748 images/874 participants	combining information about quality and scores of exudates, microaneurysms, hemorrhages, cotton-wool spots, and neovascularization
Abràmoff et al. [2], 2016	98.0	1,748 images/874 participants	integrating CNNs to detect normal retinal anatomy as well as the lesions (such as hemorrhages, exudates, and neovascularization)
Colas et al. [18], 2016	94.6	over 80,000 images/40,000 participants	using CNN to detect anomalies, diagnose the stages of DR (and referability), and provide the location of lesions
Gulshan et al. [41], 2016	99.0	over 128,175 images	ensembling 10 CNNs pretrained with ImageNet, and evaluating as referral the images classified as having moderate or worse DR, referable macular edema, or both
Gargeya and Leng [34], 2017	94.0	1,748 images/874 participants	combining CNN features and metadata features related to the retinal images, and a decision tree classifier
Quellec et al. [84], 2017	95.4	88,702 images/44,351 participants	training CNNs to detect referable DR, using a heatmap optimization procedure
Pires et al. [80], 2013	93.4	435 images	lesion-based approach with high-level features comprising scores from individual lesions, and referral decision with a SVM classifier
Pires et al. [82], 2014	94.2	435 images	lesion-based approach with high-level features comprising scores from individual lesions, and referral decision with a SVM classifier

* AUC reported by [86].

need of referral. Naqvi et al. [64] proposed a referral system for hard exudates in diabetic retinopathy using a robust mid-level representation of bag-of-visual-words (BoVW), which is reliable and easily adaptable to other lesions. However, they fall short of the aim of offering reliable referability assessment, ending up providing mostly a screening for exudates. Moreover, in the task of assessing referral, they did not follow any standard grading consensus regarding diabetic retinopathy level classification — necessary for a consistent improvement of communication among specialists [117].

Taking a different path, recent works used data mining to assess referral due to pathologies [25, 86]. Decencière et al. [25] combined visual information (retinal images) and contextual data from the individuals (e.g., patient age, weight or diabetes history) to detect retinal pathologies and to point out whether or not patients need referral to a specialist. The authors considered image quality metrics, diabetic retinopathy-related lesion information (exudates, microaneurysms, and hemorrhages), demographic and diabetes-related information.

Similarly, Quellec et al. [86] used multiple retina images and contextual information about the patient to detect abnormal retinas. Instead of detecting just one or more lesions related to a particular eye disease, they identify patients who need referral to an eye care provider, regardless of pathology. The method starts by mixing a set of retinal images and building a mosaic for each one. Then, they characterize the images using a BoVW model, extracting multi-granular histograms into a cascade of regions. After characterizing the images, their method extracts diagnosis rules relying on visual word histograms and contextual information previously collected.

In addition to those solutions, telemedicine systems also improved health care productivity and addressed the lack of access to diabetic retinopathy screening. Besides increasing local access, telemedicine programs provide risk stratification of diabetic patients so that those who require treatment can be scheduled more efficiently [108]. Automated screening systems integrated with telemedicine frameworks make diabetic retinopathy screening more accessible, efficient, and cost-effective [94] and a few such systems were proposed [1, 2, 12].

Bhaskaranand et al. [12] integrated a diabetic retinopathy screening solution that assesses severity and referral into a telemedicine system. The screening tool, called EyeArt, analyzes image quality (patients with collected ungradable images are also referred), and enhancing images so as to normalize and improve the appearance of the existing lesions. With a set of filterbank descriptors, the method identifies and describes regions with anatomical or pathological structures associated with specific lesions (microaneurysms, exudates, and hemorrhages). A supervised learning ensemble is employed at the very end for referral decision.

As exposed above, bags of visual words became a fundamental approach for image representation and is widely exploited for retinal image analysis. Rocha et al. [87] adopted a class-aware fashion (one codebook per class) especially suitable for diabetic retinopathy-related lesions. The approach was suitable to assess quality [49, 81], detect lesions [46, 64, 93], and assess referability [80, 82, 86].

Detecting lesions is usually the first stage of traditional referral assessment, and it tends to be specific for each lesion. The BoVW methodology allowed general frameworks

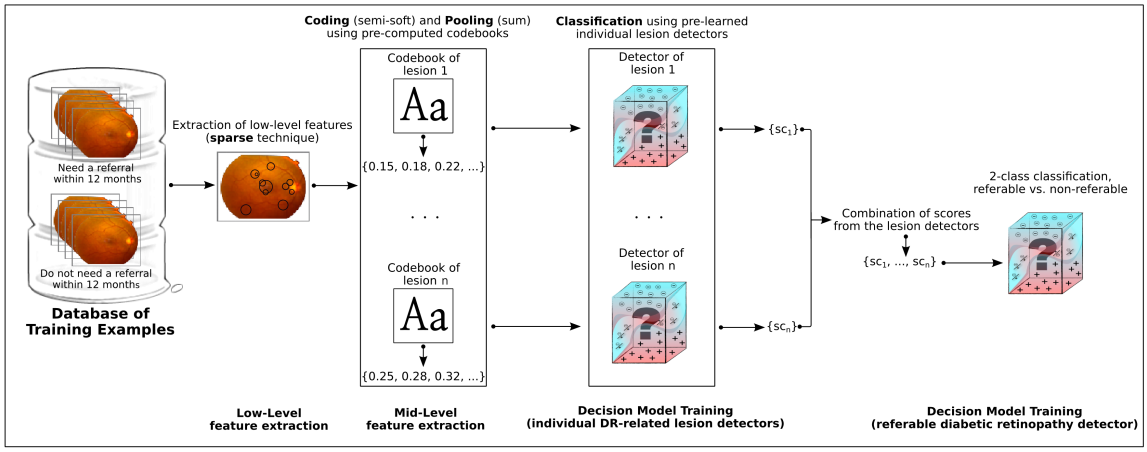


Figure 2.3: Pipeline of the lesion-based methodology for detection of referable diabetic retinopathy as described in [82].

adaptable for large classes of lesions [46, 49, 80, 93]. Pires et al. [80] applied that unified methodology to detect lesions with models based on BoVW mid-level features and SVM classifiers, and gathered the individual scores (per lesion) to referral decision making. The referability decision had an AUC of 93.4% [80].

In the first phase of the doctoral program, we improved this achievement to 94.2% by enhancing the lesion detectors with better mid-level image features [82]. Henceforward, we will refer that referability assessment proposal as state of the art and will call this as lesion-based methodology. The streamlined technique for lesion-based decision on referability is illustrated in Figure 2.3.

Abràmoff et al. [1] investigated the potential of the Iowa Detection Program (IDP) to detect referable diabetic retinopathy. The IDP is a framework for quality analysis and diabetic retinopathy lesion detection. In a per-patient setup (two images per patient), the IDP combines analysis of individual lesions, structures, and quality in a simple likelihood that encapsulates the patient's diagnostic about referable diabetic retinopathy.

In a follow-up work, Abràmoff et al. [2] integrated the IDP with a set of convolutional neural networks (CNNs) specialized on detecting hemorrhages, exudates, and neovascularization as well as normal retinal anatomy and image quality. Employing CNNs ranging from the well-known Alexnet [51] to VGG [95], the hybrid system significantly outperforms existing solutions at the task of diabetic retinopathy screening.

Unified approaches that capture discriminative patterns of distinct lesions alleviate the complexity of exploiting specific tailored visual characteristics. However, as any handcrafted technique, those approaches are subject to lose critical information that could be evinced in data-driven approaches to provide effective decisions. In this vein, showing the potential of a data-driven system over handcrafted counterparts, Gargeya and Leng [34] customized deep convolutional neural networks for extracting features to classify images into no DR vs. any stage of DR, and no DR vs. mild DR. Those features are combined with retinal image metadata for classification (original pixel height of the image, original pixel width of the image, and field of view of the original image).

Quellec et al. [84] trained CNNs to detect referable DR, using a heatmap optimization

procedure. To create heatmaps, the authors proposed a training method which involves a third pass on the CNN to propagate second-order derivatives forward. Those CNNs trained for image-level classification are also used to detect lesions related to DR (hard exudates, soft exudates, small red dots, hemorrhages).

Gulshan et al. [41] ensembled 10 CNNs with the Inception-v3 architecture [105], trained with the ImageNet dataset, to make multiple binary decisions such as (1) moderate or worse DR, (2) severe or worse DR, (3) referable diabetic macular edema, or (4) fully gradable. An image fits as referable if it fulfills criterion (1), criterion (3) or both.

Given that the doctoral program started when the state of the art for referable diabetic retinopathy detection relied upon handcrafted lesion-based methods with unified BoVW, our initial goals were centered on enhancing that scenario. Afterwards, we explored data-driven methodologies and fusion techniques.

2.5 Retinal Image Quality Assessment

The quality of the retinal images is of paramount importance for automated image analytics and a factor that successful and reliable computer-aided diagnostic models rely on. Assessing image quality has been discussed in the literature by a considerable number of authors [20, 31, 49, 52, 56, 61, 81, 89, 110] and represents an important limiting factor for automated diabetic retinopathy screening [71].

The fundus quality is subject to be reduced by artifacts such as eye lashes or dust specs on the lens, only part of the retina is seen, the image is out of focus or badly illuminated.

Image quality factors that might be considered in general includes:

- Focus: Is the focus good enough to perform adequate grading of the smaller retinal lesions such as microaneurysms?
- Illumination: Is the illumination adequate (not too dark, not too light)?
- Image field definition: Does the primary field include the entire optic nerve head and macula? Are the nasal and temporal fields adequately centered to include at least 80% of the non-overlapping portion of the field?
- Artifacts: Are the images sufficiently free of artifacts (such as dust spots, arc defects, and eyelash images) to allow adequate grading?

The most common handcrafted methods applied for classification of retinal image quality are edge intensity histograms or luminosity [52, 56] to characterize the sharpness of the image. Retinal morphology-based methods such as detection of blurring and its correlation to vessel visibility and retinal field definition [31, 110] have been applied for automatic detection of retinal image quality. Fleming et al. [31] method involves two aspects: image clarity and field definition. The clarity analysis is made upon the vasculature of a circular area around the macula. The authors conclude whether or not a given image has enough quality using as evidence the presence/absence of small vessels in the selected circular area.

Pires et al. [49, 81] proposed handcrafted techniques for analyzing image quality in two aspects: field definition and blur detection (focus). For field definition, the authors proposed the use of structural similarity to evaluate the quality of fundus images. For blur detection, the method involved the use of a series of handcrafted descriptors, each one taking full advantage of the specific variations between poor and good-quality images. The descriptors are based on area occupied by blood vessels, class-aware BoVW-based representation, and measuring similarities among original images and versions obtained by controlled blurring and sharpening operations with gaussian filters. The authors also employed fusions (early and late), and reached better performance with Meta-SVM.

Data-driven deep learning methodologies comprise the most recent techniques proposed for retinal image quality assessment [20, 61, 89].

Mahapatra et al. [61] combined unsupervised information from local saliency maps and supervised data-driven information from CNNs to analyze the quality of retinal images. The authors trained one random forest classifier for each (map-based and CNN-based) feature vector and calculated the mean of the responses.

Costa et al. [20] proposed an accountable method for quality evaluation. The method consist on classifying image patches in terms of quality, and combining the local responses to grade the entire image (the image is graded as good if there are more high-quality regions that low-quality ones). Those patch responses also compose a final heatmap that shows the regions that influences the decision.

2.6 Accountable Retinal Image Analysis

In automated diagnostic tasks, the classification accuracy is remarkably relevant, but understanding the reasons behind a computer-aided decision has become even more required and appreciated lately. Notwithstanding, most data-driven approaches are not absolutely explainable. Here we present some recent works that propose accountable methods for tasks related to diabetic retinopathy and/or performs final decisions based on local and global information. In this work, we interchange the terms heatmaps and saliency maps to express pixel importance analysis towards a decision.

To capture visual properties from retinal images, Nandy et al. [63] collect blobs from a limited set of annotated lesions and learn class-aware GMMs (control and disease), refine and combine them to provide universal GMMs. Those universal GMMs are used as prior distributions to adapt individual image's blobs distributions. Similarities between components of universal and adapted (individual) GMMs compose the feature vector of a particular image, ultimately used to train a referral model. The proposed GMM-based method achieved an AUC of 92.1% with Messidor dataset under a 4-fold cross-validation protocol.

Yang et al. [119] combine local and global mechanisms for lesion location and severity grading, respectively. The local stage has the purpose of not only detecting lesions but also reweighting the images based on a naïve strategy whereby predictions and probabilities are weights — e.g., control patches are totally removed (label 0) while exudates (label 3) are more rewarded than microaneurysms and hemorrhages (labels 1 and 2, respectively).

After reweighting, the images are used as input to the global network that grades the severity and categorizes as referable if the stage is beyond mild nonproliferative diabetic retinopathy (NPDR). The two-stage method provides an AUC of 95.9% for RDR over EyePACS test dataset.

Wang et al. [114] proposed an interpretable approach for DR screening that predicts the stage of diabetic retinopathy based on the whole image and suspicious patches. *Zoom-in-Net* mimics the clinician usual procedure to examine retinal images. The entire Zoom-in-Net comprises three sub-networks. The first one classifies retinal images producing probabilistic responses for each class, while the second receives feature maps from a specific point of the first network (before fully-connected stage) and produces scores and contextual heatmaps for each disease level. The second network is responsible for processing and combining the high-dimensional patches extracted in virtue of the heatmaps. The third network combines features from all patches by global max pooling, concatenates with image-level features from the first network and classifies the image generating class scores. The Zoom-in-Net combines (by sum) the 5-dimensional scores from each subnetwork, and later on trains an SVM for DR stage detector, whose decisions are later on converted for referral. Using Messidor dataset under a 10-fold cross-validation protocol, the authors reached an AUC of 95.7%.

An end-to-end BoVW-like methodology that bypasses the previous stage of codebook learning was proposed recently by Costa et al. [21]. The method consists on jointly training two neural networks under the same objective function and joint optimization process. The first network learns weights in order to encode local features — previously detected and described with SURF algorithm [11]. The approach aggregates those encoded features by max-pooling and uses the learned representation as input to the second network, that discriminates the image in terms of presence/absence of lesions or signs of referable/non-referable conditions. The authors enhance the model interpretability by modifying the loss function according to the class, forcing sparse representations for control images and dense representations for disease cases.

To enhance classification of DR severity, Roy et al. [88] proposed a hybrid approach that combines data-driven (global) information with BoVW-based (local) approaches by incorporating local representations encoded in terms of particular lesions. The image-based global features come from a well-performing Deep Neural Network (DNN) for severity estimation, while patch-based local features consist of non-overlapping 224×224 patches described with VGG Net [95] (pre-trained with ImageNet dataset), and encoded into discriminative and generative pathology histograms. The discriminative histogram proposed by Roy et al. is based upon encoding local patches with a dictionary created through random forest's trees [88]. After encoding into a higher dimensional and sparse space, the local features are assigned to the lesion class for which it has been classified by a pre-trained multi-lesion SVM classifier; and the sparse activations are aggregated by sum pooling. The generative histogram is based on BoVW representations by Fisher Vectors [91]. Before calculating the gradients of local patches with respect to Gaussians' mean and standard deviations, the authors reduced the local feature dimensionality by applying principal component analysis (PCA). Roy et al. concatenate global features (from the second fully-connected layer), discriminative histograms and generative histograms, and

train a random forest classifier to evaluate DR severity. By combining local and global information, the authors boost the performance from 0.81 (only global information) to 0.86 in quadratic kappa score.

Quellec et al. [84] proposed an accountable solution for automated referral assessment and automated detection of DR-related lesions. The authors trained the o_O[1] and Alexnet architectures for referable DR; and without retraining, evaluated how well the CNNs could detect lesions. The optimal checkpoints for referral and for each individual lesion (the one the provides better performance during the learning process on a dataset with manually delineated lesions) were ensembled in a patient-basis viewpoint. The ensemble corresponds to a random forest trained with responses of six different networks (checkpoints exported in different steps), for left and right eyes. By combining those responses, the method reached an AUC of 95.4% with Kaggle/EyePACS dataset.

[1]Competitor in the 2015 Kaggle Diabetic Retinopathy Detection Challenge

Chapter 3

Methodology

Our work involved three paths: Diabetic Retinopathy Lesion Detection, Referable Diabetic Retinopathy Detection and Validation with Images from Portable Devices (Sections 3.1, 3.2 and 3.3, respectively). Here, we present a brief description of each topic, including the solutions reported in papers published in a conference, journals and a book chapter during the doctoral program [75–79, 82].

3.1 Diabetic Retinopathy Lesion Detection

Assuming the adoption of diabetic retinopathy screening methodology inspired on unified approaches suitable for every lesion, it is important to remember that our current research regarding DR lesion detection relies on the following questions:

Q1.1: Can we combine the advantages of both *hard* and *soft* codings and provide a good balance for designing efficient and effective DR lesion detectors?

Q1.2: Can we preserve information in the pooling process and still obtain satisfactory results?

Q1.3: Can we improve the results employing global codebooks?

Q1.4: Can we diagnose patients in an ethnic group (e.g., isolated indigenous communities), even using models trained with data from other different ethnic groups?

Therefore, in the following sections, we delve into the methodologies we have been proposing and employing to investigate these questions. Section 3.1.1 describes the combination of the classical and the more recent coding schemes in order to provide a enhanced BoVW-based representation for retinal images (**Q1.1**). Section 3.1.2, in turn, refers to the preservation of valuable information in the pooling step (**Q1.2**). Section 3.1.3 compares the use of class-based codebooks with the use of global codebooks (**Q1.3**). Lastly, Section 3.1.4 evaluates the feasibility of diagnosing isolated communities using lesion detectors trained with data from different ethnic groups (**Q1.4**).

3.1.1 Semi-soft Coding for Lesion Detection

This section provides a complete description of the technique proposed for answering Question **Q1.1**. First a detailed overview of each conceptual aspect of technique is discussed. The scheme proposed here employs a two-tiered image representation based upon the extraction of low-level local features from the images, and then the aggregation of those local features into mid-level BoVW features. Finally, the BoVW features are used as input to a maximum-margin SVM classifier [19].

BoVW-based Representation

The mid-level BoVW representation is among the most impactful contributions of the doctoral program and is outlined here. Several BoVW-based representations have been proposed in the literature [46, 81, 87]. However, the methods discussed in these papers do not explore and compare the different possible implementations associated with BoVW-based representations nor do they present any elaborate discussion on the rationale for using the representations proposed therein.

BoVW-based representations rest upon several possible choices that have to be made for low-level feature extraction, type of codebook, coding and pooling when applying this method to image classification. The factors considered for this research are listed below and explained in the remainder of this section:

- **Low-level feature extraction**: mid-level BoVW features depend upon low-level features. The features used for low-level feature extraction have a large impact on subsequent performance of the classifier. Two low-level BoVW feature extraction possibilities (factor levels) are **sparse features**, based upon the detection of salient regions or points of interest; and **dense features**, sampled over dense grids of different scales. Section 3.1.2 provides more details regarding low-level feature extraction;
- **Choice of codebook**: "codebook learning" was performed by a k-means clustering over features chosen **at random** from a training set of images. An alternative **class-aware** factor level is also proposed;
- **Coding**: For this factor, three levels were compared:
  - **Hard assignment**: associates each descriptor fully and only to its closest codeword in the codebook (Eq. 2.1). The advantage of these schemes is the sparsity of the codes; the disadvantages are that they are subject to imprecision and noise when the descriptors fall in regions close to the limit between the codewords in the feature space. This scheme was explored in previous work for detecting DR-related lesions [46, 50, 87].
  - **Soft assignment**: there are several "soft" assignment schemes to deal with the deficiencies associated with hard assignment. The option employed here was *codeword uncertainty* [111] (Eq. 2.2), which has not been explored as a DR-related lesion detector but is generally considered the most effective for other classification tasks.

  - **Semi-soft assignment**: soft assignment solves the boundary effects of hard assignment, but creates too dense codes. A "semi-soft" scheme is often more desirable. One such scheme, designed specially for the DR-related lesion detection, is described below.

- **Pooling**: For the pooling step, both the traditional **sum**-pooling (Eq. 2.3) and the more recent **max**-pooling, described in Eq. 2.4, are employed. The pooling step is considered one of the most critical for the performance of BoVW representations, and max-pooling is considered an effective choice [8, 13, 14].

In all cases, an ℓ_1-normalization on the final BoVW vector was used.

Semi-soft Coding

The semi-soft coding tries to combine the advantages of both hard and soft assignments, i.e., avoiding the boundary effects of the former, and the dense codes of the latter. The main idea is to perform a soft assignment, but just to the codewords that are the closest to the descriptor, keeping all others at zero. This concept can be translated into many designs of which two were used for this research:

- only the closest codeword is activated;
- the activation is proportional to the inverse of the distance between the codeword and the descriptor.

Therefore, the generated codes are very sparse. On the other hand, the effect of the descriptors is "felt" even at relatively long distances (compared to exponential decay of a Gaussian kernel as in Eq. 2.2). The scheme has the advantage of not requiring parameters.

The coding function can be described as:

$$\alpha_{m,j} = \begin{cases} \frac{1}{\|\mathbf{c}_m - \mathbf{x}_j\|_2} & \text{if } m = \arg\min_k \|\mathbf{c}_k - \mathbf{x}_j\|_2 \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

Class-aware codebook

Rocha et al. [87] proposed employing a "double codebook", extending the usual scheme in a class-aware fashion, especially adapted for DR-related lesions. This is possible because, in addition to the training images being annotated for each lesion, the regions where the lesions appear are also identified (usually 2 to 5 per image from affected patients).

Using the class-aware codebook ensures a sufficient number of codewords representing the appearance of the lesion structures. Because the lesion areas are relatively small, a non-class-aware codebook tends to be dominated by codewords representing healthy regions. During the coding phase, a good codebook is important, as the local feature vectors need to be assigned to the components of the mid-level feature vector in a way that allows discriminating the positive and negative classes. Having very few codewords for the lesion structures reduces this discriminating power. Selection of feature vectors is

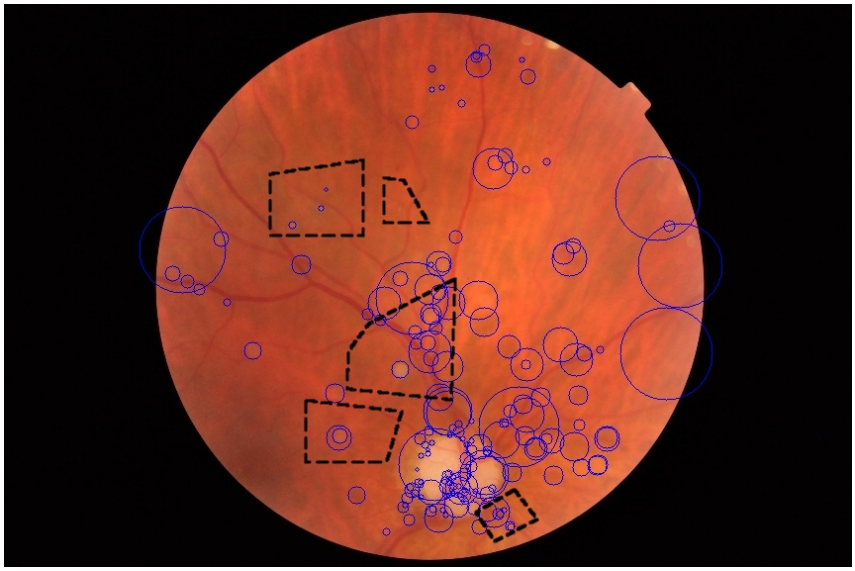


Figure 3.1: Regions of interest (dashed black regions) and the points of interest (blue circles). Points of interest falling within the regions marked by the specialist are considered for creating the class-aware codebook — half of the codebook is learned from local features sampled inside the regions marked as lesions, and half the codebook is learned from local features outside those regions.

usually employed for general-purpose visual recognition — but in those tasks, recognition does not hinge on such subtle differences, as is the case for DR-related lesions. The scheme can be employed for both dense and sparse low-level descriptors, and is illustrated for the latter case in Figure 3.1.

The class-aware scheme works by creating two independent codebooks, one from descriptors sampled from regions marked as containing lesions by the specialist, and one from descriptors outside those regions (which includes images from healthy patients). Then, two independent k-means clustering methods are performed, each with k corresponding to half the size of the desired codebook. After the clustering is finished, the two sets of centroids are simply concatenated, generating a codebook of the desired size.

3.1.2 Preserving Information for Lesion Detection

For answering Question **Q1.2**, this section describes the extraction of low-level local features from retinal images, the aggregation of those local features into mid-level BossaNova features, and the classification of those features by a Support Vector Machine (SVM) classifier [112].

