Application of Deep Learning Techniques for Biomedical Data Analysis

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Parham Khojasteh
M. Eng (Shahrood University of Technology)
B. Eng (Islamic Azad University, Karaj Branch)

School of Engineering
College of Science, Engineering and Health
RMIT University
June 2019

Declaration of Authorship

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Signed: Parham Khojasteh
Data: June 2019

Acknowledgment

I would first like to thank Prof. Dinesh Kant Kumar, who has guided and encouraged me to move During my PhD and has given me immense support as my senior PhD supervisor. I have had the pleasure of working with Dinesh and he has devoted countless hours to my development as a research scientist. Your inspiration, commitment and patience are greatly appreciated.

My thanks go to my associate supervisor Dr. Behzad Aliahmad, who provided support on several successful projects during my PhD.

I would like to thank Dr. João Paulo Papa whom I have worked closely on several projects. With his down-to-earth nature and ability to communicate complex theory, he is a joy to work with.

I would like to offer my thanks to my co-authors, Dr. Sridhar Poosapadi Arjunan, Ms. Rekha Viswanathan, Dr. Leandro Passos, Dr. Tiago Carvalho, Dr. Edmar Rezende. Their expertise and comments were valuable to the papers.

I am grateful to all my friends and colleagues in the Biosignal Lab for their valuable helps and providing inspiring research environment.

Lastly, I would like to thank my family and friends for continual encouragement and support over the last years.

Publications

1. Khojasteh, P., Aliahmad, B., Kumar, D. (2019) A Novel Color Space of Fundus Images for Reliable Automatic Exudates Detection. " Journal of Biomedical Signal Processing and Control" 10.1016/j.bspc.2018.12.004

2. Khojasteh, P., Junior, L. A., Carvalho, T., Rezende, E., Aliahmad, B., Papa, J., Kumar, D. (2018) "Exudate detection in fundus images using deeply-learnable features. "Journal of Computers in Biology and Medicine- Elsevier" 10.1016/j.compbiomed.2018.10.031

3. Khojasteh, P., Aliahmad, B., Kumar, D. (2018) Fundus Images Analysis Using Deep Features for Detection of Exudates, Hemorrhages and Microaneurysms. "Journal of BMC Ophthalmology-Springer" 10.1186/s12886-018-0954-4

4. Khojasteh, P., Aliahmad, B., S.P. Arjunan, Kumar, D. (2018) Introducing a Novel Layer in Convolutional Neural Network for Automatic Identification of Diabetic Retinopathy "IEEE EMBC 2018" 10.1109/EMBC.2018.8513606

5. Khojasteh, P., R. Viswanathan., Aliahmad, B., Kumar, D. (2018) Parkinson Disease Diagnosis on Multivariate Deep Features of Speech Signal "IEEE Life Science 2018" 10.1109/LSC.2018.8572136

6. R. Viswanathan, P. Khojasteh, B. Aliahmad, S.P. Arjunan, S. Ragnav, P. Kempster, Kitty Wong, Jennifer Nagao, D. Kumar. (2018) Efficiency of Voice Features Based on Consonant for Detection of Parkinson's Disease. "IEEE Life Science 2018" 10.1109/LSC.2018.8572266

Abstract

Deep learning and machine learning methods have been used for addressing the problems in the biomedical applications, such as diabetic retinopathy assessment and Parkinson's disease diagnosis. The severity of diabetic retinopathy is estimated by the expert's examination of fundus images based on the amount and location of three diabetic retinopathy signs (i.e., exudates, hemorrhages, and microaneurysms). An automatic and accurate system for detection of these signs can significantly help clinicians to make the best possible prognosis can result in reducing the risk of vision loss. For Parkinson's disease diagnosis, analysis of a speech voice is considered as the earliest symptom with the advantage of being non-intrusive and suitable for online applications. While some reported outcomes of the developed techniques have shown the good results and ongoing progress for these two applications, designing new algorithms is a thriving research field to overcome the poor sensitivity and specificity of the outcomes as well as the limitations such as dataset size and heuristic selection of the network parameters.

This thesis has comprehensively studied and developed various deep learning frameworks for detection of diabetic retinopathy signs and diagnosis of Parkinson's disease. To improve the performance of the current systems, this work has had an investigation on different techniques: (i) color space investigation, (ii) examination of various deep learning methods, (iii) development of suitable pre/post-processing algorithms and (iv) appropriate selection of deep learning architectures and parameters.

For diabetic retinopathy assessment, this thesis has proposed the new color space as the input for the deep learning models that obtained better replicability compared with the conventional color spaces. This has also shown the pre-trained model can extract more relevant features compared to the models which were trained from scratch. This has also presented a deep learning framework combined with the suitable pre and post-processing algorithms that increased the performance of the system. By investigation different architectures and parameters, the suitable deep learning model has been presented to distinguish between Parkinson's disease and healthy speech signal.

Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 Introduction

A combination of machine learning, image and signal processing methods have been used to address various problems in the biomedical applications, such as Diabetic Retinopathy (DR) assessment and Parkinson's Disease (PD) diagnosis (Kahai, Namuduri et al. 2006, Rahn, Chou et al. 2007, Lee, Zhou et al. 2008, Gracas, Gama et al. 2012, Zaki, Zulkifley et al. 2016). Unlike the conventional pipeline that required hand-crafted feature extraction, Deep Learning (DL) methods have overcome this issue by automatically learning informative features from the input data (Schmidhuber 2015, van Grinsven, Venhuizen et al. 2016, Kamal Maried, Abdalla Eldali et al. 2017, Tan, Fujita et al. 2017). Achieving the promising results in different applications by the DL methods has led to increasing a demand for developing new DL techniques to improve the performance of the current systems. This thesis has investigated to develop novel DL frameworks for two biomedical applications:

- DR Assessment
- PD Diagnosis

Diabetic Retinopathy Assessment. The severity of DR is currently estimated by the expert's examination of fundus images based on the amount and location of three retinopathy signs (i.e., exudates, hemorrhages, and microaneurysms) across the retina surface (Hansen, Abramoff et al. 2015). If the DR severity is misdiagnosed, it can lead to irreversible vision loss (Mohamed, Gillies et al. 2007, Shaw, Sicree et al. 2010). Consequently, an automatic and accurate system for detection of the DR signs can significantly help clinicians to make the best possible prognosis about the severity of DR and risk of vision loss. For this purpose,

DL and other machine learning methods have been employed (Hajeb Mohammad Alipour, Rabbani et al. 2012, Lazar and Hajdu 2013, Tang, Niemeijer et al. 2013).

Parkinson's Disease Diagnosis. The speech impairment in PD, as one of the earliest symptoms, has been considered for the diagnosis of the disease. It has the advantage of being non-intrusive and suitable for online applications, and hence researchers have investigated the speech features to diagnose PD (Tsanas, Little et al. 2010, Zhang, Yang et al. 2016, Pompili, Abad et al. 2017). Machine learning and speech processing techniques have been utilized to analyse the speech signals to classify between PD and control subjects.

1.2 Problem Statement

Among the three DR signs, the presence of exudate is observed as the earliest sign of DR. For automatic exudate detection, various DL and machine learning methods have been developed. However, the current methods have some drawbacks described as follow.

- Retina image analyses have been conducted using color spaces such as RGB, HSV, and LUV. Majority of the DL methods use three channels of the RGB space as the input to the networks (Gargeya and Leng 2017, Quellec, Charrière et al. 2017, Grassmann, Mengelkamp et al. 2018). However, other color spaces haven't been explored and tested with DL methods.
- Most of pervious DL frameworks are based on the Convolutional Neural Networks (CNNs) architecture. This requires identifying the suitable architecture and optimal parameters. However, there is no exact method for the selection of the network parameters (He, Zhang et al. 2016).
- Another concern in using the CNN-based methods is the need for a large dataset to train the CNN. For overcoming these limitations, there is a need to investigate different DL methods to identify a network that gives good performance and is not reliant on large datasets.

The success of diagnosis of DR requires the detection of all the three signs: exudate, hemorrhage and microaneurysm. For the simultaneous detection of all DR signs, a few studies have been conducted by employing the CNN-based methods, but their methods achieved poor performance for distinguishing between hemorrhage and microaneurysm (Tan, Fujita et al. 2017). One of the main drawbacks in these works is not using the suitable DL-based framework with appropriate pre/pro-processing methods in their algorithm. It is observed that pre-processing methods have significant effect on enhancing the contrast between the DR signs and backgrounds. Although various pre-processing techniques have been developed for the DR assessment, comparing the performance of these methods have not been tested for the DL-based methods. Additionally, the effect of post-processing methods on the output of these methods have not been investigated for automatic DR signs detection. As a result, it is essential to investigate various pre/pre-processing methods to find the suitable one for improving the performance of the CNN-based methods.

For PD diagnosis based on the speech signal, many studies have attempted to extract and analysis the time-frequency-based features, such as Jitter, Shimmer, Pitch, Harmonics to Noise Ratio, Autocorrelation, voiced and unvoiced frames (Sakar, Isenkul et al. 2013). However, there is the need for improving the performance for diagnosis. This aim of this sub-section was to investigate CNN-based models can provide informative features while the efficiency of such these models have not been tested yet.

1.3 Hypotheses

This thesis aims to investigate and develop different DL methods for assessment of DR and PD. This research has been conducted based on the following hypothesis:

i. An appropriate choice of color channels is expected to enhance the performance of the system for automatic DR signs detection.
ii. Identify a suitable DL method is able to overcome the limitations of the CNN-based methods for automatic DR signs detection.
iii. A combination of suitable pre/post-processing algorithms is expected to increase the performance of the DL-based method for automatic DR signs detection.

iv. DL methods are able to extract suitable informative features from the speech signals for distinguishing between PD and healthy individual.

1.4 Research Aim and Objectives

The aim of this research is to study and investigate the use of DL for two biomedical applications: (i)- retinal image analysis for DR assessment and (ii)- for speech analysis for PD diagnosis. The objectives of this research are to:

i. Develop a framework of DL for detection and segmentation of DR signs in the color fundus images.
ii. Investigate and compare different color spaces of fundus images for DR signs detection
iii. Compare different DL methods (CNN, pre-trained and Restricted Boltzmann Machine (RBM) models) for detecting the DR signs.
iv. Investigate different image enhancement methods to improve the performance of DL-based techniques.
v. Develop a post-processing algorithm to improve the performance of DL-based techniques.
vi. Develop a DL model for automatic and simultaneous detection of three DR signs.
vii. Propose a DL model for assessment of PD using speech signal.
viii. Investigate various DL architectures to identify suitable architecture for PD assessment.

This research has comprehensively studied and developed various DL methods for detection of DR signs and diagnosis of PD. This demonstrates the importance of combining the machine learning, image and signal processing techniques to the DL techniques and assess the impact of them on the clinical applications. This also suggests using suitable color channels and DL's architecture combined with appropriate pre/post processing algorithms can have a significant impact on the performance of the system. This thesis also delivers the effect of different designs of DL architectures on the system's performance.

1.5 Outline of the Thesis

The chapters of the thesis are organized as follow:

Chapter 1 describes the overall thesis with an introduction, problem statement, hypothesis and aims, and objectives of the work.

Chapter 2 presents a literature review of the DL methods used in this thesis.

Chapter 3 investigates and compares the performance of different color spaces of fundus images for automatic detection of exudates.

Chapter 4 investigates different DL techniques to maximize the sensitivity and specificity of the method for automatic detection of exudates.

Chapter 5 proposes a novel pre-processing layer in the CNN architecture to increase the performance of the system for DR signs detection.

Chapter 6 develops a method using the probabilistic output of the CNN model to automatically and simultaneously detect the DR signs.

Chapter 7 investigates the efficiency of the CNN model in distinguishing between Parkinson's and healthy voices.

Chapter 8 concludes the findings of the thesis and discusses future studies to be undertaken.

Chapter 2

2 Methodology Background

2.1 Overview

This chapter provides a literature review of the fundamentals of DL with a focus on the methods used in this thesis. It also provides a description of the software and hardware that is required for implementation of the algorithms. This introduces the fundamental concepts of the DL methods and comparison of their performances in the different applications of biomedical data analysis.

2.2 Deep Learning Algorithms

In common machine learning pipelines for different applications, extracting informative features are conducted by the machine learning experts on the basis of their knowledge about the target domain (Shen, Wu et al. 2017). However, DL methods have overcome this issue by learning a hierarchy of features from input data (Schmidhuber 2015, Kamal Maried, Abdalla Eldali et al. 2017). Instead of a handy-crated feature selection and extraction, the DL methods automatically learn the informative representation in a self-taught manner (LeCun, Bengio et al. 2015).

Researches have become interested in using DL methods because these methods not only require less engineering efforts but also have achieved record-breaking performances in the variety of artificial intelligence applications (Collobert and Weston 2008, Mikolov, Deoras et al. 2011, Sutskever, Martens et al. 2011, Hinton, Deng et al. 2012, Krizhevsky, Sutskever et al. 2012, Farabet, Couprie et al. 2013, Sainath, Mohamed et al. 2013, Tompson, Jain et al. 2014, Suk and Shen 2015, Szegedy, Wei et al. 2015, Zhang, Li et al. 2015, Kleesiek, Urban et al. 2016, Wu, Kim et al. 2016).

DL methods, generally, can be categorized into the five main groups (shown in Figure 2.1):

- CNNs
- Pre-trained models
- Deep neural networks (Auto-Encoders)
- Deep generative models (Restricted Boltzmann Machines (RBMs))
- Sequential models (Recurrent Neural Networks (RNNs))

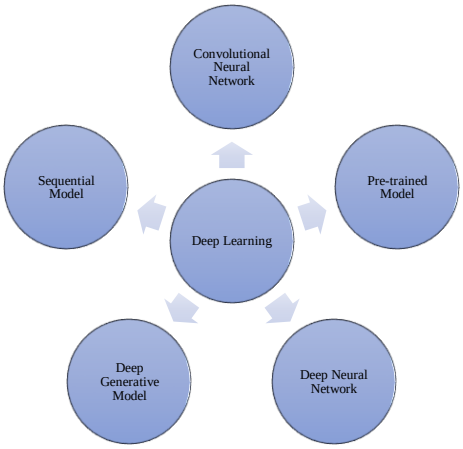


Figure 2.1. Different deep learning models.

In this thesis, three DL methods are used and described in this section:

- CNNs
- Pre-trained Models
- RBMs

2.2.1 Convolutional Neural Network

CNN architectures and their applications are explained as folow.

2.2.1.1 Convolutional Neural Network's Architecture

A CNN architecture, basically, is comprised of different layers including convolutional and pooling (down-sampling) layers that are followed by fully connected layers (similar to multi-layer neural networks). A typical CNN architecture is shown in Figure 2.2.

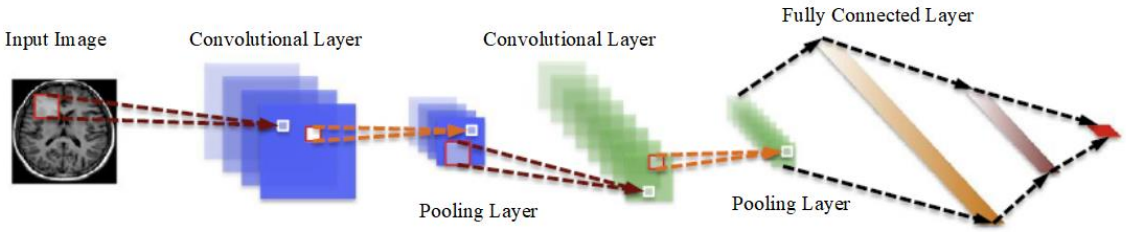


Figure 2.2. A typical CNN architecture (Zhou, Greenspan et al. 2017).

Convolutional layer

The purpose of the convolutional layer is to detect local features at different positions of the feature map of the pervious layer where a feature map is the output of one filter applied to the pervious layer. According to formula (2.1), each neuron from the same convolutional matrix generates n feature maps in the convolutional layer if there are n kernels.

$$A_i = f\left(\sum_{j=1}^{N} I_j * K_i + B_j\right) \tag{2.1}$$

In formula (2.1), I_j is j^{th} input matrix, where there is only one input matrix from the input layer, and $f(.)$ is a nonlinear activation function. k_i, B_i, $*$ indicate the i^{th} convolutional kernel matrix, bias matrix and a convolutional operation, respectively.

Pooling Layer

A pooling layer follows the convolutional layer to downsample and reduce the size of the feature map of the preceding convolution layer. This layer is responsible to progressively

reduce the spatial size of the feature map and decrease the number of computations and parameters. Another advantage of pooling layer is for transformation invariance over small spatial shifts in the input. Among different types of the pooling layers, Max-Pooling is the most privilege method which is widely used in the CNN's architectures. This layer selects the maximum number of each stride. Figure 2.3 shows an example of the transformation invariance feature of the Max-Pooling operation on the two different feature maps. While the blue matrix by the size of 2×2 in the feature maps is different, the output for both are same because the blue matrix in feature map B is the horizontally flipped version of the feature map A.

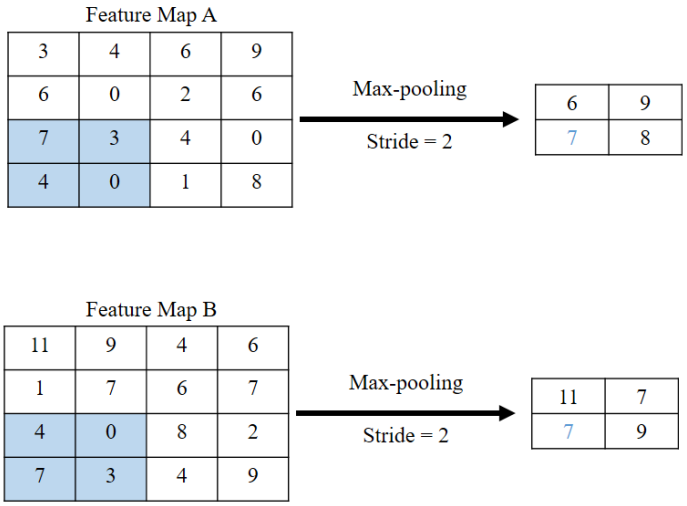


Figure 2.3. Max-Pooling processing on two different feature maps. While the blue matrixes are different, the output are same due to transformation invariance feature.

Fully Connected layer

Fully connected layers have the characteristic as same as the multi-layer neural networks (Abbod, Catto et al. 2007, van Gerven and Bohte 2017). The input units are the last proceeding feature map which is reshaped to a vector followed by the hidden layers and the output units are mostly the number of target classes which varies from one application to another one.

In a CNN architecture, updating learnable parameters (weights and biases) require a training phase which is described as follow.

Training Convolutional Neural Networks

To train a CNN model, the back-propagation algorithm is used to update the network parameters (weights and biases). Assuming that $\theta = \{W_i, b_i\}$ defined as network parameters, where w and b correspond to weight and bias in the convolutional and fully connected layers, respectively. For the training process, the loss function of L_c is defined as follow:

$$L_c = -\frac{1}{|C|}\sum_{i=1}^{|C|} \ln\left(p\left(D^i|C^i\right)\right) \tag{2.2}$$

Where $|C|$ represents the number of items in the training data, C^i and D^i denote the i^{th} training sample and its label, respectively. To update θ parameters, stochastic gradient descent (SGD) (Pang, Yu et al. 2017) method is used. The θ in each iteration is updated by (2.3):

$$\theta(p+1) = \theta(p) - \gamma\frac{\partial L_c}{\partial \theta} + \Delta\theta(p) - \mu\gamma\theta(p) \tag{2.3}$$

where γ, ϑ and μ denote learning rate, momentum rate and weight delay rate, respectively. One of the main challenges in training a deep CNN architecture is the backpropagated error message that becomes inefficient due to vanishing gradient after repeated multiplications which is called 'Vanishing Gradient' (Schmidhuber 1992). This causes the slower learning process in the layers close to the input layer compared to ones are closer to the output layer. To tackle this issue, some techniques are proposed as follow:

- Rectified Linear Unit
- Batch Normalization
- Dropout

Rectified Linear Unit

A neuron activation function is a step after the convolutional and fully connected layer. In the multi-layer neural networks, the hyperbolic tangent and sigmoid are the two most common non-linear activation functions methods which are used. Using such these functions, however, lead to decreasing the classification accuracy in the deep CNNs because of the saturation characteristics which could make the output gradient drop close to zero (Li, Cai et al. 2014). To address this issue, non-linear non-saturating functions, such as Rectified Linear Unit (RELU) and maxout (Goodfellow, Warde-Farley et al. 2013) functions, are used in the CNN architectures. The RELU function can vanish the gradient problem (Hinton and Salakhutdinov 2006, Glorot and Bengio 2010), and consequently, leads to improving the learning speed. The RELU is given by formula (2.4).

$$F(x) = \max(0, x) \tag{2.4}$$

Batch Normalization

Due to the change in the network parameters during the training phase, it's observed that the change in the distribution of the network activations could cause longer training time. To address this problem, Ioffe and Szegedy (Ioffe and Szegedy 2015) presented a batch normalization method by applying the normalization for each mini-batch and backpropagating the gradients through the normalization parameters such as scale and shift. This layer follows the Max-Pooling layer to refine the pooling layer. Each unit of feature map is normalized by 2.5 :

$$\hat{u}_i = \frac{u_i - \bar{\mathrm{E}}[u_i]}{\sqrt{Var\,[u_i]}} \tag{2.5}$$

Where i stands for an index of the units (u_i) in the layer. In the next step, a pair of learnable parameters α_i and β_i are introduced to scale and shift the normalized values as follow:

$$y_i = \alpha_i\, \hat{u}_i + \beta_i \tag{2.6}$$

The output of equation 2.6 (y_i) then is fed into the following layer.

Dropout

The aim of dropout technique is to randomly deactivate a fraction of the units, e.g., 50%, in a network on each training iteration. This method prevents a complex co-adaptation among units that helps the CNN to avoid overfitting. The temporal and random removal of unites in the training phase can be assumed to train the different networks with sharing connections weights. This method is used in the training phase, while all units must be activated in the test phase. Figure 2.4 shows a neural network before and after applying the dropout, where the circles with a cross symbol inside denote deactivated units.

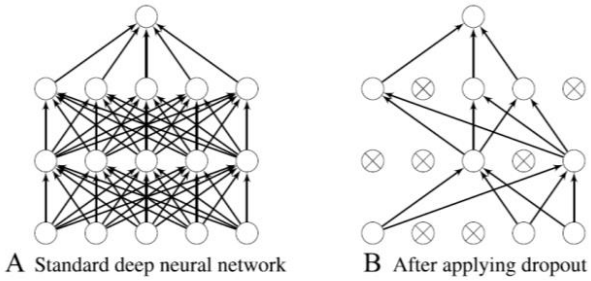


Figure 2.4. Comparison of a neural network before and after applying dropout method, where the circles with a cross symbol inside denote deactivated units (Zhou, Greenspan et al. 2017).