Low-level Local Feature Extraction

Typically, local feature extraction includes two steps: feature detection and feature description. The former aims at finding a set of interest points, or salient regions in the image that are invariant to a range of image transformations. The latter step aims at obtaining robust local descriptors from the detected features. In the doctoral program, we extract Speeded-Up Robust Features (SURF) local descriptors [11].

Two types of local feature extraction can be distinguished [109]: (i) *sparse*, based upon the detection of salient regions or interest points, or (ii) *dense*, where patches of fixed size are placed on a regular grid over multiple scales. For sparse feature detection, the SURF [11] is used. SURF sensitivity parameters are pre-tuned to detect, on average, 400 interest points per retinal image.

For dense features, patches are selected on a dense grid using radii of 12, 19, 31, 50, 80, 128 pixels. These radii are used both as scale and as the vertical/horizontal sampling steps of the grid.

SURF is used to create a feature vector for each detected point of interest. The algorithm is parameterized to operate on twice the image resolution and to extract 128-dimensional extended feature vectors instead of the default 64-dimensional feature vectors.

Once extracted, these points need to be analyzed and filtered so as to select the ones most appropriate for detecting DR lesions. For that, we transform the low level features into mid-level ones creating a two-tiered representation scheme. These low-level local features are used also for referral decisions, as discussed in Section 3.2.1.

Mid-level Feature Extraction: BossaNova Representation

As explained in 2.1.1, BossaNova is a recent mid-level representation that follows the BoVW formalism [97], but keeps more information during the pooling step. Recall that in a BoVW model, pooling is the step responsible for aggregating different features activating the same visual word onto a final summarized feature vector.

In summary, by using a histogram of distances to capture the relevant information, the BossaNova approach remains very flexible and keeps the representation compact. In comparison to the BoVW representation, BossaNova significantly outperforms BoVW on many challenging image classification benchmarks [8]. Considering those results, we have chosen the BossaNova approach for mid-level features given that it takes into account some spatial relationship between features which we believe would be important for DR lesion detection.

In our experiments, we kept the default BossaNova parameter values the same as in [8] ($B = 2$, $\alpha_m^{min} = 0.4\sigma_m$, $\alpha_m^{max} = 2.0\sigma_m$, $s = 10^{-3}$), except for the number of visual codewords M, where we considered {1,000, 4,000}.

3.1.3 Class-based Scheme vs. Global Dictionary

For answering Question **Q1.3**, this section compares the traditional global dictionary and the class-based scheme. In our work we have used the *class-based* scheme (also referred as class-aware scheme 3.1.1) for image representation, which performs well for retinal images. The class-based scheme, proposed by Rocha et al. [87], creates two independent codebooks, one from descriptors extracted from retinal images with the lesion present, and one from descriptors extracted from images of healthy retinas. Then, two independent k-means clustering methods are performed, each with k corresponding to half the size of the desired codebook. After the clustering process, the two sets of centroids are concatenated, generating a codebook of the desired size.

This class-based scheme is compared to the global dictionary scheme when applying the BossaNova approach in which the clustering is performed only once with the desired codebook size.

3.1.4 Diabetic Retinopathy Screening for Isolated Indigenous Communities

The current generation of computer-based classifiers requires training images from the same population that is being tested, preferably from the same retinal camera and with the same resolution in addition to specific pre-processing of images for each type of lesion and is therefore not ideal for multi-lesion detection [30, 35, 47] in diverse populations. In this section, we present the methodology employed for answering Question **Q1.4**.

Automated computer detection of multiple lesions presented as single lesions or in combination of different lesions associated with DR from digital images has the potential to further improve access to primary care-based screening for Aboriginal and Torres Strait Islander peoples with diabetes by introducing an automated process for initial classification of images into those that require and those that do not require referral or further investigation [80]. This saves the time-consuming process of evaluating every image by specialists or trained primary care physicians. Using an automated classifier, specialists and primary care physicians can better use their time for patient consultation and reviewing images that have been identified abnormal by the automated screening process.

The methodology employed for image representation in this step is the Bag of Visual Words (see Section 2.1), widely explored by the computer vision community [80, 82].

The objective of this study is to demonstrate that the BoVW-based automated diabetic retinopathy classification system can identify abnormal retinal images in an Aboriginal and Torres Strait Islander population.

3.2 Referable Diabetic Retinopathy Detection

Most existing art focuses on the detection of DR lesions using either visual characteristics specific to each type of lesion [29, 30, 32, 35, 66, 67] or unified DR-lesion detectors [46, 49, 75, 80, 82, 87].

It is much harder (and polemic) to decide automatically to refer or not the patient to the ophthalmologist [1, 80, 82, 101]. Automated referral is a hot topic, because DR risk assessment is complex, based not only on the presence of lesions and their evolution, but also on subtle hints revealed during examination and anamnesis. We argue, however, that automated or semi-automated decision of referable cases can have a huge impact on the management of care, reducing the specialist's workload while still attending to the patients in need. While agreeing that face-to-face consultations with a specialist are always ideal, we stress that many communities simply lack the luxury of offering them to every suspect case.

Assuming the interest in proposing effective referability assessment methods and preserving relevant information, we recap the questions crafted to guide the researches within the scope of referral decisions:

Q2.1: Can we forgo the detection of individual DR lesions and still have an effective referral decision?

Q2.2: Can sophisticated handcrafted mid-level features improve direct referral decision?

Q2.3: Can we confirm the suitability of handcrafted direct referral in a second, independent, dataset?

Q2.4: Can we learn in the data-driven manner highly abstract features that leverage referable DR detection without the need of manual feature engineering?

Q2.5: Can we diagnose retinal images collected under different acquisition conditions with data-driven models?

Q2.6: Can we transfer knowledge acquired with a different but similar task in the context of diabetic retinopathy screening?

Q2.7: Can we provide an effective and accountable data-driven solution for referable diabetic retinopathy detection?

Q2.8: Can we enhance the performance of the referral model by applying two-tiered data-driven image representation?

Q2.9: Can we combine global data-driven and local saliency-oriented two-tier representations for more accurate decision-making?

In the following sections, we present the methodologies proposed to investigate these questions. Section 3.2.1 evaluates if forgoing the detection of lesions is practicable for pursuing accurate referral decisions (**Q2.1**). Section 3.2.2 presents the proposal for improving the direct referral approach using sophisticated mid-level features (**Q2.2**). Section 3.2.3 describes the scheme to evaluate the fitness of the proposed approach in a public dataset (**Q2.3**). Section 3.2.4 refers to outstanding data-driven strategy for assessing need of face-to-face consultation (**Q2.4**). Section 3.2.5 exposes the methodology behind the cross-dataset evaluation (**Q2.5**). Section 3.2.6 describes the investigation regarding transferring knowledge from a similar but different domain to the referral context (**Q2.6**). Section 3.2.7 illustrates the methodology employed to measure pixel importance and explain what the complex model is learning indeed (**Q2.7**). Section 3.2.7 details the process of creating a two-tier mid-level representation based upon contextual information previously learned by the data-driven model (**Q2.8**). Finally, Section 3.2.9 presents the investigation regarding the fusion of global- and local-based image characterization for enhancing final decision (**Q2.9**).

3.2.1 Beyond Lesion-based Diabetic Retinopathy: a Direct Approach for Referral

Currently, automated referral decisions combine separate DR-lesion classifiers into a final decision. Those models are complex, cumbersome to implement, and often have limited

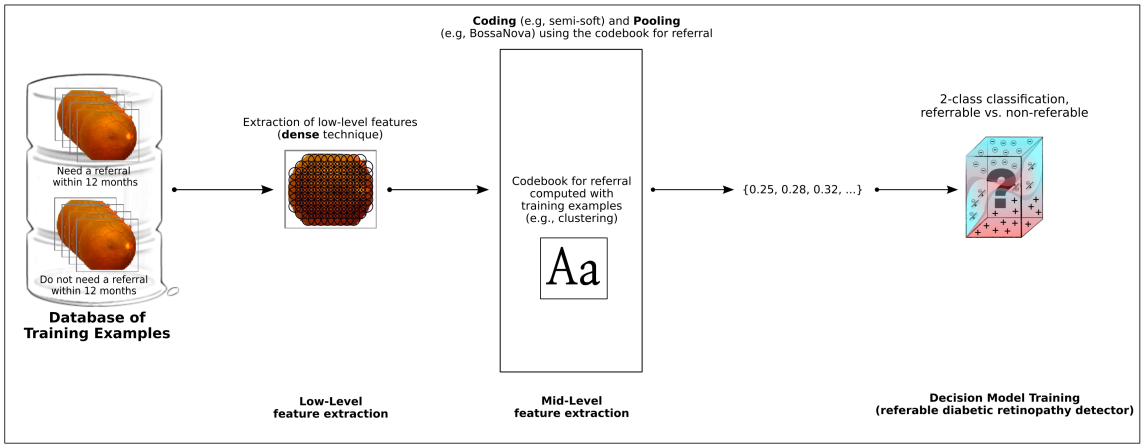


Figure 3.2: Pipeline of the methodology for direct diabetic retinopathy referral assessment proposed in this work.

accuracy. For answering Question **Q2.1**, we take a different approach with an effective method for directly assessing the referability of patients (completely forgoing the identification of specific types of lesions).

The automated referral assessment proposed in the doctoral program dispenses with the intermediate stage of detecting DR lesions. That direct approach has both theoretical and practical motivations.

Lesion-based referral decisions loses critical information on the interface between the lesion-specific classifiers and the referability classifier. Often, the referral classifier receives just a vector of classification scores, one per lesion classifiers. This is unfortunate, because cogent information is lost, like the number, intensity, and even position of the lesions in the retina. One can create *ad hoc* schemes to transfer those data between the classifiers, but, it may be simpler to forgo the lesion classifiers altogether, and just provide the retinal images, with all cogent information, directly to the referability classifier.

Giving the whole image to the referability classifier has also practical advantages. The classical scheme involves implementing, debugging, training, and testing several lesion classifiers, and then one additional layer to combine the results and make the referral decision. Although using a unified approach for the lesion detectors [75, 80, 82] simplifies the task, it remains much more complex than implementing, training, and testing a single model.

The streamlined technique for direct decision on referability is illustrated in Figure 3.2.

For low-level feature extraction, the method selects patches on a dense grid using diameters of 12, 19, 31, 50, 80, 128 pixels. The selected patches are described with SURF [11] in 128 dimensions. The low-level features are then integrated into a single feature vector using mid-level features.

Bag of Visual Words is employed for mid-level features. In the codebook learning step, while the lesion-based approach uses pre-computed codebooks for each individual lesion detector [82], the direct methodology uses k-means with Euclidean distance over a sample of low-level features. The codewords are the centroids of k-means, keeping the class-aware scheme proposed in [87] (half of the codebook from descriptors sampled from referable

Table 3.1: Retinopathy grade criterion used for MESSIDOR annotation.

Grade	Criterion
0	(μA = 0) AND (H = 0)
1	(0 < μA <= 5) AND (H = 0)
2	((5 < μA < 15) OR (0 < H < 5)) AND (NV = 0)
3	(μA >= 15) OR (H >= 5) OR (NV = 1)

μA: number of microaneurysms
H: number of hemorrhages
NV = 1: neovascularization - NV = 0: no neovascularization

images, and half from nonreferable images). The method trains just one decision model for referability.

3.2.2 Direct Referral with Sophisticated Mid-level Features

For the mid-level features, we employ simple BoVW as a baseline and explore two recent alternatives, BossaNova [8] and Fisher Vector [74] (Section 2.1), for improving direct referral decisions and answering Question **Q2.2**.

For comparison purposes, the same low-level features are used in this step. In the codebook learning step, BossaNova also uses centroids of k-means, while in Fisher Vector, the codebook learning employs (class-agnostic) Gaussian Mixture Models that is intrinsic to the representation.

3.2.3 Direct Referral in a Public Dataset

In the doctoral project, for answering Question **Q2.3** we also evaluate the fitness of the proposed direct referral methodology in a public, independent dataset.

The Messidor dataset[1], detailed in Chapter 4, comprises images cropped in order to establish that the relevant circular area of the retina has a radius similar to the DR2 database. Although it has not been graded for referable diabetic retinopathy, the dataset was annotated with two significant criteria: retinopathy grade (see Table 3.1) and risk of macular edema.

The severity of the diabetic retinopathy is largely used as guideline for the frequency of examinations [1, 33]. Although a frequent consultation is recommended for patients with moderate or severe nonproliferative DR [33], for those without DR lesion or with just microaneurysms the annual incidence of progression is low [10]. In these situations, longer intervals between examinations may be recommended (one year for diabetics). Hence, based on the original annotations about DR severity, the guidelines of periodic referrals and the opinion of an expert about the criterion employed in the Messidor dataset, we switched the grading into referable or nonreferable. Given that the presence of just microaneurysms does not suggest a referral in less than one year, the grades 0 and 1 (including also no risk of macular edema) are considered as nonreferable, while grades 2 and 3 (and also apparent macular edema) as referable, resulting in 688 negative and 512 positive images.

[1]kindly provided by the MESSIDOR program partners: see `http://www.adcis.net/en/Download-Third-Party/Messidor.html` (accessed February 13, 2019)

3.2.4 A Data-Driven Approach to Referable Diabetic Retinopathy Detection

In this section, for answering Question **Q2.4** we present our deep learning-based solution for diabetic retinopathy screening and highlight a series of approaches explored to achieve a robust and effective framework. An important landmark for automated diabetic retinopathy detection was a recent competition promoted at Kaggle[2] by California Health Foundation, with images provided by EyePACS, a platform for retinal screening. The dataset, comprising more than 88 thousand samples (see Section 4), was the largest publicly available dataset of retinal images at the time. The aim of the competition was classifying the images into five degrees of severity, ranging from 0 (no sign of diabetic retinopathy) until 4 (proliferative diabetic retinopathy). However, here we investigate and propose binary decisions of referability, while the competition aimed the task of severity classification. While from the human point of view there's a relatively direct map between the two tasks, from a Machine Learning point of view, limiting to binary classification has theoretical and practical advantages.

Solutions presented at the Kaggle competition have to improve the target metrics in short time. Rigorous validation of factors leading to performance is less important than quickly improving metrics. Our aim here is opposite: we are less interested in shaving tenths of percents from the classification error, and more interested in evaluating the cost-benefit of each choice and novel contributions.

Our Architecture

Aiming both at effectiveness and efficiency, we present a streamlined architecture for retinal image analysis, which bears some resemblance with two main networks in prior art (namely o_O, a key competitor in the 2015 Kaggle Diabetic Retinopathy Detection Challenge, and VGG, a key competitor for natural image classification in the ImageNet challenge) but with key insights and differences. The solution we propose, which is depicted in Fig. 3.3, is significantly novel both in practical (CNN architecture) and methodological (scientific procedure leading to the solution) terms. The proposed architecture has 16 weight layers, with about 10 million parameters.

In terms of arrangements of convolutional and pooling layers, our architecture resembles the VGG-16 [95] except that we reduce by half the number of units on the first four out of five convolutional blocks and use strides when appropriate, focusing on efficiency and aiming at working with higher resolution images. We use very small receptive-field (3×3) convolutions. Pooling layers separate sequences of two or three convolutional layers. The first pair of convolutional layers starts with 32 kernels/filters. When pooling layers take place reducing drastically the feature map sizes, the convolutional layers double the number of kernels. Additionally, we stride in two the first and third convolutional layers across their respective inputs to reduce more aggressively the initial layers and work with higher dimensional images. In the fully-connected stage, similarly to the o_O team, we apply RMSPool with both pool size and stride of (3, 3), and Feature Pool with pool size

[2] `https://www.kaggle.com/c/diabetic-retinopathy-detection` (accessed February 13, 2019)

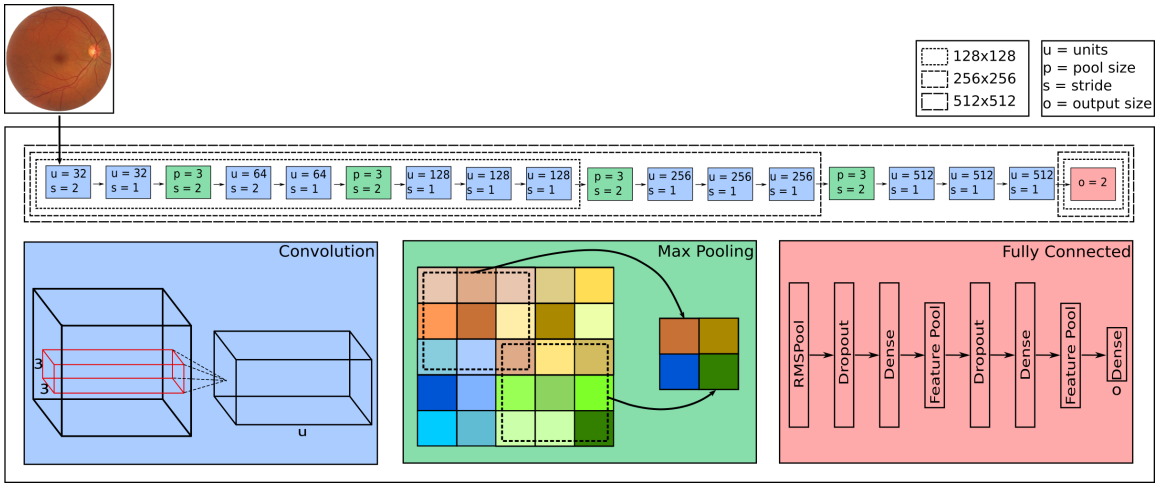


Figure 3.3: The proposed solution decides to refer the patient directly from the pixels of the retinal exam, without preliminary feature extraction or lesion detection. From an initial "basic" configuration, we propose and evaluate improvements step-by-step. The image is best viewed in electronic form.

of (2, 2). The units of the 1024-unit hidden dense layers employ drop-out with a fixed probability of 0.5. Our final architecture, in this case, is a hybrid one (not presented before), which brings to bear essential ideas for an efficient and effective analysis of retinal images. For instance, in the original VGG-16 design, almost 90% of the parameters reside on the fully-connected layers. In our new arrangement, convolutional layers represent 75% of the parameters while fully-connected layers represent only 25%, a remarkable feat for efficient implementation.

Our idea here is creating a network as similar as possible to VGG-16 [95] in terms of convolutional and pooling arrangements, but also similar to o_O's ones in terms of the fully-connected sequence. Comparing our convolutional/pooling arrangement with one of the o_O's architectures, we essentially moved one layer from the coarse stage to finer one, resulting in more parameters. Above all else, we tried to preserve the structure that enabled the use of smaller networks for multi-resolution training (see Section 3.2.4).

After each convolutional and dense layer (except the last one), we use leaky rectifier units (leaky RELU with alpha = 0.01) that applies a small negative slope to address the shortcomings of the simple rectified linear unit ("dying" RELU), while accelerates the convergence of the gradient in comparison with conventional activation functions.

We optimize the CNN with Nesterov momentum [103] that, contrasting with standard stochastic gradient descent, is capable of accelerating convergence in regions of low-curvature. As well as classical momentum, Nesterov momentum accelerates gradient descent by accumulating a velocity vector in directions of persistent reduction in the objective [103]. While conventional momentum computes the gradient update from the current position θ_t, Nesterov momentum first computes $\theta_t + \mu v_t$ (a partial update to θ_t). This allows Nesterov changing velocity in a more responsive and stable way, especially for higher values of μ. Given a objective function f, Nesterov momentum is written as:

$$\begin{aligned} v_{t+1} &= \mu v_t - \varepsilon \nabla f(\theta_t + \mu v_t) \\ \theta_{t+1} &= \theta_t + v_{t+1} \end{aligned} \tag{3.2}$$

where $\varepsilon > 0$ is the learning rate, $\mu \in [0, 1]$ is the momentum coefficient, and $\nabla f(\theta_t + \mu v_t)$ is the gradient at the partially updated θ_t.

We defined a fixed schedule of 250 epochs for training, starting with a learning rate of $3{\times}10^{-3}$ in the first 150 epochs, decreasing it to $3{\times}10^{-4}$ in the following 70 epochs and finally to $3{\times}10^{-5}$ until the end.

We exploited different alternatives for optimization (algorithms and hyper-parameters). Comparing to o_O's proposal, we kept the same optimization schedule although with a different network architecture as that schedule also worked for our problem. Moreover, here we tackle the problem with a classification point of view rather than a regression one and apply cross-entropy instead of mean squared error as objective function.

Henceforward, we review and describe, from a rigorous scientific point of view, a set of approaches proposed by the o_O team and new contributions of our own. We use such approaches to improve progressively our "basic" framework. We anticipate that we cannot analyze the following variations (or improvements) independently since it embraces coarser approaches — essential to the convergence of the CNN — that must be kept to analyze finer approaches.

Resampling and Data Augmentation

Convolutional Neural Networks thrive with hundreds of thousands, up to several million learning samples for training. However, contrasting to general-purpose object recognition, medical tasks count on relatively small annotated datasets. In the rare situations where we can count on relatively large medical image datasets, the data tends to be very unbalanced, with the overwhelming portion of the images corresponding to the control group (healthy patients).

Data augmentation can improve the learning process [51, 120]. The augmentation can consist of image perturbations by geometric (e.g., zoom, translations, rotations, cropping) or photometric (e.g., histogram equalizations, contrast enhancements) transformations.

In the technique evaluated here, we propose a data augmentation method to simultaneously address the small sample size and the class unbalancing problems. The aim of data augmentation here is not just inflating the training set for each epoch, but also dealing with the under-sampled classes by dynamically applying random transformations to their samples, while keeping the classes balanced.

The set of operations (or perturbations) we consider comprises zoom, flipping, rotations, translations, stretching, and color augmentation. The perturbations are dynamically performed right before submitting an image to the network, bypassing the need for saving numerous versions of each image. We exploit zooms, rotations, and translations by randomly choosing a variable into a predefined interval (for instance, we apply rotations between 0 and 360 degrees, zooms between $1/1.15$ and 1.15, and translations between -40 and 40).

The color augmentation, proposed by Krizhevsky et al. [51], consists on changing pixel

intensities of RGB channels by adding multiples of principal components (PCA) found in the training set, with magnitudes proportional to the corresponding eigenvalues times a random variable drawn from a Gaussian with mean zero and standard deviation 0.5. The resampling ensures that all classes will be represented equally. The number of randomly perturbed versions (data augmentation) for each class depends on the balance weights, that is inversely proportional to the number of images for each class.

Multi-Resolution Training

Poor initialization of network weights leads to poor local minima and, consequently, to an ineffective solution. Additionally, training large CNNs from scratch requires a very large dataset. To address those shortcomings, we propose a multi-resolution training strategy that consists on training simplified variants of the CNN — requiring less training samples — and using the learned parameters as starting point for next stages. Those variants have less layers but preserve the number of units of each layer of the original network.

We train reduced versions of the entire network (fewer convolution layers) using smaller images, and preinitialize larger networks with the learned parameters. Simonyan and Zisserman [95] employed a similar approach, training the network with the shallowest configuration and after initializing the first four convolutional and the last three fully-connected layers of deeper configurations with the learned parameters.

Fig. 3.3 shows in dashed lines the composition of the original and the two smallest networks. Initially, using the same images resized to 128×128 pixels and the same ground-truth, we train a small variant of the network that does not have the two last pooling layers and convolution trios (sequences of three convolution layers). Note that this is a simplified version that comprises just seven convolution layers with the same number of parameters of the corresponding layers in the original network. Thereafter, using images resized to 256×256 pixels, we produce another simplified network based on the complete architecture, but without the last pooling and the last convolution trio (with ten convolution layers), and use the parameters learned with the previously trained network with 128×128-pixel images to initialize the weights before training. Subsequently, we initialize the first ten convolution layers with the learned parameters and train the architecture using 512×512 pixels images, our final image dimensions of interest.

Robust Feature-Extraction Augmentation

Deep learning-based methods provide us with an end-to-end learning process: the learning models receive raw images as inputs and produce probabilities as outcomes (resulting classes), after an extensive and strongly abstract learning highway. In this case, image representation and pattern recognition are performed together, enabling us to extract features in any layer before the one responsible for the final decision, and use that information in a posterior decision process. This feature extraction procedure is highly flexible as it allows us to use different machine learning algorithms.

To exploit that flexibility, we extract features in a different pathway. Following the o_O team's proposal, we apply pseudo-random data augmentation and create n versions for each image, both from training and test sets. Pseudo-random augmentation ensures

that the same sets of perturbations are always applied to all images. The final feature vector for each image is achieved by concatenating mean and standard deviation of the n individual feature vectors from the respective n image versions:

$$\mathbf{x_i} = [\mu_i, \alpha_i], \tag{3.3}$$

where $\mathbf{x_i}$ represents the image i. We created 20 image versions ($n = 20$) in our experiments and extracted features in the last pooling layer of the CNN. The whole process is data-driven in the sense that all features are extracted directly from the data with no human intervention.

Per Patient Analysis

As our ultimate goal is checking whether or not a patient needs to see a doctor within 12 months, and not merely pinpointing the presence of lesions in his/her retinas, whenever photographs of the two retinas are available, we leverage this additional imagery to mak the final referral decision.

To provide an outcome for each retina, we concatenate features of the both and include a binary indicator variable that refers to left or right. The feature vector for a retina is created as follows:

$$\mathbf{x_{retina}} = [\mu_{retina}, \mu_{retina'}, \alpha_{retina}, \alpha_{retina'}, \delta_{right}] \tag{3.4}$$

where $retina'$ represents the complementary retina.

In addition to combining information and diagnosing retinas individually, we go beyond and assign to the patient the response of the retina that presents the highest risk (highest probability of needing referral).

3.2.5 Cross-Dataset Validation Protocol

After refining the automated solution for referable diabetic retinopathy screening from scratch (last question), we investigate the performance over distinct datasets. In this section, we investigate the performance of the proposed method when training with the Kaggle/EyePACs dataset and testing it with Messidor-2 and DR2, which have very different acquisition conditions. Basically, we use the CNN trained with data augmentation and multi-resolution training (Section 3.2.4) to extract features for the test sets (robust feature-extraction augmentation), and test the features with the classification that provides referral decisions. We emphasize again that one of the most valuable advantages of extracting features (in this case, robust feature-extraction augmentation) is that it provides flexibility to choose different machine learning algorithms. Henceforward, we use two algorithms: Neural Network and Random Forest. Therefore, for responding Question **Q2.5**, we follow a challenging cross-dataset validation protocol with the best solution we found in the previous section that was trained with Kaggle/EyePACs data.

We highlight that the solution incorporates the proposed data augmentation for training, multi-resolution training, and robust feature-extraction augmentation steps. We also exploit the per-patient information whenever we have access to images of both eyes.

3.2.6 Knowledge transfer for Diabetic Retinopathy Screening

As expounded in Section 2.2.1, we generally employ transfer learning when there are not enough training samples in the target domain or when there is already a reasonable solution for a related (source) problem. Additionally, most of the existing works assume that the source and target domains are related to each other [70].

In this section, for answering Question **Q2.6**, our proposed domain adaptation corresponds to transfer knowledge from a domain also related to retinal image analysis, in which the source is severity of diabetic retinopathy (a 5-class decision problem), and the target is referability of diabetic retinopathy (a 2-class decision problem). We explore transfer learning in its two scenarios — feature extraction and fine-tuning — since we intend to compare data-driven approaches with previous work that employed handcraft methods with relatively small datasets under cross-validation protocol.

3.2.7 Accountable Referable Diabetic Retinopathy Detection

For answering Question **Q2.7**, in this section we present a robust and self-explainable detector of diabetic retinopathy referability. The method adopts a deep Convolutional Neural Network trained in an end-to-end fashion for screening of diabetic retinopathy (referral).

In the next steps, we briefly describe concepts related to accountability, term that is being widely discussed in artificial intelligence community nowadays. In sequence, we describe the purely end-to-end data-driven approach we propose.

Accountable Machine Learning

The adoption of machine-learning techniques in supporting automated DR screening brings an essential question: how these so-called "black boxes" results can be explained? In this work, we have used a *post-hoc* interpretation [58], which explains predictions to ophthalmologists but hides the computational details of the technique.

Considering our data-driven approach through deep CNN, this explanation can be done using saliency maps. The map highlights input local regions that influence the results based on the output gradient. However, they are often noisy and difficult to interpret by human eyes. Smilkov et. al [98] discuss this problem comparing some strategies used to create saliency maps. In this work, we have used the Guided Backpropagation [102] strategy, which ignores negative values in the network backward flow. In this way, it underlines positive neuron contributions to the gradients and attenuates negative ones in the ReLU functions.

The extracted saliency maps represent sharper visualizations of the activated screening images, which are pivotal for accountability. In addition, our proposed approach takes advantage of these maps by extracting saliency-oriented local features, described later on.

Global data-driven approach

Our data-driven approach comprehends essentially a deep CNN model for automated DR screening. The network design we employ herein is the Inception-ResNet-v2 [104],

previously trained with ImageNet dataset. We keep using the cross-entropy loss function with two neurons in the last layer, corresponding to positive and negative categories of referral.

In this case, we are transferring knowledge learned on a source task to improve the learning process in a target task. As exposed in Section 2.2.1, transfer learning is normally sought when the (target) training set is not large, there exists a reasonable solution for a related (source) problem, and both problems are similar. Since the source and target domains here are very different, we fine-tune the model with a considerable high learning rate (we are not concerned so much about distorting pre-trained parameters and need to practically train from scratch).

Before actually training the entire network, we trained only the last layer, keeping most of the parameters frozen and avoiding immediate destruction of learned patterns. Afterwards, we extended the backpropagation to all the network, and optimized the model with the RMSProp algorithm [107]. We employed an initial learning rate of 0.01 and decayed it with stochastic gradient descent with warm restarts [59], an aggressive learning rate reduction combined with periodic restarts. The reduction relies on cosine function, and the i-th restart takes place after 3.0×1.5^i epochs to a learning rate of 0.01×0.8^i. Increasing the learning rate at each restart allows us to skip of possible local minima and continue exploring the loss.

To avoid overfitting, we use a weight decay of 0.0004 and keep the dropout with fixed probability of 0.8 (i.e., preserving 80% of the weights in the training stage). Regarding data augmentation, we employ both geometric and photometric transformations. Geometric perturbations comprise zoom, flipping, rotations, translations, and stretching, and are performed by randomly choosing a variable into a predefined interval (for instance, we apply rotations between 0 and 360 degrees, zoom between $\frac{1}{1.15}$ and 1.15, and translations between -40 and 40). We apply color augmentation for geometric transformations, that consists on adding, to each training images, multiples of principal components previously found on the set of RGB pixel values throughout the training set. The magnitudes are proportional to the corresponding eigenvalues times a random variable drawn from a Gaussian distribution with mean zero and standard deviation 0.5.

After optimizing the CNN, we can use the model to diagnose new images as well as to extract saliency maps that highlight pixel importance for the decision made. The saliency maps provide intuitive explanations about particular decisions. In this work, we use the guided backpropagation [102] method to acquire pixel importance.

3.2.8 Saliency-Oriented Data-Driven Approach to Diabetic Retinopathy Detection

In this section, we extend the global data-driven and self-explainable referable DR detector, presented in Section 3.2.7, to involve also local two-tier saliency-oriented representations. For answering Question **Q2.8**, we use locally significant image regions to capture evidence that might be stressed to enhance the model. Note that the current stage requires a reasonably well-optimized model trained in the previous stage (last section).

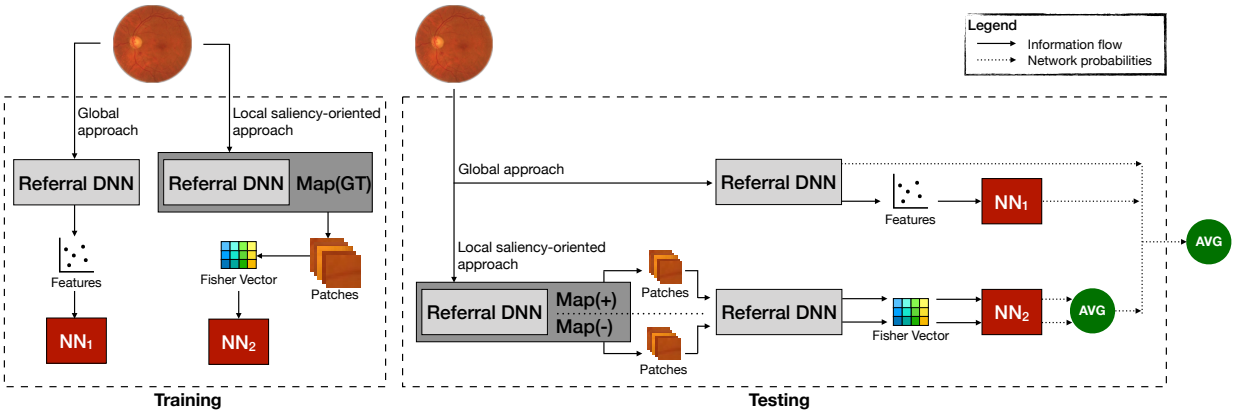


Figure 3.4: Overview of the proposed method. **Training:** two Neural Networks (NN) are trained; the first based on features from a trained referral deep neural network and a second one based on lesion patches. **Testing:** the testing phase combines results of the two trained neural networks plus the probabilities of the already trained referral deep neural network.

In the next step, we describe the saliency-oriented data-driven local methodology that reinforces the performance of the purely global approach as well as the understanding of the solution. Fig. 3.4 depicts an overview of the proposed solution.

Local saliency-oriented data-driven approach

One of the main novelties in this work is our saliency-oriented local representation methodology that relies on heatmaps to gather significant regions (from the previous data-driven global decision) for enhancing the pipeline and providing a more robust screening method.

In this section, we describe the patch extraction protocol under an image preprocessing viewpoint, briefly detail the encoding technique (fisher vector), and describe the methodology for local feature representation.

Patch extraction

After we pass the original image into the deep network, and propagate back the pixel importance for the decision taken to generate a saliency map (see Fig. 6.11), we process the map and capture coordinates that are sequentially used to capture regions that could be relevant to enhance decision.

The saliency map has the same dimensions of the image. As we intend to capture importance and preserve locality, we initially convert the 3D tensor to a grayscale 2D tensor by summing up the activations per channel. In general, heatmaps are subject to visual noise. To reduce undesirable effects in region selection, we apply a threshold with binarization purposes. Other recent alternatives such as adding noise to reduce noise could be explored [98], but a single threshold was a reasonable choice towards efficiency. We filtered the maps with a threshold of 150, a good trade-off for removing noise and preserving small activations. In sequence, we invert and erode the binary structure (basic mathematical morphology operation) in order to extend chunks and connect close components.

After processing the saliency map, we identify contours in the 2D structure and capture their respective coordinates. To preserve the aspect ratio and produce visible regions towards boundaries (e.g., keep the boundaries of lesions), we square the regions and enlarge them using a factor inversely proportional to the original patch size. That operation doubles the height and width for small regions (microaneurysm candidates) and extends by 10% the dimensions of large regions (in general blood vessels or possible connected large lesions). Smallest regions (in general microaneurysm) are enlarged more than largest regions. In Fig. 3.5, we show a fundus image superposed with its saliency map, and respective significant regions extracted based on pixel importance for data-driven referral decision.

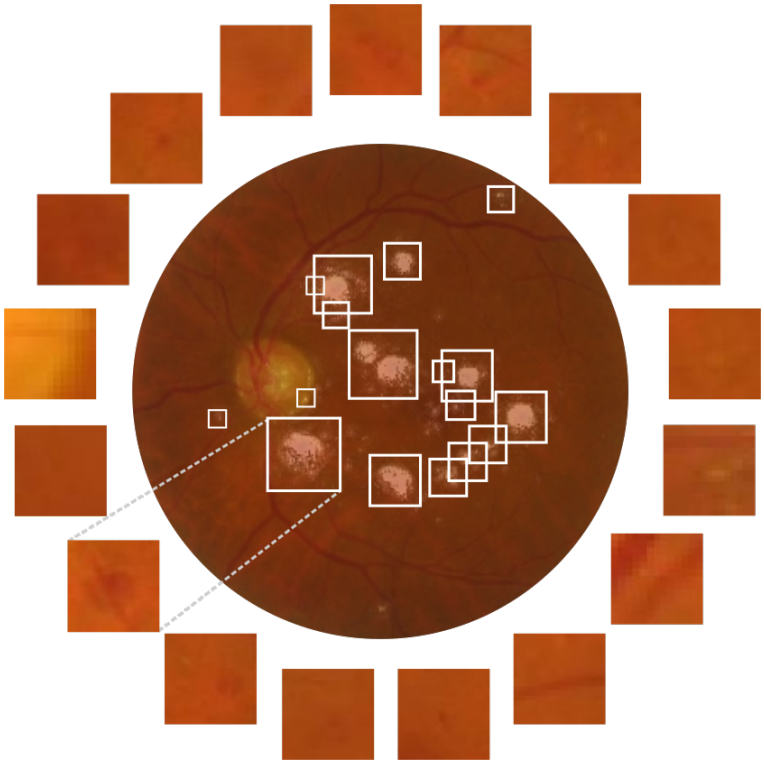


Figure 3.5: Saliency-oriented squared patches from which we extract local descriptors. We enlarge patches in a controlled design taking into account the region sizes: e.g., small regions (in general microaneurysms) are enlarged more than large regions (e.g., soft and hard exudates).

Fisher Vector encoding

Once we have the patches, a Fisher Vector encoding strategy is used to capture their local descriptors by pooling patches features [72, 73]. By combining generative and discriminative techniques, we rank low-level patches descriptions based on their deviation from a GMM (generative model) by calculating the patch gradient with respect the model parameters.

Integration

After training the deep model for referable diabetic retinopathy detection, generating the saliency maps for interpretation and local representation, and encoding the multiple patch-based features, we train a shallow neural network to take a novel complementary decision regarding need of consultation.

To avoid the new model to be strictly dependent on the decisions performed by the baseline CNN model, we extract two different, separate mid-level representations for the test sets — one for each class. We extract those maps by guided backpropagation, each of which guided by one specific class/neuron. Given that the groundtruth is known, we can fully use it to extract a unique saliency map and respective mid-level representation for the training set. As exposed in Fig. 3.4, for inference, the final local saliency-oriented decision is taken by averaging the two per-class responses.

Per Patient Analysis

As highlighted in Section 3.2.4, the more data available, the more confident and effective the learned model. One requirement for a robust data-driven model is having available a large amount of data, except when transferring parameters previously learned from a different, but similar task. Since our purpose is examining whether or not a patient needs to see a doctor within one year, we could substantially leverage the accuracy of the model once photographs of the two retinas are available. As such, we combine image information to provide final patient responses both in feature level and score level. When the method involves feature extraction, we concatenate features of both and include a binary indicator variable that refers to left or right (See Equation 3.4). Regarding score level, we assign to the patient the response of the retina that presents the highest probability of needing referral.

3.2.9 Fusion of Global Data-driven and Local Saliency-Oriented Features to Diabetic Retinopathy Detection

For answering Question **Q2.9**, we combine global data-driven and two-tiered local saliency-oriented responses for enhancing diabetic retinopathy screening. In this work, we perform late fusion by averaging the three softmax probabilities: (1) from the CNN, (2) from the shallow neural network trained/tested with data-driven features, and (3) from the shallow neural network trained/tested with the mid-level representations.

Contextualizing with State of the Art

Our accountable solution encompasses saliency-oriented data-driven local patches, encoded with Fisher Vectors, combined with a global data-driven representation. In this section, we compare the proposed method with recent works related to retinal image analysis (not necessarily regarding referral decisions), presented in the Section 2.6. Table 3.2 summarizes the comparison regarding accountability, combining local and global information, contextual patch extraction, and mid-level representation.

Table 3.2: Contrasting recent similar works with ours

Work	AML	GLI	CPE	MLR
Yang et al. 2017 [119]	✓	✓		
Wang et al. 2017 [114]	✓	✓	✓	
Costa et al. 2018 [21]	✓			✓
Roy et al. 2017 [88]	✓*	✓		✓
Quellec et al. [84]	✓		✓	
Pires et al. (Ours) [79]	✓	✓	✓	✓

AML: Accountable machine learning GLI: Global and local information
CPE: Contextual patch extractor MLR: Mid-level image representation
* Partially accountable

Yang et al. [119] apply a two-stage data-driven method to detect lesions and identity severity of diabetic retinopathy. Although using a naïve pixel reweighting approach, the automated diagnosis of DR severity is regarded as accountable because input images are weighted based on lesion type and location, aspects that altogether determine the diabetic retinopathy stage. The authors also used local and global information. Even they are directly used for different purposes (local features for lesion detection and global features for severity analysis), the local information has meaningful importance on severity diagnosis. Note, however, that the patch extraction is performed in a sliding-window fashion rather than contextual, and the regions are used directly to take individual decision regarding lesions instead of being encoded into a rich representation for severity/referral decisions.

The Zoom-in-Net methodology [114] involves three sub-networks with particular functions. The approach is accountable in the sense that it generates heatmaps that represent pixel importance. Similar to us, the authors also use those heatmaps for region extraction, and combine those regions in the third subnetwork. Those local information are also combined with global information. Note, however, that the method only combines features of all the patches with a global pooling, instead of encoding those information into a richer and contextual-aware representation. The final result is achieved by combining three complete models, aggregating the three best performing results.

Costa et al. [21] proposed a handcrafted BoVW-like methodology that simultaneously learn encoding and classification steps. The approach is accountable as it can pinpoint regions on the image that triggered the diagnosis decision, evenly trained with weakly labeled data (image-level annotation regarding presence/absence of lesions). Note, however, that the regions are entirely detected by the SURF algorithm. The method does not necessarily extract patches in a contextual manner, but describe non-contextual interest points and encode them into a learned mid-level representation. The method does not take into account global representation.

The hybrid method proposed by Roy et al. [88] combines discriminative and generative local lesion-based representations with a global data-driven representation. Although individual patches are classified according to the presence of lesion (in a certain step of discriminative representation pipeline), it is not projected back onto the image to show why the decision regarding DR stage was taken (e.g., one specific region was classified as neovascularization and that justifies the severity decision). Therefore, Roy et al.'s

approach is partially accountable. The adopted patch extraction is not dependent on pixel importance or presence of lesion, for instance, but used in a posterior lesion detection and encoding (mid-level representation).

Quellec et al. [84] proposed a solution that jointly detects referable DR and what influences the decision in the pixel level (in terms of lesion or other biomarkers). The accountable solution also produces heatmaps with an additional strategy that reduces drafting artifacts during training. Although also using heatmaps in order to detect lesions at the lesion level (as well as image level), the approach does not necessarily combines global and local information. Indeed, the lesion detection had an important impact on the final referable DR performance, but it was taken by ensembling different models specialized in detecting lesions but without retraining for it. "Different" models consist on parameters that came up with the same architecture, but captured under distinct training steps. Additionally, the method provides the use of heatmaps to extract regions with potential candidates to lesions, and use them to detect/identify the lesion itself. They neither encode nor combine with global information to enhance the detection of referable DR.

3.3 Validation with images from portable devices

We reiterate that the current work validates the images from portable devices only in terms of quality assessment since the dataset for presence/absence of referable diabetic retinopathy is under preparation.

Good quality retinal images are essential for providing reasonable and robust machine learning models (training) and for receiving reliable diagnostic responses (test). For clinical decisions, the quality is merely mandatory. Hence, before diagnosing images from low-cost devices, the images must beforehand be evaluated if it has enough quality.

Assuming the interest in integrating our software solutions with simple retinal imaging devices, we recap the question crafted to guide the researches within the scope of validation over a new dataset captured with portable devices:

Q3.1: Can we assess the quality of images from portable devices with data-driven models trained with images from high-cost instruments?

Therefore, in the following section we describe the methodologies we have been employing to investigate that question. Section 3.3.1 reports the approaches to investigate and develop models for evaluating the quality of retinal images (**Q3.1**).

3.3.1 Quality Assessment of images from portable devices

In this section, for responding Question **Q3.1** we aim at evaluating in real-time whether the captured fundus image is proper to higher-level decisions such as need of consultation within one year. If the image does not suffice in quality, a new one will be collected by holding the portable device in same position. As long as the exhausting process of image acquisition and grading is being conducted in hospitals, we postpone the stage 2 of the

mentoring program (see Section 1.2) that aims at adapting lesion detection and need of referral solutions for the images acquired by Phelcom. The third stage of the partnership involves:

- Investigate the feasibility of implementing the techniques developed in IC/Unicamp and published in the literature on the image acquisition equipment produced by the company.

In the third stage, we address to investigate and implement efficient and effective smart quality assessment models that should be incorporated into portable imaging devices such as the Phelcom. We keep on analyzing the focus factor (blurring detection) since the test set is not properly annotated in terms of field definition aspect. We have concluded the second series of experiments (stage 3 of the partnership) in the scope of quality assessment [81].

We've performed a series of experiments — exploring several alternatives in terms of data augmentation, regularization, optimization, weight initialization and transfer learning — and throughout the training epochs we use the DR2 dataset to pick up the set of parameters that provided a better performance. In order to investigate the feasibility of implementing decision models on the image acquisition equipment produced by Phelcom, and already considering worse-case scenarios such as remote areas that have no internet connection, we explored mobile models. Our experiments are based on the MobileNetV2, the architecture that has the current state-of-the-art performance of mobile models [92]. MobileNetV2 has a good trade-off between accuracy, and number of operations, as well as the number of parameters. By fine-tuning the MobileNetV2 and discarding batch normalization, we reduce the memory footprint (size of the checkpoint) from 98MB to 50MB.

Self-annotation procedure

One of the main challenges the scientific community faces to investigate and implement robust methods for computer-aided decisions is the limitation of data annotated for the purpose. To address that limitation, we explore the scientific machine learning manner of self-annotation, that consists of inferring the new data (in the same context) with a partially-trained model, and attributing labels according to the scores and adopted confidence level. That process could be interactive (re-annotating the new data in different moments/epochs) or not (just one self-annotation step). In our experiments, we pre-trained a deep learning model for quality assessment with InceptionResnetV2 architecture (a costlier hybrid Inception version [104]) with the DR1 dataset, and use the temporary model to re-annotate the Kaggle/EyePACs dataset. The Kaggle/EyePACs dataset comprises +88k images taken under a variety of imaging conditions (a large portion of the data has limitations in terms of focus and illumination factors), and is annotated for diabetic retinopathy stage and need of referral (see Chapter 4). We annotate the Kaggle/EyePACs images using a model that achieved nearly 90% of AUC over DR2 dataset, and applying the confidence level of 80% (images with score above 80% are labeled as good, and the ones with score below 20% are considered as poor). We perform the re-annotation just once (one iteration), and come up with 10,402 poor images and 34,300 good images.

Chapter 4

Experimental Protocol

In this chapter, we describe the datasets (Section 4.1), validation protocols (Section 4.2), and metrics (Section 4.3) adopted in this thesis.

4.1 Datasets

In our research, we are considering eight different retinal image datasets annotated by medical specialists:

- **DR1 dataset**, provided by the Department of Ophthalmology, Federal University of São Paulo (Unifesp), Brazil. Each image was manually annotated by three medical specialists and all the images in which the three annotations agree were kept in the final dataset. The images were captured using a TRC-50X (Topcon Inc., Tokyo, Japan) mydriatic camera with maximum resolution of one megapixel (640×480 pixels) and a field of view (FOV) of $45°$. The dataset is annotated in image level for quality and the commonest DR lesions, and additional auxiliary coordinates that delimit lesions. The DR1 dataset comprises 5,776 images, including poor- and good-quality images, with or without lesion annotations.

- **DR2 dataset**, provided by the Department of Ophthalmology, Federal University of São Paulo (Unifesp), Brazil. The images were annotated by two medical specialists (none of them worked on the DR1 dataset). The dataset was captured using a TRC-NW8 retinograph with a Nikon D90 camera, creating 12.2 megapixel images, which were then reduced to 867×575 pixels for accelerating computation. The dataset is annotated in image level for quality, presence of DR lesions, and need of referral. The DR2 dataset comprises a total of 920 images.

- **Messidor dataset**, captured in three different French ophthalmologic departments. There are three subsets, one for each department. The images were captured using a Topcon TRC-NW6 non-mydriatic retinograph with a $45°$ field of view, at the resolutions of $1,440 \times 960$, $2,240 \times 1,488$ or $2,304 \times 1,536$ pixels. The dataset is annotated in image level for retinopathy grade (0 to 3) and risk of macular edema. The Messidor dataset has 1,200 images.

- **Messidor-2 dataset**, an extension of the Messidor dataset, is a collection of diabetic retinopathy examinations, each of which consisting of two macula-centered eye fundus images (one per eye) [24, 85]. These images were captured with a Topcon TRC NW6 non-mydriatic fundus camera with a 45° field of view. Images from Messidor-2 were independently graded in image level by three board-certified retinal specialists from all subjects according to the ICDR severity scale and a modified definition of macular edema (ME) [1, 2]. The Messidor-2 dataset contains 1,748 images for 874 examinations (two images per patient).

- **Inala dataset**, from the Inala Indigenous Health Service. Captured with a Canon CR-DGi IOS 30D (Canon Australia Pty. Ltd. Sydney) non-mydriatic digital retinal camera. The camera settings were $15 \times 22.5mm$, 8.5 megapixels (Mp) and 45° field of view. The dataset is annotated in image level according to presence/absence of red and bright lesions. The Inala dataset comprises 30 images.