2.2.1.2 Application of Convolutional Neural Networks

Different CNN architectures and number of parameters have resulted in presenting various networks used in addressing different problems. The main CNN's applications can be categized into:

- Image classification, segmentation and registration
- Object detection
- Speech analysis
- Text detection and recognition
- Action and pose estimation

In this thesis, image classification and segmentation and speech analysis are used.

Image Classification

In 1990, the first version of the CNN model was introduced by LeCun et al. for hand-written digits recognition (Cun, Boser et al. 1990). The main purpose of this network (LetNet) was to classify handwritten digits into the ten classes. LetNet with five layers could not achieve a promising result due to the lack of a large number of data sets for training the model. In 2012, Krizhevsky et al. (Krizhevsky, Sutskever et al. 2012) presented the 'AlexNet' with eight layers including five convolutional and pooling layers and three fully connected layers (shown in Figure 2.5) which won the completion of ImageNet challenge (Russakovsky, Deng et al. 2015) and attracted lots of attention from the other researchers in machine learning and computer vision field.

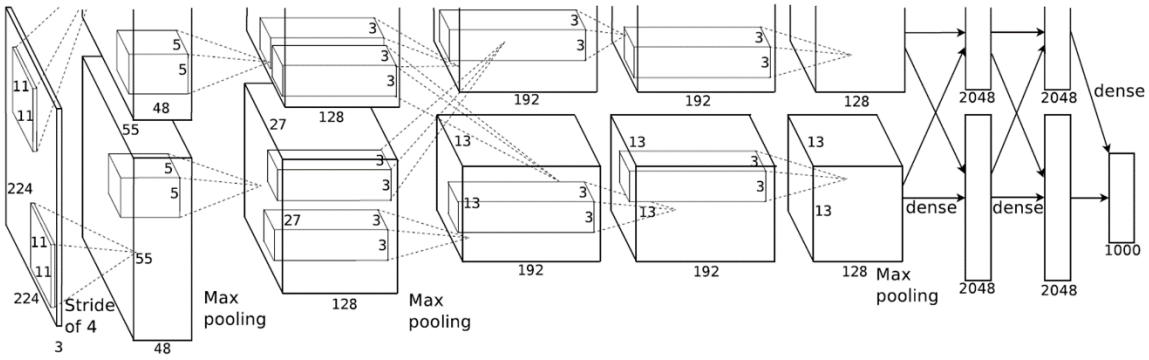


Figure 2.5. AlexNet Architecture (Krizhevsky, Sutskever et al. 2012).

The results of the AlexNet showed a decent accuracy although training these numbers of weights took lots of time. For addressing this issue, the ability of graphical processing units (GPUs) for parallel processing were used to decrease the training time while it took a long time just by using central processing units (CPUs). Currently, both GUPs and CUPs are managed to do some parts of the process together. By using smaller kernel size in the convolutional layer, a similar function can be represented with fewer parameters. Another advantage of such networks is having a lower memory footprint which makes a reliable and feasible deployment of these models on some devices such as esdge devices. In 2014,

Simonyan and Zisserman (Simonyan and Zisserman 2014) proposed a CNN architecture (VGGNet) with 19 layers that won the ImageNet challenge of 2014. Presenting the deeper models continued until 'GoogleNet' was proposed by Szegedy et al. (Szegedy, Liu et al. 2014). Beside using deeper architecture, this model introduced inception blocks (Lin, Chen et al. 2013). Unlike the previous networks (AlexNet and VGGNet), the inception block in GoogleNet extracts features with the different kernel sizes in the inception block that then they are concatenated at the end of the inception block. Comparison of different topologies used in AlexNet and VGGNet with GoogleNet are shown in Figure 2.6. This image shows that the AlexNet employs different kernel sizes (11×11, 5×5 and 3×3) and VGGNet uses the small kernel size (3×3). However, GoogleNet utilizes the inception module was able to extract features by different kernel sizes.

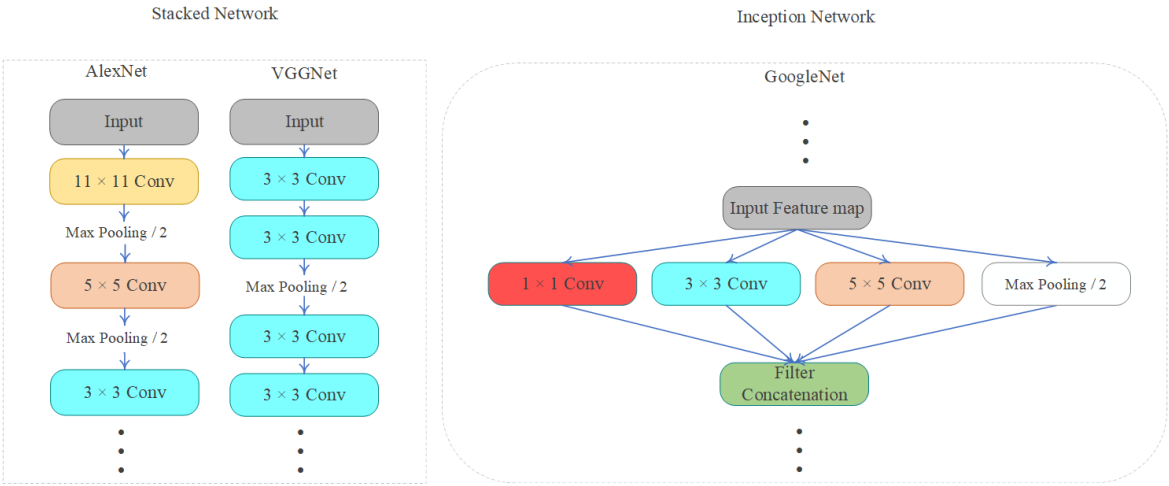


Figure 2.6. Comparison of the three different topologies used in AlexNet, VGGNet and GoogleNet.

After presenting GoogleNet, the researchers tried to develop a deeper network to increase the performance of the system. However, it was observed that increasing the number of layers was led to decreasing the accuracy due to the vanishing gradient problem that even was not compensated even by using the RELU and dropout layers. In 2015, the solution was devised by He et al. (He, Zhang et al. 2016) by introducing residual connections (He, Zhang et al. 2016). In the residual setup, this does not only pass the output of convolutional layer (F(x)) to the next layer, but this also adds up the input of convolutional layer (x) and then

pass (F(x) + x) to the next layer (Figure 2.7). Expanding this principle into the entire network resulted in introducing the new architecture ‘ResNet’ that won ImageNet completion in 2015.

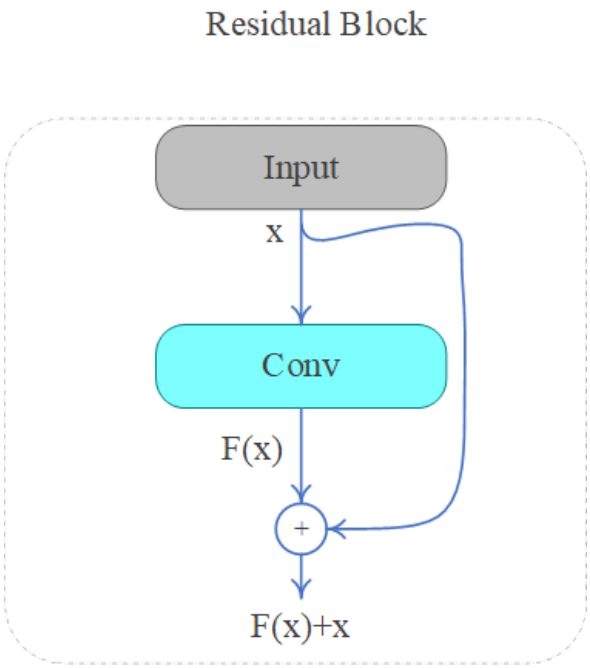


Figure 2.7. Residual block. (Conv: convolutional layer)

In Figure 2.8, the 34-layer residual network is compared with 34-layer plain network and VGG-19.

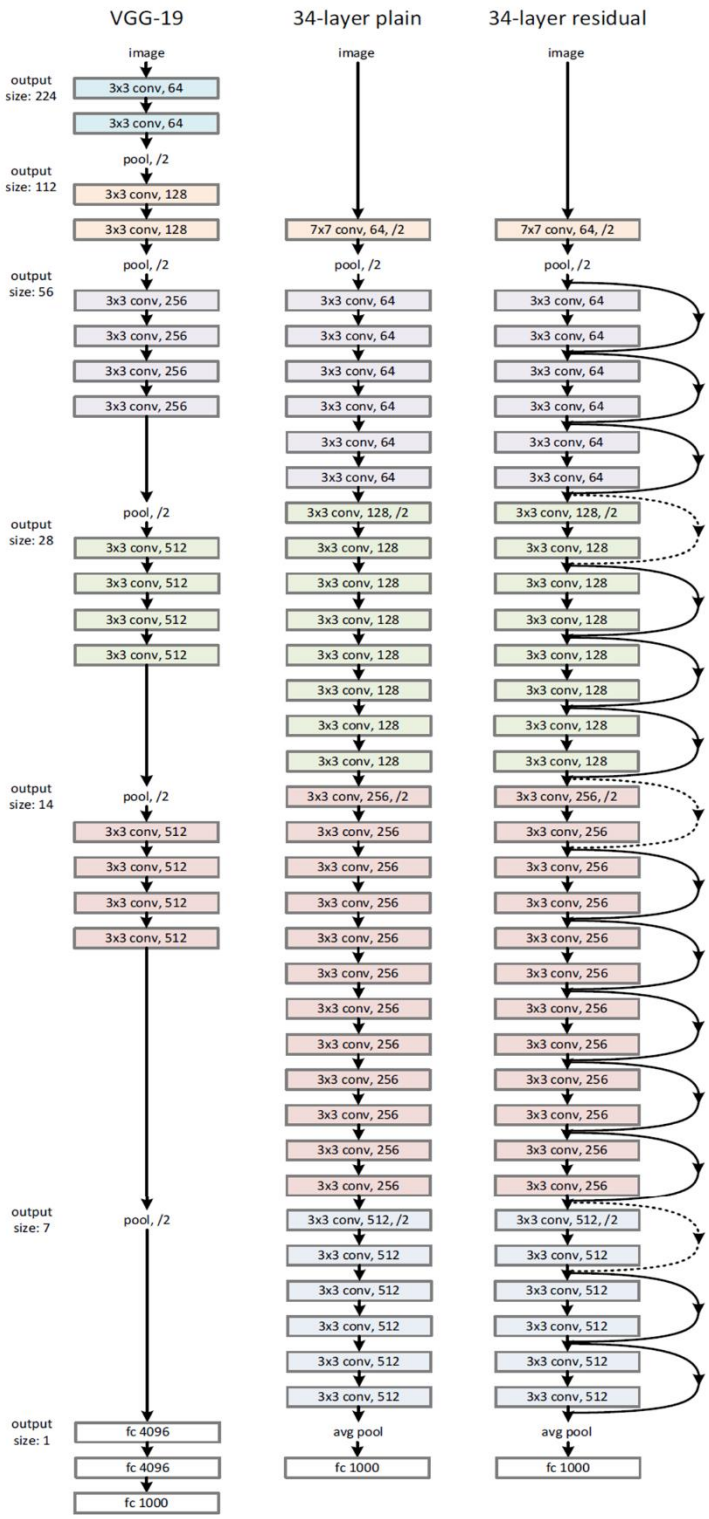


Figure 2.8. Comparison of the 34-layer residual network with 34-layer plain network and VGG-19 (He, Zhang et al. 2016).

Image Segmentation

One of the drawback of using a fully convolutional neural network (fCNN) for the segmentation of the entire image is losing resolutions due to the pooling layers. 'Shift-and-stitch' is one of the several techniques proposed to prevent this resolution decrease (Shelhamer, Long et al. 2017). The fCNN was applied to a shifted version of the input image and by stitching the result together, one obtains a full resolution version of the final output, minus the pixels lost due to the 'valid' convolutions.

Using an upsampling path was become popular among the proposed CNN-based methods for the segmentation purpose (Long, Shelhamer et al. 2015). In 2015, Ronneberger et al. (Ronneberger, Fischer et al. 2015) proposed 'U-net' architecture comprising a fCNN followed by an upsampling section which increased the image size (shown in Figure 2.9). The authors combined it with skip-connections to directly connect opposing expanding and contracting convolutional layers. Milletari et al. (Milletari, Navab et al. 2016) extended this U-Net layout that incorporates ResNet and a Dice loss layer that could directly minimize segmentation error.

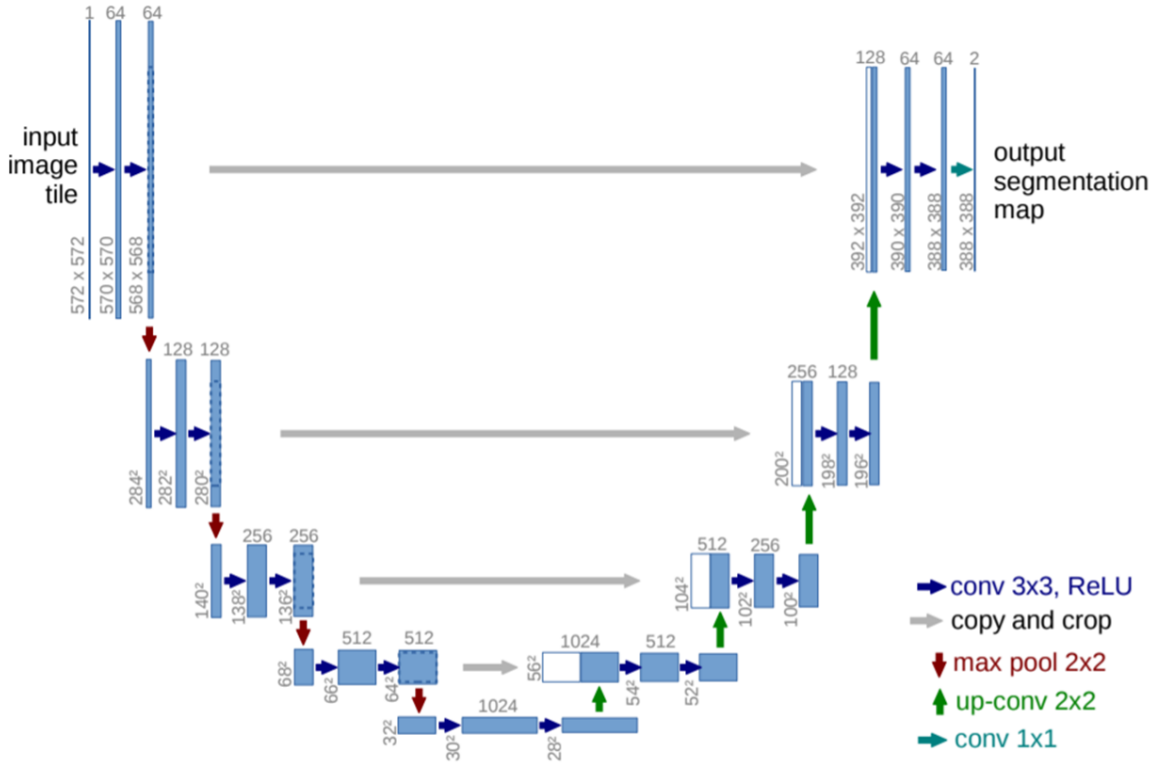


Figure 2.9. U-net architecture for segmentation application (Ronneberger, Fischer et al. 2015).

As another example of a CNN-based model for segmentation purpose, mitosis detection for screening the breast cancer was one the first work in this area (Ciresan, Giusti et al. 2013). For pixel classification, they used 'sliding-window' approach that each pixel was presented by the neighborhood pixels. The developed CNN architecture could win the ICPR competition in 2012. However, a drawback of this naïve method is producing huge overlaps for analyzing the pixels as well as variability of the window sizes. Small patches can introduce just local information of the image while the larger ones are the presenter of global characteristics. However, both local and global features can play an important role in some applications (Shyu, Brodley et al. 1998, Murphy, Torralba et al. 2006). To address this issue, Zho and Jia (Zhao and Jia 2016) proposed a multiscale CNN model for brain tumor segmentation. Instead of one fixed patch size, they trained three networks with the three different patch sizes (48×48, 28×28 and 12×12) and the output was yielded by a

combination of the three networks (Figure 2.10). In comparison to the traditional one-pathway architecture, their algorithm was more accurate and robust.

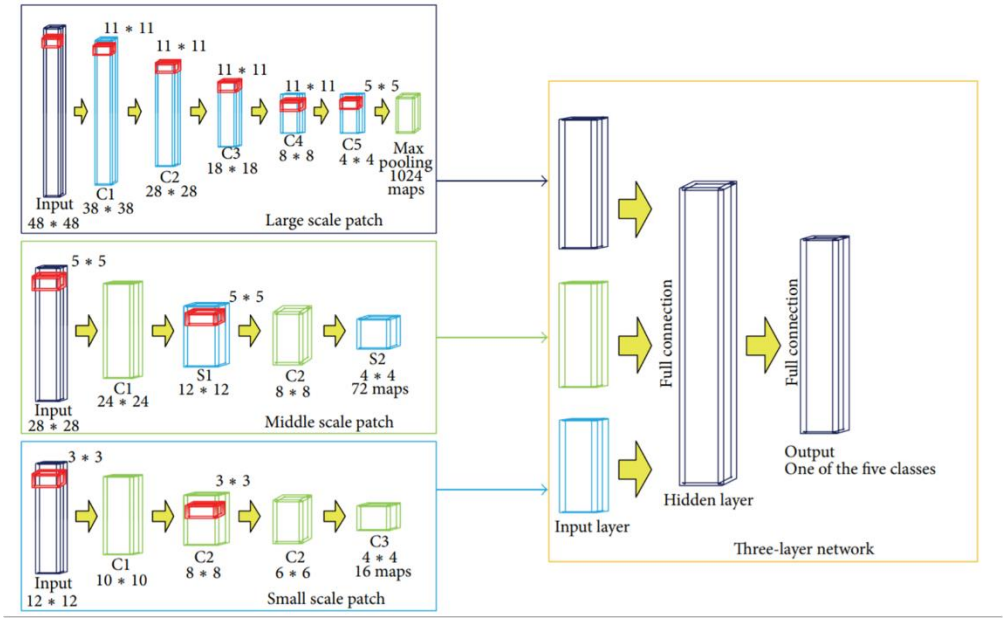


Figure 2.10. The architecture of multiscale three-layer neural network (Zhao and Jia 2016).

As a segmentation application of the CNN-based method for retinal image analysis, Li et al. (Li, Feng et al. 2016) turned the segmentation task into vessel mapping problem by using a 5-layer deep neural network for segmentation of the vessels in retinal fundus images. In 2016, Liskowski and Krawiec (Liskowski and Krawiec 2016) proposed a structure prediction scheme to highlight the context information with a 7-layers CNN architecture without any pooling layer. Similarly, Fu et al. (Fu, Xu et al. 2016) with a combination of the 7-layer CNN and conditional random field presented a recurrent neural network to model long-range pixel interactions. In 2018, a combination of the wavelet transform with a CNN model was presented to overcome the variability of width and direction of the vessel structure in the retina (Oliveira, Pereira et al. 2018). As a result of features of CNN-based model described above, the main advantages and drawbacks of these models can be concluded as follow (Pak and Kim 2017, Batmaz, Yurekli et al. 2018, Khan and Yairi 2018, Nash, Drummond et al. 2018).

Advantages:

- State-of-the-art performance in a variety of applications.
- Automatically salient features extraction from different types of data.
- The ability to perform function operation rapidly on multiple core GPUs up to 5000% over CPU-only implementations.

Drawbacks:

- Require large datasets to train the model from scratch (small dataset could just train the limited number of the neurons which not reflecting the high-level representations).
- Evaluating and tuning lots of parameters in order to get high performance, such as number of the layers, kernel size, filters, learning rate and type of activation function.

2.2.2 Pre-trained Model

As mentioned in section 2.2.1.2, training a CNN model from scratch requires a large dataset. However, most of the medical datasets are typically small (hundreds/thousands of samples) compared to millions of images in the computer vision application (Russakovsky, Deng et al. 2015). Therefore, using pre-trained models (Transfer Learning) has become popular among the researchers, especially in medical image applications.

Transfer learning is a method that transfer the knowledge from the source domain to the target domain (Figure 2.11).

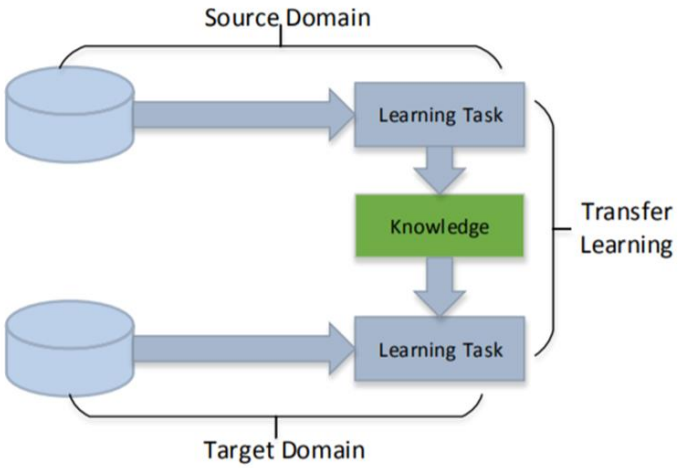


Figure 2.11. Transfer learning process (Tan, Sun et al. 2018).

There are two strategies for transfer learning:

- Using a pre-trained network as a feature extractor
- Fine-tuning a pre-trained network

Compare to the fine-tunning method, the extra benefit of using a pre-trained network as a feature extractor is not requiring a training stage and tunning network's parameters that is easily plugged into the existing image analysis pipelines. To compare the performance of these two strategies, different studies were investigated (Azizpour, Razavian et al. 2015, Gulshan, Peng et al. 2016, Kim, Corte-Real et al. 2016, Tajbakhsh, Shin et al. 2016, Esteva, Kuprel et al. 2017, Tiulpin, Thevenot et al. 2018). For instance, Li et al. (Li, Pang et al. 2017) used three different approaches for classification of DR: (i) find-tunning, (ii) transfer learning and (iii) full training. Their results showed that the transfer learning method outperformed the other approaches.

For different applications, different pre-trained models were used. For classification of tissues in histological images, Mazo et al. (Mazo, Bernal et al. 2018) used four different pre-trained models (VGG-16, VGG-19, GoogleNet and ResNet) which initially trained on the ImageNet dataset (Russakovsky, Deng et al. 2015). It is worth to mention that each network has a different architecture with a different number of parameters (Table 2.1).

Table 2.1. Detail of the different architectures used in (Mazo, Bernal et al. 2018).

Item	VGG-16	VGG-19	GoogleNet	ResNet
General				
Parameters	15.5 M	20.8 M	24.1 M	25.9 M
Channels	3	3	3	3
Input size	100×100	100×100	150×150	200×200
Number of layers				
Convolutional	13	16	94	53
Max- Pooling	5	5	4	1
Fully connected	3	3	3	3
Presence of modules				
Batch normalization	No	No	Yes	Yes
Residual connection	No	No	No	Yes

Abbasi-Sureshjani et al. (Abbasi-Sureshjani, Dashtbozorg et al. 2018) used the ResNet architecture for extracting a feature from retinal fundus images for DR assessment.
It can be seen that different pre-trained models have been used in different applications. Therefore, the success of the transfer learning method (either fixed feature extractor or fine-tuning) is subjected to the application and the type of network.

2.2.3 Restricted Boltzmann Machine

The first version of the Restricted Boltzmann Machine (RBM) models was presented by Smolensky in 1986 (Smolensky 1986) and it was then represented based on the fast learning algorithm by Hinton (Hinton 2012). An RBM model is an unsupervised method and energy-based stochastic neural networks composed of two different layers: (i) Visible layer and (ii) Hidden layer. These layers include a different number of visible (V) and hidden (h) units,

respectively. A weight matrix ($W_{m\times n}$) stands for the connection weights between the visible and hidden layer with m and n units, respectively. This connection is bidirectional, so given an input vector V that one can obtain the latent feature representation h and vice versa (Figure 2.12)

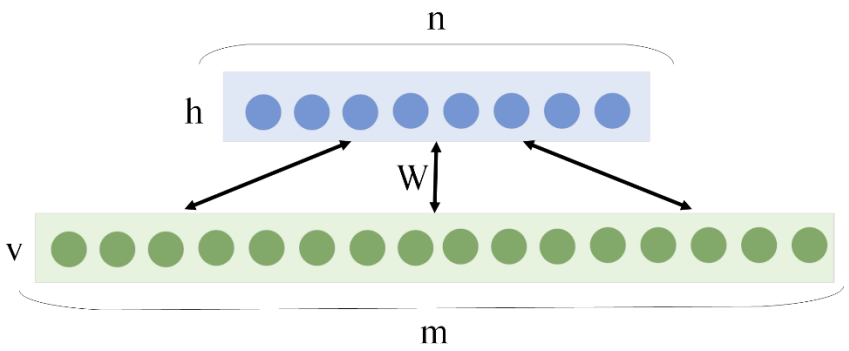


Figure 2.12. An example of the RBM architecture.

An energy function of $E(v,h)$ is defined and minimized by:

$$E(v,h) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n} v_i h_j W_{ij} \tag{2.7}$$

Where a and b are the biases of visible and hidden unites, respectively.
The probability of joint configuration (v,h) is computed as follow:

$$P(v,h) = \frac{1}{z} e^{-E(v,h)} \tag{2.8}$$

Where Z is the normalization factor computed over all possible configurations including hidden and visible unites called the partition function. The marginal probability of the visible vector is given by:

$$P(v) = \frac{1}{z}\sum_{h} e^{-E(v,h)} \tag{2.9}$$

As the activations of the visible and hidden units are independent, $P(v|h)$ is leading to following conditional probability:

$$P(v|h) = \prod_{i=1}^{m} P(v_i|h) \tag{2.10}$$

$$P(h|v) = \prod_{j=1}^{n} P(h_j|v) \tag{2.11}$$

Where:

$$P(v_i = 1|h) = \emptyset\left(\sum_{j=1}^{n} W_{ij}h_j + a_i\right) \tag{2.12}$$

$$P(h_j = 1|v) = \emptyset\left(\sum_{i=1}^{nm} W_{ij}v_j + b_j\right) \tag{2.13}$$

Note that $\emptyset(.)$ stands for the logistic-sigmoid function and assume that $\theta = (W, a, b)$ are the RMB's learnable parameters. The aim is maximizing the product of the probabilities given by the training data set (D) as follow:

$$argmax_\theta \prod_{v \in D} P(v) \tag{2.14}$$

To solve the equation 2.14, the contrastive divergence method was proposed by Hinton (Hinton 2002) and it was then developed and expanded in 2012 (Montavon, Orr et al. 2012). The next generation of the RBM models is Deep Belief Networks (DBNs) (Bengio, Lamblin et al. 2006, Hinton, Osindero et al. 2006) and Deep Boltzmann Machines (DBMs) (Salakhutdinov and Larochelle 2010) which had the similar application with the slight

difference from the original RMB model. One of the challenges in using the RBM models for image application was that the original RBM model was only suitable for the binary image processing. In order to deal with the real images (gray scale and color images), a series of RBM variants are put forward, such as Gaussian-Binary RBM (GRBM) (Lee, Grosse et al. 2009, Cho, Ilin et al. 2011), the covariance RBM (cRBM) (Ranzato, Krizhevsky et al. 2010), the mean and covariance RBM (mcRBM) (Ranzato and Hinton 2010) and the spike-slab RBM (ssRBM) (Courville, Bergstra et al. 2011, Courville, Bergstra et al. 2011, Goodfellow, Courville et al. 2012, Goodfellow, Courville et al. 2012, Courville, Desjardins et al. 2014).

The original and developed RBM models are commonly used for various applications in image processing and machine learning. Lee et al. proposed a method by a combination of a CNN and RBM model that could learn the two-dimensional structure information of images and realized classification task (Lee, Grosse et al. 2009, Lee, Yan et al. 2009, Norouzi, Ranjbar et al. 2009, Lee, Hong et al. 2011). It is worth to mention that the application of the RBM models is not limited in the image analysis and these such models are widely used in analyzing the sequential data for speech and metadata analysis (Sutskever and Hinton 2007, Sutskever, Hinton et al. 2009, Swersky, Tarlow et al. 2012, Mittelman, Kuipers et al. 2014, Reed, Sohn et al. 2014, Montúfar, Ay et al. 2015, Montúfar and Morton 2015).

2.3 Software and Hardware

As mentioned in section 2.2, it can be concluded that the success of the DL methods is due to the two main factors (Hinton and Salakhutdinov 2006, Vincent, Larochelle et al. 2010, Srivastava, Hinton et al. 2014):

- Accessing the big dataset
- Advances in high-tech GPUs

The widespread availability of GPUs and GPU-computing libraries (CUDA, OpenCL) has the main contribution in developing the DL methods. With current high-tech GPUs, deployment of the DL methods becomes roughly 10-30 times faster than on CPUs.

Another reason for popularity in using the DL methods is the availability of the open-source software package. It allows to user to implement ideas at a high level rather than worrying about efficient implementation. The most popular packages are:

- Caffe (Jia, Shelhamer et al. 2014) provides C++ and python interface, developed by UC Berkeley.
- Tensorflow (Abadi, Agarwal et al. 2016) comes with C++ and python interface which was developed by Google research.
- Theano (Bastien, Lamblin et al. 2012) provides a python interface, developed by MILA lab in Montreal.
- Torch (Collobert, Kavukcuoglu et al. 2011) provides the Lua and python interface which wad developed by Facebook AL research.

In this thesis, The Caffe platform is used for implementing the DL algorithms.

2.4 Conclusion

This chapter has described the literature review of the methodology used in this thesis. The general concept of DL methods including the three main DL approaches are explained in detail: (i) CNNs, (ii) Transfer Learning method and (iii) RBMs. The pros and cons of these methods are discussed and compared together. A review of the literature on successful applications of the DL methods has been explained. This chapter also covers the detail of the required software and hardware for implementing the DL algorithms.

Further to the investigation of the current DL approaches, this thesis has proposed the novel methods to improve the performance of the current DL-based method in the biomedical applications as follow:

i. Color space analysis of fundus images for automatic exudate detection (Chapter 3)
ii. Investigation different DL methods for automatic exudate detection (Chapter 4)

iii. A novel pre-processing layer in convolutional neural network for automatic identification of diabetic retinopathy (Chapter 5)

iv. Analysis of deep probabilistic features for detection of exudates, hemorrhages and microaneurysms (Chapter 6)

v. Parkinson's disease diagnosis based on multivariate deep features of speech signals (Chapter 7)

vi. Conclusion and future work (Chapter 8)

Chapter 3

3 Color Space Analysis of Fundus Images for Automatic Exudates Detection

3.1 Overview

This chapter has compared the performance of different color spaces of fundus images for automatic detection of exudates. A convolutional neural network was employed to assess the performances of different color spaces generated by orthogonal transformation of the original colors in red/green/blue (RGB) space. Experiments were conducted on two publicly available databases: 1- DIARETDB1 and 2- e-Ophtha. Based on the experimental results, this chapter has proposed a new color space of fundus images with three channels: (i) second eigenchannel of the RGB space, (ii) hue and (iii) saturation channels of Hue/Saturation and Intensity (HSI) space. This achieved an accuracy, sensitivity and specificity of 98.2%, 0.99 and 0.98, respectively. Twenty times 20-fold cross validation technique confirmed that proposed color space obtained higher replicability compared with conventional color spaces.

3.2 Introduction

DR is a common cause of vision impairment in the world population (Abràmoff, Reinhardt et al. 2010, Mookiah, Acharya et al. 2013, Leontidis, Al-Diri et al. 2017) and presence of exudates on retina has been found to have impact on vision loss (Kaur and Mittal 2018). However, vision loss can be prevented in 50% of patients if DR is diagnosed and treated in time (Ege, Hejlesen et al. 2000, Hsu, Pallawala et al. 2001, Hove, Kristensen et al. 2004).

DR diagnosis requires detection of exudates on the retina which is performed by visual examination of eye fundus images. However, this is a time-consuming task and the outcomes are dependent on expertise of the examiner. While fully automatic analysis of retinal images for exudate detection is highly desirable, significant variations in shape, size and texture of the exudates makes this a challenging task.

A number of automatic exudate detection methods have been developed which can be divided into two main categories: (i) Morphological and (ii) Machine learning based methods (Walter, Klein et al. 2002, Fleming, Philip et al. 2007, Niemeijer, van Ginneken et al. 2007, Jaafar, Nandi et al. 2010, Ali, Sidibé et al. 2013, Harangi and Hajdu 2014, Naqvi, Zafar et al. 2015, Pereira, Gonçalves et al. 2015, Zaki, Zulkifley et al. 2016). Sopharak et al. (Sopharak, Uyyanonvara et al. 2008) employed morphological operations to detect exudates based on the intensity channel (I) of HSI space. Sánchez et al. (Sánchez, García et al. 2009) used a statistical mixture model-based clustering for dynamic thresholding of the exudate pixels. García et al. (García, Sánchez et al. 2009) compared three different classification methods to identify the candidate pixels from the green channel of retinal images which include the multilayer perception (MLP), radial basis function (RBF) and support vector machine (SVM). Giancardo et al. (Giancardo, Meriaudeau et al. 2012) used the green channel and the intensity channel from eye retinal images to detect the exudate. In 2014, Zhang et al. (Zhang, Thibault et al. 2014) identified exudates from the green channel of fundus images using random forest classifier after image normalization and de-noising. In 2017, Fraz et al. (Fraz, Jahangir et al. 2017) used combination of morphological reconstructions, Gabor filter banks and a bootstrap decision tree for multiscale segmentation of exudates. However, most of these methods suffer from poor sensitivity and accuracy, especially near the vascularized region.

DL techniques have delivered promising results in various computer vision applications (Long, Shelhamer et al. 2015, Fu, Xu et al. 2016, Charron, Lallement et al. 2018, Chen, Papandreou et al. 2018, Fu, Liu et al. 2018), including retinal image analyses (van Grinsven, Venhuizen et al. 2016, Tan, Fujita et al. 2017, Khojasteh, Aliahmad et al. 2018, Khojasteh, Aliahmad et al. 2018). Prentašić and Lončarić (Prentašić and Lončarić 2015) trained a ten-layered Convolutional Neural Network (CNN) to detect exudate pixels on RGB retinal

images. However, the results suffered from the poor sensitivity of 0.77. In 2016 Perdomo et al. (Perdomo, Otalora et al. 2016) used CNN and achieved a high sensitivity of 92%, but the specificity of the method was poor (40.6%). Yu et al. (Yu, Xiao et al. 2017) and Prentašić et al. (Prentašić and Lončarić 2016) developed different CNN architectures to distinguish between exudate and non-exudate pixels. Their algorithm achieved sensitivity of 0.88 and 0.78 on a pixel level, respectively. In 2019, Khojasteh et al. (Khojasteh, Júnior et al. 2019) compared performance of a CNN-based model with other DL methods (pre-trained networks and Discriminative Restricted Boltzmann Machines) for exudates detection and they obtained accuracy of 89.1% for the CNN model. Literature indicates that DL approaches are promising for automated exudate from the eye-fundus images but require further research.

Retina image analyses have been conducted using color spaces such as RGB, HSV and LUV, and most of them have employed green channel of the RGB space because this has been found to have the highest contrast (Pachiyappan, Das et al. 2012). However, discarding other channels could lead to loss of information (Unnikrishnan, Aliahmad et al. 2013). Majority of the earlier DL methods have used three channels of the RGB space as the input of the networks (Gargeya and Leng 2017, Quellec, Charrière et al. 2017, Grassmann, Mengelkamp et al. 2018). To the best of our knowledge, no study has been reported a comparison between different color spaces.

We hypothesis that appropriate choice of color channels will enhance performance of the system for detection of exudates using a CNN-based model. This chapter reports a comparison of using different combination of color channels to identify a combination that gives the best performance of automatic exudate detection using a CNN model. Experiments were conducted using publicly available dataset to identify the combination of color channels that yielded the best performance measured based on sensitivity, specificity and accuracy, and 20-fold cross validation was used to test the replicability of the spaces.

The remainder of this chapter is organized as follows. Section 3.3 presents the methodology regarding the techniques used in this paper, sections 0 and 3.5 discuss the material and experiment setup adopted for exudate identification. Tables showing the results are in

Section 3.6, and the observations have been discussed in section 3.7. Finally, the conclusions are drawn, and limitations listed in the section 3.8.

3.3 Methodology

There are two sub-sections to the methodology: (i) Color space identification and (ii) CNN-based model for exudate detection. Three conventional color spaces of fundus images (RGB, HSI and LUV), orthogonal transformed version of these spaces and combinations of these were investigated. The images were segmented to obtain two patch groups were prepared in all the spaces belonging to (i) Exudate and (ii) Non-exudate regions. These were then used to train and test the CNN to detect the exudates and the performance was the basis for determining the most suitable set of color channels. Twenty times 20-fold cross validation approach was used and the performance of the different combination of color channels were compared. The details are described below.

3.3.1 Color Space Identification

Original images in RGB space were transformed into HSI and LUV spaces as reported in (Sopharak, Uyyanonvara et al. 2008, Welfer, Scharcanski et al. 2010). Principal component analysis (PCA) was later applied on the images for the three color spaces. The rationale behind the transformation is based on the study by Unnikrishnan et al. (Unnikrishnan, Aliahmad et al. 2013) which showed that application of PCA on the RGB space can provide eigenchannels with better contrast between background and foreground objects compared to the green channel. In this study, we used PCA transformation on each color space (RGB, HSI and LUV) and the procedure is described below.

PCA on an Image

By applying PCA on an image with three channels, three eigenchannels (principle components) are created. It is assumed that intensity of a pixel (P) for an image (I) is presented by the equation 2.1:

$$P_{(x,y)} = \big(I_{ch1}(x,y), I_{ch2}(x,y), I_{ch3}(x,y)\big); 1 \le x \le m, 1 \le y \le n \qquad (3.15)$$

where I_{chZ} corresponded to intensity of Z^{th} ($1 \le Z \le 3$) channel (I_{ch}) and (x, y) is the location of the pixel. To apply PCA on an image, each channel is reshaped to a vector (v_z) with size of [1, $m \times n$] and, consequently, that creates matrix V = [v_1 v_2 v_3] where v_z is described in 3.16:

$$V = \begin{bmatrix} v_1 & v_2 & v_3 \\ I_{ch1}(1,1) & I_{ch2}(1,1) & I_{ch3}(1,1) \\ I_{ch1}(1,2) & I_{ch2}(1,2) & I_{ch3}(1,2) \\ \vdots & \vdots & \vdots \\ I_{ch1}(1,n) & I_{ch2}(1,n) & I_{ch3}(1,n) \\ I_{ch1}(2,1) & I_{ch2}(2,1) & I_{ch3}(2,1) \\ I_{ch1}(2,2) & I_{ch2}(2,2) & I_{ch3}(2,2) \\ \vdots & \vdots & \vdots \\ I_{ch1}(2,n) & I_{ch2}(2,n) & I_{ch3}(2,n) \\ \vdots & \vdots & \vdots \\ I_{ch1}(m,1) & I_{ch2}(m,1) & I_{ch3}(m,1) \\ I_{ch1}(m,2) & I_{ch2}(m,2) & I_{ch3}(m,2) \\ \vdots & \vdots & \vdots \\ I_{ch1}(m,n) & I_{ch2}(m,n) & I_{ch3}(m,n) \end{bmatrix} \qquad (3.16)$$

In the next step, PCA was applied on V to produce eigenvector matrix EG = [eg_1 eg_2 eg_3], where eg_z corresponds to Z^{th} eigenvector and T_{chZ} (m,n) stands for value of PCA transformation (equation 3.17).

$$EG = \begin{bmatrix} eg_1 & eg_2 & eg_3 \\ T_{ch1}(1,1) & T_{ch2}(1,1) & T_{ch3}(1,1) \\ T_{ch1}(1,2) & T_{ch2}(1,2) & T_{ch3}(1,2) \\ \vdots & \vdots & \vdots \\ T_{ch1}(1,n) & T_{ch2}(1,n) & T_{ch3}(1,n) \\ T_{ch1}(2,1) & T_{ch2}(2,1) & T_{ch3}(2,1) \\ T_{ch1}(2,2) & T_{ch2}(2,2) & T_{ch3}(2,2) \\ \vdots & \vdots & \vdots \\ T_{ch1}(2,n) & T_{ch2}(2,n) & T_{ch3}(2,n) \\ \vdots & \vdots & \vdots \\ T_{ch1}(m,1) & T_{ch2}(m,1) & T_{ch3}(m,1) \\ T_{ch1}(m,2) & T_{ch2}(m,2) & T_{ch3}(m,2) \\ \vdots & \vdots & \vdots \\ T_{ch1}(m,n) & T_{ch2}(m,n) & T_{ch3}(m,n) \end{bmatrix} \quad (3.17)$$

Each eigenvector was reshaped back to the original image size [m,n] to obtain an image in the space corresponding to PCA of original color space. Figure 2.13 shows the schematic of the transformation method. After applying this to the three color spaces (RGB, HSI and LUV), it resulted in three new spaces. The total of six color spaces (three original and three PCA) were used for further analysis.

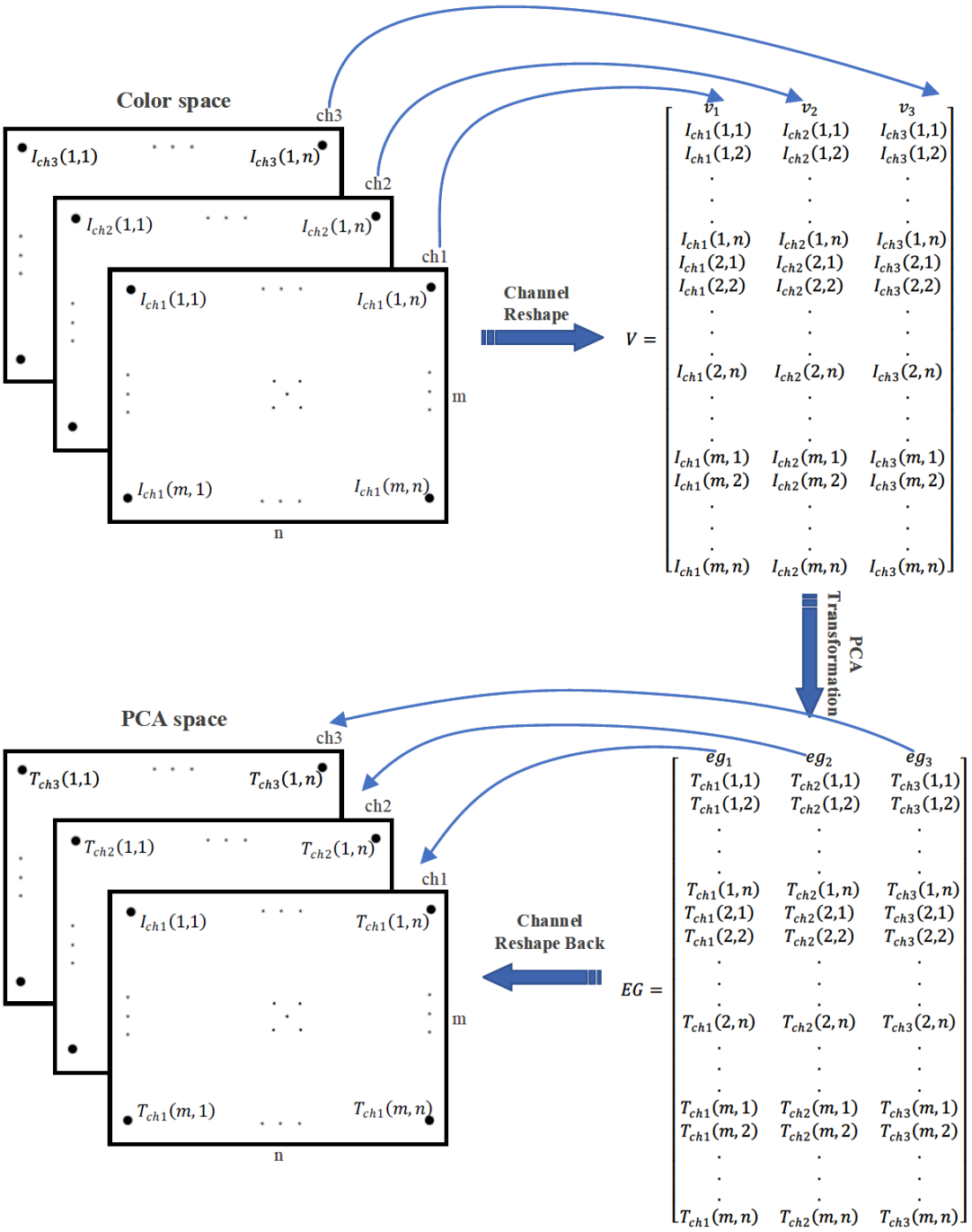


Figure 2.13.Schematic of PCA transformation procedure on an image.