- **Kaggle/EyePACS dataset**, provided by EyePACS, a free platform for retinopathy screening. It is a large set of high-resolution retinal images, taken under a variety of imaging conditions. For every subject, the dataset contains two images corresponding to the left and the right fields. The resolution ranges from 320×211 pixels to 5184×3456 pixels. The dataset is annotated in image level in five classes (0 to 4) according to the DR stage. The Kaggle/EyePACS dataset comprises a total of 88,702 images.

- **IDRiD dataset**, is a dataset captured by a retinal specialist at an Eye Clinic located in Nanded, Maharashtra, India, and provided by a recent competition of segmentation and grading of diabetic retinopathy. IDRiD images were acquired using a Kowa VX-10 alpha digital fundus camera with 50° field of view (FOV), and have a resolution of $4,288 \times 2,848$. The entire dataset is annotated in image level regarding severity (according to the ICDR scale), and a portion of the data has additional pixel-level annotations about signs of DR. The IDRiD dataset comprises a total of 516 images.

- **Phelcom dataset**, collected with the Phelcom Eyer in the Barretos Cancer Hospital, Pio XII Foundation, Barretos, Brazil. The dataset is annotated in image level for quality assessment. The Phelcom dataset has 600 eye-fundus images.

Both DR1 and DR2 datasets are publicly available under accession number 10.6084 and URL `http://dx.doi.org/10.6084/m9.figshare.953671`. The datasets were collected in different environments with different cameras, at least one year apart and in different hospitals.

The Messidor dataset, kindly provided by the MESSIDOR program partners, is also available for the scientific community: `http://messidor.crihan.fr`. In Messidor, the images are annotated not only for the presence of the lesions, but also for the severity, evaluating the number of microaneurysms and hemorrhages (red lesions), the presence or absence of neovascularization (not evaluated in this work), and the proximity of the exudates to the macula. In order to make the cross-dataset classification possible, and

the joint statistical analysis of the two sets of experiments (DR2 and Messidor) feasible, we proposed correspondences in the annotations.

The Messidor-2 dataset, kindly provided by the LaTIM laboratory (see `http://latim.univ-brest.fr`) and the Messidor program partners (see `http://messidor.crihan.fr`), has a mean κ value among the three experts is 0.822. The reference standard for referable diabetic retinopathy is available for researchers[1].

The dataset from the Inala Indigenous Health Service involves the use of images of Aboriginal and Torres Strait Islander peoples, a vulnerable group, and will require ethical oversight to be shared[2].

The Kaggle/EyePACS dataset is originally graded as DR stages, and we convert the labels (source domain) to referral necessity (target domain), following the International Clinical Diabetic Retinopathy recommendations (ICDR) [117]: tagging as non-referable only those patients with no diabetic retinopathy signal or mild non-proliferative diabetic retinopathy (NPDR). Patients with moderate, severe, or proliferative DR must be referred (note that the images are not labeled for macular edema). The conversion of labels is not algorithmic, but it is done manually before training the classifiers. The dataset is available in `https://www.kaggle.com/c/diabetic-retinopathy-detection/data`.

The IDRiD images are graded regarding disease severity level and diabetic macular edema. Part of the dataset, 81 color-fundus images, has pixel-level annotations of lesions such as microaneurysms, cotton-wool spots, hard exudates and hemorrhages. Some images contain multiple lesions. The IDRiD is publicly available in `https://idrid.grand-challenge.org`.

Tables 4.1, 4.2 and 4.3 show annotations occurrences for the datasets, in terms of lesion detection, need of referral and quality grading, respectively.

Table 4.1: Annotation occurrences regarding lesions for the datasets

Lesion	DR1	DR2	Messidor	Inala	IDRiD*
Hard Exudates (HE)	234	79	654	5	81
Superficial Hemorrhages (SH)	102	—	—	—	81
Deep Hemorrhages (DH)	146	—	—	—	80
Red Lesions (RL)**	—	98	226	12	—
Cotton-wool Spots (CS)	73	17	—	—	40
Drusen (D)	139	50	—	—	—
Other lesions, excluding above	—	71	—	—	—
All lesions***	482	149	654	12	81
Normal (no lesions)	595	300	546	18	0

* Herein we consider only the images with pixel level annotation. The total dataset, including the image-level grading data, has 516 images.

** "Red Lesion" is a more general annotation that encompasses both SH and DH, besides microaneurysms.

*** The lesions do not sum to this value because an image can present different types of lesion at once.

In DR1, images with red lesions are annotated with the specific tags deep and superficial hemorrhage. A few images are not only labeled in image level, but also have the

[1] `http://www.medicine.uiowa.edu/eye/abramoff/` (accessed February 13, 2019)

[2] Please contact Prof. Anderson Rocha (`anderson.rocha@ic.unicamp.br`) and Prof. Geoffrey Spurling (`g.spurling@uq.edu.au`) for an application to receive these images.

Table 4.2: Annotation occurrences regarding referral for the datasets

Grading	DR2	Messidor-2	Kaggle/EyePACS*
Negative	337	1,368	28,253
Positive	98	380	6,873

* Distribution of the training set (35,126 images).
These images are originally rated only by DR stages.

Table 4.3: Annotation occurrences regarding quality for the datasets

Grading	DR1*	DR2*	Kaggle/EyePACS**	Phelcom
Poor quality	1,392	194	10,402	200
Good quality	1,300	466	34,300	400

* Poor quality category involves only the blurred images (images of periphery are not considered herein.)

** With one-step self-annotation procedure (see Section 3.3.1).

locations of lesions (in terms of image coordinates), essential in our context of lesion-aware mid-level representation.

In DR2 and Inala, only the general red lesion tag is employed. The DR2 dataset has an additional annotation indicating the need for referral by the patient for follow-up by an ophthalmologist in the following 12 months.

The Kaggle/EyePACS dataset comprises 88,702 images, of which 35,126 are designed for training and 53,576 for testing. For responding the Question **Q2.7**, we use the entire Kaggle/EyePACs (+88k images) as training set.

4.2 Validation Protocol

We use three validation protocols: training and testing from the same dataset (without intersection), the 5×2-fold cross-validation and the cross-dataset validation protocols.

In the first protocol, we perform training and testing operations with different parts of the dataset in a hold-out fashion. The idea here is finding the best configuration of our method in a dataset that already provides a clear division of training and testing (Kaggle/EyePACS). We use this protocol for investigating and responding the Question **Q2.4**, in which we propose our first data-driven model for referable diabetic retinopathy detection.

The 5×2-fold cross-validation protocol consists of repeating by five times the process of two-fold cross validation [26] in which we randomly separate the samples into two groups balanced by class, and use one of them for training and the other for testing. We perform two experiments per step, with the groups switching roles. We use this protocol to compare to previous work, mainly evaluations with the DR2 dataset. We use 5×2-fold cross-validation protocol for investigating and responding the Questions **Q2.1**, **Q2.2** and **Q2.3**, in which we propose the handcrafted direct referral assessment; and for responding the Question **Q2.6**, in which we explore transferring data-driven knowledge acquired with a different task in the context of diabetic retinopathy screening.

Finally, the cross-dataset protocol is the strictest and therefore closer to real-world operational conditions, in which we train and test the classifiers on different datasets collected in very different environments with different cameras, at least one year apart and in different hospitals. This protocol plays an important role in the design, since in clinical practice, rarely the analyzed images will have the same image specification (camera, resolution, operator, FOV) as the images used for training the classification method. Most experiments in this work are performed using the cross-dataset protocol: (1) all the experiments regarding DR lesion detection; (2) all the data-driven experiments regarding referable DR detection (except the one about transfer learning, for purpose of comparison with handcrafted methods); and (3) all the experiments about quality assessment.

4.3 Metrics

To quantify precisely the performance of the proposed method and enable reliable comparisons, we employ receiver operating characteristic curves (ROCs) [27], which plot the compromise between specificity (few false positives) and sensitivity (few false negatives). To quantify performance as a single scalar, the area under the ROC curve (AUC) is applied. Since the classifier can trade specificity for sensitivity, the AUC gives a better overall performance measure than any particular point of the specificity-sensitivity metrics.

Chapter 5

Results: DR Lesion Detection

In this chapter, we present the results for each question regarding detection of diabetic retinopathy lesions. For this topic, we have published three papers: two were in the PLOS One journal [78, 82] for responding questions *Q1.1* and *Q1.4*, respectively; and one was at the IEEE Engineering in Medicine and Biology Society (EMBC'14) conference for responding questions *Q1.2* and *Q1.3* [75].

5.1 Semi-soft Coding for Lesion Detection

The experiments for *proposing a good balance for designing efficient and effective lesion detectors* were performed using a cross-dataset protocol. The entire DR1 dataset was employed as the training dataset. The DR2 and Messidor datasets were then employed for testing.

The detailed results are presented in Tables 5.1 and 5.2, which show the AUCs obtained for each lesion with the DR2 and Messidor datasets.

Table 5.1: AUCs in %, for Training with DR1, Testing with DR2

	Sparse features			Dense features		
	Hard	Semi-soft	Soft	Hard	Semi-soft	Soft
Hard Exudates (HE)	93.1	**97.8**	95.5	94.5	95.6	95.6
Red Lesions (RL)	92.3	**93.5**	87.1	89.1	90.6	89.9
Cotton-wool Spots (CS)	82.1	**90.8**	84.9	84.5	90.4	90.3
Drusen (D)	66.5	82.8	62.6	**84.1**	82.5	75.5

Table 5.2: AUCs in %, for Training with DR1, Testing with Messidor

	Sparse features			Dense features		
	Hard	Semi-soft	Soft	Hard	Semi-soft	Soft
Hard Exhudates (HE)	64.4	70.3	66.2	**70.5**	70.0	70.0
Red Lesions (RL)	77.4	83.1	76.6	**85.2**	85.1	82.5

The results suggest that the best configuration of the BoVW for each lesion (and dataset) are the proposed semi-soft coding on sparse features, except for the drusen,

where semi-soft coding performs better with dense features. We believe that the good performance of dense features on Messidor is due to the presence of very challenging images (patients with very early DR signs, showing very few lesions). However, the results show that the semi-soft coding scheme works well on the Messidor dataset when associated either with sparse features or dense features.

Such local case-by-case analysis, however, fails to account for random effects. A less naïve analysis must take into account all results across BoVW parameters, datasets and lesions. The goal in DR classification is to obtain the overall best configuration for the BoVW, if such configuration can be found with confidence. The DR2 and Messidor datasets provide different annotation standards, with the former having annotations for all four levels of lesions, but the latter having annotations only for hard exudates (HE) and red lesions (RL). This presents a challenge for performing (and to interpreting) such unbalanced experimental designs and separate balanced studies were performed: one considering only DR2 and all four lesions; and another for both test sets, but with only HE and RL lesions.

The box-plot in Figure 5.1 illustrates, for each treatment, how much it improves or decreases the performance of the detection of the lesions, in comparison to the other treatments. As the lesions and datasets vary widely in difficulty, and we are interested in determining a treatment (combination of factor levels) that performs globally better than the others, we analyzed the normalized impact on the AUC of each factor. In order to do that, for each combination of lesion–dataset, we normalized the AUCs (subtracting the mean and dividing by the standard deviation of AUCs for that combination). More formally, the procedure takes each specific lesion ℓ, computes the mean AUC μ_ℓ for all treatments on that lesion, computes the standard deviation of those AUCs σ_ℓ, and then, if the AUC of a specific treatment on that lesion is β_ℓ, the normalized AUC will be $v_\ell = (\beta_\ell - \mu_\ell)/\sigma_\ell$. Therefore, Figure 5.1 shows, graphically, those standardized effects. The correct interpretation of the box-plot shows, for example, that the treatment "sparse–semi-soft" is, on average for all lesions on DR2, one standard deviation above the mean of AUCs obtained by all treatments, i.e., $avg_\ell[v_\ell] \sim 1$.

The synergy between sparse feature extraction and semi-soft coding for DR-lesion classification can be better appreciated in the box-plot of Figure 5.1. Remark that most combinations of feature extraction and coding function have a wide distribution of standardized effect, meaning that they improve the detection of some lesions at the cost of decreasing the performance of others. In contrast, sparse feature extraction and semi-soft coding offer consistently improved results.

In order to obtain quantitative results, we have also performed a factorial ANOVA that formalizes the same experimental design used on Figure 5.1. The following factors (and levels) were employed:

(1) low-level feature extractor (Sparse, Dense),

(2) coding (Soft, Semisoft, Hard), and

(3) test dataset (DR2, Messidor).

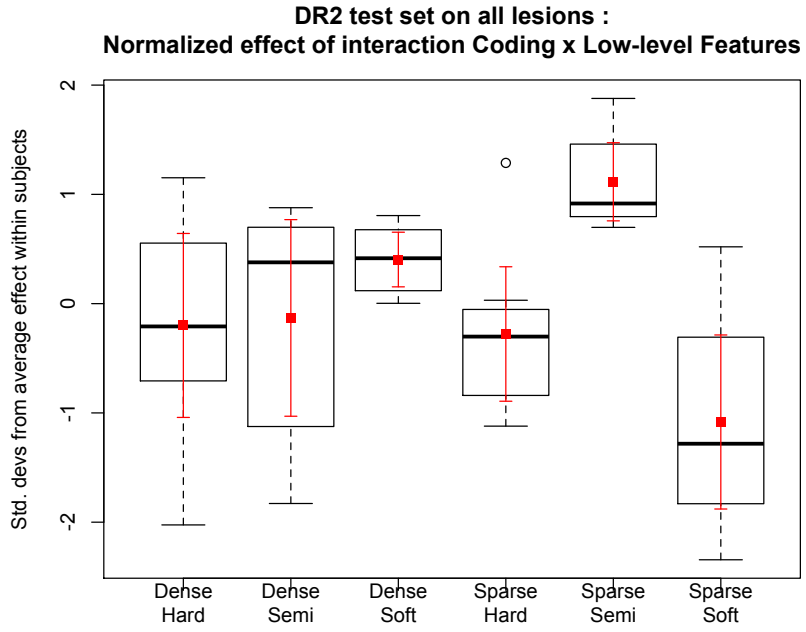


Figure 5.1: Standardized AUCs per lesion, for six combinations of feature extraction and coding (horizontal axis). In the box-plots (black), the whiskers show the range up to 1.5× the interquartile range, and outliers are shown as small circles. Averages (small squares) and 95%-confidence intervals (error bars) are also shown, in red, for the same data. The strong synergy between sparse feature extraction and semi-soft coding is evident: it has consistently improved results for all lesions, while the other combinations improve the results of some lesions at the cost of decreasing it for other lesions (as shown by the spread of the standardized effects in the vertical axis). This plot is based on a balanced design with the DR2 dataset and all lesions, another balanced design with both datasets and two lesions show similar results.

with repeated measures for each lesion (HE, RL, CS, D) and all errors measured within-subjects (the subjects are each individual combinations of lesion and dataset). To remove the strong scaling effect of the lesions and datasets, each subject was independently standardized by subtracting the average and dividing by the standard deviation, as explained above.

The analysis on the DR2 subset indicated an important interaction effect between the choice of Low-level Features and Coding ($p = 0.007$). The main effect of Coding alone just fails significance ($p = 0.062$), and all other effects and interactions are non-significant. These factors have a significant interaction effect due to the two low-level feature extractors providing better results with different coding schemes (Table 5.1). The analysis on the other data subset, with both test datasets and only HE and RL, shows similar results, with significantly better outcomes for the sparse+semi-soft combination ($p = 0.011$).

Therefore, we conclude that *combining advantages of both hard and soft codings provides a good balance for designing efficient and effective DR-related lesion detectors*. Hence, the answer for *Q1.1* is *yes*.

5.2 Preserving Information for Lesion Detection

The experiments for *investigating whether keeping information still provides satisfactory results* were performed using a cross-dataset protocol. The DR1 dataset was used as training set, whilst the DR2 dataset was used as the test set in our experiments. To quantify performance as a single scalar, all the results are reported as the area under the receiver operating characteristic curve (AUC-ROC).

In this work, we demonstrate the methodology described in Section 3.1.2 for the detection of hard exudates and red lesions. The experiments were performed with two distinct codebook sizes: 1,000 and 4,000.

Figure 5.2 shows the ROC curves with their respective AUCs for the detection of hard exudates and red lesions employing the class-based scheme. The results are contrasted with those obtained in [80] (showed in the figures as "Sparse/Hard" and "Sparse/Soft") using the BoVW approach with 500 codewords for sparse low-level technique and class-based scheme (the mismatch in codebook size is due to the fact the previous art performed better with smaller codebooks [87]). As the work [80] already comprises the best results reported in [87], we opted for not repeating them here.

For the detection of hard exudates, initial results using the sparse low-level feature extraction but employing larger codebooks, showed that the BossaNova provides a better representation and a better accuracy. The proposed new method provides an AUC of 95.9%, compared to an AUC of 95.6% obtained by the BoVW with soft-max coding/pooling. When the dense extraction step is used, the difference between BossaNova and BoVW increases: the best result was with 4,000 codewords, with an AUC of 96.4% compared to the BoVW with an AUC of 95.6%, which represents an error reduction of over 18%.

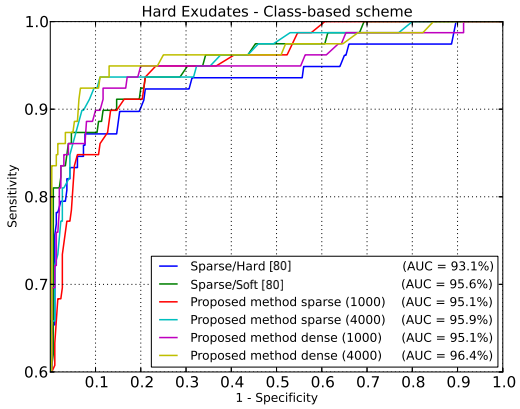
For the detection of red lesions, BossaNova with the sparse low-level feature extraction technique did not provide a significant advantage over the Bag of Visual Words approach based on the hard-sum coding/pooling. AUC of 91.9% was achieved, against 92.3% obtained with BoVW [80]. However, once again, the dense extraction shows its superiority when applied with the BossaNova mid-level feature extraction, presenting an AUC of 93.5% using a codebook of size 4,000, compared to the BoVW with an AUC of 92.3%, which represents an error reduction of over 15%.

Both for hard exudates and for red lesions detection, the results outperformed previous methods showing the importance of preserving some relationship between the detected features in retinal images instead of just throwing them away as previous BoVW-based solutions have done.

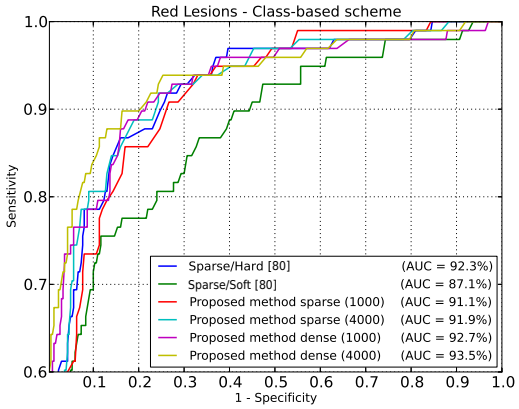
Therefore, we conclude that *preserving information in the pooling process still produces satisfactory results*. Hence, the answer for *Q1.2* is *yes*.

5.3 Class-based Scheme vs. Global Dictionary

The experiments for *investigating whether global codebooks is better than the class-aware approach* were performed with BossaNova for detection of hard exudates and red lesions. The codebook sizes were also 1,000 and 4,000, for comparison purposes.



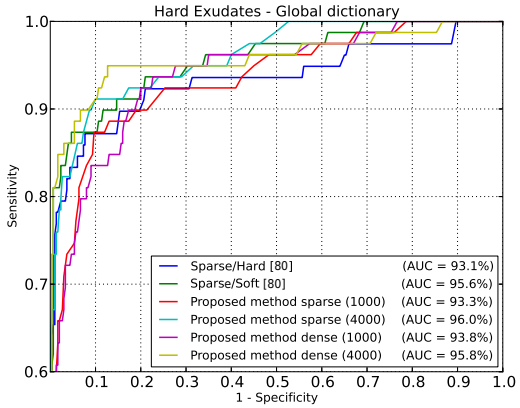
(a) Results for Hard Exudates



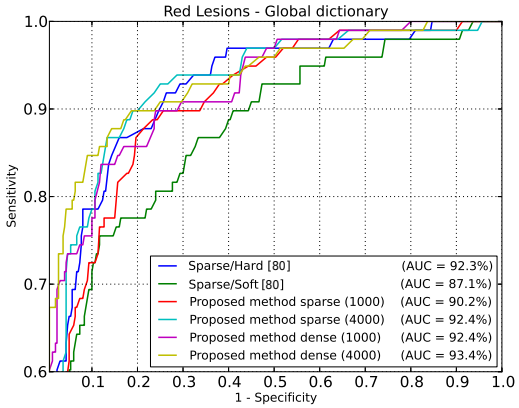
(b) Results for Red Lesions

Figure 5.2: ROC results using class-based sampling for hard exudates detection (a) and red lesions detection (b), for codebooks of sizes 1,000 and 4,000. Both for exudates and red lesions, the best configuration is the new technique, using BossaNova, dense low-level features, and large codebooks (4,000). For hard exudates, the error reduction compared to the prior art is over 18% for the best proposed method; while for red lesions it is over 15% for the best proposed method.

Figure 5.3 shows the ROC curves with their respective AUCs for the detection of hard exudates and red lesions employing the global dictionary approach. In comparison with the results illustrated in Figure 5.2, we can note that, both for exudates and for red lesions, the class-based approach that have been used by our research team [46, 80, 81, 87] is more robust for retinal image representation than the global codebook approach.



(a) Results for Hard Exudates



(b) Results for Red Lesions

Figure 5.3: ROC results using the global codebooks for hard exudates detection (a) and red lesions detection (b), for codebooks of sizes 1,000 and 4,000.

Therefore, we conclude that *global codebooks do not improve the results achieved using the class-aware approach*. Hence, the answer for *Q1.3* is *no*.

5.4 Diabetic Retinopathy Screening for Isolated Indigenous Communities

The experiments for *investigating the diagnostic of patients in different ethnic groups* aims at verifying if it is possible to train classifiers for DR lesion detection using images from a given population strata and test such classifiers with completely different images coming from another population strata [78].

We turn our attention to showing that the developed lesion detectors have good generalization power when training on one dataset and testing on another. In this case, we train the lesion detectors using the DR1 dataset, which comprises mostly Caucasian peoples and test such classifiers with images from an Aboriginal and Torres Strait Islander population. The use of the visual word dictionary on this Aboriginal and Torres Strait Islander population sample achieved an AUC of 97.8%, with a sensitivity of 100.0% and specificity of 88.9% for bright lesion detection, as shown in Figure 5.4.

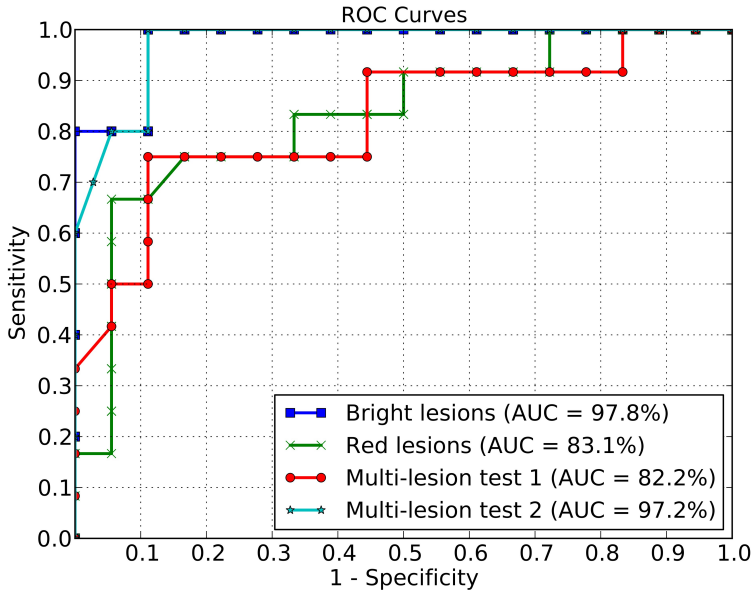


Figure 5.4: DR Screening for isolated indigenous communities: Red lesion, bright lesion, and multi-lesion detection.

For red lesion detection, the classifiers achieved an AUC of 83.1% (Figure 5.4), with 67.0% sensitivity and 95.0% specificity.

The multi-lesion classifier combined the red and white lesion classifier outcomes. The image is classified as normal if both the red and bright lesion classifiers agree it is normal, which minimizes the classification of false-negatives.

For identification of multiple lesions, two different test sets were chosen. The first set contained a mix of 18 normal images and 12 images with red, bright, or both types of lesions (5 images), representing what may be seen in a screening program. An accuracy of 82.2% was achieved with a sensitivity of 75.0% and specificity of 88.9%. The ROC curve is represented in Figure 5.4 as *Multi-lesion test 1*.