Introducing New Color Space

The different color spaces were ranked based on the CNN performance measured using accuracy of classification. The new color space was identified experimentally by combining sets of three high ranking channels and comparing their performances. To reduce the number of combinations, a selection rule was formulated to combine only those channels with the low similarities measured using structural similarity index method (SSIM) (Wang, Bovik et al. 2004) (equation 3.18). This was calculated between each channel of the image against rest of the channels and ranked based on their similarity with the target channel.

$$SSIM(x,y) = \frac{(2\alpha_x\alpha_y + c_1)(2\sigma_{xy} + c_2)}{(\alpha_x^2 + \alpha_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3.18}$$

where α_x, α_y, σ_x, σ_y and σ_{xy} are the local means, standard deviations and cross-covariance for images *x* and *y*, respectively. $c_1 = (0.01 \times L)^2$ and $c_1 = (0.03 \times L)^2$, where L is dynamic range of the pixel-values. The SSIM is a value between -1 and +1, +1 indicates that the two images are identical. By comparing all the possible pairs from the nominated channels, the channels were ranked and the three channels with lowest similarity index were identified to form the new space. In the next step, the CNN was trained and tested to distinguish between exudates and non- exudates patches.

3.3.2 CNN-Based Method

To assess the performance of different color spaces for exudate detection, a CNN model was developed, and its architecture is described in Table 2.2.

Table 2.2. Architecture of the proposed CNN. I - Input image, C - Convolutional layer, MP - Max pooling layer, FC - fully-connected, FM - Feature map, CH - Channels, Neurons-N.

Layer	Type	Feature Maps and Neurons	Filter Size	Weights
0	I	3CH-25×25	-	-
1	C	16FM-23×23N	3×3	448
2	MP	16FM-12×12N	2×2	-
3	C	16FM-10×10N	3×3	2,320
4	C	16FM-8×8N	3×3	2,320
5	MP	16FM-4×4N	2×2	-
6	C	16FM-2×2N	3×3	2,320
7	MP	16FM-1×1N	2×2	-
8	FC	100N	1× 1	202

In this architecture, four convolutional layers were used with 16 feature maps in each layer of size 3 × 3 pixels. Rectified linear unit (ReLU) was used as the neuron activation function to ensure that the output was positive. Max-pooling layers of the size 2 × 2 pixels were used after the first, third and fourth convolutional layer. For updating network parameters, backpropagation training approach and stochastic gradient descent (SGD) was used.

3.4 Material

Two publicly available databases were used in this chapter and are described below:

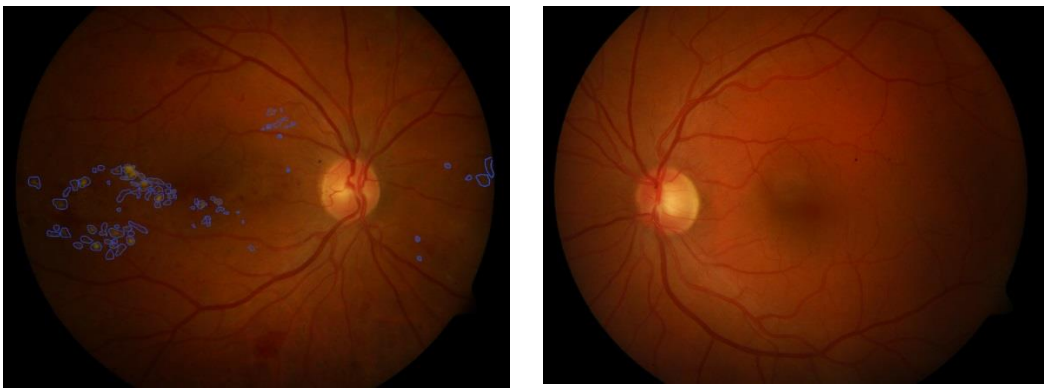
DIARETDB1

The DIARETDB1 database (Kauppi, Kalesnykiene et al. 2007) consists of 89 color retinal images of the size 1500 × 1152. All images were taken by digital fundus camera with a 50

degree field of view and there are differences between the quality of the different images. The exudates have been manually annotated and evaluated by four independent medical experts. In this work, hard and soft exudates were grouped together and considered as a single class. The Figure 2.14 on the left shows an example of a manually segmented retinal image of a diabetic patient with exudates marked in blue while on the right shows the retinal image of a healthy subject with no sign of exudate. It can be seen that the size of the exudate's spots varies from small specks to large patches.

e-Ophtha

This database comprises of 47 color retina images and have been provided with manual annotated exudates (Decencière, Cazuguel et al. 2013). In this database, the image sizes vary from 1400 × 960 to 2544 × 1696 pixels, and hence all images were resized to the size of images in the DIARETDB1 database (1500 × 1152 pixels) using average size of the optic disk (OD) as the reference.



(a) (b)

Figure 2.14. Samples from DIARETDB1 database: (a) Image from a patient where exudates are highlighted in blue, and (b) Image from a healthy subject without any exudate.

3.5 Experimental Setup

The experimental setup for this chapter contains three main sections: (i) Color space transformation, (ii) Patch preparation and (iii) Network setup.

3.5.1 Color Space Transformation

As described in section 3.3, the retinal images were first transformed from RGB to LUV and HSI spaces and PCA was then applied on each space to obtain three sets of eigenchannels: (i) PCA-RGB, (ii) PCA-LUV and (iii) PCA-HSI. An example of retina image in RGB space, its eigenchannels after color transformation are shown in Figure 2.15. Exudate spots were marked by yellow color on each eigenchannels for visual comparison. From the Fig. 3, it is observed that the second eigenchannel for each color representation indicates the highest contrast between exudate spots and background while the third and first eigenchannel include lower contrast detail.

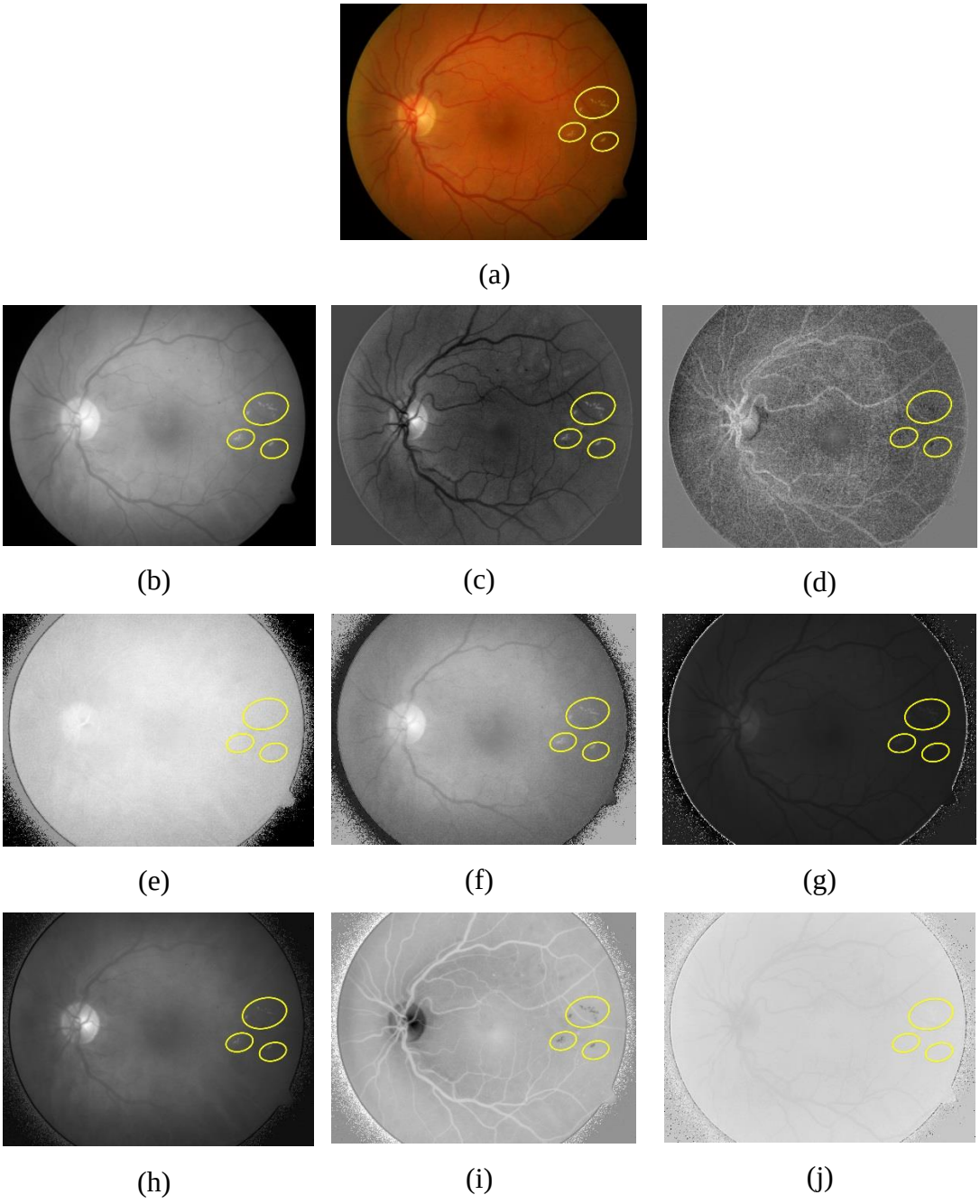


Figure 2.15. Eigenchannels of RGB, HSI and LUV: (a) Original RGB image; (b) 1st Eigenchannel of RGB; (c) 2nd Eigenchannel of RGB; (d) 3rd Eigenchannel of RGB; (e) 1st Eigenchannel of HSI; (f) 2nd Eigenchannel of HSI; (g) 3rd Eigenchannel of HSI; (h) 1st Eigenchannel of LUV; (i) 2nd Eigenchannel of LUV; (j) 3rd Eigenchannel of LUV.

3.5.2 Patch Preparation

The patch size was selected such that it enclosed only one exudate spot. The smallest spot that had been annotated corresponded to the patch size of 25 × 25 pixels which was selected for further analysis. This resulted in total of 90800 exudate patches from the two databases. To obtain a balanced dataset, similar numbers of patches (i.e. 100000) with no exudate signs were extracted and referred to Non-exudate patches. The Non-exudate patches contained vessels, background tissue and optic nerve head. There was no overlap between adjacent patches. Figure 2.16 and Figure 2.17 show the example of some exudate and non-exudate patches.

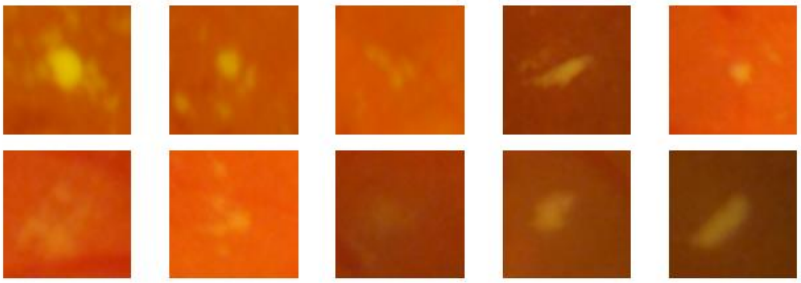


Figure 2.16. Examples of Exudate patches.

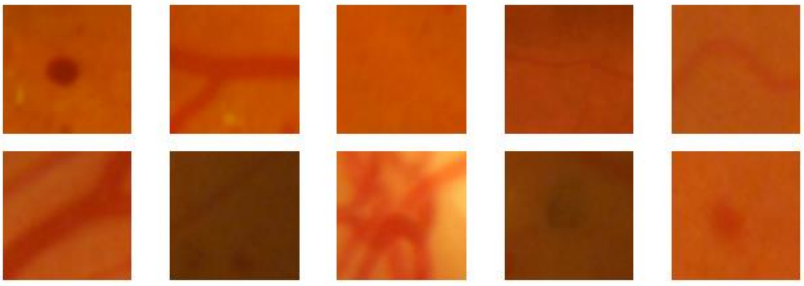


Figure 2.17. Examples of Non-Exudate patches.

3.5.3 Network Setup

The CNN parameters used in this chapter are summarized in Table 2.3. Training of the network requires identifying the maximum number of epochs to ensure that the network is not over-trained.

Table 2.3. CNN parameters for train and test phases.

CNN Parameters	
Learning Rate	0.01
Momentum	0.9
Gaussian Weight Filters	0.01
Training Batch Size	128
Validation and Test Batch Size	32
Solver Method	SGD
Gamma	0.1
Policy of the SGD	Step-Down
Step size of SGD	33
Number of training epochs	30

It is also essential that the number of epochs is sufficient, so that the network is trained such that the error is minimized. The number of epochs were obtained by plotting the error and the number of epochs which has been shown in Figure 2.18.

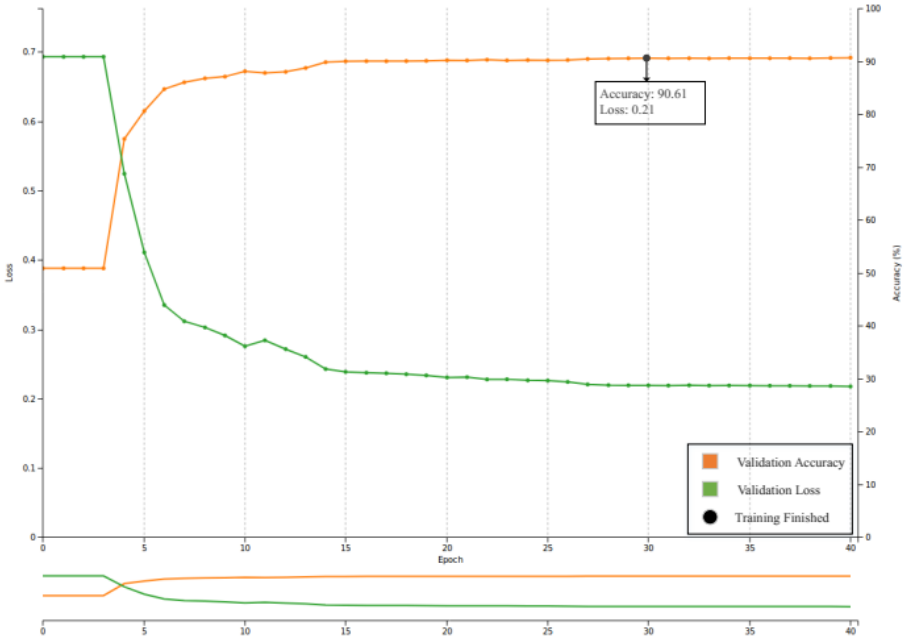


Figure 2.18. Accuracy and loss values over 40 training epochs for the RGB space. A saturation can be observed after the 30^{th} epoch.

From this Figure 2.18, it is observed that for the RGB space, the accuracy monitored over 40^{th} epochs shows that the average accuracy reached 88% after 10^{th} epoch, then gradually improved to 91% after 30^{th} epochs where it was saturated. As a result, the filter weights of the 30^{th} epoch were fixed and used for the test phase. This procedure was performed for each color spaces and the suitable number of epochs were identified for each. The implementation of the CNN was performed on the caffe platform (Jia, Shelhamer et al. 2014) with GeForce GTX 1070 graphics processing unit which took total of 150 seconds for the training phase.

3.6 Results

Performance of the algorithm was evaluated by four parameters: (i) accuracy (ACC), (ii) sensitivity (SN), (iii) specificity (SP) and (iv) positive predictive value (PPV) (Fraz, Jahangir et al. 2017). To consider a reliable evaluation of the spaces, the experiments were

validated using 20-fold cross validation approach. Mean of the results was calculated and are shown in Table 2.4.

Table 2.4. Performance of the six spaces for the detection of the exudates.

Space	ACC	SN	SP	PPV
RGB	90.07%	0.91	0.90	0.90
HSI	97.62%	0.97	0.98	0.97
LUV	89.21%	0.89	0.88	0.87
PCA-RGB	96.16%	0.95	0.96	0.96
PCA-HSI	95.48%	0.94	0.96	0.96
PCA-LUV	87.75%	0.89	0.86	0.85

The best performance among the six spaces was obtained by HSI space with 97.62% accuracy and 0.98 sensitivity. The next rank went to PCA-RGB space with 96.16% accuracy and 0.96 sensitivity. Based on these results, HSI and PCA-RGB space were identified as the most suitable spaces for representing the images for automated recognition of exudates.

3.6.1 New Color Space

The two spaces (HSI and PCA-RGB) consisting of total size channels achieved the best performances for exudate identification and these were used to create a new color space. To assess the similarity between the six channels of HSI and PCA-RGB spaces, mean of the SSIM was calculated for all the patches as shown in Figure 2.19 where PCA1, PCA2 and PCA3 correspond to the first, second and third eigenchannel of the PCA transformation, respectively.

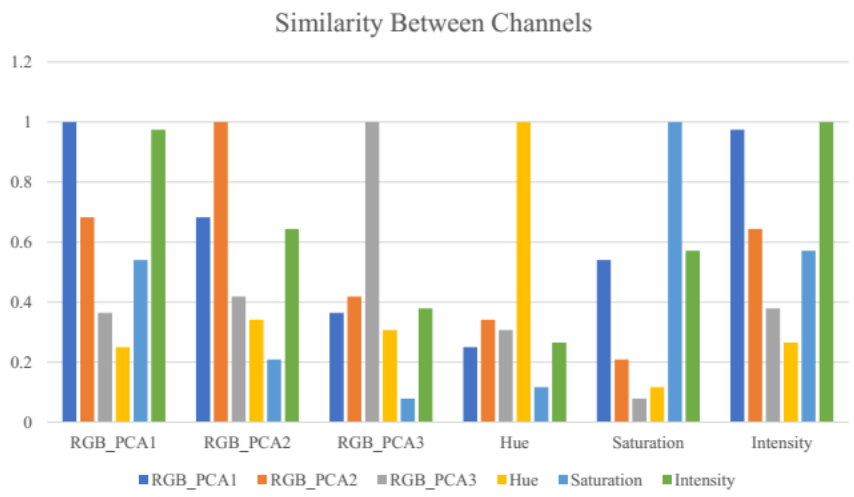


Figure 2.19. SSIM index between channels of HSI and PCA-RGB spaces.

Figure 2.19 shows that the highest similarity was between PCA2-RGB and intensity (0.97). The second rank went to similarity between PCA2-RGB and PCA1-RGB (0.70). Consequently, PCA2-RGB was selected among these three channels as a representative channel. As the PCA3-RGB contains uncorrelated noise reported by Unnikrishnan et al. (Unnikrishnan, Aliahmad et al. 2013), it was excluded from the analysis. Consequently, the new space was formed by three channels: (i) PCA2-RGB, (ii) hue and (iii) saturation which we have called *"PHS"* space.

Different sample selection of the subsets (train and test sets) from the all samples leads to the differences in the performance of the algorithm. Therefore, it is important to test the repeatedly of the CNN-model to assess the reliability of the different color spaces and the range of potential accuracies. For this purpose, the experiments for the proposed color space (PHS) and the two-color spaces which achieved the best accuracies (HSI and PCA-RGB) were compared together using 20-fold cross validation approach with 20 runs. Accuracy over the twenty runs for the three spaces are shown in Figure 2.20 and Figure 2.21.

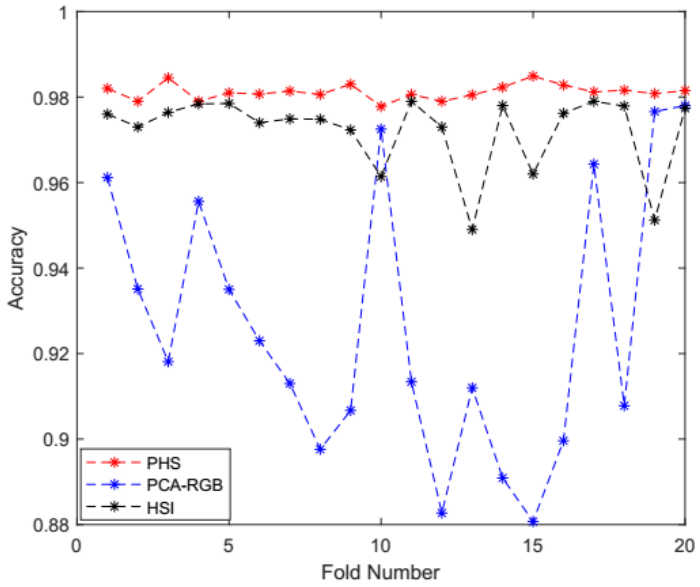


Figure 2.20. Comparing replicability of the three spaces (PHS, HSI and PCA-RGB) using 20-fold cross validation approach with 20 runs.

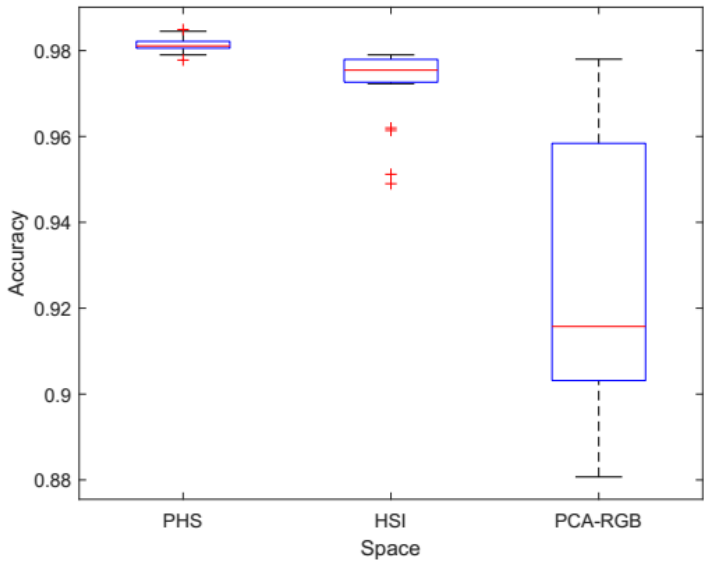


Figure 2.21. Variation of the accuracy changes for the three spaces (PHS, HSI and PCA-RGB) using 20-fold cross-validation approach with 20 runs.

It can be seen that PHS space obtained highest accuracy (98.5%) and most repeatable results with less than 0.7% variability between highest and lowest accuracy. Other color spaces were less accurate and had higher variability, with HSI's accuracy varying between 94.9% to 97.85% and PCA-RGB between 88.1% and 97.3%.

3.7 Discussion

This chapter has shown the importance in the choice of color space for representing the eye-fundus images for exudate detection. It has also identified a new color space which gives higher accuracy when used the CNN model for automatic detection of exudates from eye-fundus images.

In this chapter, a nine-layered CNN was developed and trained to distinguish between Exudate and Non-exudate patches. Patch size of 25 × 25 pixels pixel was found to cover the smallest exudate and used for the analysis. This, however, is different from some previous studies who determined the size based on the ratio of the average optic disc (OD) size but failed to detect the small exudates.

This chapter compared different color representations to detect the exudate using the CNN-model: (i) three conventional color spaces (RGB, HSI and LUV) and (ii) PCA transformation of the three spaces (PCA-RGB, PCA-HSI and PCA-LUV). The six color channels of these spaces were then combined based on information rule and then tested using the CNN-model. It was observed that PCA transformation leads to reduction in the performance of HSI and LUV spaces but improvement for RGB space.

The distribution of the pixels in the patches were assumed to be Gaussian and they were labelled as exudate and non-exudate using Gaussian Mixture Models (GMM) (Balafar 2014) (Figure 2.22).

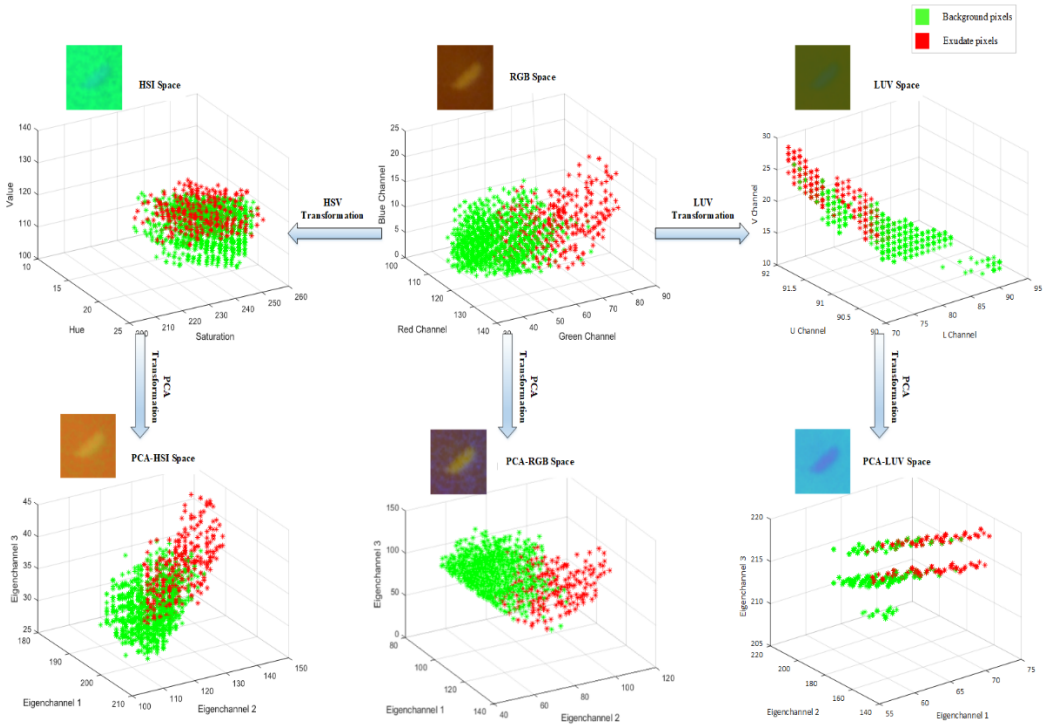


Figure 2.22. Distribution of the exudate and background pixels of an exudate patch sample in different color spaces.

The Gaussian models were presented by (μ_1,Σ_1) and (μ_2,Σ_2) where μ and Σ correspond to mean and covariance of each Gaussian models. The Euclidean distance between the means (D_μ) and covariances (D_Σ) of the two Gaussian models was considered as a factor to measure the overlap between the two models. These two parameters were calculated for 1000 exudate patches which were obtained randomly from the datasets and the difference in these parameters were calculated for each space and shown in Table 2.5.

Table 2.5. Changes of D_μ and D_Σ from a color space to transformed orthogonal space (↑ = increase, ↓ = decrease).

	RGB to PCA-RGB	HSI to PCA-HSI	LUV to PCA-LUV
D_μ	7% ↑	4% ↓	2% ↓
D_Σ	15% ↑	10% ↓	4% ↓

It is seen that PCA transformation increases the overlap between two Gaussian models for HSV and LUV spaces, reducing the classification performance. However, the overlap for the PCA-RGB space was lower.

The SSIM index was used to assess the similarity between color channels and to identify the new color space. The results showed that combination of second eigenchannel of PCA-RGB, Hue and Saturation channels (PHS) achieved higher accuracy with lower variability between repetitions compared to other color spaces. The results have been compared with the previous works reported in the literature as shown in Table 2.6. It can be seen that the CNN-model using "PHS" achieved significantly better accuracy (98.2%) and specificity (0.96) than the recent work by Fraz et al. (Fraz, Jahangir et al. 2017). Our method also achieved significantly better sensitivity (0.99) compared to the work by Jaafar et al. (Jaafar, Nandi et al. 2010) (0.89). Most of the previous works in Table 2.6 used the average metric for testing their method on one dataset. In this work, we created train and test set from both datasets together based on by twenty-fold cross validation technique. We then reported average performance of the test set that is one strength of our work.

This is the first study that has investigated the reliability and repeatability of the proposed space using 20-fold cross validation in 20 runs of the CNN-model. The results showed that the "PHS" space not only outperformed the conventional spaces, the results were consistent over multiple iterations.

Another strength of this work is that it was validated on two databases simultaneously to obtain a reliable assessment of the performance of different spaces.

Another advantage of the proposed algorithm is using a simple CNN architecture designed with approximately 7500 parameters that is significantly less than its network counterparts (i.e. AlexNet, GoogleNet, Residual networks). As a result, this has less computational complexity and high-speed response specially for training phase which is taken just 150 seconds. One limitation of this study is that the proposed space was tested for the application of exudate detection and this was not applied for other applications.

Table 2.6. Comparison the result of proposed method with previous works reported in the literature. The symbol '-' stands for unreported results. DR1 and EO corresponding to DIARETDB1 and e-Ophtha databases, respectively.

Technique	Color Space	Database	ACC	SN	SP	PPV
Proposed Method	PHS	DR1_EO	98.2%	0.99	0.96	0.97
Fraz et al. (Fraz, Jahangir et al. 2017)	RGB	DR1	87%	0.92	0.81	0.90
Walter et al. (Walter, Klein et al. 2002)	LUV	DR1	-	0.76	-	0.59
Jaffar et al. (Jaafar, Nandi et al. 2010)	RGB	DR1	99.0%	0.89	0.99	-
Welfer et al. (Welfer, Scharcanski et al. 2010)	HSI	DR1	-	0.70	0.98	0.92
Harangi et al. (Harangi and Hajdu 2014)	RGB	DR1	82%	0.86	-	0.84
Harangi et al. (Harangi and Hajdu 2014)	RGB	DR1	-	0.73	-	0.69
Zhang et al. (Zhang, Thibault et al. 2014)	HSI	EO	-	0.74	-	-

3.8 Conclusion

This chapter has shown that the choice of color space is an important factor when performing fundus image analysis. It is seen that HSI and PCA-RGB outperformed other spaces. This work has also developed a new color space where the Intensity channel of the HSI space was replaced with the second eigenchannels of the RGB space to propose the new “PHS” space. The results show that the PHS space achieved the most reliable and repeatable results compared to the conventional spaces.

Chapter 4

4 Investigation Different Deep Learning Methods for Exudate Detection

4.1 Overview

This chapter describes the investigation of different DL techniques to maximize the sensitivity and specificity of the methods for automatic detection of the exudates. I have compared multiple DL methods, and both supervised and unsupervised classifiers for improving the performance of automatic exudate detection, i.e., CNNs, pre-trained Residual Networks (ResNet-50) and RBMs. The results show that ResNet-50 with Support Vector Machines outperformed other networks with an accuracy and sensitivity of 98% and 0.99, respectively. This shows that ResNet-50 can be used for the analysis of the fundus images to detect exudates.

4.2 Introduction

Different DL methods have been used for automatic exudate detection. In 2015, Prentašić and Lončarić (Prentašić and Lončarić 2015) developed a ten-layered CNN to detect exudates. However, it had low sensitivity (0.77). Perdomo et al. (Perdomo, Otalora et al. 2016) used a different CNN architecture and achieved a sensitivity of around 0.90 but the specificity was poor (0.40). While Yu et al. (Yu, Xiao et al. 2017) and Prentašić et al. (Prentašić and Lončarić 2016) achieved a reasonable sensitivity (0.88), their method required manual pre-processing steps for optic disk removal and vessel segmentation. Tan et al. (Tan, Fujita et al. 2017) overcame the shortcomings of manual feature extraction and pre-processing, and their sensitivity was 0.87.

CNN-based methods do not require hand-crafted feature extraction, but it is necessary to identify suitable architectures and optimal parameters. As there is no exact method for selection of the network parameters such as the number of convolutional layers and filter sizes, these have been determined heuristically (He, Zhang et al. 2016). The size of the datasets is another limitation of using such networks because the full training of these requires large volumes of data. For overcoming these limitations, there is a need to investigate different DL methods and identify a network that gives good performance and is not reliant on large datasets.

The aim of this chapter is to identify a suitable DL method for good a performance to overcome the limitations mentioned above. We have compared the performance of multiple DL methods, and both, supervised and unsupervised classifiers. The CNN model, pre-trained residual networks (ResNet-50) with a supervised classifier and RBMs were investigated. The sensitivity, specificity and accuracy of the proposed methods were compared with previous works using two publicly available databases: (i) DIARETDB1 and (ii) e-Ophtha. The novelty of this chapter is that it has compared the performance of different DL methods for detection of exudates in the fundus images.

The remainder of this chapter is organized as follows. Section 4.3 presents the theoretical background of the techniques used in this chapter. Section 4.4 describes the experiments conducted to compare the different deep-learning methods adopted for exudate detection. The results are presented in section 4.5 and discussed in section 4.6 followed by conclusions and future works in section 4.7.

4.3 Theoretical background

This chapter has considered three DL techniques (i) CNN, (ii) pre-trained residual network with a classifier, and (iii) DRBM-based model. This section presents a brief description of these approaches.

4.3.1 Convolutional Neural Networks

The network parameters, i.e., connection weights and biases, are θ, and θ^t is the updated θ at time step t. Parameters η, α and β denote the learning rate, momentum, and weight decay, respectively. Figure 2.23 shows the hierarchical architecture of the proposed network where *I, C, FM, MP, NM, FC* denote the input image, convolutional layer, feature map, max pooling, normalization layer, and fully connected layer, respectively. SGD was used to update network parameters.

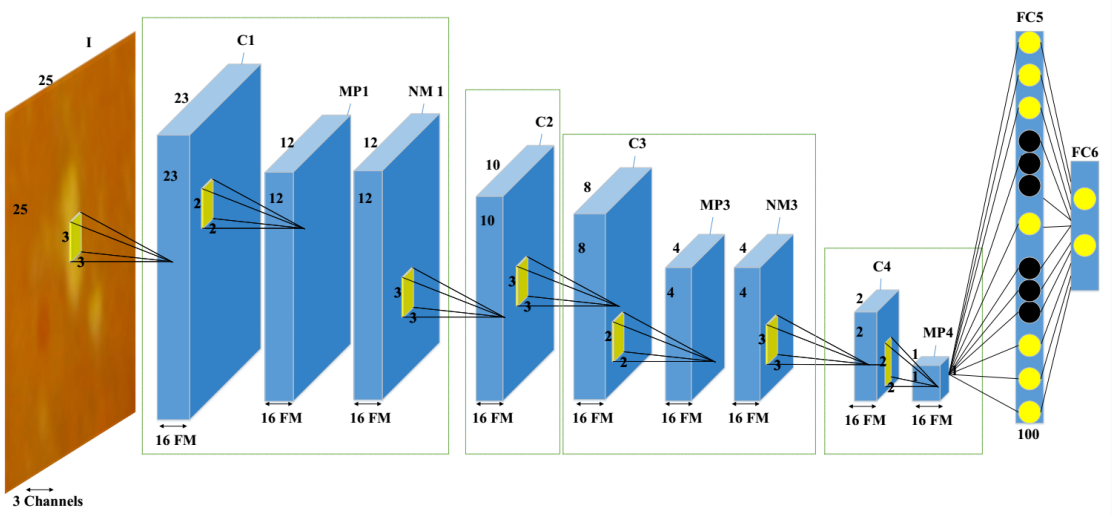


Figure 2.23. Architecture of the proposed CNN used in this work.

4.3.2 Deep Residual Networks

In this section, Residual Network (ResNet-50) (He, Zhang et al. 2016) which was pre-trained using ImageNet dataset (Russakovsky, Deng et al. 2015) has been used but the associated fully-connected softmax layer was replaced with supervised classifiers. Three classifiers were tested and have been described below.

The convolutional layers had filters of sizes 1×1, 3×3 and 1×1. The overall architecture followed two design rules: (i) for the same output size, the layers should have the same number of filters and (ii) if the feature map size is halved, the number of filters should be doubled. Max-pooling layers were performed directly by convolutional layers with the

stride of 2 followed by batch normalization after each convolution operation and before ReLU (Nair and Hinton 2010) activation function.

The rules for the residual block shortcuts were based on the relative dimension of input and output. When the input and output had the same dimension, an identity shortcut was applied. When the difference in the dimension of the input and output increased, the shortcuts, there were two options: (i) extra zero entries were padded for increasing dimensions (notice that this option introduces no extra parameter), and (ii) the projection shortcut was used to match dimensions (performed by 1x1 convolutions). For both options, when the shortcuts go through the feature maps of two sizes, they are performed with a stride of 2.

The ResNet-50 model was pre-trained for the object detection task on the ImageNet 2012 dataset with 1.28 million images from 1,000 classes. We then used the first 49 layers of ResNet-50 as feature extractors to generate deeply-learnable features which were then fed to the classifier. The weights learned from the first step were kept the same, and we added an extra layer composed of different supervised classifiers which are described in the next section. Three different classifiers were considered: Support Vector Machine (SVM), Optimum-Path Forest (OPF), and *k*-Nearest Neighbours (KNN). These have been discussed below.

Support Vector Machines

SVM is a non-probabilistic, binary, linear classifier based on statistical learning theory. It uses kernels for transforming the feature space such that after transformation the samples from two different classes are linearly separable. It uses soft margins to maximize the inter-class distance.

Optimum-Path Forest

Optimum-Path Forest (OPF) (Papa, Falcao et al. 2009, Papa, FalcãO et al. 2012, Papa, Fernandes et al. 2017) is a graph-based supervised approach whose training samples are represented as nodes in a graph, and the distance between them corresponds to the weight of the two nodes. The training step consists of finding the most representative samples (i.e.,

the prototypes) for each class, thus obtaining the optimal partition (according to some path-cost function) of the graph representing this training set. The classification of a test sample is achieved by computing its optimum-paths to the whole training graph and assigning to it the label of the root (i.e., prototype) whose path has strongest connections.

k-Nearest Neighbours

In the k-Nearest Neighbours (*k*-NN), data is classified based on the majority votes of its neighbours (Altman 1992). The object is assigned to the class that is most frequent in its *k*-sized neighbourhood. In the simplest form, if *k*=1, then the object is assigned to the class of its nearest neighbour. In this chapter, we have considered different values of *k*.

4.3.3 Discriminative Restricted Boltzmann Machines

In this section, a modified RBM (named Discriminative Restricted Boltzmann Machine (DRBM)) which includes classification is used (Larochelle and Bengio 2008). This comprises one additional input layer, i.e., the label layer, which contains the label of each input sample using one-hot encoding. This additional layer is connected to both, the hidden and visible layers, but with a different set of weights.

4.4 Material and Methodology

In this chapter, two publicly available databases were used: (i) DIARETDB1 and (ii) e-Ophtha (described in Chapter 3) and preparation of the dataset is described as follow.

4.4.1 Data Preparation

The size of the exudates varies, and thus the patch size required to box these can be very different. Figure 2.24 shows a variation of patch sizes corresponding to all extracted patches, where the *X* and *Y* axes correspond to the length and width of a patch. It shows that after ignoring the outliers, the range of exudate patch varies from (25 × 25) to (286 × 487). To

have one size of the patch for the analysis of the image, this was selected to be the smallest patch corresponding to the smallest pathological sign identified by the examiners. This is similar to the work reported by Shan and Li (Shan and Li 2016) and Cao et al. (Cao, Shan et al. 2017) on the same database (DIARETDB1).

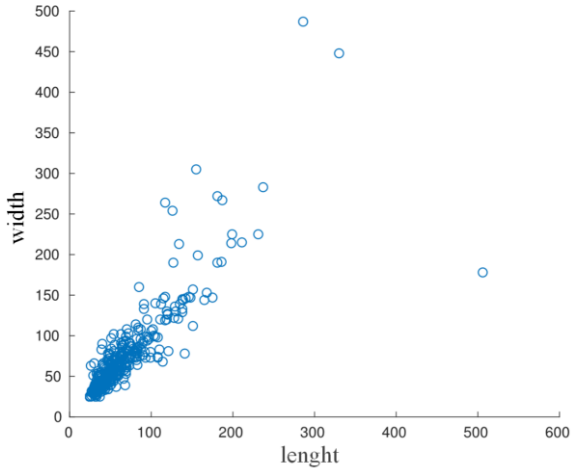


Figure 2.24. Samples of exudate patches

In this chapter, we used patches of size 25 × 25 with three color channel (i.e. Red, Green, Blue (RGB)) which were labelled in two groups: (i) Exudate and (ii) Non-exudate. As a result, 67,600 and 23,200 exudate patches were manually extracted from the DIARETDB1 and e-Ophtha databases, respectively. To obtain a balanced dataset, 70,000 and 25,000 non-exudate patches were extracted from the regions without exudate from the two databases. Non-exudate patch group contained vessels, background tissues and optic nerve heads. All patches were extracted without any overlap.

4.4.2 Methodology

Figure 2.25 shows the flowchart of the tasks. Pre-labelled exudate and non-exudate patches were used for training and testing the three deep-learning methods: (i) CNN, (ii) ResNet-50 + classifier (three different types), and (iii) DRBMs. In the ResNet-50, the softmax layer of the original architecture was replaced with three different classifiers: OPF, SVM, and *k*-NN. Ten-fold cross-validation with ten runs technique was used to evaluate each method.

.

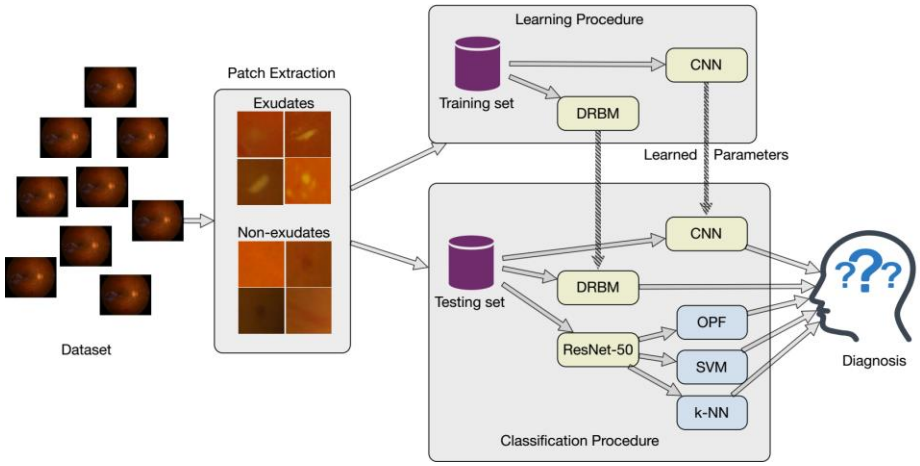


Figure 2.25. Flowchart for detecting exudate and non-exudate patches.

4.4.2.1 Network Design

In this section, the network setup and corresponding results are presented

Network Setup

CNN Model. The CNN parameters were obtained empirically as follows: learning rate = 0.01, momentum = 0.9, and the variance corresponding to the Gaussian filters for the convolutional layers was set to 0.01. Optimization was performed using stochastic gradient descent with the step-down policy and step size of 33. The batch sizes for training, validation, and test phases were 128, 32 and 32, respectively.

The CNN was implemented on Caffe platform (Jia, Shelhamer et al. 2014) with a Geforce GTX 1070 graphics card. The optimum number of epochs was identified after evaluating the network during the training procedure for 40 epochs. Figure 2.26 shows the convergence curve during training using the DIARETDB1 database.

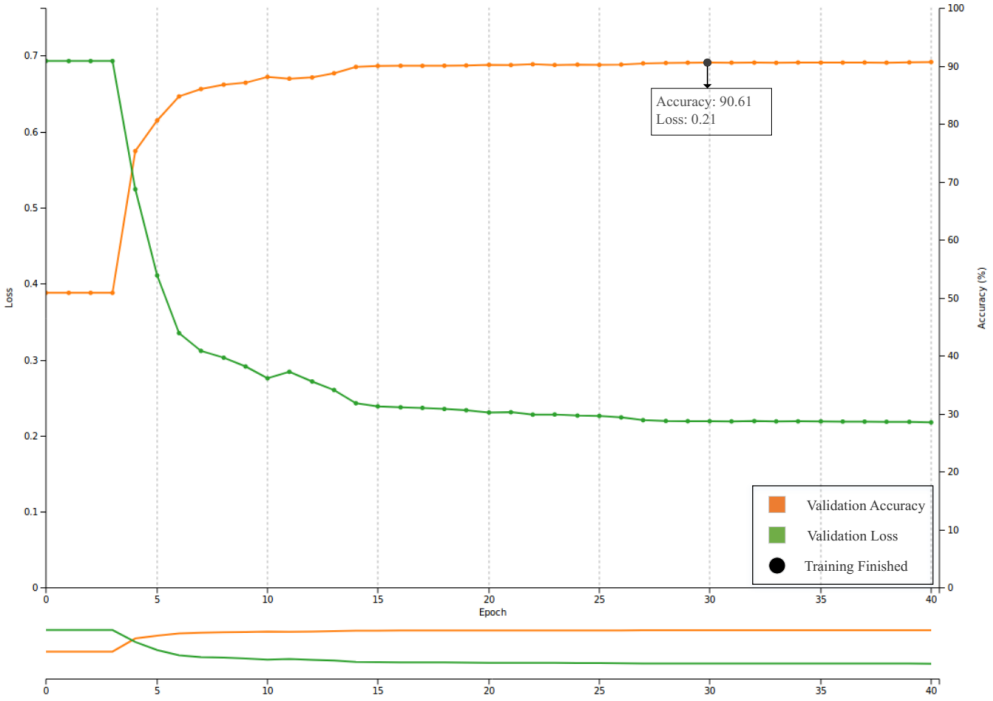


Figure 2.26. Accuracy and loss changes over 40 training epochs concerning the DIARETDB1 database. A saturation is observed after the 30th epoch.