The second test compared 18 normal images and 5 images with both lesions (red and bright lesions) present. The result achieved an accuracy of 97.2%, associated with a sensitivity of 100.0% and specificity of 88.9%. The ROC curve is represented in Figure 5.4 as *Multi-lesion test 2*.

Therefore, we conclude that *patients in isolated indigenous communities can be diagnosed using lesion detectors trained with data from different ethnic groups*. Hence, the answer for *Q1.4* is *yes*.

Comparison with the State of the Art

For any automated computer-based screening to be clinically useful, it needs to meet standards such as the St Vincent declaration, which recommends a sensitivity of 80% and specificity of 95% in agreement with the British Diabetic Association. The Australian National Health and Medical Research Council (NHMRC) guidelines for primary health care require 60% sensitivity with 90-95% specificity [15, 62, 116]. The current study investigated a multi-lesion classifier that does not require pre- or post-processing, and is robust against differences between training and testing images as well as differences in ethnicity and meets the NHMRC standards.

Our result for determining red lesions in Aboriginal and Torres Strait Islander images outperforms previously reported results within the acceptable range for Australian screening programs [62], although the test image battery was smaller compared to previous reported studies and the test images were of higher resolution than the training images. In addition, the number of training images with only red lesions was smaller than the training set for bright lesions.

Bright lesion detection for diabetic retinopathy using the visual word dictionary was highly accurate being equal to or improving on previous studies. Our results indicated 100% sensitivity and specificity of 88.9% despite the use of a generic training set which included no Indigenous images nor pre- or post-processing of any kind. By comparison, a recent study by Niemeijer et al. achieved a sensitivity of 95.0% and specificity of 88.0%, for the detection of bright lesions [67], albeit in a non-Aboriginal cohort.

Ideally, automated multi-lesion detection systems should be able to detect specific lesions present in any combination and at any time of retinopathy progression. Only a few studies have reported on multi-lesion detection in diabetic retinopathy. Abràmoff et al. combined existing software from diverse laboratories and achieved a sensitivity of 84.0% with 64.0% specificity (accuracy 87.0%) [3]. A larger cohort improved the accuracy to 90.0%. Fleming et al. reported that inclusion of exudate and blot hemorrhages detection improved the overall sensitivity of detecting referable retinopathy from 94.9% to 96.6% [28]. The exudate detection algorithm proposed by Fleming and collaborators is based on the microaneurysm detector previously reported and adds white lesion detection [68]. Separate pre-processing of images with white and red lesions being detected in sequential steps as well as requiring prior identification and removal of poor quality images is required by this software. A more recent study by Abràmoff et al. also discussed referable retinopathy accuracy but did not discuss bright or red lesion detection accuracy nor multi-lesion detection [1].

Chapter 6

Results: Referable DR Detection

In this chapter, we present the results for each question regarding referable diabetic retinopathy detection. For this topic we have one paper published in the IEEE Journal of Biomedical and Health Informatics (J-BHI) [76] for responding questions *Q2.1*, *Q2.2* and *Q2.3*. We also have one paper published in the Elsevier Artificial Intelligence in Medicine (AIIM) [77] for responding questions *Q2.4*, *Q2.5* and *Q2.6*. We were also invited to publish a chapter in the book *Photo Acoustic and Optical Coherence Tomography Imaging: An Application in Ophthalmology* [79]. The manuscript responds questions *Q2.7*, *Q2.8* and *Q2.9*.

6.1 Beyond Lesion-based Diabetic Retinopathy: a Direct Approach for Referral

The experiments for *evaluating whether the proposed direct referral leads to better decisions than the previous lesion-based schemes* were performed with the 5×2-fold cross-validation protocol using the DR2 dataset.

In our experiments, we extracted visual codebooks of {1,000, 2,000} visual codewords.

In order to investigate the hypothesis that *lesion detection is nonessential for an effective referral assessment*, we use exactly the same mid-level features both for the traditional lesion-based method and the current approach. Both employ the BoVW with semi-soft coding explained in [82].

Figure 6.1 shows the results on the DR2 dataset for both methodologies: lesion-based [82] (AUC = 94.2%) and the best direct-referral. Direct referral performs better with 2,000 codewords, reaching an AUC of 94.7%.

The results obtained with BoVW, shown in Figure 6.1, validate our hypothesis that detection of individual DR lesions is not necessary to provide effective referral decisions.

Therefore, we conclude that *forgoing the detection of individual DR lesions still provides effective referral decisions*. Hence, the answer for *Q2.1* is *yes*.

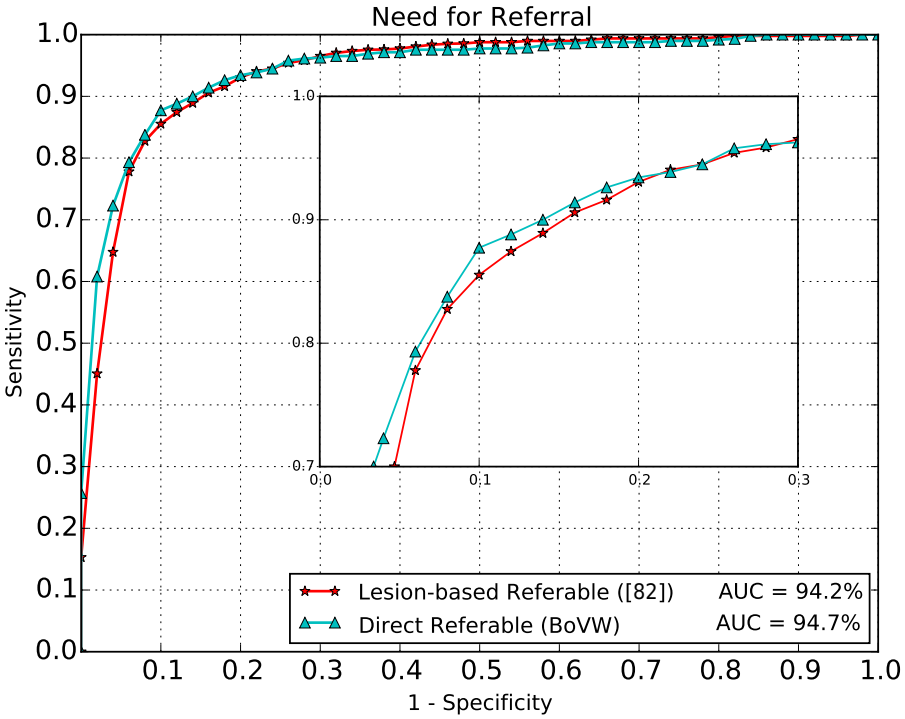


Figure 6.1: ROC results for direct assessment for need of referral using BoVW mid-level characterization approach. The experiments were performed for hypothesis validation.

6.2 Direct Referral with Sophisticated Mid-level Features

In the experiments for *evaluating whether sophisticated mid-level features improve direct referral decision*, the results obtained in Section 6.1 will act as a baseline. We evaluate two recent mid-level features: BossaNova [8] and Fisher Vector [74].

We extracted visual codebooks of {1,000, 2,000} visual codewords. Except for the number of visual codewords, we kept the default BossaNova parameter values the same as in [8]. For Fisher Vector, we used GMM with {128, 256} Gaussians after reducing the dimensionality of the SURF descriptors to 64 by applying Principal Component Analysis (PCA), as suggested in [91].

Figure 6.2 shows the best results achieved with each mid-level feature. While BoVW achieved its best result with 2,000 codewords (AUC = 94.7%), BossaNova reached the best AUC using 1,000 codewords (AUC = 95.7%). Finally, Fisher Vector obtained the best result using just 128 Gaussians (AUC = 96.4%).

The results presented in Figure 6.2 express how accurate are the referral decisions by direct assessment, emphasizing that richer representation approaches yield better results for referable DR detection. We highlight that the Fisher Vector approach outperforms the traditional BoVW, reducing the classification error by over 30% (5.3% to 3.6%).

Therefore, we conclude that *sophisticated mid-level features improve direct referral decisions*. Hence, the answer for *Q2.2* is *yes*.

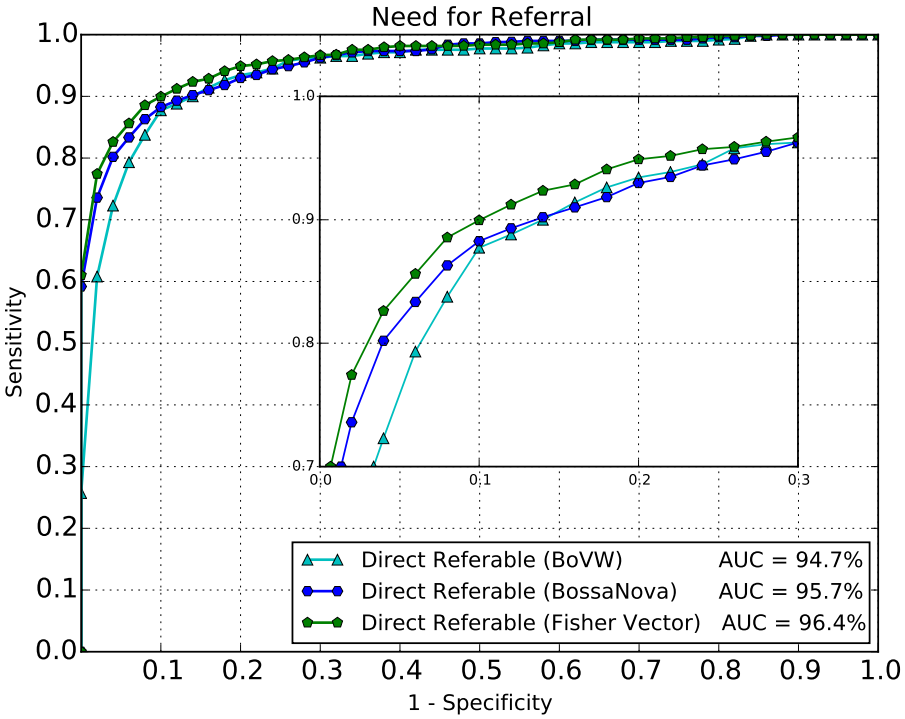


Figure 6.2: ROC results for direct assessment for need of referral using advanced mid-level characterization approaches. The experiments were performed for improvement of the method.

Statistical Analysis

In this section we explore the significance of the previous results.

Each single experiment requires picking a large number of parameters (mid-level representation, codebook size), that may have interactions. In order to investigate if the choice of mid-level feature would still appear significant when considering the totality of experiments performed, we applied a factorial analysis of variance (ANOVA) [45, chap. 22], with a block design using the folds as blocks, on the DR2 dataset. Since the AUC is a rate and behaves very non-linearly at the extremes of the [0-1] scale, we employ the more linear "log odds" scale (logit). We lessened the nuisance effect of the choice of the training set, by subtracting the global average of each fold from the results relative to that fold. In Table 6.1, the statistical results reinforce the importance of the choice mid-level representation (p-value < 0.001). Note that the mid-level is responsible for more than 40% of the variation (see the column 'Sum of squares').

6.3 Direct Referral in a Public Dataset

The experiments for *confirming the suitability of direct referral in an independent dataset* aims at reinforcing the direct approach. We use the Messidor dataset as benchmark.

Once again, we perform this experiment with both the current direct referral and the previous lesion-based approaches. Figure 6.3 depicts the results reached with the

Table 6.1: Partial view of the ANOVA table. We omit the second-order interactions since none of them were significant. The choice of mid-level representation explains the non-random variation, as seen in the Sum of Squares column.

Parameter	Degrees of freedom	Sum of squares	Mean square	F value	p-value	
mid-level	2	59.77	29.885	20.787	2.02×10^{-7}	***
codebook	1	2.57	2.568	1.786	0.187	
residuals	54	77.64	1.438			
total	59	143.86				

Significance codes: *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05

lesion-based method, as well as the best results achieved with the direct approach for each mid-level representation. For BoVW, the best result was obtained with a codebook of size 2,000 (AUC = 79.1%). BossaNova reached its best AUC using 2,000 codewords (AUC = 85.6%). For Fisher Vector, the best result was achieved using 256 Gaussians (AUC = 86.3%).

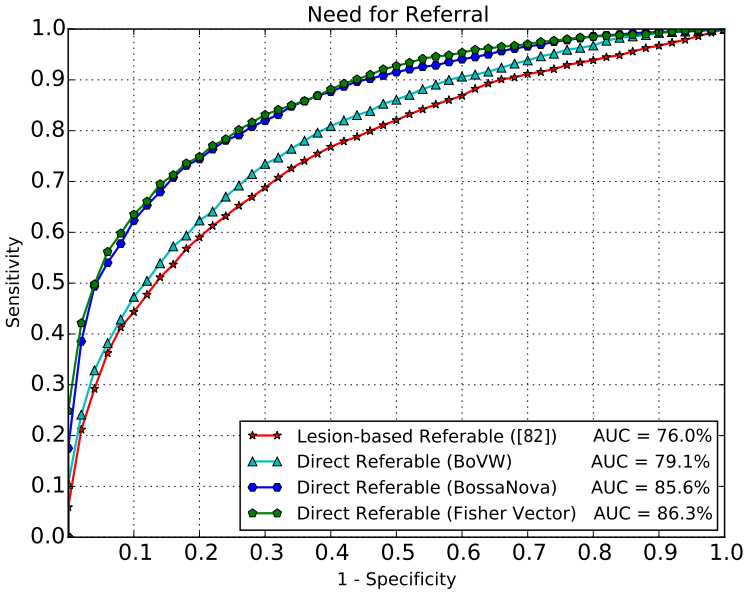


Figure 6.3: ROC results for direct assessment for need of referral using advanced mid-level characterization approaches. The experiments were performed for emphasizing the fitness of the method in the Messidor dataset.

While the lesion-based method obtained an AUC of 76.0%, we achieve the promising result of 86.3% using the current method that does not depend of lesion detection. The results reveal Messidor as much more challenging dataset than DR2. Nevertheless, the relative performance of the techniques confirm the interest of the direct referral choice, which, in this case, appear prominently better.

Therefore, we conclude that *the direct referral is suitable also for independent datasets*. Hence, the answer for *Q2.3* is *yes*.

6.4 A Data-Driven Approach to Referable Diabetic Retinopathy Detection

The experiments for *learning data-driven models that leverage referable diabetic retinopathy detection without manual feature engineering* start from scratch and refine the solution according to the performance for referral assessment.

In this section we present the first model which we will refer to as baseline, and investigate some hypothesis aiming at progressively improving the approach before advancing to the following steps and answering the latter questions. We use the protocol of training and testing with the Kaggle/EyePACs dataset following the original splits exposed in Chapter 4.

In this first version of the solution, we still do not employ any data augmentation technique nor use the augmented feature extraction policy discussed in Section 3.2.4. We perform a naïve data balancing through sample removal for the most favored classes, repeating the process in each epoch. Such initial baseline method leads to an AUC of 71.6%. Starting with this baseline, we now break the Question **Q2.4** and pose a series of research questions in order to evaluate possible improvements and design decisions. For reference, questions Q1–Q4 refer to results depicted in Fig. 6.4.

Q1: *Is data augmentation essential to train the proposed CNN?* To investigate the first question, we applied geometric image perturbations and color augmentation, always aiming at balancing the classes, as detailed in Section 3.2.4. After augmenting the training set, the CNN reached an AUC of 93.1%. The data augmentation remarkably improved the initial results, showing that it is critical to choose a good policy of data augmentation/balancing.

Q2: *Is the multi-resolution training important to train with larger images?* To investigate this question, we start the data-driven process by training reduced versions of the architecture, adapted to images of lower resolutions, as explained in Section 3.2.4. The weight initialization slightly boosted the AUC to 93.9%, and showed that it was essential for the convergence of the CNN since deeper networks require larger datasets, and is a satisfactory option to provide an effective solution.

Q3: *Is the robust feature-extraction augmentation satisfactory?* Here we extracted features from augmented versions of each image and created a final feature vector by concatenating mean and standard deviation of the image versions, as exposed in Section 3.2.4. With the augmented features, which feed an extra neural network of two hidden layers for referral decision, we achieved an AUC of 94.6%, a satisfactory improvement.

Q4: *Is the per-patient analysis important to provide more robustness?* Since the Kaggle/EyePACs dataset contains images from the left and right eyes for each patient, we contrasted per-image decision (each image analyzed independently) to per-patient decision (aggregating features of both eyes). In the system we advance, the patient need of referral is the highest resulting classification probability between the two eyes (see Section 3.2.4). The per-patient analysis improved results considerably, leading to a 95.5% AUC (95% CI: 95.1% – 95.8%).

After testing each progressive enhancement, the final solution comprises a (1) CNN trained with data augmentation; (2) employing weight initialization with parameters from

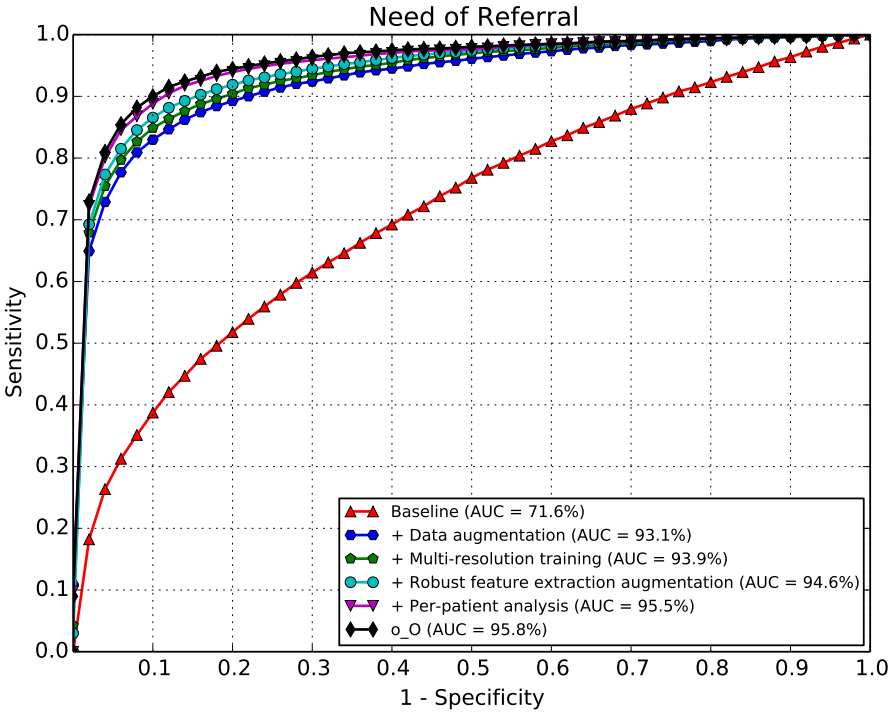


Figure 6.4: ROC for referral assessment on the Kaggle/EyePACs dataset (with the official competition splits). The baseline CNN results are compared to the progressive proposed improvements, showing that those are advantageous for the task. We also compare with the original o_O's solution (ensemble of six classifiers).

smaller networks; (3) using robust feature-extraction augmentation and training a decision classifier (Neural Network) on top of these features. In addition, we also consider (4) diagnosing patients with images from both eyes. We compare our method with the original o_O's proposal. Using an ensemble of **six** networks (two physical networks with three distinct sets of parameters) and a per-patient analysis, the o_O solution yields an AUC of 95.8% AUC (95% CI: 95.5% – 96.1%), while our method, which only relies on one network instead of six, yields and AUC of 95.5%.

Therefore, we conclude that *data-driven models leverage referable diabetic retinopathy detection*. Hence, the answer for *Q2.4* is *yes*.

6.5 Cross-Dataset Validation Protocol

The experiments for *diagnosing retinal images collected under different acquisition conditions* consists on validating the solution of last section with distinct datasets in a challenging cross-dataset validation. In this section we provide more comparisons regarding efficiency and effectiveness and a deeper cost-benefit analysis of our claims.

After refining the automated solution for referable diabetic retinopathy screening from scratch, we investigate the performance over distinct datasets. Basically, we use the CNN trained with data augmentation and multi-resolution training to extract features for the test sets (robust feature-extraction augmentation), and test the features with the classifi-

cation that provides referral decisions. We emphasize again that one of the most valuable advantages of extracting features (in this case, robust feature-extraction augmentation) is that it provides flexibility to choose different machine learning algorithms. Henceforward, we use two algorithms: Neural Network and Random Forest. Therefore, for research Question **Q2.5**, we follow a challenging cross-dataset validation protocol with the best solution we found in the previous section that was trained with Kaggle/EyePACs data.

We highlight that the solution incorporates the proposed data augmentation for training, multi-resolution training, and robust feature-extraction augmentation steps. We also exploit the per-patient information whenever we have access to images of both eyes.

In this section, we assess the possibility of training a retinal image diagnosis system with one set of images and test it using images collected under very different acquisition conditions. Here, we use Kaggle/EyePACs dataset images to train the expert CNN and then use it as a feature extractor for DR2 and Messidor-2. For DR2, the per-patient analysis is not feasible as the dataset does not have two images per individual.

Fig. 6.5 depicts the ROC curves achieved with the two considered classifiers in the cross-dataset validation protocol with the DR2 or Messidor-2 datasets for testing. The Neural Network-based classifier yields the best result for testing with DR2 dataset, with an AUC of 96.3% (95% CI: 93.8% – 98.1%). With Random Forest, in turn, the AUC is 96.1% (95% CI: 93.1% – 98.0%). The results show that the models learned with the Kaggle/EyePACs dataset images (with higher variance) produced relevant results with a very different dataset (DR2).

We now turn to the case in which we extract features from the Messidor-2 dataset using network and parameters learned with the Kaggle/EyePACs dataset and test the method with the individual decision models (Neural Network and Random Forest) on Messidor-2 dataset. As the Messidor-2 dataset also provides pairs of images from left and right eyes for each patient, we employ the per-patient analysis herein. The Neural Network classifier reached the best result, resulting in an AUC of 98.2% (95% CI: 97.4% – 98.9%) in a per-patient analysis. The Random Forest algorithm achieved an AUC of 97.9% (95% CI: 97.0% – 98.6%). These results show the solution has a remarkable performance also with Messidor-2 dataset, even considering the challenging cross-dataset validation protocol.

These results corroborate the hypothesis that *it is possible to train a robust data-driven solution to precisely pinpoint diabetic retinopathy referral needs, independently of the ethnicity, operators, and camera settings of the training set of images*. Hence, the answer for *Q2.5* is *yes*.

Comparison with Previous Work on Messidor-2 Dataset

Just for the sake of completeness, we now compare our work (solution explained in Section 6.4) with Abràmoff et. al [1, 2]. Note, however, that this comparison is not totally direct as both methods use different training sets. Considering the Messidor-2 dataset on a per-patient basis, Abràmoff et. al [1] reported an AUC of 93.7%, and after replacing most of the feature detectors with CNNs trained to detect features, further improved to 98.0% (95% CI: 96.8% – 99.2%) [2].

In turn, the method described here and trained with Kaggle/EyePACs data produces

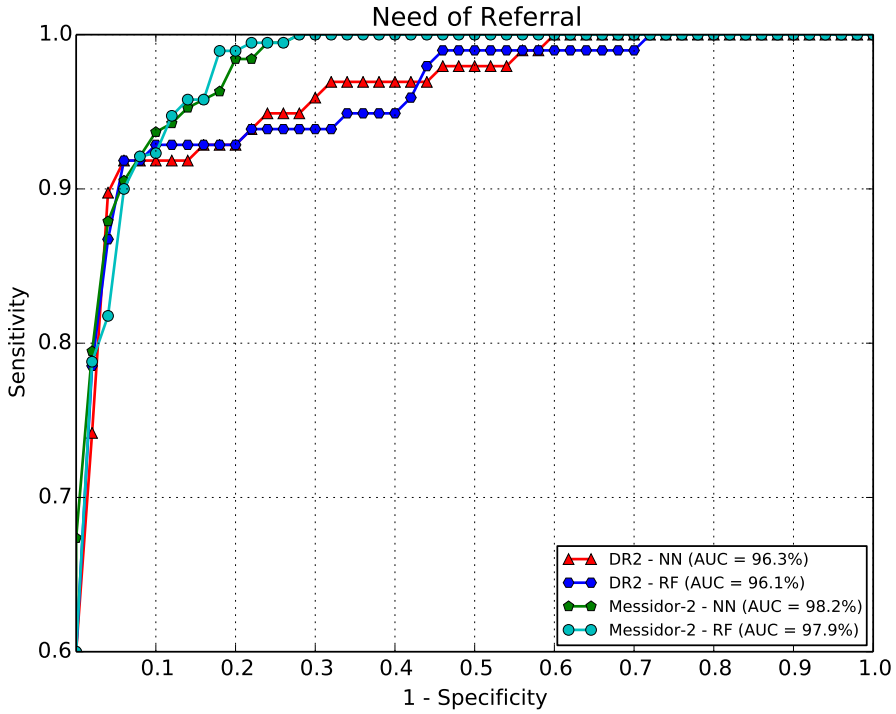


Figure 6.5: ROC results for referral assessment in a cross-dataset protocol (training with Kaggle/EyePACs; and testing with either DR2 or Messidor-2). The cross-dataset protocol is the most strict and realistic one, as the dataset present different acquisition characteristics (operators, equipment, population of patients, etc.)

equally remarkable results for the Messidor-2 dataset with an AUC of 98.2% (95% CI: 97.4% – 98.9%), reinforcing the fact that detecting diabetic retinopathy lesions is not essential for a reliable and effective diabetic retinopathy screening. Note here that, for the method in this work, the CNN was trained and optimized using only Kaggle/EyePACs data, and Messidor-2 data was never used for optimizing nor training the CNN, showing the robustness of the method.