From Figure 2.26, it is observed that there was a significant improvement in the accuracy and reduction in loss after the 5^{th} epoch. The accuracy saturated around the 30^{th} epoch, where its value reached 90.6%. Thus, the number of epochs was fixed to 30 for training and corresponding weights were used for the test phase.

Deep Residual Networks. The patches corresponding to exudate and non-exudate groups were resized to 224 × 224 to match the input image size required by Resnet-50. Feature reduction was applied to the 2048 dimensional output feature vector of the last convolution layer using Principal Component Analysis (PCA) (Sarkar, Saha et al. 2014) for dimensionality reduction and 100 dimensional features were obtained. Three classifiers were investigated in this study: OPF, SVM, and *k*-NN. The classifier parameters were optimized using a grid-search approach. The results obtained from each classifier were compared with the ones obtained by Resnet-50 with softmax layer.

DRBM. To apply an image to DRBM model, the intensity value of the pixels in the RGB space (i.e., pixels from each patch) were normalized to the range of [0,1]. The three channels of each patch were concatenated to one vector which was the input for the DRBM model. Contrastive Divergence training algorithm using single sampling for each iteration was used and trained with mini-batches of size 100 and 200 epochs. The meta-parameters such as learning rate and the number of hidden units were optimized using a grid-search as shown in Figure 2.27. It was observed that the higher accuracies (values in dark red) were located in the central region of the heat map and thus the number of hidden neurons was selected to be in the range [3,000- 4,000] and learning rate in the range [0.1- 0.17].

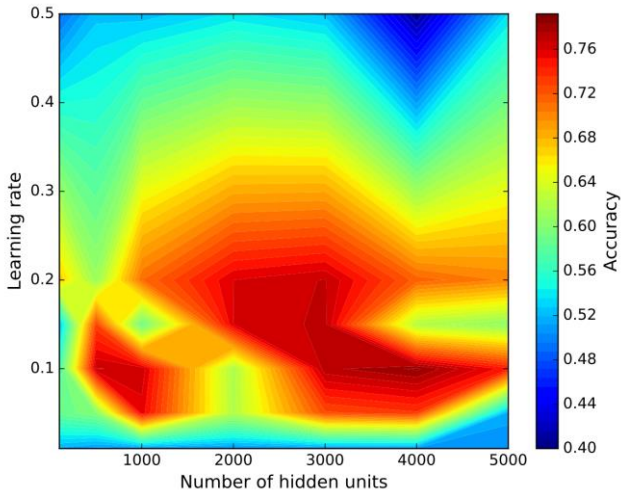


Figure 2.27. Heat map obtained from the grid-search showing the best combination of learning rate and the number of hidden units.

4.5 Results

The performance of the different methods was evaluated using the measures described by Loong (Loong): overall accuracy (ACC), sensitivity (SN), and specificity (SP). The mean results obtained from 10-fold cross-validation approach with 10 runs corresponding to DIARETDB1 and e-Ophtha databases are presented in Table 2.7 and Table 2.8, respectively.

Table 2.7. Performance of the proposed methods for DIARETDB1 database: overall accuracy, sensitivity, and specificity.

Technique	ACC	SN	SP
CNN	90.6%	0.90	0.91
ResNet-50+KNN	97.1%	0.98	0.95
ResNet-50+OPF	95.7%	0.93	0.99
ResNet-50+SVM	98.2%	0.99	0.96
DRBM	76.37%	0.70	0.78

Table 2.8. Performance of the proposed methods for e-Ophtha database: overall accuracy, sensitivity, and specificity.

Technique	ACC	SN	SP
CNN	89.1%	0.89	0.91
ResNet-50+KNN	96.0%	0.98	0.91
ResNet-50+OPF	94.7%	0.90	0.99
ResNet-50+SVM	97.6%	0.98	0.95
DRBM	70.37%	0.70	0.78

From Table 2.7 and Table 2.8, it is seen that the performances for both the databases were similar. For DIARETDB1 database, “Resnet-50 + SVM” achieved the best sensitivity and accuracy of 0.99 and 98.2%, respectively. “Resnet-50 + OPF” obtained the highest specificity (0.99) compared to “Resnet-50 + SVM”, “Resnet-50 + KNN” and CNN model with the specificity of 0.96, 0.95 and 0.91, respectively. The “Resnet-50 + SVM” model also performed the best for e-Ophtha database. It is observed that the Residual Networks outperformed the CNN and DRBM models.

Since the best results were achieved through “Resnet-50 + SVM”, this was used to compare with works reported in the literature (Table 2.9). For the DIARETDB1 database, the proposed approach (i.e., “Resnet-50 + SVM”) achieved significantly better sensitivity compared to worked by Jaafar et al. (Jaafar, Nandi et al. 2010) (0.99 vs 0.89), while both

methods obtained similar accuracies. It also achieved significantly better accuracy (98.2%) and specificity values (0.96) compared with Fraz et al. (Fraz, Jahangir et al. 2017). For the e-Ophtha database, the proposed method outperformed the recent work by Mo et al. (Mo, Zhang et al. 2018) with 0.18 improvement in sensitivity, but there was 0.04 decrease in the specificity.

Table 2.9. Comparison of the proposed approach against other works reported in the literature. The symbol `-' stands for unreported results.

Database	Technique	ACC	SN	SP
DIARETDB1	Proposed Method	98.2%	0.99	0.96
	Fraz et al. (Fraz, Jahangir et al. 2017)	87%	0.92	0.81
	Jaafer et al. (Jaafar, Nandi et al. 2010)	99.0%	0.89	0.99
	Welfer et al. (Welfer, Scharcanski et al. 2010)	-	0.70	0.98
	Walter et al. (Walter, Klein et al. 2002)	-	0.76	-
	Harangi et al. (Harangi and Hajdu 2014)	82%	0.86	-
	Harangi et al. (Harangi and Hajdu 2014)	-	0.73	-
e-Ophtha	Proposed Method	97.6%	0.98	0.95
	Mo et al. (Mo, Zhang et al. 2018)	-	0.92	-
	Das et al. (Das and Puhan 2017)	-	0.85	-
	Imani et al. (Imani and Pourreza 2016)	-	0.80	0.99
	Liu et al. (Liu, Zou et al. 2017)	-	0.76	-

4.6 Discussion

This chapter has investigated three DL techniques for the detection of exudates using fundus images. The results show that there is a significant difference in the results (i.e., accuracy, sensitivity, and specificity). The accuracy obtained using "Resnet-50 + SVM" (i.e. 98%) was the highest among all methods with sensitivity and specificity values as of 0.99 and 0.96, respectively. The performance achieved by "Resnet-50 + SVM" was significantly better than the other methods considered in this study, as well as those reported in the literature. However, there is potential for further improvements with computationally efficient classifiers, and the specificity may be improved using approaches such as Twin SVM.

The CNN-based model requires investigation on finding suitable parameters, which is one of the challenging tasks when using such models. Learning rate is important for the network convergence and hence the effect of different learning rates, ranging from 0.0001 to 0.2, was investigated while momentum and variance were fixed to 0.01. Figure 2.28 shows the accuracy of the network corresponding to different learning rates, where the best accuracy (90.6%) corresponded to learning rate of 0.01, which was similar to that reported in Alexnet experiments (Krizhevsky, Sutskever et al. 2012). It was also found that learning rates less than 0.001 and larger than 0.1 resulted in the network not converging and were not evaluated further. Because the proposed CNN architecture was a modified version of the Alexnet, other parameters (i.e., batch size, momentum, and variance) were set as the same as of the Alexnet model.

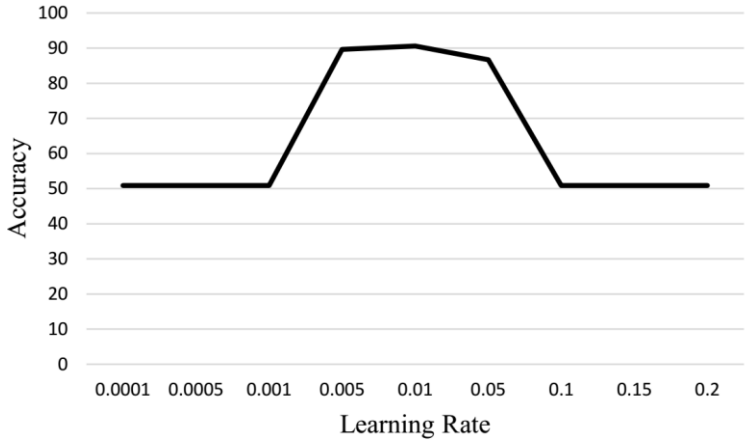


Figure 2.28. Change of accuracy of the network corresponding to different learning rates.

The results show that DRBMs achieved an accuracy of only 73%. This poor performance can be attributed to a number of reasons such as preparation of data, and the shallow architecture. Normalized vectors used as the input of the DRBM were obtained by the concatenation of the three color channels (i.e., red, green, and blue) to obtain a single vector. This may not be the best method for representation of an image as an input of the DRBM.

In this chapter, the DRBMs were used without additional layers such as used for Deep Belief Networks and Deep Boltzmann Machines. Thus, the tested network was comparatively shallow, while the deeper networks would be capable of extracting more detailed and intrinsic information from the data.

This chapter has shown that the selection of the parameters for CNN and DRBM models is challenging because there are no direct rules. Additionally, a limited number of samples for training purposes might have an impact on the performance of the network. One advantage of using “Resnet-50” is that it does not require a large training dataset.

Another point that deserves attention concerns the OPF classifier. Although it did not outperform SVM relating to the overall accuracy, it obtained the best specificity (i.e. 0.99). Additionally, OPF is parameterless and much faster for training should be explored in the future.

The results obtained in this chapter show that pre-trained deep-learning methods have the potential for detecting exudates in retinal images. This can significantly help the experts to

obtain better, faster and more accurate estimation of exudate spots on the retina. The choice of the network parameters has been found to be important for obtaining good results and our experiments suggest the use of “Resnet-50 + SVM”.

4.7 Conclusion

The novelty of this chapter is that it has compared different deep-learning approaches for automatic detection of exudate based on the experimentally obtained accuracy, sensitivity, and specificity. It has shown that “Resnet-50 + SVM” is the best among these for automatic detection of exudates in the fundus images, especially the sensitivity, which is important for medical diagnostics. This method also has the advantage of incorporating pre-trained feature extraction layers which does not require large datasets for training.

This chapter has shown that the outcome of deep-learning approach is independent on the choice of the parameters. The next steps are to investigate other deep neural networks such as the Generative Adversarial Networks. It is also important to determine the most appropriate method for presenting the data to the networks. Data augmentation through synthetic image generation and DBMs could be considered to build generative models that can better aggregate different color channels than their concatenation.

Chapter 5

5 A Novel Pre-processing Layer in Convolutional Neural Network for Automatic Identification of Diabetic Retinopathy

5.1 Overview

This chapter has proposed a novel framework for the convolutional neural network architecture by embedding a pre-processing layer followed by the first convolutional layer to increase the performance of the network. Two image enhancement techniques i.e. 1- Contrast Enhancement 2- Contrast-limited adaptive histogram equalization were separately embedded in the proposed layer and the results were compared. For identification the DR signs (exudates, hemorrhages and microaneurysms), the proposed framework achieved the total accuracy of 87.6%, and 83.9% for the contrast enhancement and contrast-limited adaptive histogram equalization layers, respectively. However, the total accuracy of the convolutional neural network alone without the pre-processing layer was found to be 81.4%. Consequently, the new convolutional neural network architecture with the proposed pre-processing layer improved the performance of convolutional neural network.

5.2 Introduction

Severity of DR is currently estimated by the expert's examination of fundus images based on the amount and location of three retinopathy signs (i.e. exudates, hemorrhages and microaneurysms) across the retina surface (Hansen, Abramoff et al. 2015). An automatic and accurate system for detection of the DR signs can significantly help clinicians make the best possible prognosis about the severity of DR and risk of vision loss. For this purpose,

DL and other machine learning methods have been widely employed (Hajeb Mohammad Alipour, Rabbani et al. 2012, Lazar and Hajdu 2013, Tang, Niemeijer et al. 2013).

To increase the performance DR detection algorithms, different image pre-processing techniques have been employed mainly to enhance the contrast of retina images (van Grinsven, van Ginneken et al. 2016, Tan, Fujita et al. 2017, Kaur and Mittal 2018). Hajeb et al. (Hajeb Mohammad Alipour, Rabbani et al. 2012) used Contrast-limited adaptive histogram equalization (CLAHE) as the image enhancement technique. To detect the DR signs, they employed curvelet transform to extract features from color fundus images and fed them into a support vector machine. Agurto et al. (Agurto, Murray et al. 2010) applied multi-scale amplitude-modulation-frequency-modulation (AM-FM) on color fundus images to extract texture features to distinguish between DR and Normal signs. In 2016, Grinsven *et al.* (van Grinsven, van Ginneken et al. 2016) proposed a nine-layer CNN only for hemorrhage detection. As the pre-processing technique, they applied the contrast enhancement (CE) method on original color fundus images. For exudates detection, they achieved areas under the receiver operating characteristic (ROC) curve of 0.97. In 2017, Tan *et al.* (Tan, Fujita et al. 2017) used a CNN for automatic segmentation of exudates, hemorrhages and microaneurysms. For the image enhancement, the original retina images were first converted from RGB to LUV color space and then the L channel was corrected for local contrast and uneven illumination.

For automatic DR sign detection, image enhancement techniques were widely used to increase the performance of conventional machine learning and image processing techniques. However, no study has embedded that in CNN architecture as a separate layer for enhancing the input images. In this chapter, we present a novel framework and architecture for CNN by embedding a pre-processing layer (PPL) for automatic identification of exudates, hemorrhages and microaneurysms using color fundus images. The novelty of this work is the application of a pre-processing layer in the CNN architecture for the first time for detection of DR.

5.3 Materials

DIARETDB1 dataset was used as a benchmark for detection of DR signs (Kauppi, Kalesnykiene et al. 2007). The annotation of DR signs (exudates, hemorrhages and microaneurysms) were manually performed and evaluated by four independent medical experts. In this chapter hard and soft exudates were grouped together and considered as a single class.

5.4 Methodology

From the original retina images, four patch groups of size 25×25 pixels corresponding to exudates, hemorrhages and microaneurysms and backgrounds were manually extracted. For the classification of these four groups, a CNN architecture was proposed by designing a PPL before the first convolutional layer. As a pre-processing method, CLAHE and CE were individually implemented in the PPL. The CNN architecture and pre-processing techniques are discussed in the following sections.

5.4.1 Convolutional Neural Network Architecture

A new CNN architecture with 11 layers was developed. We proposed PPL as the first layer followed by first convolutional layer in the network. In the proposed architecture, four convolutional layers were designed with 16 kernels of size 3×3 pixels. Three max pooling layers with kernel size 2×2 pixels were utilized after the first, the third and the fourth convolutional layers. Three normalization layers were applied after each max pooling layer. From the last max pooling layer, 16 features were extracted and fed into 100 neurons of the fully connected layer which is shown in Figure 5.1. The backpropagation algorithm with stochastic gradient descent (SGD) was used for the training process (Pang, Yu et al. 2017).

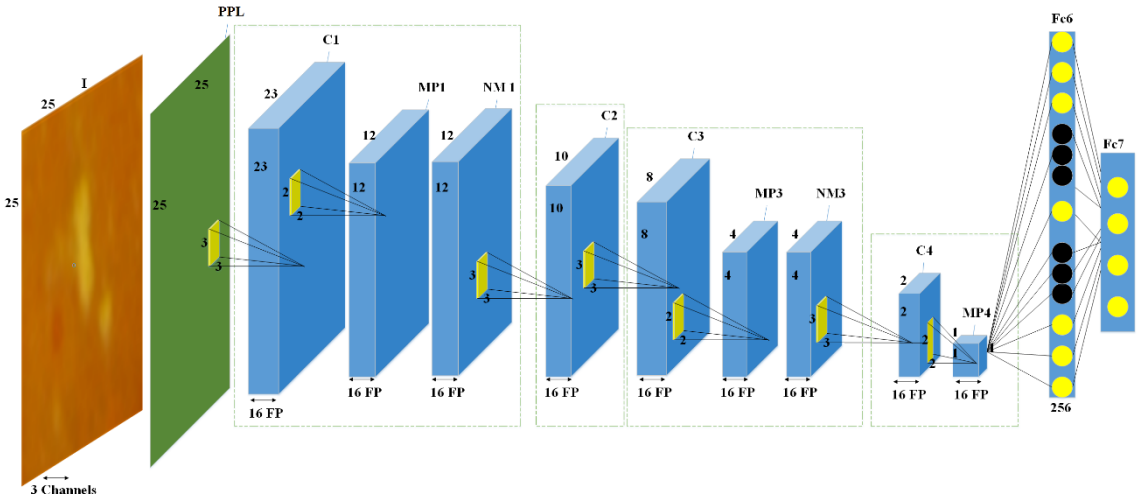


Figure 5.1. Architecture of the proposed CNN- I: input image, PPL: pre-processing layer C: convolutional layer, FP: feature map, MP: max pooling, FC: fully-connected layer.

5.4.2 Image Enhancement Techniques

As CLAHE and CE methods showed a significant improvement in contrast of the retinal images (Hajeb Mohammad Alipour, Rabbani et al. 2012, van Grinsven, van Ginneken et al. 2016), we individually implemented these two techniques in the PLL. With respect to CLAHE method, pixel values of a small region of the input image are transformed to a particular distribution, in this chapter Rayleigh distribution was used (Hajeb Mohammad Alipour, Rabbani et al. 2012). This is then followed by combining the neighboring region using bilinear interpolation to avoid introduction of artificially induced boundaries. The CLAHE technique was applied to the neighbourhood size of 25 × 25 pixels with the bottom histogram clipping limit set to 0.8 %.

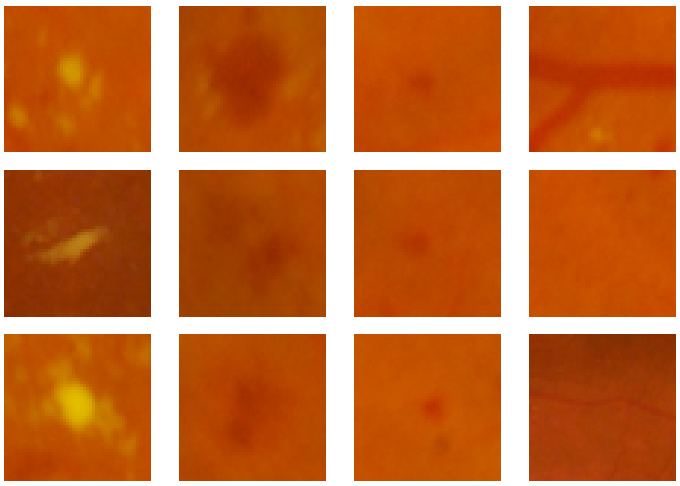
By applying the CE method on the original image $I(x,y)$, I_{CE} is obtained as follows (van Grinsven, van Ginneken et al. 2016):

$$I_{CE} = \alpha I(x,y) + \beta G(x,y;\sigma) * I(x,y) + \mu \qquad (19)$$

where * represents the convolution operator. $G(x, y; \sigma)$ is a Gaussian filter with the scale σ. Parameter values of the α, β, σ and μ were empirically chosen as 4, -4, $300/30$ and 128, respectively. In this method, the local average color with β coefficient was subtracted from the original image (multiplied by a coefficent) and the local average was mapped to 50% gray scale level (μ).

5.5 Experiment

From the 89 retinal images, a number of patches corresponding to four DR classes (exudate, hemorrhage, microaneurysm and background) were manually extracted which are shown in Figure 5.2.



(a) (b) (c) (d)

Figure 5.2. Patch examples corresponding to four classes: (a) exudate; (b) hemorrhage; (c) microaneurysm; (d) background.

A patches was centered by the target class of each group. The background patches were randomly extracted from the retinal regions that did not have any DR sign. The patch size was found empirically as the optimum size covering DR signs. To achieve reliable results, there was no overlap between patches of the test set. Table 5.1 summarizes the number of

patches corresponding exudate (EX), hemorrhage (HM), microaneurysm (MA) and background (BK).

Table 5.1. Statistics of the patches- exudate (EX), hemorrhage (HM), microaneurysm (MA) and background (BK).

	EX	HM	MA	BK
Training	47320	30643	12723	35013
Validation	10140	6566	2726	7503
Test	10140	6567	2727	7502
Total Number	67600	43776	18176	50018

For training the CNN, the parameters were empirically set as follows:
The learning rate and the momentum were set to 0.01 and 0.9 respectively. The weight of Gaussian filters for all convolutional layers was set to 0.01. The SGD method was used as the solver with step-down policy and with a step size of 33. The Gamma value was set to 0.1 for each iteration. The training batch size was fixed to 128 images and for both validation and test processes it was set to 32.

5.6 Results and Discussion

In this section, performances of three CNN frameworks were assessed: 1- without PPL (No-PPL), 2- with PPL implemented by CLAHE and 3- with PPL implemented by CE. The training process of three CNNs during 100 epochs is shown in Figure 5.3.

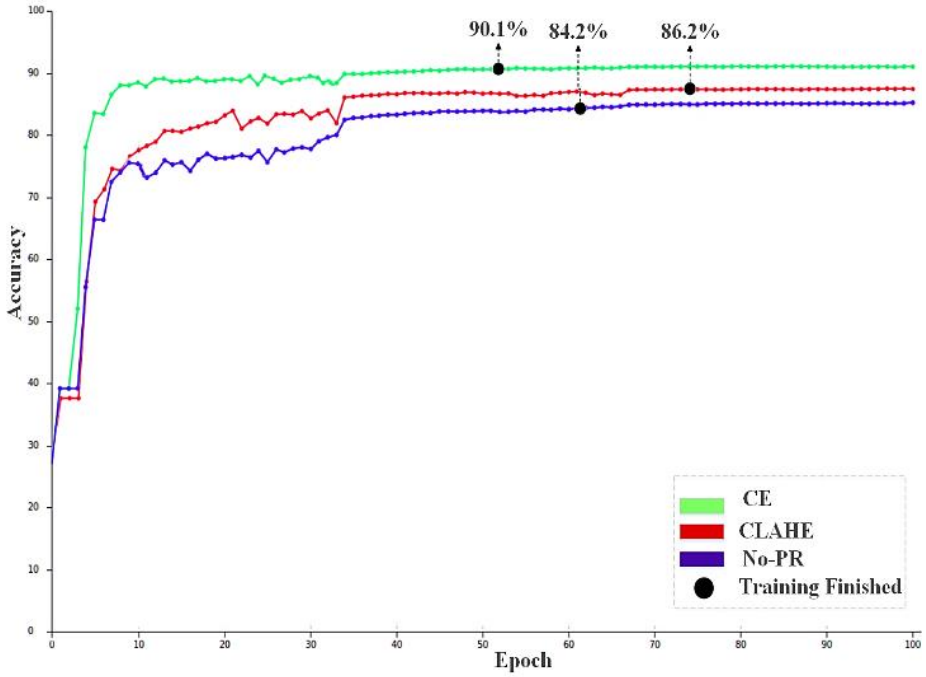


Figure 5.3. The training process for three designed CNNs. After 53, 65 and 74 epochs, the training process stopped for CE, CLAHE and No-PPL, respectively.