Comparison with o_O's solution

For the sake of comparison, we adapted the o_O solution for a two-class classification problem and evaluate its performance for referable DR detection in terms of efficiency and effectiveness. To do so, we replaced the last one-neuron fully-connected layer by another with two neurons (one per class), and tackled the problem with a classification point of view rather than regression (cross-entropy as objective function, instead of mean squared error). We recall that the o_O solution consists on a ensemble of six different models, trained with features extracted from two different CNNs.

Table 6.2 reports the time and memory required for inference. We performed the tests using one GeForce GTX TITAN X. We simulated a real-time diagnostic environment in which 50 patients are screened for referral after capturing their fundus images (left and right), totaling one hundred images. In terms of time consumption, we report the "real" time, that encompasses loading all required libraries, loading parameters of the CNNs,

pseudo-augmenting and describing the images and, finally, the inference part with higher probability among the eyes. For the memory footprint, we consider the disk space required to keep the CNN parameters and of the two-hidden-layer neural networks in memory.

Table 6.2: Efficiency: Time and memory comparisons.

Work	Time (s)	Memory (MB)
o_O	295	285
Ours	**60**	**56**
Improvement	4.91×	5.08×

Fig 6.6 depicts the results achieved with the current work and o_O proposal, both using the cross-dataset validation protocol over DR2 and Messidor-2 datasets. The DR2 results correspond to diagnosing one image at a time while Messidor-2 are for patients analysis. For DR2, the o_O's ensemble reached an AUC of 96.1% (sens. = 86.7%, spec. = 95.5%), while we achieved an AUC of 96.3% (sens. = 90.8%, spec. = 95.5%). For Messidor-2, o_O reached 97.9% of AUC (sens. = 98.4%, spec. = 79.5%), and we reached 98.2% (sens. = 95.8%, spec. = 83.3%).

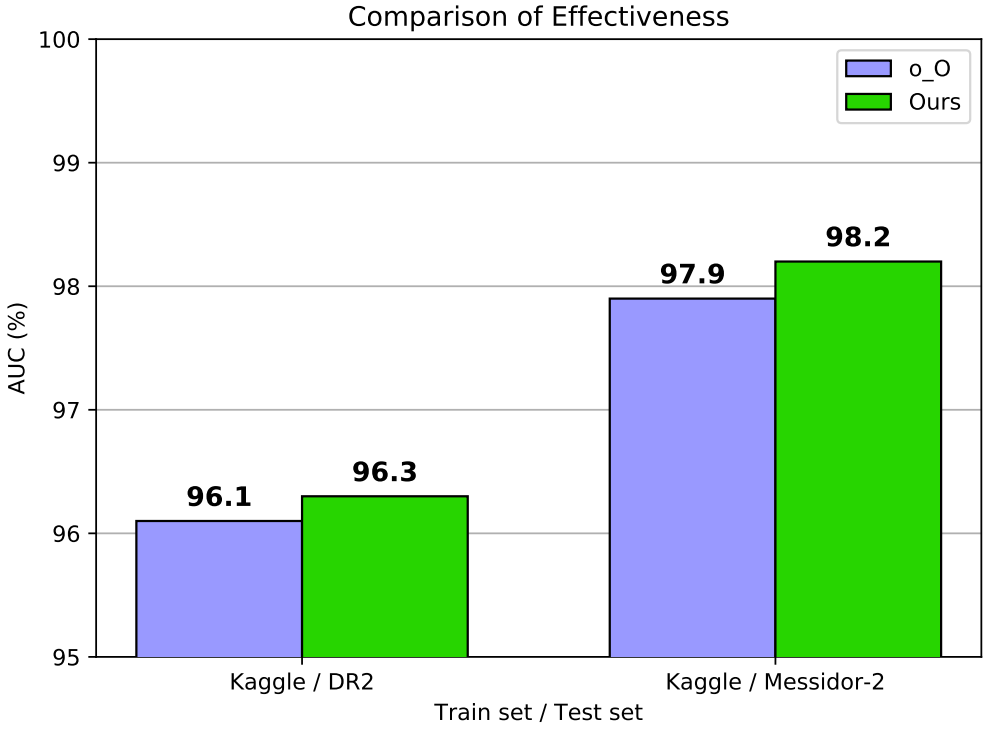


Figure 6.6: Comparison of our solution with the o_O's method on a cross-dataset validation protocol in terms of effectiveness (quality of referral assessments).

The proposed solution yields an improvement close to 5× in terms of efficiency and memory footprint when compared to o_O's method. Furthermore, we have a slightly superior performance in terms of effectiveness, especially considering the difficult cross-dataset validation setup.

6.6 Knowledge transfer for Diabetic Retinopathy Screening

After testing the proposed screening solution under different conditions, we now *investigate the performance of the transfer learning concept to improve decisions about the need of referral.*

The transfer learning field is conducted as a domain adaptation, which generally arises when the goal is learning an effective model from a source task on a different, but related, target task. In this case, we use the same dataset, but adapt the model from a different but related problem: from severity to referability analyses of diabetic retinopathy.

For this experiment, we do not use the CNN trained in Section 6.4 to explore the transfer learning concept. Rather, as we want to evaluate the potential of using transfer learning from a source problem to a target problem, we initially train, from scratch, a CNN with the same architecture (except that the decision layer has five outputs) to assess severity of diabetic retinopathy incorporating the improvements — data augmentation and multi-resolution training — we discussed in Section 6.4. We also apply robust feature-extraction augmentation, but just for referral assessment when transfer learning comes into play. Note that the robust feature-extraction augmentation is not necessary for the source problem, since diagnosing severity is beyond the scope of the current work. Again, we use the per-patient protocol just whenever it is possible.

We explore transfer learning in its two setups — feature extraction and fine-tuning — explained in Section 2.2.1. We start the experiments with the DR2 dataset in a strict 5×2-fold cross-validation protocol using the same dataset splits reported in [76, 80, 82]. Fig. 6.7 depicts the results obtained with the two transfer learning setups and DR2 dataset. Looking at the classification algorithms individually, we note that Random Forests considerably outperforms the Neural Network classifier. Additionally, the fine-tuning setup excels considerably the feature extraction, reaching an AUC of 98.0%.

We also evaluate the effectiveness of transfer learning with the Messidor-2 dataset. Fig. 6.8 shows the ROC curves achieved with Messidor-2 dataset with transfer learning on per-image and per-patient analyses. Observing the results, we note that Random Forest is superior in the two transfer learning scenarios on a per image or integrated per-patient diagnostic.

When evaluating each eye individually (Fig. 6.8(a)), we achieve an AUC of 95.3% when we use the frozen CNN to extract features, and improve to 96.0% when we tune the parameters to the target problem. Diagnosing the patient instead of giving a score for each eye is promising, as Fig. 6.8(b) shows. The performance using fine-tuning with Random Forest is slightly higher than just using feature extraction: 98.3% over 98.2%.

The results reported herein make it clear that *transfer learning with feature extraction is promising for referable diabetic retinopathy detection, and fine-tuning has the potential to enhance considerably the effectiveness of the solution.* The result also confirms that a patient-basis diagnostic decision is more effective than just a single-image based decision. Hence, the answer for *Q2.6* is *yes*.

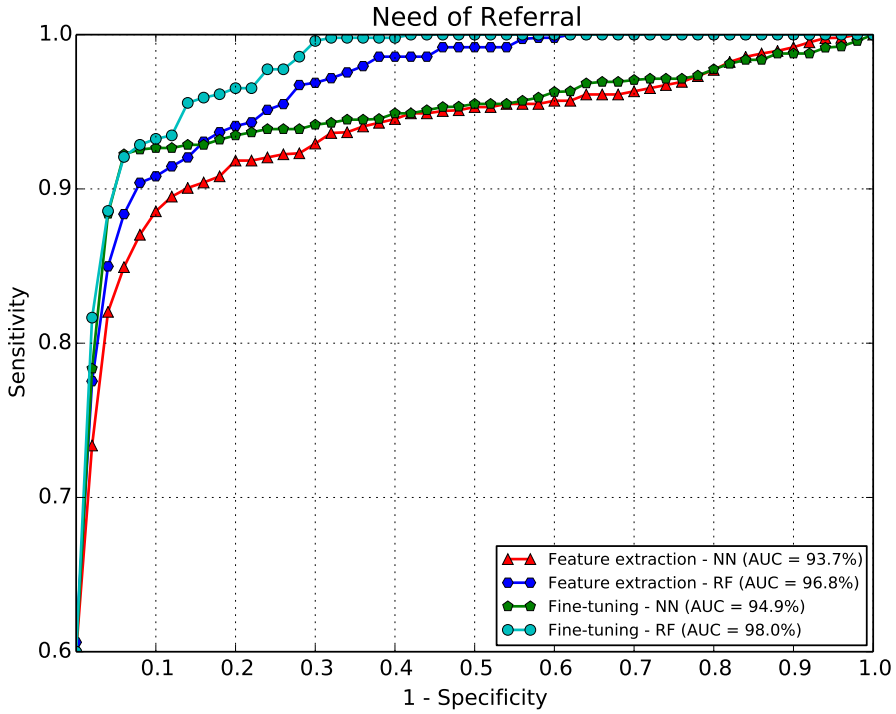
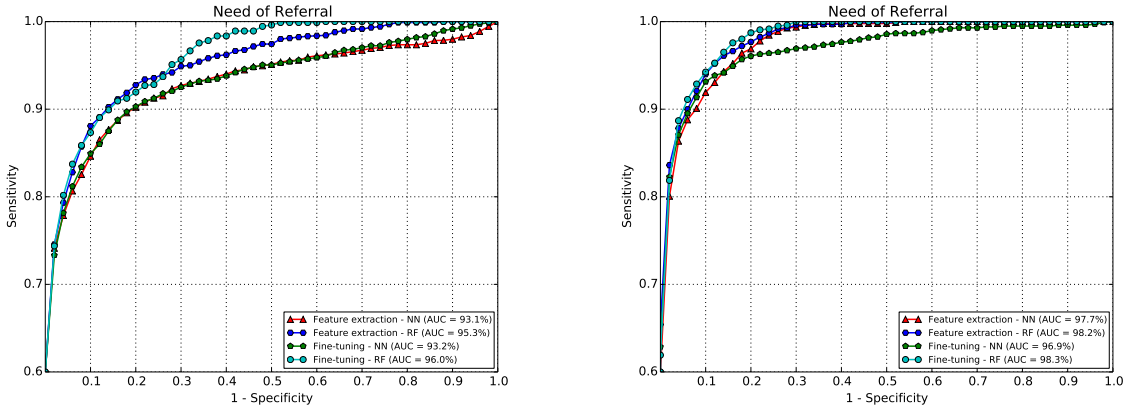


Figure 6.7: ROC results for referral assessment evaluating different transfer learning schemes over DR2 dataset. Two classifiers (Neural Network and Random Forest) are used to make the final decision, with fine-tuning and without it (using the CNN for "feature extraction"). The best technique employs Random Forests with fine-tuning.

Comparison with Related Methods on DR2 Dataset

In this section, we compare our solution with prior work [80, 82] that employed the 5×2-fold cross-validation protocol over the same DR2 dataset. We also compare to our previous proposal based on handcrafted approaches [76], exposed in Section 6.2.

Those researches have proposed general frameworks adaptable to large classes of lesions, and recently bypassed the lesion detection and evaluated directly the referability of diabetic retinopathy. Initially, we provided referral decision with an AUC of 93.4% [80], further improving it to 94.2% by enhancing the lesion detectors with better mid-level image features [82]. Finally, bypassing lesion detection and directly training custom-tailored referral classifiers, we achieved an AUC of 96.4% [76]. Putting in context, the data-driven method described here outperforms all previous solutions with an AUC of 98.0% when using the concept of transfer learning by fine-tuning discussed in Section 3. We reduce the classification error by over 44% over the current state of the art [76], 65% over the solution that applied enhanced lesion detectors [82], and 70% over the first referral proposal with DR2 that depends on explicit information of lesions [80].



(a) Per-Image analysis

(b) Per-Patient analysis

Figure 6.8: ROC results for referral assessment using transfer learning over Messidor-2 dataset, showing that per-patient analysis (b), which combines both eyes, systematically outperforms per-image analysis (a), with independent decisions for each eye.

6.7 Accountable Referable Diabetic Retinopathy Detection

The experiments for *investigating the proposal of an accountable and effective data-driven solution for referable diabetic retinopathy detection* are performed with the Inception-Resnet convolutional neural network [104], whose parameters were previously optimized with ImageNet[1]. We then adapt such parameters for two-class screening of diabetic retinopathy. The weight adaptation starts by training only parameters that are indeed being optimized from scratch (last layer) for a few iterations, to avoid losing patterns previously captured. Thereafter, we propagate the training efforts to the entire network.

In this section, we use the entire Kaggle/EyePACs dataset (88,702 images) for training, except a portion of 10% reserved for validation. We achieved the optimal performance over the validation set (10% of the Kaggle/EyePACS dataset) in terms of AUC after 26 epochs. From then on, kept training the network but the result did not improve. The performance was reached after four learning rate restarts (see Section 3.2.7 for more details regarding learning rates).

After training the model for referable DR screening, we investigate the performance over distinct datasets. The training was performed with the Kaggle/EyePACS dataset, while the test was carried out with Messidor-2 and DR2, which have different acquisition conditions. Basically, we evaluate the global data-driven method in two different setups: simply passing the test set once to obtain responses (softmax probabilities), and passing both training and test sets to extract feature and posteriorly training and validating a particular classifier. We emphasize that one of the most valuable advantages of extracting features is that it provides flexibility to choose different machine learning algorithms later on.

[1]The parameters are available for research purpose under URL `https://github.com/tensorflow/models/tree/master/research/slim` (accessed February 13, 2019)

Fig. 6.9 depicts the ROC curves for training with Kaggle/EyePACS dataset and testing with DR2. Results were achieved in the per-image scenario and two different protocols adopted for global information (softmax probabilities and feature extraction). The CNN alone provided an AUC of 93.73% (95% CI: 89.9% – 96.9%). By feeding the neural network with global features extracted from the CNN, we reach an AUC of 95.8% (95% CI: 93.5% – 97.7%), reducing the classification error by 33% over the softmax version.

The results show the models learned with the Kaggle/EyePACS dataset images (with higher variance) produced relevant results with a very different dataset (DR2).

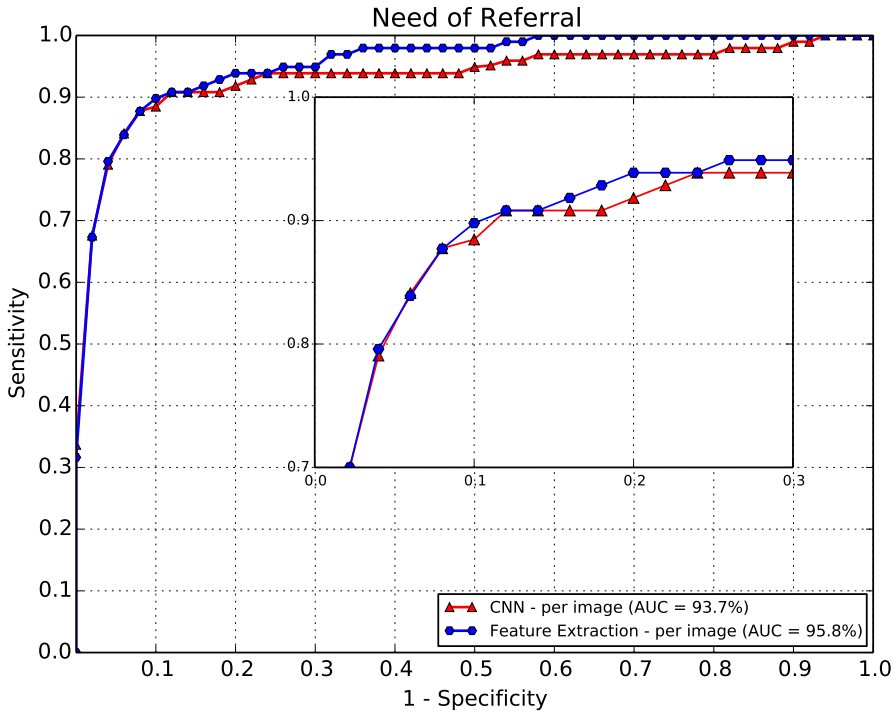


Figure 6.9: ROC results for referral assessment using the proposed global data-driven approach in a cross-dataset protocol: training with Kaggle/EyePACS and testing with DR2.

Fig. 6.10 depicts cross-dataset ROC curves and respective AUCs for testing with Messidor-2 dataset. Using softmax probabilities (the chance of needing consultation), both the image and patient analysis provide AUC of 98.3%[2]. By extracting features and using a new shallow neural network for decision making, we reach an AUC of 97.6% (95% CI: 96.8% – 98.3%) for per-image decision, while diagnosing the patient yields an AUC of 98.5% (95% CI: 97.8% – 99.0%). We emphasize that all results whose approach requires feature extraction, either on a per-image and per-patient basis, involves combining features from left and right eyes. The difference is on measuring the performance with the decisions themselves or attributing the maximum score to the patient.

As our proposal resides on providing a data-driven referable DR detector that is not only effective and accurate, but also self-explainable; we use the guided-backpropagation

[2]AUC of 98.3% (95% CI: 97.8% – 98.8%) for image analysis and 98.3% (95% CI: 97.6% – 99.0%) for patient analysis.

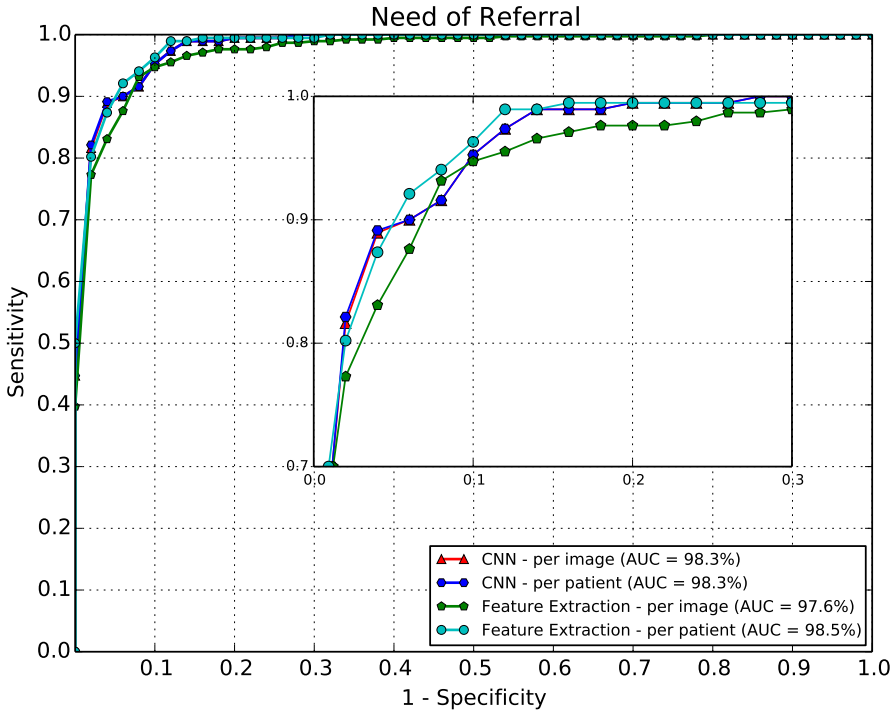


Figure 6.10: ROC results for referral assessment using global data-driven approach in a cross-dataset protocol: training with Kaggle/EyePACS and testing with Messidor-2.

technique to extract pixel importance for the decision taken by the deep-learning model. Fig. 6.11 depicts a retinal image, its respective saliency map towards importance analysis for referral decision and their superposition. In a clinical situation, the superposition could be presented to the ophthalmologists/nurses as well as to the patients to clarify the reasons behind the automated decision.

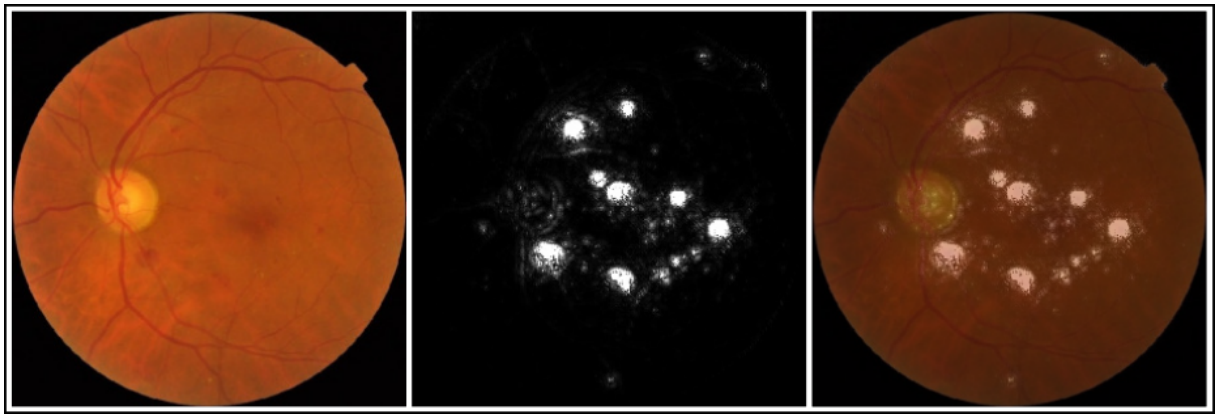


Figure 6.11: Retinal image (left); respective saliency map extracted with guided back-propagation (middle); and the superposition of the map with the image, highlighting important regions and providing an explanation of the reasons behind the decision made by the model (right).

Therefore, we conclude that *effective solutions for referable DR detection can be proposed with purely end-to-end data-driven approach, and heatmaps reflecting pixel impor-*

tance turn the method self-explainable. Hence, the answer for *Q2.7* is *yes*.

6.8 Saliency-Oriented Data-Driven Approach to Diabetic Retinopathy Detection

The experiments for *enhancing the global model with a two-tiered data-driven image representation* explore heatmaps that express pixel importance for extracting regions of interest. Those regions are represented with the same CNN, encoded to provide an additional and complementary model for referable DR detection. We use Fisher Vectors for mid-level representation, with lesion-aware GMMs constructed with two datasets that have regions annotated by experts (DR1 and IDRiD). We evaluate the performance through an analysis per image and per patient, when possible.

Fig. 6.12 shows results achieved with the mid-level representation for testing with DR2 and Messidor-2. For the former, only images are analyzed individually, while the latter consider the two scenarios. Fisher Vector provides an AUC of 97.3% (95% CI: 95.6% – 98.6%) with DR2. Regarding Messidor-2, features extracted through saliency maps activations give an AUC of 97.0% (95% CI: 96.0% – 97.9%) per image, that is considerable improved to 98.7% (95% CI: 98.1% – 99.3%) when left- and right-eye responses are combined.

In both DR2 and Messidor-2, local-based results are significantly superior to the global-based ones, showing the novel approach has potential, and that the operation of emphasizing areas that the CNN model could not assimilate sufficiently is promising.

Therefore, we conclude that *saliency maps represent a powerful mechanism for comprehension of complex models and, additionally, for robust two-tiered image representation*. Hence, the answer for *Q2.8* is *yes*.

6.9 Fusion of Global Data-driven and Local Saliency-Oriented Features to Diabetic Retinopathy Detection

The experiments for *improving the performance of the model by combining global data-driven and local saliency-oriented two-tier representations* involve the combination of responses we achieve with both local and global techniques. In this work, we perform late fusion by averaging the three softmax probabilities: (1) from the CNN, (2) from the shallow neural network trained/tested with data-driven features, and (3) from the shallow neural network trained/tested with the mid-level representations. We present results for DR2 only for analyzing images, and consider image and patient scenarios for Messidor-2 dataset.