The accuracy associated with validation set was calculated by mean of four pre-class accuracies. To avoid overfitting, the training process stopped after epoch 53, 65 and 74 for CE, CLAHE and No PLL, respectively. The CNN with CE, CLAHE and No PLL achieved an accuracy of 90.1%, 86.2% and 84.2%, respectively. Additionally, the training process approximately took eight minutes on the Geforce GTX 1070.The Caffe platform was used for implementation of the algorithm (Jia, Shelhamer et al. 2014).

The test set was used for evaluating the performance of the three CNNs. For classification of exudates (EX), hemorrhages (HM) and microaneurysms (MA) and backgrounds (BK), three parameters including accuracy (ACC), sensitivity (SN) and specificity (SP) (Dice 1945) were calculated to evaluate the CNN's performances. To obtain reliable results, the experiments were repeated twenty times by random selection of training, validation and testing sets and the mean of the parameters are shown in Table 5.2.

Table 5.2. Performance of three proposed CNNs- exudate (EX), hemorrhage (HM), microaneurysm (MA) and background (BK).

CNN design		EX	HM	MA	BK
No-PPL	SN	0.90	0.75	0.79	0.88
	SP	0.93	0.93	0.96	0.95
	ACC	89.9%	81.3%	64.9%	89.5%
PPL with CLAHE	SN	0.90	0.77	0.82	0.94
	SP	0.94	0.94	0.96	0.97
	ACC	90.2%	83.3%	67.8%	94.4%
PPL with CE	SN	0.91	0.80	0.84	0.98
	SP	0.94	0.94	0.97	0.99
	ACC	91%	82.5%	78.8%	98.4%

The proposed CNN without PPL achieved sensitivity of 0.79 for the microaneurysm (MA) detection while with CLAHE and CE it obtained sensitivity of 0.82 and 0.84, respectively. Almost the similar increasing trend achieved for hemorrhage (HM) detection by using CLAHE and CE. The accuracy of background (BK) detection after applying CE and CLAHE increased by 9% and 5%, respectively, compared to the No-PPL architecture. Mean of the four pre-class accuracies was calculated as the total accuracy for three CNN architectures and is shown in Figure 5.4.

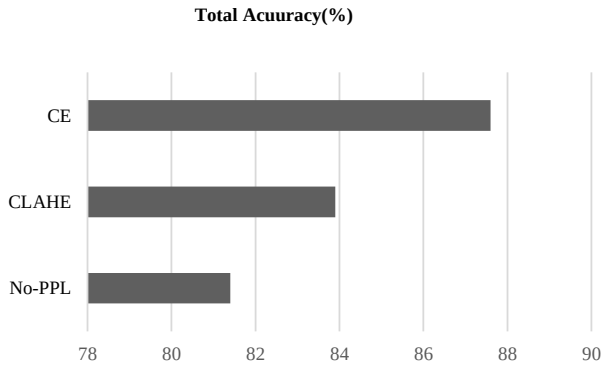


Figure 5.4. Total accuracy of the classification corresponding to three CNNs.

The CNN with No-PPL achieved the total accuracy of 85% which was increased to 87% and 90% by using CLAHE and CE design, respectively.

According to the Table 5.2 and Figure 5.4, it can be seen that the CNN's performance for classification of DR signs was increased by the proposed architecture, especially for identification of MA and HM.

5.7 Conclusion

In this chapter, CNN-based techniques have been developed for identifying DR using color fundus images. In this chapter, a novel CNN architecture was proposed by embedding a PPL as the first layer of the network for enhancing input images. Three CNN architectures were assessed in this work: 1- without PPL 2- with the PPL implemented by CLAHE and 3- with PPL implemented by CE. Three major DR signs including EX, HM and MA were tested. The study found that the proposed PPL effectively improved the performance of the system for detection of DR signs compared to the conventional CNN without the PPL. In comparison to the CLAHE, the CE was found more suitable for detection of DR signs.

Chapter 6

6 Analysis of Deep Probabilistic Features for Detection of Exudates, Hemorrhages and Microaneurysms

6.1 Overview

This chapter has proposed a method using probabilistic output from a convolution neural network to automatically and simultaneously detect exudates, hemorrhages and microaneurysms. The method was evaluated using two approaches: patch and image-based analysis of the fundus images on two public databases: DIARETDB1 and e-Ophtha. The novelty of the proposed method is that the images were analyzed using probability maps generated by score values of the softmax layer instead of the use of the binary output. The sensitivity of the proposed approach was 0.96, 0.84 and 0.85 for detection of exudates, hemorrhages and microaneurysms, respectively when considering patch-based analysis. The results show overall accuracy for DIARETDB1 was 97.3 % and 86.6 % for e-Ophtha. The error rate for image-based analysis was also significantly reduced when compared with other works. This also obtained accuracy and sensitivity which were significantly better than the reported studies and makes it suitable for automatic diabetic retinopathy signs detection.

6.2 Introduction

DR is diagnosed by visual examination of retinal images to detect three most common pathological signs i.e. (i) exudate (ii) hemorrhage and (iii) microaneurysm (Hansen, Abramoff et al. 2015). However, this is a manual time-consuming procedure and outcomes are subjective and dependent on expertise, thus, there is potential bias of the examiner. The

diagnosis can be performed by analysis of color fundus images or fluorescein angiograms (FA) to identify pathological signs. Although FA enables better differentiation between microaneurisms and micro hemorrhages, due to its invasive nature along with costs and the risk of allergic reactions, fundus images are the preferred modality. For automatic detection of pathological signs, most computer-based studies have developed algorithms for the automatic analysis of the fundus images with the aim to make the diagnosis more objective and easier to access by people in remote communities. However, this is very challenging because of variation in size, color, texture and shape of these signs (Figure 5.5).

In computer-based methods, detection of exudate, hemorrhage and microaneurysm can either be done separately for each sign (Walter, Klein et al. 2002, Osareh, Mirmehdi et al. 2003, Niemeijer, van Ginneken et al. 2005, Fleming, Philip et al. 2006, Walter, Massin et al. 2007, Quellec, Lamard et al. 2008, Sopharak, Uyyanonvara et al. 2008, Welfer, Scharcanski et al. 2010, Bae, Kim et al. 2011, Giancardo, Meriaudeau et al. 2012, Lazar and Hajdu 2013, Tang, Niemeijer et al. 2013, Akram, Tariq et al. 2014, Grinsven, Ginneken et al. 2016, Shan and Li 2016, van Grinsven, Venhuizen et al. 2016, Fraz, Jahangir et al. 2017) or all signs simultaneously (Gardner, Keating et al. 1996, Sinthanayothin, Boyce et al. 2002, Agurto, Murray et al. 2010, Garcia, Lopez et al. 2010, Acharya, Ng et al. 2012, Roychowdhury, Koozekanani et al. 2014, Imani, Pourreza et al. 2015, Tan, Fujita et al. 2017). For exudate detection, Sánchez et al. (Sánchez, García et al. 2009) used a statistical mixture model-based clustering for dynamic thresholding to separate exudate from background. The algorithm obtained sensitivity of 90.2 % and 96.8 % for lesion and background, respectively. Giancardo et al. (Giancardo, Meriaudeau et al. 2012) proposed a method based on color and wavelet decomposition features from exudate candidates to train classifiers. They achieved the best result using support vector machine (SVM) classifier with areas under the receiver operating characteristics (AUC) between 0.88 and 0.94, depending on different datasets. In 2017, Fraz et al. (Fraz, Jahangir et al. 2017) developed a method to detect exudate based on the multiscale segmentation. They used combination of morphological reconstructions and Gabor filter banks for feature extraction followed by bootstrap decision tree for classification of exudate pixels. In 2018, Kaur and Mittal (Kaur and Mittal 2018) used a dynamic thresholding method for detection of exudate boundaries.

The algorithm obtained sensitivity of 88.85 % and 94.62 % in lesion-based and image-based, respectively.

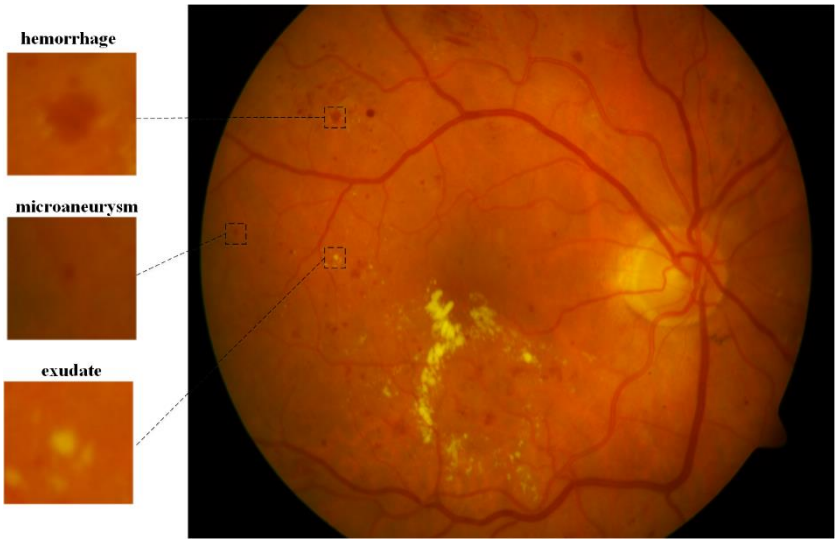


Figure 5.5. Example of Retina Images containing three DR sings. This image shows an entire retina image with haemorrhage, microaneurysm and exudate labelled by graders, and which was then cropped to illustrate individual patch.

Hemorrhage detection was reported by Tang et al. (Tang, Niemeijer et al. 2013) who divided the image into small sub-images (also called splats) for extracting splat features such as texture, splat area, and color. They evaluated their method based on patch and image level analysis and obtained AUC 0.96 and 0.87, respectively. For automatic detection of microaneurysm, Walter et al. (Walter, Massin et al. 2007) used morphological operations and kernel density estimation to extract a feature vector applied to a KNN, Gaussian and Bayesian risk-minimizing classifiers; their method achieved an accuracy of 88.5%.

In the past few years, DLGrinsven et al. (Grinsven, Ginneken et al. 2016) presented Convolutional Neural Network (CNN) architecture for detecting hemorrhage with nine layers trained by the selective misclassified negative samples. Their algorithm obtained AUC of 0.89 and 0.97 for two different datasets. In 2016, Shan and Li (Shan and Li 2016) used a patch-based analysis method to detect microaneurysm and applied a stacked sparse auto-encoder to distinguish between those two groups and they obtained 91.38 % accuracy.

The success of diagnosis of DR requires the detection of all the three signs: exudate, hemorrhage and microaneurysm. While some of the studies reported earlier achieved acceptable performance for detection of single pathological sign, they were not suitable for identification of all the three signs simultaneously. Agurto et al. (Agurto, Murray et al. 2010) used multiscale amplitude-modulation-frequency-modulation (AM-FM) method for extracting texture features from segmented retinal images to differentiate between groups with and without DR. To distinguish between these two groups, they computed distance metrics between the texture features. While they identified the segments with DR signs, the method did not discriminate between the three DR signs, which is essential for treatment planning. In 2017, Tan et al. (Tan, Fujita et al. 2017) proposed a ten layers CNN architecture for DR sign detection. Their proposed network achieved a sensitivity of 0.87 for exudate detection, but this was only 0.62 and 0.46 for detection of hemorrhage and microaneurysm, respectively. Another limitation of this chapter was that they detected individual patches but did not consider the entire image which may explain the poor sensitivity due to misclassification of the background (with no pathological sign). Table 1 compares performance of the pervious methods for detection of exudate, hemorrhage and microaneurysm.

Table 5.3. Comparison between performance of the pervious methods for detection of exudate, hemorrhage and microaneurysm.

Methodology	Exudate		Hemorrhage		Microaneurysm	
	sensitivity	specificity	sensitivity	specificity	sensitivity	specificity
Tan et al. (Tan, Fujita et al. 2017)	0.87	0.98	0.62	0.98	0.46	0.97
Sinthanayothin et al. (Sinthanayothin, Boyce et al. 2002)	0.88	0.99	0.77	0.88	0.77	0.88
Grandet et al. (Gardner, Keating et al. 1996)	0.94	-	0.89	-	-	-
Naqvi et al.(Naqvi, Zafar et al. 2015)	0.92	0.81				
Walter et al. (Walter, Klein et al. 2002)	0.92	-	-	-	-	-
Fraz et al.(Fraz, Jahangir et al. 2017)	0.92	0.81	-	-	-	-
Sopharak et al. (Sopharak, Uyyanonvara et al. 2008)	0.82	0.99	-	-	-	-
Prentašić et al.(Prentašić and Lončarić 2015)	0.78	-	-	-	-	-
Welfer et at. (Welfer,	0.7	0.98	-	-	-	-

Scharcanski et al. 2010)						
Niemeijer et al. (Niemeijer, van Ginneken et al. 2005)	-	-	0.31	-	0.31	-
Fleming et al. (Fleming, Philip et al. 2006)	-	-	-	-	0.63	-
Walter et al. (Walter, Massin et al. 2007)	-	-	-	-	0.88	-
Garcia et al. (Garcia, Lopez et al. 2010)	-	-	0.86	-	0.86	-
Quellec et al. (Quellec, Lamard et al. 2008)	-	-	-	-	0.89	-
Bae et al. (Bae, Kim et al. 2011)	-	-	0.85	-	-	-
Walter et al. (Walter, Massin et al. 2007)	-	-	-	-	0.88	-

The patch-based analysis has been commonly used for CNN-based retinal image analysis (Shuang Yu, Di Xiao et al. 2017, Tan, Fujita et al. 2017). However, this approach can lead to disparity in the size of the sign due to patch size (Tan, Fujita et al. 2017), and the inexact evaluation because of the focus on the pathological signs without considering the neighbourhood and the background. While there are studies that have separated the background from the microaneurysm, and there are other studies that have accurately contoured the exudate, these perform analysis for one sign rather than all three signs. Such an approach can lead to the detection with overlaps between the three signs. Another

shortcoming is that while there are a number of isolated techniques that perform image enhancement, detect the presence of DR signs and perform processing to contour the signs, there is no framework that covers all the aspects.

In this chapter, the framework for a complete CNN-based system has been described for automatic and simultaneous detection and segmentation of exudate, hemorrhage and microaneurysm from fundus images. A ten-layered CNN architecture was designed and trained using images with annotated patches corresponding to the three signs and the background (No-sign) which was then used to obtain probability maps corresponding to each category (i.e. three sign and background). A post-processing algorithm was developed to differentiate pixels corresponding to a type of pathology from similar-looking cluttered pixels. Receiver Operating Characteristic (ROC) curve analysis was used to find a suitable threshold for differentiating between different types of pathologies This proposed framework was evaluated for both, patch and image-based analysis. Two publicly available databases were used, one was used for training while both were used for evaluation of the proposed method. The performance of the algorithm with and without probabilistic analysis was measured by taking the mean accuracy of ten repetitions.

6.3 Material

In this chapter, two public databases were used: 1- DIARETDB1, 2- e-Ophtha with total of 284 fundus images. 75 images from DIARETDB1 were used for patch-based analysis, while 209 images were used for image-based analysis.

DIARETDB1. Out of this database, 75 images were used for training the CNN while the remaining 14 images were used for testing and validating the performance of this method. In the training data, exudate, hemorrhage and microaneurysm were manually contoured by an experienced grader.

E-Ophtha. This is made up of two subsets: (i) “e-Ophtha EX” which contains 47 color retina images with annotated exudate, (ii) “e-Ophtha MA” which has 148 color retina images with annotated microaneurysm (Decencière, Cazuguel et al. 2013). In this database, there is a

variation in the size and resolution of the images, ranging from 1440×960 to 2544×1696 pixels. All images were resized to the size of the DIARETDB1 (1500×1152 pixels).

6.4 Methodology

The proposed framework consists of two main phases: 1) patch-based and 2) image-based analysis. The images were enhanced and then segmented in patches which were manually annotated and used to train the CNN. This trained CNN was used to analyze the other images for each pixel and a probability map was created using with which the locations of the pathological signs were identified. These images were processed to remove the isolated signs because these were noise and the spread of the signs which occurs during the earlier stages. The resultant images were compared with the manually annotated images to determine the accuracy of this method. An overview of the proposed method is shown in Figure 5.6 and the steps are described below.

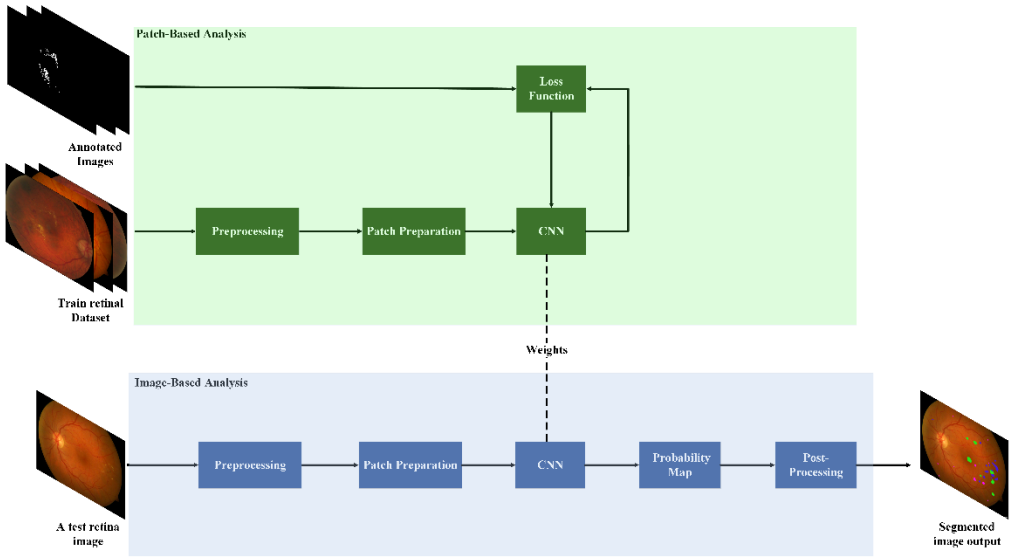
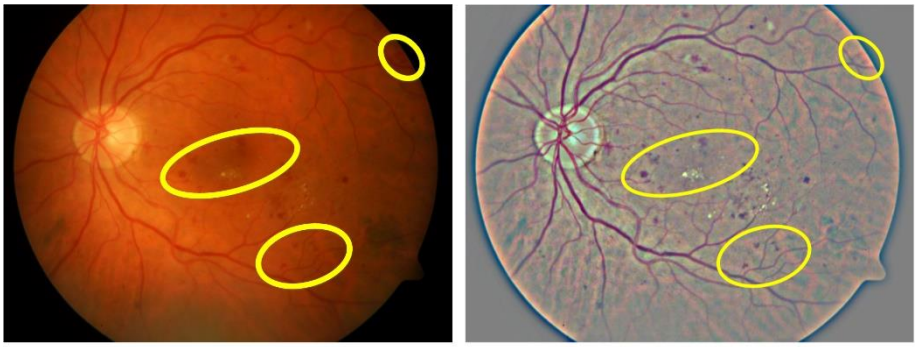


Figure 5.6. Overview of the proposed framework contains two main phases: 1) patch-based and 2) image-based analysis. The patch-based section corresponds to training and testing a CNN model to discriminate between the different DR signs. Image-based analysis of the entire image generates probability maps for each sign.

6.4.1 Pre-processing

CE technique was used in this chapter to enhance the contrast between three DR pathological signs and background. The result of image enhancement has been shown in Figure 5.7 that revealing that some new lesions can be singularized by image enhancement, as specified by the yellow marks.



(a) (b)

Figure 5.7. Applying the image enhancement technique on an example retina image. (a) Original retina image; (b) After image enhancement. This shows that some new lesions can be singularized by image enhancement shown by yellow annotations).

6.4.2 Convolutional neural network

The enhanced images were segmented into patches of size of S×S which were labelled based on the ground truth images corresponding to the three pathological signs: exudate, hemorrhage, microaneurysm and background (without any pathological sign). These patches were the input to the CNN which was trained against the target labels. The choice of CNN architecture and the parameters were as same as section 3.3.2.

6.4.3 Image analysis

In this chapter, pixel-based analysis of the image was performed by taking a patch of size S×S centered around pixel (x_i, y_i). This patch is the input to trained CNN which gives membership probabilities (range 0 to 1) at location (x_i, y_i) for the three pathological signs: i.e. exudate, hemorrhage and microaneurysm (shown by $P_{E,xi,yi}$, $P_{H,xi,yi}$ and $P_{M,xi,yi}$). Consequently, three probability maps for the image are created and the scheme of this mapping process is shown in Figure 5.8.

To identify the signs, a threshold was determined for each of the probability maps. This threshold (Th) was obtained by maximizing the receiver operating characteristics curve and used to binarize each probability map and obtain a binary map corresponding to the three

signs. Overlaps were avoided by ranking the points with overlap based on the probability values.

One difficulty that is faced by such methods is the appearance of redundant boundaries and cluttered pixels (False positive pixels) around the segmented signs. To overcome this shortcoming, three morphological operations: closing, opening and erosion were performed with masks of size 5×5, 5×5 and 4×4 pixels, respectively (Zana and Klein 1997, Soille 2003). This was followed by a rule based post-processing where signs with area of less than $\frac{S^2}{4}$ were removed.