Fig. 6.13 depicts ROC curves and respective AUCs for late fusion. We observe the combination of global and local information provides an AUC of 96.3% for DR2 dataset (95% CI: 93.7% – 98.3%). For Messidor-2, in turn, we achieved an AUC of 98.3% (95%

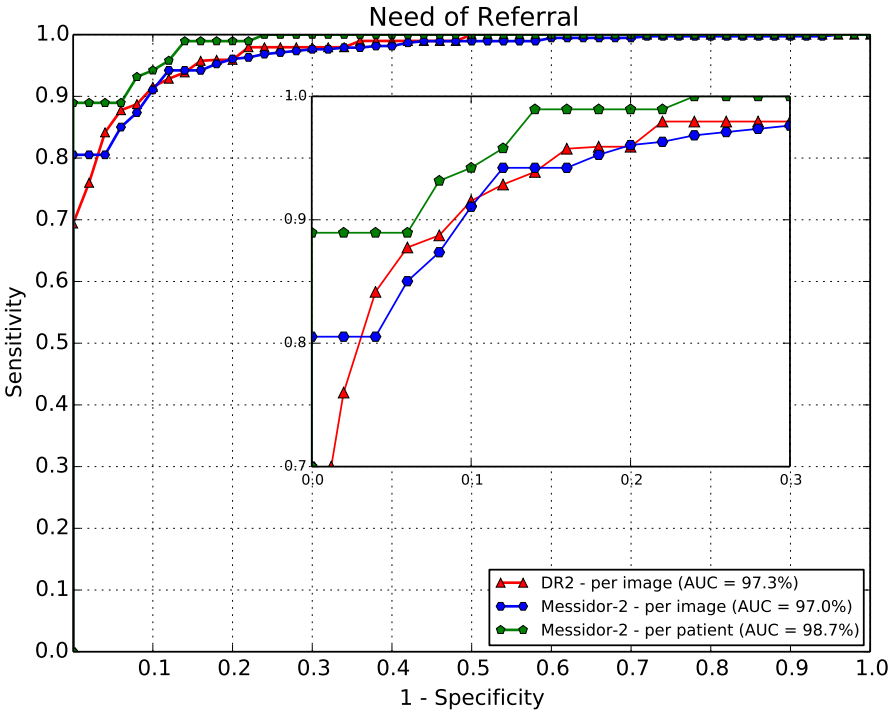


Figure 6.12: ROC results for referral assessment using the proposed local saliency-oriented data-driven approach in a cross-dataset protocol: training with Kaggle/EyePACS and testing with DR2 and Messidor-2.

CI: 97.8% – 98.8%) in the image-based setup, and improved the approach performance to 98.7% (95% CI: 98.1% – 99.2%) by combining eye responses and diagnosing patients.

We note the fusion does not outperform the local-based information (AUC=97.3%) for DR2 dataset, since the performance of the CNN alone achieved an AUC of 93.7%. Possibly, the result did not improve since left and right eyes play an important role in any approach.

The improvement is more evident for Messidor-2. In terms of image analysis, the AUC with fusion is the same as the AUC by the CNN: 98.3%. However, the method improves the classification accuracy from 85.2% to 89.9%. When we diagnose patients, the improvement is even more noticeable. By applying the late fusion of global and local information, in turn, the AUC is equivalent to the one achieved by Fisher Vector encoding (AUC=98.7%). However, the fusion not only increases the accuracy — from 88.3% to 89.5% — but also reduces both false positive and false negative rates (one false negative for ninety-one false positives).

Therefore, we conclude that *the fusion of global data-driven and local saliency-oriented two-tier representations is promising, and the improvement is more relevant when at least one image per eye is available.* Hence, the answer for *Q2.9* is *yes*.

Comparison with State of the Art

Just for the sake of completeness, we now compare our results with all the previous work that has used DR2 and/or Messidor-2 datasets for testing in prior art. The DR2 is widely

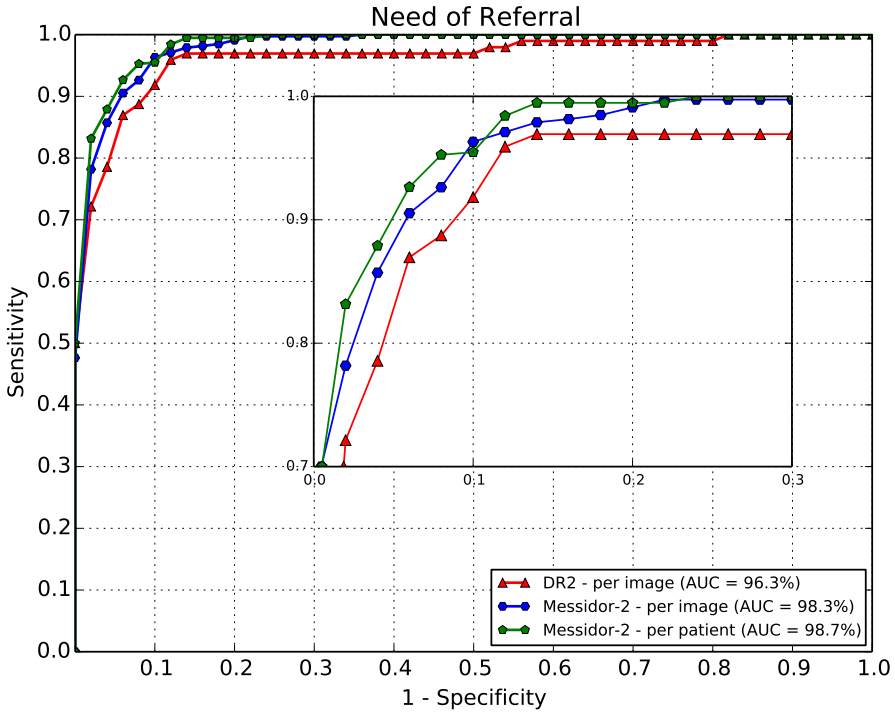


Figure 6.13: ROC results for referral assessment using both the proposed global data-driven approach and local saliency-oriented approach in a cross-dataset protocol: training with Kaggle/EyePACS and testing with DR2 and Messidor-2.

used for referral assessment [76, 80, 82], however it has mostly been applied under the 5×2-fold cross-validation protocol. Here we compare the performance with our recent work, presented in Section 6.5, that also employs the cross-dataset validation policy [77]. In this work, we have reached an AUC of 97.3% by using mid-level representation, and 96.3% using the average of all individual responses. In turn, in Section 6.5, we achieved also 96.3%. Contrasting the current local-based approach with the previous result, we reduce the classification error by 27%. We reinforce that, although we use here a larger training set (the entire kaggle/EyePACS dataset), the result in Section 6.5 [77] is provided by a robust feature extraction augmentation with a convolutional neural network trained under a multi-resolution training procedure. While we train the CNN with images in 299×299, previous work has performed the optimization and validation with images of 448×448 pixels.

In turn, we also compare the performance with previous work that have validated models with the Messidor-2 dataset [1, 2, 77]. Note, however, that this comparison is not totally direct as the methods use different training sets. Considering only the per-patient scenario, Abràmoff et. al [1] reached an AUC of 93.7% and further improved this result significantly to 98.0% [2] (95% CI: 96.8% – 99.2%). More recently, we proposed a data-driven method [77], presented in Section 6.5, reporting an AUC of 98.2% (95% CI: 97.4% – 98.9%), with a CNN with multi-resolution training and extracting global data-driven features in an augmented fashion. Herein we provide a remarkable improvement by reaching an AUC of 98.7% (95% CI: 98.1% – 99.2%), showing the robustness of applying local

saliency-oriented region characteristics to reinforce the learning process the network has acquired globally. We reduce the classification error by 28% over the proposal presented in Section 6.5 [77], and 35% over the solution proposed by Abràmoff et. al [2].

Chapter 7

Results: Quality Assessment

In this chapter, we present the results for the question regarding the integration of our software solutions with simple retinal imaging devices.

7.1 Quality Assessment of images from portable devices

The experiments for *assessing the quality of images from portable devices with data-driven models trained with images from high-cost instruments* were performed using a cross-dataset protocol — training with the DR1 dataset and validating with Phelcom dataset. In practical terms, that protocol puts in an equal footing the Phelcom portable device and the classical, high-cost retinographs.

In the current stage of the project, we used the entire DR2 dataset for validation. As well as in the first stage of the project, we report the results with area under ROC curve (AUC) and accuracy.

The first round of experiments for the third stage of the partnership consists of transferring knowledge learned with ImageNet dataset to the blur detection problem. Note, however, that we bypass the batch normalization technique — that did not show be favorable in our situation — which does not characterize that procedure as a real knowledge transferring.

By training the MobileNetV2 with the DR1 dataset, and picking up the set of parameters that gave the best accuracy with DR2, we achieved an AUC of 91.14% with an accuracy of 87.50% over the Phelcom dataset (specificity of 74.00% with sensitivity of 94.25%). Table 7.1 shows the confusion matrix regarding the best result in the first round (specificity of 84.0% with sensitivity of 87.5%).

These results are promising since we've surpassed the performance reached with Inception-Resnet architecture (AUC of 91.03% and accuracy of 86.33%) using much fewer parameters.

The second round of experiments aims at investigating if the use of more data could improve the effectiveness of the quality assessment network. In order to investigate the hypothesis, we explored the self-annotation procedure for labeling the kaggle/EyePACs dataset to pre-train the mobile network; and posteriorly fine-tuned the model with the

Table 7.1: Round #1: Validating retinal images captured with Phelcom devices for quality assessment. The model was trained with the DR1 dataset.

		Actual class	
		poor	good
Predicted class	poor	148	23
	good	52	377

DR1 dataset.

After training the MobileNetV2 network with the self-annotated dataset and selecting the best parameters (based on performance with DR2), we tested it with Phelcom dataset, reaching an AUC of 92.60% with an accuracy of 87.00%. In sequence, we fine-tuned the model with DR1 dataset and pick up the best parameters in virtue of DR2 performance. The method enhanced substantially the results, providing an AUC of 93.6%, with 87.33% of accuracy (specificity of 86.00% with sensitivity of 88.00%). Table 7.2 shows the confusion matrix regarding the best result in the second round.

Table 7.2: Round #2: Validating retinal images captured with Phelcom devices for quality assessment. The model was trained with the DR1 dataset and self-annotated Kaggle/EyePACS dataset.

		Actual class	
		poor	good
Predicted class	poor	172	48
	good	28	352

Figure 7.1 shows the ROC curves that highlight the potential of mobile architectures to address the problem of quality assessment, and the promising benefit of applying self-annotation to increase the dataset.

Therefore, we conclude that *mobile data-driven models trained with images from high-cost instruments can evaluate the quality of images taken with low-cost devices*. Hence, the answer for *Q3.1* is *yes*.

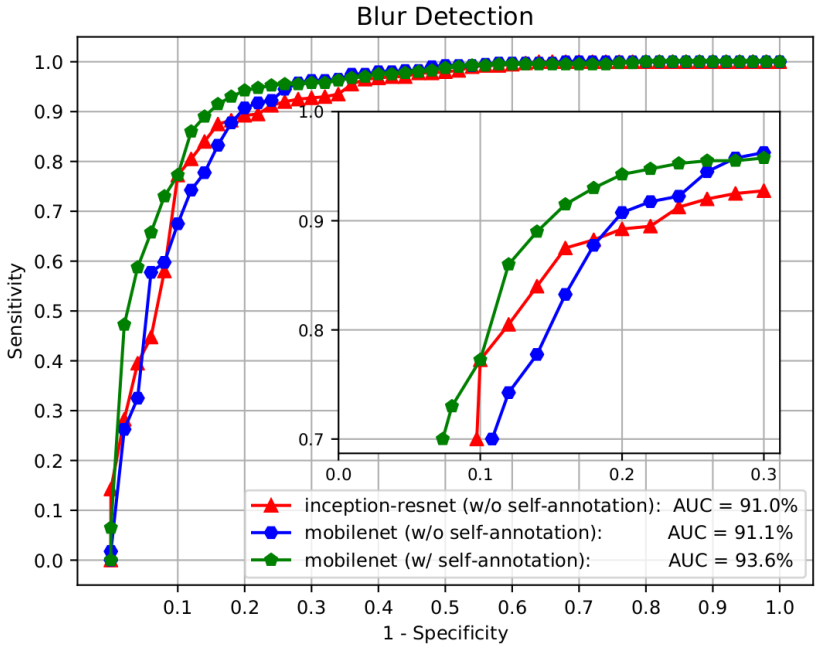


Figure 7.1: AUCs in % for quality assessment, validating retinal images captured with Phelcom devices.

Chapter 8

Conclusions

In this work, we presented image analytics solution for diabetic retinopathy detection, that encompasses *DR Lesion Detection*, *Referable DR Detection* and *Validation with Images from Portable Devices*. The work incorporates both handcrafted and data-driven approaches, on bottom-up and top-down scenarios, one-stage (global based) and two-tiered (local based) representations, and ends up by combining methodologies proposed at the initial stages and the final stages of the doctoral program, compounding a sophisticated accountable and saliency-oriented mid-level representation.

Prior research on automated screening of diabetic retinopathy has followed the natural route of identifying DR lesions and gathering all the extracted high-level information to evaluate referability and/or disease severity. That scenario inherits the common and natural bottom-up fashion: *lesion first, referral later*. Before the explosion of data-driven methodologies, we proposed an approach for direct referral assessment that contradicted previous beliefs regarding the mandatory necessity of detecting lesions, suggesting that the loss of information in the interface between the two-stage classification is detrimental to accurate diagnostic. The adoption of deep-learning strategies came to confirm our previous conclusion, dramatically improving the performance and becoming the state of the art for automated DR screening. With the recent demand for accountable solutions — not only robust and accurate, but also self-explainable — the research has moved toward an opposed *top-down* direction: *referral first, lesion later*. As the models do not require lesion information for deciding about referable DR, pinpointing pathologies or anatomical parts through pixel importance turn out a fashionable artifice for understanding the reasons behind a computer-aided screening.

Throughout the research, we explored prevailing and ongoing computer-vision and machine-learning methodologies, and proposed novel and advanced retinal image analytic approaches, with invaluable impacts both in biomedical and technical contexts.

In terms of clinical and biomedical impacts, we highlight the proposal of a "polemic" method to automatically refer or not the patient to the ophthalmologist, bypassing the cumbersome task of learning individual lesion detectors. We also designed an effective and efficient data-driven model for the binary task of referral/non-referral, motivated by eventual integration with portable low-cost retinographs. Finally, we combined decisions from local saliency-oriented and global data-driven image representations, with a notable performance in a particular dataset (two images per patient), reducing both false positive

and false negative rates contrasting with the local mid-level representation.

In terms of technical and scientific impacts, we explored advances of computer vision and machine learning thoroughly also proposing important methods. In a summarized timeline, we started with local-based two-tiered mid-level representations, explored global data-driven approaches with deep learning, and combined both to yield an accountable, mid-level, and data-driven rich representation. One of the major impacts was the proposal of a novel handcrafted coding method, semi-soft, that combines advantages of both hard and soft codings. We also crafted a new deep-learning architecture, inspired on two proposals in recent competitions, as well as performed a procedural investigation for analyzing, from a rigorous scientific point of view, the advantages and disadvantages of a series of novel and consolidated techniques. Among those methods, we highlight the multi-resolution training and the robust feature extraction augmentation. Finally, we proposed a hybrid computer-aided diagnostic model that relies on global data-driven decision and local mid-level saliency-oriented decision. Our saliency-oriented mid-level technique can be viewed as less human-centered two-tiered approach since it does not require local lesion annotation.

The thesis was organized in a series of research questions which are adequately answered through investigations in order to evaluate possible improvements and design decisions. The question-answers design is based upon a rigorous investigation of the techniques explored herein, measuring how much they improve the solution. We reported the advances of this research in top publication venues:

1. Automatic Diabetic Retinopathy Detection using BossaNova Representation
- Conference: IEEE Engineering in Medicine and Biology Society (EMBC'14)
- Date of Publication: November 6, 2014
- Context: Diabetic Retinopathy Lesion Detection
- Questions: *Q1.2*, and *Q1.3*

2. Advancing Bag-of-Visual-Words Representations for Lesion Classification in Retinal Images
- Journal: PLoS ONE
- Date of Publication: June 2, 2014
- Context: Diabetic Retinopathy Lesion Detection
- Questions: *Q1.1*

3. Automated Multi-Lesion Detection for Referable Diabetic Retinopathy in Indigenous Health Care
- Journal: PLoS ONE
- Date of Publication: June 2, 2015
- Context: Diabetic Retinopathy Lesion Detection
- Questions: *Q1.4*

4. Beyond Lesion-based Diabetic Retinopathy: a Direct Approach for Referral
- Journal: IEEE Journal of Biomedical and Health Informatics (J-BHI)
- Date of Publication: November 05, 2015

- Context: Referable Diabetic Retinopathy Detection
- Questions: *Q2.1*, *Q2.2* and *Q2.3*

5. A Data-driven Approach to Referable Diabetic Retinopathy Detection
- Journal: Elsevier Artificial Intelligence in Medicine (AIIM)
- Date of Publication: March 26, 2019
- Context: Referable Diabetic Retinopathy Detection
- Questions: *Q2.4*, *Q2.5* and *Q2.6*

6. An Accountable Saliency-Oriented Data-Driven Approach to Diabetic Retinopathy Detection
- Book: Photo Acoustic and Optical Coherence Tomography Imaging: An Application in Ophthalmology
- Date of Publication: 2019
- Context: Referable Diabetic Retinopathy Detection
- Questions: *Q2.7*, *Q2.8* and *Q2.9*

8.1 Diabetic Retinopathy Lesion Detection

The state of the art in BoVW methods for DR lesion detection was advanced by extending possible combinations of applying BoVW for detecting DR-related lesions in retinal images. We explored several combinations of handcrafted alternatives for the extraction of low-level features, and the creation of mid-level representations pointing out important choices when designing a unified framework for detecting DR lesions.

One of the contributions in this work was the proposal of a new semi-soft coding scheme, which explores the advantages of the most traditional hard-sum coding (sparse coding) as used in prior work for DR lesion detection and soft assignments (which better deal with imprecision and noise). As we show in the experiments, with ANOVA, the semi-soft coding associated with sparse feature extraction provides a good balance for designing an efficient and effective DR-related lesion detector.

Besides proposing a new coding approach to retinal images analysis, we also explored the BossaNova, a mid-level feature extraction technique that consists of an improvement in the pooling stage. Both for hard exudates and for red lesions detection, the results outperformed previous methods showing the importance of preserving some relationship between the detected features in retinal images instead of just throwing them away as previous BoVW-based solutions have done.

As another contribution of this work, we mention the description and the use of a cross-training methodology with results obtained from training with images from a particular ethnic group, and testing with a different population: Aboriginal and Torres Strait Islander people. This policy has important health screening and treatment impacts as the program can be trained at metropolitan clinics and taken to rural and remote areas for live screening and have a high degree of accuracy — typically, training images are not always available from remote communities in enough numbers to train the classifier and many towns comprise a very heterogeneous ethnic population mix.

8.2 Referable Diabetic Retinopathy Detection

In this research, we prioritized investigations of need of consultation, helping to decide who should be referred to the ophthalmologist for further examination. We started by detecting referable DR based on responses from lesion classifiers, by elaborating fusion technique based on meta-classification (which seeks a pattern based upon the classification score confidences returned by each individual lesion detector).

We also proposed a novel approach to decide, directly from the retinal images and without preliminary lesion detection, whether or not a patient will need to be referred to an ophthalmic specialist within a year. This decision to forgo specific DR-lesion detection has both theoretical motivations (making the referral decision using all information present in the image instead of just lesion scores) and practical advantages (much simpler to implement, test, and deploy). We highlight that, in the time we proposed the methodology and before the adoption of data-driven approaches, direct assessment was new for referable DR, and had not been developed before.

We emphasize the novelty of using cutting-edge mid-level representations (BossaNova and Fisher Vector), over the traditional BoVW approach. The experiments and statistical analysis confirm that the choice of the mid-level representation is critical. The best result for direct referral, reached by the Fisher Vector approach, clearly outperforms the traditional lesion-based method by more than two percentage points, reducing the classification error by almost 40% (from 5.8% to 3.6%). We also conclude that the direct referral is suitable also for independent datasets.

In this work, we also presented a data-driven solution for referable diabetic retinopathy detection. In contrast to works we inspire on (proposed in recent competitions), here we take into account not only the aim of improving classification accuracies, but we also take into consideration two important issues for real-world deployment: the (computational and implementation) complexity of the solution, and its ability to generalize under the stricter cross-dataset protocol. In addition to the solution itself, we also offer a novel research methodology regarding a procedural investigation approach. Our evaluation shows CNNs performance can be boosted by a set of directives. First, good data augmentation is essential for robust decision. Also, a robust feature-extraction augmentation improves performance considerably, while allowing for a diverse choice of machine learning algorithms at the final decision layer. The experiments also show that it is possible (and advantageous) to train high-resolution networks using the weights of low-resolution ones as initialization.

As expected, more information about the patients translates to better-informed decisions for referral. Thus, per patient diagnosis is more effective than per image, even under a challenging cross-dataset protocol.

Another novel aspect of our work is the investigation of the capability of transfer learning in the context of diabetic retinopathy screening, in order to compare data-driven approaches with previous work that employed handcraft methods with relatively small datasets (DR2 and Messidor-2) under cross-validation protocol. The model trained (fine-tuned) with DR2 clearly outperforms the method of direct referral over the same dataset, reducing the classification error by over 44%. This is in agreement with recent studies in

the literature: a novel family of data-driven methods is the state of the art for diabetic retinopathy screening.

In this work, we also presented an accountable and robust framework for automated screening of diabetic retinopathy. A series of works in prior art have been focused on accurate data-driven approaches to effective diagnostic, even using a single deep convolutional neural network or ensembling a set of models. However, the interpretability of those models — which has become a requirement in order to understand the reasons behind a decision — is frequently disregarded. In this vein, we proposed the use of saliency maps whose objective is twofold: highlighting regions that potentially influence the decision taken, and capturing regions of interest that could be leveraged for the final model response.

Purely data-driven CNNs naturally perform global evaluations, by receiving the entire image and assigning decision probabilities. Herein, we use global information to extract two data-driven responses: the softmax probability itself as well as scores coming from a neural network that receives pre-softmax features.

One of the main novelties in the current work is the breakdown of the global data-driven scenario. The pipeline involves extracting saliency-oriented regions of interest and combining those information through Fisher Vector. Exploring the encoded contextual local-based representations, we reduce the classification error by 27% contrasting with our previous global data-driven method testing with DR2 under cross-dataset validation protocol. By testing with Messidor-2, the robust local saliency-oriented region characteristics had a remarkable improvement, reducing the classification error by 28% over our previous global data-driven proposal. By enhancing considerably the performance both over DR2 and Messidor-2, in comparison with the strict global method, we showed that the guided mid-level representation, which emphasizes areas that the CNN model could not assimilate sufficiently, is promising.

8.3 Validation with images from portable devices

One of the main aims at this work was integrating our solutions regarding image analytics for diabetic retinopathy detection with simple and portable retinal imaging solutions. Our purpose was combining our software with portable and low-cost image acquisition devices.

Our intention was validating retinal images from portable devices with our methods. As the images collected in time for validation during the doctoral program were only graded in terms of quality assessment, we evaluate if they were properly ready for a computer-aided DR screening (if they encompass the minimum required quality).

By fine-tuning the MobileNetV2, that has good trade-off between accuracy and number of operations and parameters, we achieved a promising result that surpassed the Inception-Resnet architecture. We also figured out that the self-annotation procedure — for labeling a large dataset to pre-train the mobile network — enhanced substantially the results.

8.4 Future Work

In closing this work, we would like to emphasize important open questions to be explored in image analytics for diabetic retinopathy detection.

Once the performance of combining global data-driven and local saliency-oriented characteristic depends on the robustness of the baseline CNN model, possible future work comprises exploring higher image resolutions, possibly exploring strategies such as multi-resolution training [77] to easily capture very small lesions and subtle image details.

Moving towards an opposed direction comparing to the natural advance of research in diabetic retinopathy diagnostic represents a strong route. Instead of detecting lesions and using the assembled information to decide about disease stages or referral need, in a bottom-up manner; it would involve identifying referable DR and using pixel importance for pointing out and recognizing lesions or anatomical retina parts, as a top-down approach.

Another trend is identifying the DR severity degree of a patient — further classifying the images as related to DR cases in early, mild, proliferative and severe stages — and exploring saliency-oriented representation in order to perform an extensive comparison with the prior art.

Possible future work also encompasses exploring more deeply the iterative self-annotation procedure, by dynamically training the model interchanging the train set per epoch: with expert-annotated (DR1) and model-annotated (kaggle/EyePACs) datasets.

Additionally, investigating whether the quality-aware photometric data augmentation is a strategy capable of enhancing performance. We believe that applying a little of blurring in poor images and sharpening in good images might highlight the expected aspects into retinal images that were not properly annotated by experts.

Finally, future work also comprehends validating the referral solutions with fundus images captured via mobile devices, and intensifying investigations of models that efficiently trade off between effectiveness (accuracy) and time consumption/memory footprint, in order to bypass cloud services and embed into portable retinal cameras.

Bibliography

[1] Michael D Abràmoff, James C Folk, Dennis P Han, Jonathan D Walker, David F Williams, Stephen R Russell, Pascale Massin, Beatrice Cochener, Philippe Gain, Li Tang, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmology*, 131(3):351–357, 2013.

[2] Michael D Abràmoff, Yiyue Lou, Ali Erginay, Warren Clarida, Ryan Amelon, Jim Folk, and Meindert Niemeijer. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest. Ophthalmol. Vis. Sci.*, 57(13):5200–5206, 2016.

[3] Michael D. Abràmoff, Meindert Niemeijer, Maria S.A. Suttorp-Schulten, Max A. Viergever, Stephen R. Russell, and Bram van Ginneken. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care*, 31(2):193–198, 2008.

[4] Michael D Abràmoff and Maria S A Suttorp-Schulten. Web-based screening for diabetic retinopathy in a primary care population: the eyecheck project. *Telemed J E Health*, 11:668–674, 2005.

[5] U. R. Acharya, C. M. Lim, E. Y. K. Ng, C. Chee, and T. Tamura. Computer-based detection of diabetes retinopathy stages using digital fundus images. *Journal of Engineering in Medicine*, 223:545–553, 2009.

[6] Balint Antal and Andras Hajdu. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Transactions on Biomedical Engineering*, 59(6):1720–1726, 2012.

[7] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo. BossaNova at ImageCLEF 2012 Flickr Photo Annotation Task. In *Working Notes of the Conference and Labs of the Evaluation Forum*, 2012.

[8] Sandra Avila, Nicolas Thome, Matthieu Cord, Eduardo Valle, and Arnaldo de A Araújo. Pooling in image representation: The visual codeword point of view. *Computer Vision and Image Understanding*, 117(5):453–465, 2013.

[9] Ricardo Baeza-Yates and Berthier Ribeiro Neto. *Modern Information Retrieval*, volume 1. Addison Wesley, 1999.

[10] Timothy Batchelder and Michael Barricks. The wisconsin epidemiologic study of diabetic retinopathy. *Archives of ophthalmology*, 113(6):702–703, 1995.

[11] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[12] Malavika Bhaskaranand, Jorge Cuadros, Chaithanya Ramachandra, Sandeep Bhat, Muneeswar Nittala, Srinivas Sadda, and Kaushal Solanki. EyeArt + EyePACS: Automated retinal image analysis for diabetic retinopathy screening in a telemedicine system. *Ophthal. Med. Image Anal. Sec. Intl. Workshop*, pages 105–112, 2015.

[13] Y-Lan Boureau, Francis Bach, Yann LeCun, and Jean Ponce. Learning mid-level features for recognition. In *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2010.

[14] Y-Lan Boureau, Jean Ponce, and Yann LeCun. A theoretical analysis of feature pooling in visual recognition. In *Intl. Conference on Machine Learning*, pages 111–118, 2010.

[15] British Diabetic Association. *Retinal photography screening for diabetic eye disease*. BDA, London, 1997.

[16] European Political Strategy Centre. The age of artificial intelligence. towards a european strategy for human-centric machines, March 2018. [Online; accessed 17-October-2018].

[17] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. In *British Machine Vision Conference (BMVC)*, 2011.

[18] E. Colas, A. Besse, A. Orgogozo, B. Schmauch, N. Meric, and E. Besse. Deep learning approach for diabetic retinopathy screening. *Acta Ophthalmologica*, 94, 2016.

[19] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[20] Pedro Costa, Aurélio Campilho, Bryan Hooi, Asim Smailagic, Kris Kitani, Shenghua Liu, Christos Faloutsos, and Adrian Galdran. Eyequal: Accurate, explainable, retinal image quality assessment. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on*, pages 323–330. IEEE, 2017.

[21] Pedro Costa, Adrian Galdran, Asim Smailagic, and Aurélio Campilho. A weakly-supervised framework for interpretable diabetic retinopathy detection on retinal images. *IEEE Access*, 6:18747–18758, 2018.

[22] Kate Crawford and Ryan Calo. There is a blind spot in AI research. *Nature News*, 538(7625):311, 2016.

[23] Michael J. Cree, E. Gamble, and D. J. Cornforth. Colour normalisation to reduce inter-patient and intra-patient variability in microaneurysm detection in colour retinal images. In *Workshop on Digital Image Computing*, pages 163–168, 2005.

[24] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Anal. Stereol.*, 33(3):231–234, 2014.

[25] Etienne Decencière, Guy Cazuguel, Xiwei Zhang, Guillaume Thibault, J-C Klein, Fernand Meyer, Beatriz Marcotegui, Gwénolé Quellec, Mathieu Lamard, Ronan Danno, et al. Teleophta: Machine learning and image processing methods for teleophthalmology. *Ingénierie et Recherche Biomédicale*, 2013.

[26] Thomas G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10:1895–1923, 1998.

[27] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

[28] A. D. Fleming, K. A. Goatman, S. Philip, G. J. Williams, G. J. Prescott, G. S. Scotland, P. McNamee, G. P. Leese, W. N. Wykes, P. F. Sharp, and J. A. Olson. The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy. *British Journal of Ophthalmology*, 94(6):706–711, 2010.

[29] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp. Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE Transactions Medical Imaging*, 25:1223–1232, 2006.

[30] Alan D. Fleming, Keith A. Goatman, Sam Philip, Gordon J. Prescott, Peter F. Sharp, and John A. Olson. Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts. *British Journal of Ophthalmology*, 94(12):1606–1610, 2010.

[31] Alan D Fleming, Sam Philip, Keith A Goatman, John A Olson, and Peter F Sharp. Automated assessment of diabetic retinal image quality based on clarity and field definition. *Investigative Ophthalmology & Visual Science*, 47(3):1120–1125, 2006.

[32] Alan D. Fleming, Sam Philip, Keith A. Goatman, Graeme J. Williams, John A. Olson, and Peter F. Sharp. Automated detection of exudates for diabetic retinopathy screening. *Physics in Medicine and Biology*, 52(24):7385–7396, 2007.

[33] Donald S Fong, Lloyd Aiello, Thomas W Gardner, George L King, George Blankenship, Jerry D Cavallerano, Fredrick L Ferris, and Ronald Klein. Retinopathy in diabetes. *Diabetes care*, 27(suppl 1):s84–s87, 2004.

[34] R. Gargeya and T. Leng. Automated identification of diabetic retinopathy using deep learning. *Ophthalmol.*, 124(7):962–969, 2017.

[35] Luca Giancardo, Fabrice Meriaudeau, Thomas Karnowski, Yaqin Li, Kenneth Tobin, and Edward Chaum. Microaneurysm detection with radon transform-based classification on retina images. In *Intl. Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5939–5942, 2011.

[36] Luca Giancardo, Fabrice Meriaudeau, Thomas P Karnowski, Yaqin Li, Seema Garg, Kenneth W Tobin, and Edward Chaum. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis*, 16(1):216–226, 2012.

[37] Luca Giancardo, Fabrice Mériaudeau, Thomas P Karnowski, Kenneth W Tobin, Yaqin Li, and Edward Chaum. Microaneurysms detection with the radon cliff operator in retinal fundus images. In *SPIE Medical Imaging*, pages 76230U–76230U. International Society for Optics and Photonics, 2010.

[38] Diane M Gibson. The geographic distribution of eye care providers in the united states: Implications for a national strategy to improve vision health. *Preventive Medicine*, 73:30–36, 2015.

[39] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2nd edition, 2006.

[40] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *Intl. Conf. Learn. Representations*, 2015.

[41] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[42] Ribhi Hazin, Marcus Colyer, Flora Lum, and Mohammed K Barazi. Revisiting diabetes 2000: challenges in establishing nationwide diabetic retinopathy prevention programs. *American Journal of Ophthalmology*, 152(5):723–729, 2011.

[43] International Diabetes Prevalation. IDF diabetes atlas, sixth edition. http://www.idf.org/diabetesatlas, Accessed: 2015-09-18.

[44] Tommi S. Jaakkola and David Haussler. Exploiting generative models in discriminative classifiers. In *Advances in Neural Information Processing Systems*, pages 487–493, Cambridge, MA, USA, 1999. MIT Press.

[45] Raj Jain. *The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling.* Wiley professional computing. Wiley, 1991.

[46] Herbert Jelinek, Ramon Pires, Rafael Padilha, Siome Goldenstein, Jacques Wainer, and Anderson Rocha. Data fusion for multi-lesion diabetic retinopathy detection. In *IEEE Computer-Based Medical Systems*, pages 1–4, 2012.

[47] Herbert F Jelinek, Michael J Cree, Jorge JG Leandro, João VB Soares, Roberto M Cesar, and A Luckie. Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy. *Journal of the Optical Society of America A*, 24(5):1448–1456, 2007.

[48] Herbert F. Jelinek, Michael J. Cree, David Worsley, Alan P. Luckie, and Peter Nixon. An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice. *Clinical and Experimental Optometry*, 89(5):299–305, 2006.

[49] Herbert F. Jelinek, Ramon Pires, Rafael Padilha, Siome Goldenstein, Jacques Wainer, and Anderson Rocha. Quality control and multi-lesion detection in automated retinopathy classification using a visual words dictionary. In *Intl. Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5857–5860, 2013.

[50] Herbert F. Jelinek, Anderson Rocha, Tiago Carvalho, Siome Goldenstein, and Jacques Wainer. Machine learning and pattern classification in identification of indigenous retinal pathology. In *Intl. Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5951–5954, 2011.

[51] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[52] Marc Lalonde, Langis Gagnon, Marie-Carole Boucher, et al. Automatic visual quality assessment in optical fundus images. In *Proceedings of vision interface*, volume 32, pages 259–264. Ottawa, 2001.

[53] Istvan Lazar and Andras Hajdu. Retinal microaneurysm detection through local rotating cross-section profile analysis. *IEEE Transactions on Medical Imaging*, 32(2):400–407, 2013.

[54] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10), 1995.

[55] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[56] Samuel C Lee and Yiming Wang. Automatic retinal image quality assessment and enhancement. In *Medical Imaging 1999: Image Processing*, volume 3661, pages 1581–1591. International Society for Optics and Photonics, 1999.

[57] Yaqin Li, Thomas P. Karnowski, Kenneth W. Tobin, Luca Giancardo, Scott Morris, Sylvia E. Sparrow, Seema Garg, Karen Fox, and Edward Chaum. A health insurance portability and accountability act–compliant ocular telehealth network for the remote diagnosis and management of diabetic retinopathy. *Telemedicine and e-Health*, 17(8):627–634, 2011.

[58] Zachary C Lipton. The mythos of model interpretability. *Queue*, 16(3):30, 2018.

[59] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[60] J. Macqueen. Some methods for classification and analysis of multivariate observations. In *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[61] Dwarikanath Mahapatra, Pallab K Roy, Suman Sedai, and Rahil Garnavi. Retinal image quality classification using saliency maps and cnns. In *International Workshop on Machine Learning in Medical Imaging*, pages 172–179. Springer, 2016.

[62] Paul Mitchell, Suriya Foran, and Jim Foran. *Guidelines for the management of diabetic retinopathy.* National Health and Medical Research Council, 2008.

[63] Jay Nandy, Wynne Hsu, and Mong Li Lee. An incremental feature extraction framework for referable diabetic retinopathy detection. In *Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on*, pages 908–912. IEEE, 2016.

[64] S. Naqvi, M. Zafar, and Ihsan ul Haq. Referral system for hard exudates in eye fundus. *Comput. Biol. Med.*, 64:217–235, 2015.

[65] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE Intl. Conf. Comput. Vis. Pattern Recog.*, pages 427–436, 2015.

[66] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Li Qin, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. D. Abràmoff. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 29(1):185–195, 2010.

[67] Meindert Niemeijer, Bram van Ginneken, Stephen R. Russell, Maria S. A. Suttorp-Schulten, and Michael D. Abràmoff. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative Ophthalmology & Visual Science*, 48(5):2260–2267, 2007.

[68] JA Olson, PF Sharp, KA Goatman, et al. The role of automated grading of diabetic retinopathy in a primary care screening programme. *Edinburgh: Chief Scientist Office*, 2006.

[69] World Health Organization. *Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005*. World Health Organization, 2006.

[70] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22(10):1345–1359, 2010.

[71] Niall Patton, Tariq M Aslam, Thomas MacGillivray, Ian J Deary, Baljean Dhillon, Robert H Eikelboom, Kanagasingam Yogesan, and Ian J Constable. Retinal image analysis: concepts, applications and potential. *Progress in retinal and eye research*, 25(1):99–127, 2006.

[72] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007.

[73] Florent Perronnin and Diane Larlus. Fisher vectors meet neural networks: A hybrid classification architecture. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3743–3752, 2015.

[74] Florent Perronnin, Jorge Sánchez, and Thomas Mensink. Improving the fisher kernel for large-scale image classification. In *European Conference on Computer Vision*, pages 143–156. Springer, 2010.

[75] Ramon Pires, Sandra Avila, Herbert F. Jelinek, Jacques Wainer, Eduardo Valle, and Anderson Rocha. Automatic diabetic retinopathy detection using bossanova representation. In *Intl. Conference of the IEEE Engineering in Medicine and Biology Society*, pages 146–149, 2014.

[76] Ramon Pires, Sandra Avila, Herbert F. Jelinek, Jacques Wainer, Eduardo Valle, and Anderson Rocha. Beyond lesion-based diabetic retinopathy: a direct approach for referral. *IEEE Journal of Biomedical and Health Informatics*, 21(1):193–200, 2017.

[77] Ramon Pires, Sandra Avila, Jacques Wainer, Eduardo Valle, Michael D Abramoff, and Anderson Rocha. A data-driven approach to referable diabetic retinopathy detection. *Artificial Intelligence in Medicine*, 96:93–106, 2019.

[78] Ramon Pires, Tiago Carvalho, Geoffrey Spurling, Siome Goldenstein, Jacques Wainer, Alan Luckie, Herbert F. Jelinek, and Anderson Rocha. Automated multi-lesion detection for referable diabetic retinopathy in indigenous health care. *PLoS ONE*, 10(6):e0127664, 06 2015.

[79] Ramon Pires, Alexandre Ferreira, Sandra Avila, Jacques Wainer, and Anderson Rocha. An accountable saliency-oriented data-driven approach to diabetic retinopathy detection. In to be published, editor, *Photo Acoustic and Optical Coherence Tomography Imaging: An Application in Ophthalmology*, chapter 10. Elsevier, Elsevier, 2019.

[80] Ramon Pires, Herbert Jelinek, Jacques Wainer, Siome Goldenstein, Eduardo Valle, and Anderson Rocha. Assessing the need for referral in automatic diabetic retinopathy detection. *Transactions on Biomedical Engineering*, 60(12):3391–3398, 2013.

[81] Ramon Pires, Herbert F. Jelinek, Jacques Wainer, and Anderson Rocha. Retinal image quality analysis for automatic diabetic retinopathy detection. In *IEEE Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 229–236, 2012.

[82] Ramon Pires, Herbert F. Jelinek, Jacques Wainer, Eduardo Valle, and Anderson Rocha. Advancing bag-of-visual-words representations for lesion classification in retinal images. *PLoS ONE*, 9(6):e96814, 06 2014.

[83] Frédéric Precioso and Matthieu Cord. Machine learning approaches for visual information retrieval. In *Visual Indexing and Retrieval*, SpringerBriefs in Computer Science, pages 21–40. Springer New York, 2012.

[84] G. Quellec, K. Charrièreb, Y. Boudib, B. Cochenerc, and M. Lamardc. Deep image mining for diabetic retinopathy screening. *Medical Image Analysis*, 39:178–193, 2017.

[85] G. Quellec, M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener, and C. Roux. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Transactions on Medical Imaging*, 27(9):1230 –1241, 2008.

[86] Gwenolé Quellec, Mathieu Lamard, Ali Erginay, Agnès Chabouis, Pascale Massin, Béatrice Cochener, and Guy Cazuguel. Automatic detection of referral patients due to retinal pathologies through data mining. *Medical Image Analysis*, 29:47–64, 2016.

[87] Anderson Rocha, Tiago Carvalho, Herbert Jelinek, Siome Goldenstein, and Jacques Wainer. Points of interest and visual dictionaries for automatic retinal lesion detection. *IEEE Transactions on Biomedical Engineering*, 59(8):2244 – 2253, 2012.

[88] Pallab Roy, Ruwan Tennakoon, Khoa Cao, Suman Sedai, Dwarikanath Mahapatra, Stefan Maetschke, and Rahil Garnavi. A novel hybrid approach for severity assessment of diabetic retinopathy in colour fundus images. In *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 1078–1082. IEEE, 2017.

[89] Sajib Kumar Saha, Basura Fernando, Jorge Cuadros, Di Xiao, and Yogesan Kanagasingam. Deep learning for automated quality assessment of color fundus images in diabetic retinopathy screening. *arXiv preprint arXiv:1703.02511*, 2017.

[90] Clara I Sánchez, María García, Agustín Mayo, María I López, and Roberto Hornero. Retinal image analysis based on mixture models to detect hard exudates. *Medical Image Analysis*, 13(4):650–658, 2009.

[91] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision (IJCV)*, 105(3):222–245, 2013.

[92] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[93] Désiré Sidibé, Ibrahim Sadek, and Fabrice Mériaudeau. Discrimination of retinal images containing bright lesions using sparse coded features and svm. *Computers in Biology and Medicine*, 62:175–184, 2015.

[94] Dawn A Sim, Pearse A Keane, Adnan Tufail, Catherine A Egan, Lloyd Paul Aiello, and Paolo S Silva. Automated retinal image analysis for diabetic retinopathy in telemedicine. *Current Diabetes Reports*, 15(3):1–9, 2015.

[95] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Intl. Conf. Learn. Representations*, 2015.

[96] Chanjira Sinthanayothin, James F. Boyce, Tom H. Williamson, Helen L. Cook, Evelyn Mensah, Shantanu Lal, and David Usher. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Medicine: a Journal of the British Diabetic Association*, 19(2):105–112, 2002.

[97] Josef Sivic and Andrew Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. In *IEEE Intl. Conference on Computer Vision*, pages 1470–1477, 2003.

[98] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.

[99] Joao V. B. Soares, Jorge J. G. Leandro, Roberto M. Cesar, Herbert F. Jelinek, and Michael J. Cree. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222, 2006.

[100] Akara Sopharak, Bunyarit Uyyanonvara, Sarah Barman, and Thomas H. Williamson. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics*, 32:8, 2008.

[101] Enrique Soto-Pedre, Amparo Navea, Saray Millan, Maria C Hernaez-Ortega, Jesús Morales, Maria C Desco, and Pablo Pérez. Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists' workload. *Acta Ophthalmologica*, 2014.

[102] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

[103] Ilya Sutskever, James Martens, George E Dahl, and Geoffrey E Hinton. On the importance of initialization and momentum in deep learning. In *Intl. Conf. Machine Learn.*, pages 1139–1147, 2013.

[104] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[105] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[106] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *Intl. Conf. Learn. Representations*, 2014.

[107] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[108] Kevin Tozer, Maria A Woodward, and Paula A Newman-Casey. Telemedicine and diabetic retinopathy: Review of published screening programs. *J Endocrinol Diab*, 2(4):1–10, 2015.

[109] T. Tuytelaars. Dense interest points. In *Computer Vision and Pattern Recognition*, pages 2281–2288, 2010.

[110] Dumskyj Usher, M Dumskyj, Mitsutoshi Himaga, Tom H Williamson, Sl Nussey, and J Boyce. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. *Diabetic Medicine*, 21(1):84–90, 2004.

[111] Jan van Gemert, Cor J. Veenman, Arnold W. M. Smeulders, and Jan-Mark Geusebroek. Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1271–1283, 2010.

[112] V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[113] Vision Problems in the U.S. Prevalence of adult vision impairment and age-related eye disease in America. Accessed: 2015-02-12.

[114] Zhe Wang, Yanxin Yin, Jianping Shi, Wei Fang, Hongsheng Li, and Xiaogang Wang. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 267–275. Springer, 2017.

[115] Daniel Welfer, Jacob Scharcanski, and Diane Ruschel Marinho. A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images. *Computerized Medical Imaging and Graphics*, 34(3):228–235, 2010.

[116] WHO. *Diabetes care and research in Europe: the St Vincent Declaration. Diabetes mellitus in Europe: a problem for all ages in all countries, a model for prevention and self-care. WHO Regional Office for Europe and International Diabetes Federation.* BDA, St Vincent (Aosta), Italy, 1989.

[117] C Wilkinson, F Ferris III, R Klein, P Lee, C Agardh, M Davis, D Dills, A Kampik, R Pararajasegaram, and J Verdaguer. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmol.*, 110(9):1677–1682, 2003.

[118] Tien Y. Wong, Chui Ming Gemmy Cheung, Michael Larsen, Sanjay Sharma, and Rafael Simó. Diabetic retinopathy. *Nature Reviews Disease Primers*, 2(16012):1–17, 2016.

[119] Yehui Yang, Tao Li, Wensi Li, Haishan Wu, Wei Fan, and Wensheng Zhang. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 533–540. Springer, 2017.

[120] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Proc. Eur. Conf. Comput. Vis.*, pages 818–833, 2014.

[121] Bob Zhang, Xiangqian Wu, Jane You, Qin Li, and Fakhri Karray. Hierarchical detection of red lesions in retinal images by multiscale correlation filtering. In *SPIE medical imaging*, pages 72601L–72601L. International Society for Optics and Photonics, 2009.

Appendix A

Copyright Permissions

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Feb 16, 2019

This Agreement between University of Campinas -- José Ramon Pires ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4527590816892
License date	Feb 14, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Disease Primers
Licensed Content Title	Diabetic retinopathy
Licensed Content Author	Tien Y. Wong, Chui Ming Gemmy Cheung, Michael Larsen, Sanjay Sharma, Rafael Simó
Licensed Content Date	Mar 17, 2016
Licensed Content Volume	2
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Image Analytics Techniques for Diabetic Retinopathy Detection
Institution name	University of Campinas
Expected presentation date	Dec 2018
Order reference number	Figure 4
Portions	Figure 4 : Clinical signs of diabetic retinopathy on fundoscopic examination.
Requestor Location	University of Campinas Av. Albert Einstein, 1251 Cidade Universitária CEP: 13083-852 Campinas, São Paulo 13083-852 Brazil Attn: University of Campinas
Billing Type	Invoice
Billing Address	University of Campinas Av. Albert Einstein, 1251 Cidade Universitária CEP: 13083-852 Campinas, Brazil 13083-852 Attn: University of Campinas
Total	0.00 USD
Terms and Conditions	

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Feb 14, 2019

This Agreement between University of Campinas -- José Ramon Pires ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4477020829962
License date	Nov 27, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	Deep learning
Licensed Content Author	Yann LeCun, Yoshua Bengio, Geoffrey Hinton
Licensed Content Date	May 27, 2015
Licensed Content Volume	521
Licensed Content Issue	7553
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Image Analytics Techniques for Diabetic Retinopathy Detection
Institution name	University of Campinas
Expected presentation date	Dec 2018
Order reference number	Figure 2
Portions	Figure 2: Inside a convolutional network. The Figure is in page 438 (page 3 of the paper).
Requestor Location	University of Campinas Av. Albert Einstein, 1251 Cidade Universitária CEP: 13083-852 Campinas, São Paulo 13083-852 Brazil Attn: University of Campinas
Billing Type	Invoice
Billing Address	University of Campinas Av. Albert Einstein, 1251 Cidade Universitária CEP: 13083-852 Campinas, Brazil 13083-852 Attn: University of Campinas
Total	0.00 USD
Terms and Conditions	

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g.

Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Copyright Clearance Center RightsLink®

Home | Account Info | Help

IEEE Requesting permission to reuse content from an IEEE publication

Title: Automatic Diabetic Retinopathy detection using BossaNova representation

Conference Proceedings: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

Author: Ramon Pires; Sandra Avila; Herbert F. Jelinek; Jacques Wainer; Eduardo Valle; Anderson Rocha

Publisher: IEEE

Date: 26-30 Aug. 2014

Copyright © 2014, IEEE

Logged in as:
José Ramon Pires
University of Campinas
LOGOUT

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Copyright Clearance Center RightsLink®

Home | Account Info | Help

IEEE Requesting permission to reuse content from an IEEE publication

Title: Beyond Lesion-Based Diabetic Retinopathy: A Direct Approach for Referral
Author: Ramon Pires
Publication: Biomedical and Health Informatics, IEEE Journal of
Publisher: IEEE
Date: Jan. 2017
Copyright © 2017, IEEE

Logged in as:
José Ramon Pires
University of Campinas
LOGOUT

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com