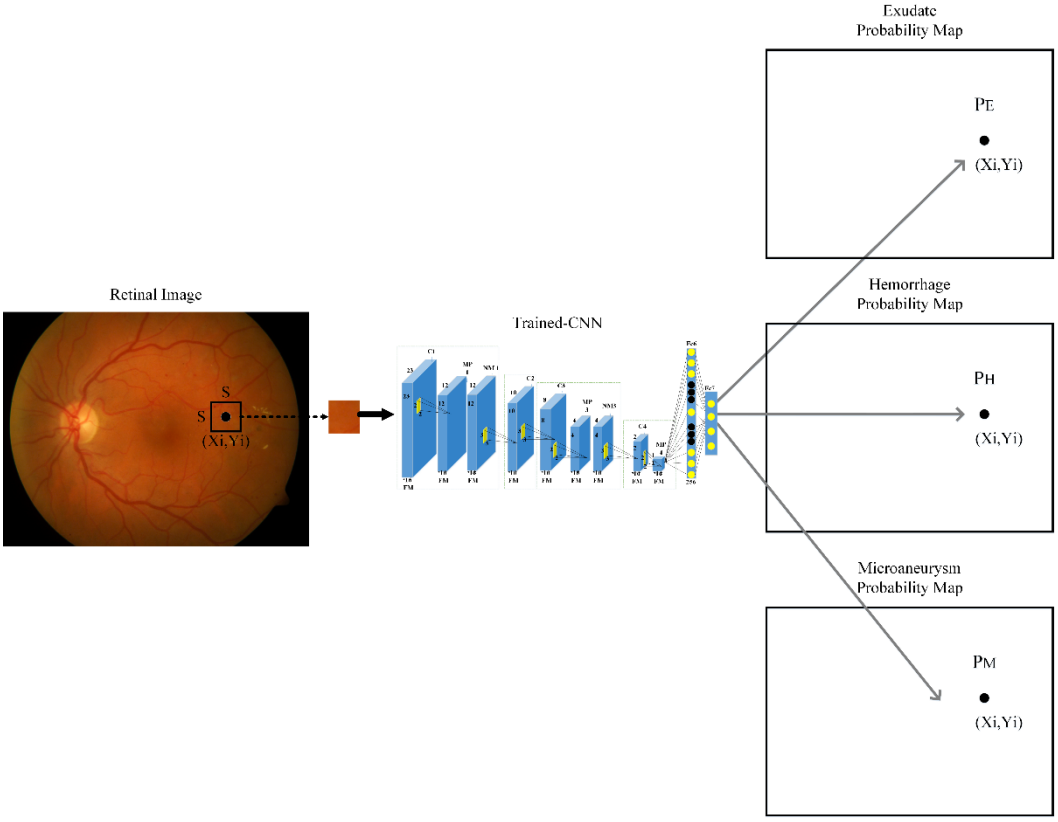


Figure 5.8. Process of generating three probability maps corresponding to exudate, hemorrhage and microaneurysm from a retina image. By taking a patch of size S×S centered around pixel (x_i, y_i), each patch is fed to the trained CNN that determines the membership probabilities at location (x_i, y_i) for the three pathological signs: i.e. exudate, hemorrhage and microaneurysm (shown by $P_{E,xi,yi}$, $P_{H,xi,yi}$ and $P_{M,xi,yi}$).

6.4.4 Validation Parameters

The performance was evaluated based on false positive (FP), true positive (TP), true negative (TN) and false negative (FN) rates (Dice 1945) (Table 5.4).

Table 5.4. Validation parameters

Parameter	Equation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Error rate	$\frac{FP + FN}{TP + TN + FP + FN}$
Positive predict value (PPV)	$\frac{TP}{TP + FP}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$

6.5 Experiments

6.5.1 Data preparation

The image was segmented into patches by the size of $S \times S$, with $S = 50,$ which was determined based on the smallest pathological signs in these images. Patches corresponding to the signs were manually extracted from 75 retina images of the DIARETDB1 database and used for the training the network. These resulted in 22719, 18882 and 17824 patches for exudate, hemorrhage and microaneurysm and 50518 patches with no pathological signs (No-Sign). The No-Sign patches contained vessels, background tissue and optic nerve head. There was no overlap between each to adjacent patch. To increase the robustness of the algorithm, data augmentation was performed using both horizontal and vertical filliping and

rotating (Roth, Lee et al. 2015, Pang, Yu et al. 2017). Figure 5.9 shows patch examples corresponding to four classes and Table 3 summarizes the number of patches considered for the training (75%), validation (15%) and testing (15%) CNN.

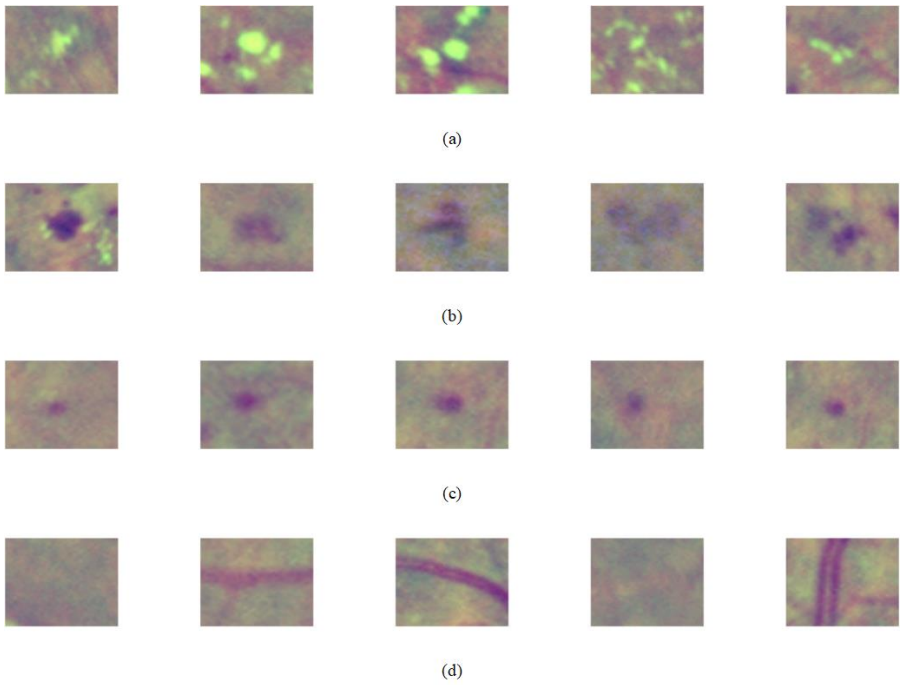


Figure 5.9. Patch examples corresponding to the four classes; (a) exudate. (b) hemorrhage. (c) microaneurysm. (d) no-sign.

Table 5.5. Statistics information of sign patches.

	Exudate	Haemorrhage	Microaneurysm	No-Sign
Training	15646	13339	12477	35013
Validation	3353	2859	2674	7503
Test	3720	2684	2333	8002
Total Number	22719	18882	17484	50518

6.5.2 Image Analysis

The test image set of DIARETDB1 and all images of e-Ophtha were used to evaluate the performance of the proposed method using image-based analysis. These images were analyzed (section 6.4.3) and the probability map was created of the all pixels in the image which resulted in three probability maps corresponding to exudate, hemorrhage and microaneurysm. Figure 5.10 shows an example with the three probability maps. Figure 5.11 shows the images after applying post-processing. It can be seen that the algorithm's outcome accurately segmented the actual pixel's signs from the all pixels which were assigned as potential pixels for the signs with different probability.

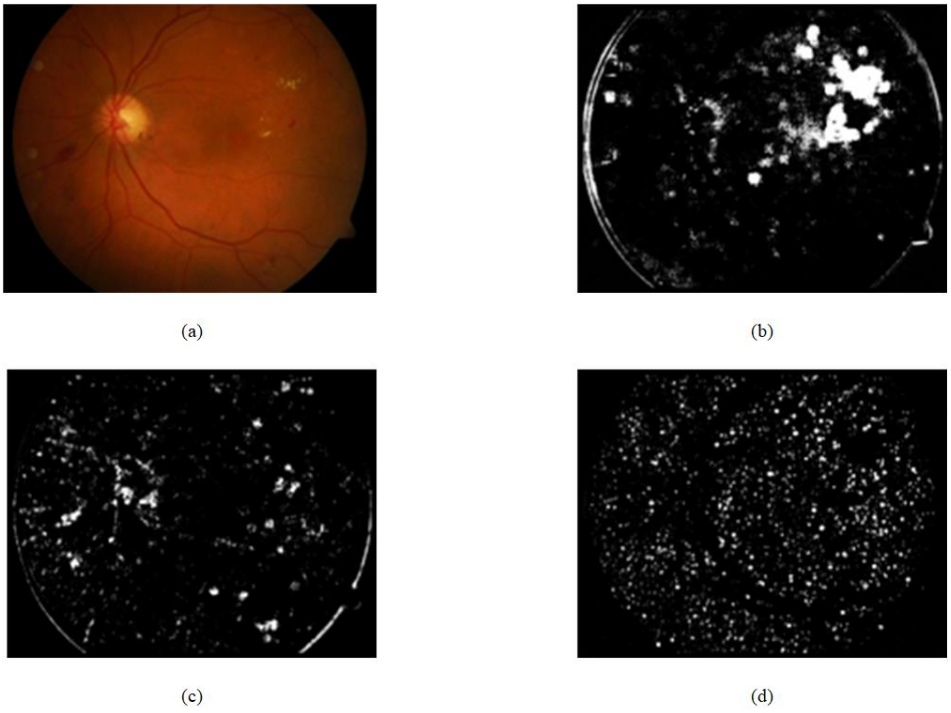


Figure 5.10. Three probability maps were generated from an example retina image: (a) original retina image; (b) Exudate probability map; (c) Hemorrhage probability map; (d) Microaneurysm probability map. Colorbar shows the severity level of a pixel belong to the sign that is ranging between 0 to 1 corresponding to blue to red color.

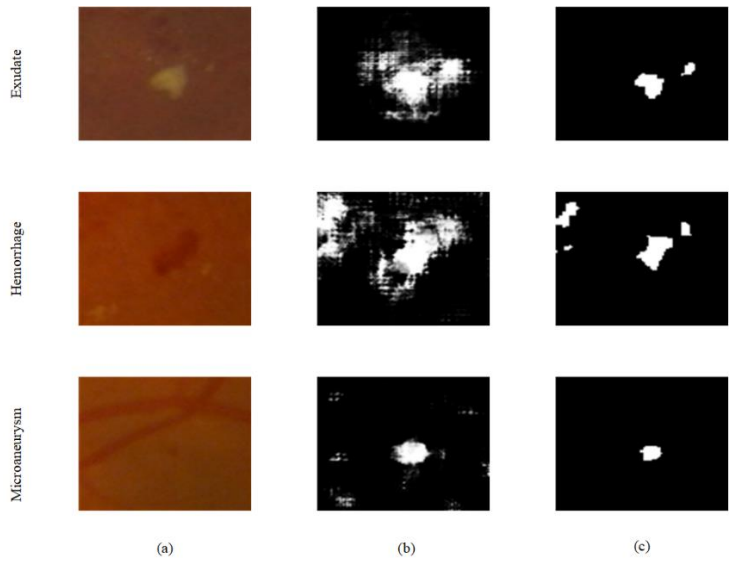


Figure 5.11. Three examples of pathological signs before and after post-processing. (a) Original image. (b) Probability map corresponding to the sign. (c) Image output after post-processing

6.6 Results

For the patch-based evaluation, the mean results of ten repetitions for the training are described in Table 5.6 and Figure 5.12 shows the ROC curve for the CNN performance. Table 5.6 shows the sensitivity, specificity and accuracy for the proposed method. The best results were for the exudates with sensitivity, specificity and accuracy of 0.96, 0.98 and 0.98, respectively, while that for hemorrhages was 0.84, 0.92 and 0.90, and 0.85, 0.96 and 0.94 for microaneurysm.

Table 5.6. Sensitivity, specificity, accuracy and PPV of the proposed method in patch-level evaluation for detection of exudate, hemorrhage and microaneurysm.

	Exudate	Hemorrhage	Microaneurysm	No-Sign
Accuracy	0.98	0.90	0.94	0.96
Sensitivity	0.96	0.84	0.85	0.95
Specificity	0.98	0.92	0.96	0.97
PPV	0.94	0.85	0.83	0.96

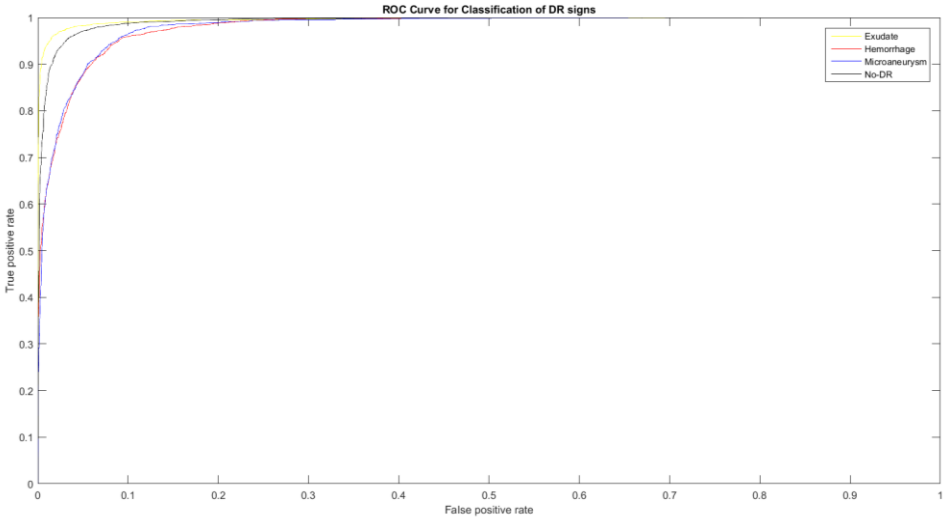


Figure 5.12. ROC curve corresponding classification of the four classes (exudate, hemorrhage, microaneurysm and no-sign). This shows that the CNN-model obtained the highest AUC for detection of exudate (yellow color) and almost same AUC for detection hemorrhage and microaneurysm.

For image-level evaluation, performance of the proposed method was compared to the method which used the binary outputs of the network for both datasets and shown in Figure 5.13. It is observed that for DIARETDB1, the proposed method achieved the accuracy of 0.96, 0.98 and 0.97 and error rate of 3.9 %, 2.1 % and 2.04 % for segmentation of exudate, hemorrhage and microaneurysm, respectively which shows that this technique outperforms techniques reported in literature. Similarly, there was significant improvement for exudate and microaneurysm detection in the e-Ophtha dataset with accuracy of 0.88, and 3.0 and error rate of 4.2 % and 3.1 %, respectively. Figure 5.14 shows example of a retinal image with pathological signs detected by the proposed algorithm.

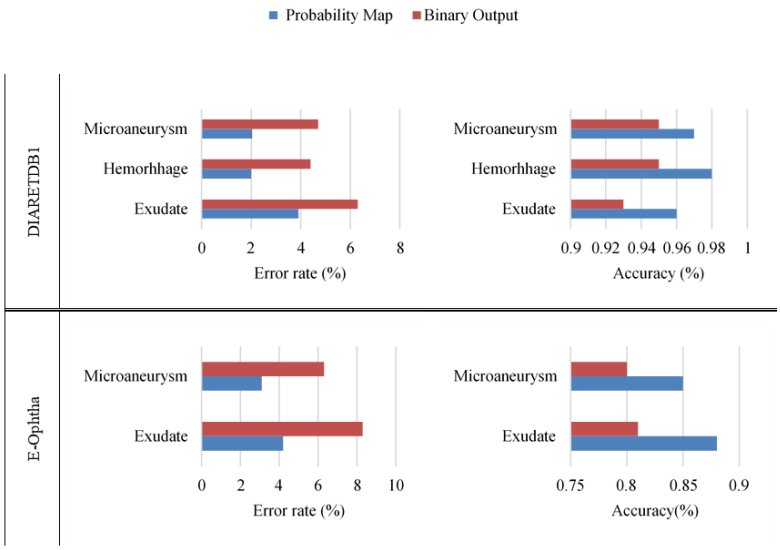


Figure 5.13. Performance of proposed framework for the sign detections using two databases (DIARETDB1 and e-Ophtha) compared to the method with binary outputs of the network.

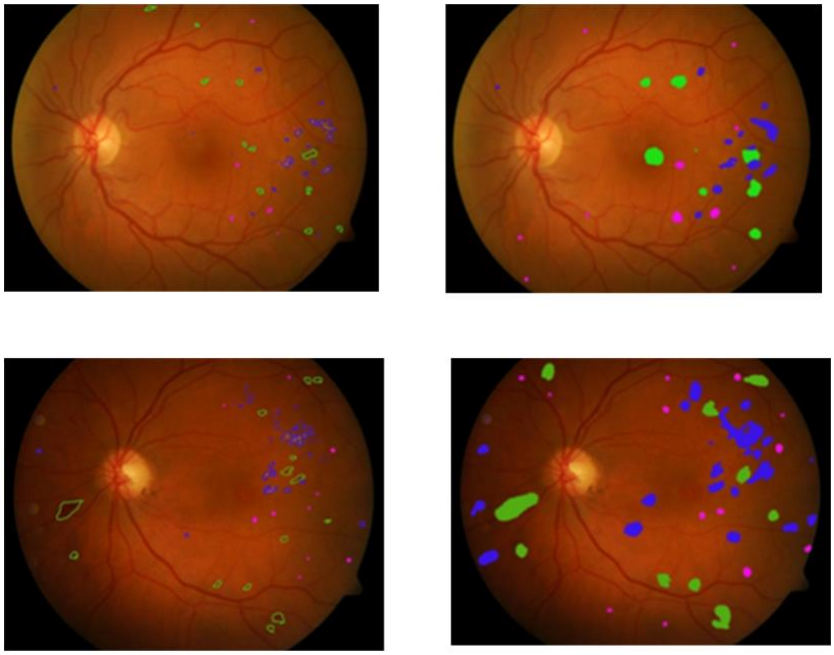


Figure 5.14. Segmentation output image of the example retina image. (a) Manually annotated images that exudate, hemorrhage, and microaneurysm signs marked by blue, green and pink color, respectively. (b) Segmented output by the proposed algorithm.

6.7 Discussion

This chapter has presented a CNN-based framework to analyze the retina fundus images for detection of pathologic signs indicative of DR: exudate, hemorrhage and microaneurysm. The images were first pre-processed to enhance the contrast and then segmented in patches which were then manually annotated and used for training the CNN network. This network was then used to determine the probability for each pixel to belong to the four classes of exudate, haemorrhage, microaneurysm, and background (no pathologic sign). The resultant probability map was then used to determine the locations of all the three types of pathological signs corresponding to DR. The isolated signs and the spread due to convolution were automatically removed in a post-processing step described earlier.

The results show that there was a difference in the accuracy, sensitivity and specificity when using the two databases: DIARETDB1 and e-Ophtha which could be because the CNN was trained using only DIARETDB1. Compared to previous works in which the two databases were used (Table 5.3), the performance of the proposed approach was higher. It also observed that average sensitivity and specificity for detecting exudates (0.96 and 0.98) is higher than for hemorrhage and microaneurysm. According to Table 5.3, most of the previous studies suffer from poor sensitivity, particularly for discrimination between hemorrhages and microaneurysms. Comparing our results with the work by Tan et al (Tan, Fujita et al. 2017) shows that our method achieved significantly better sensitivity for detection of hemorrhage (0.84 vs 0.62) and microaneurysm (0.85 vs 0.46), although the specificity is similar. Our method also obtained better performance for both, sensitivity and specificity, for detection of the three DR signs when compared to the work by Sinthanayothin et al. (Sinthanayothin, Boyce et al. 2002).

Our method simultaneously detects the three pathological signs with improved performance compared to previous studies where only one sign was considered. This makes it suitable for more reliable detection of DR because when the signs are identified individually, there is the potential error of identifying the same region for multiple signs. This method performs comprehensive analysis and detects all the three signs simultaneously. The other study that

attempted the simultaneous detection of the three signs was by Tan et al. (Tan, Fujita et al. 2017) which suffered from poor performance.

One innovation of this method is the use of score values obtained from the softmax layer instead of using the binary output of the network. This results in the generation of the probability map of the locations of the pathological signs on the image, which with suitable post-processing reduces the error rate in the size of the signs.

The first significant strength of this chapter is that we considered two different publicly available databases, with the training done on one and the testing on both with comparable results. The second strength of this chapter is that fundus images were analysed using both, patch and image-based analysis, and the results show that this method is significantly better than other studies. The third strength is that this method simultaneously identifies the three different pathological signs on the images which makes it suitable for automatic detection of diabetic retinopathy because when the signs are identified individually, there is potential error when the same region is identified for multiple signs.

A limitation of this method is that it is unable to differentiate between hemorrhages and microaneurysms if there is an overlap between these. This is also a limitation of the dataset because overlaps in the original images have not been labelled. Another limitation is that the database of 284 images was imbalanced with very few images with hemorrhages. There is the need for further testing of this method for databases belonging to different demographics to determine the suitability for different societies.

6.8 Conclusion

This chapter reports a CNN based framework for the analysis of retinal images to detect the three major signs of diabetic retinopathy: exudates, hemorrhages and microaneurysms. The novelty of this system is that it uses the softmax output of the layers to generate the probability map for the three pathologic signs of DR which is then used to segment the fundus image and identify the signs. The system was trained using one dataset and tested on two datasets which shows the universality of the approach. The results show that such a

system can be used for automatic analysis of fundus images for the detection of diabetic retinopathy without requiring a large dataset for training the network.

Chapter 7

7 Parkinson's Disease Diagnosis Based on Multivariate Deep Features of Speech Signals

7.1 Overview

In this chapter, we have tested the effectiveness of deep convolutional neural network (DCNN) in distinguishing between Parkinson's and healthy voices using spectral features from sustained phoneme /a/ (as pronounced in "car"). Various designs of the DCNN architecture were investigated on raw pathological and healthy voices of varying lengths. This chapter also investigated the effect of parameters such as frame size, number of convolutional layers and feature maps as well as topology of fully connected layers on the accuracy of the classification outcome. The best network achieved accuracy of 75.7% corresponding on 815 ms of data for distinguishing between Parkinson's and healthy samples. This chapter has demonstrated that online speech recording has the potential for being used to screening people for Parkinson's disease.

7.2 Introduction

PD is a neurodegenerative disorder that is characterised by motor and non-motor features including tremor, rigidity, bradykinesia, cognitive impairment, sleep disturbances and depression (Watson and Leverenz 2010, DeMaagd and Philip 2015). In clinical practice, Unified Parkinson's disease scale (UPDRS) is used for PD diagnosis which is subjective and dependent on the expertise of the clinician. It also requires the patients' presence in the clinic, and the tests are time consuming for the patient and the clinicians. There is an urgent need for alternate diagnostic methods that could be performed online.

Associated with PD speech disorder referred to as Hypokinetic Dysarthria. The speech of the PD patient is less intelligible, breathy and soft. The speech impairment in PD is one of the earliest symptoms and has been considered for diagnosis of the disease. It has the advantage of being non-intrusive and suitable for online applications and hence researchers have investigated the speech features to diagnose PD (Tsanas, Little et al. 2010, Zhang, Yang et al. 2016, Pompili, Abad et al. 2017). It also provides a platform for objective monitoring of the progression of the disease.

Many studies have utilised time-frequency based features extracted from the speech signal to classify between PD and control subjects (Rahn, Chou et al. 2007, Lee, Zhou et al. 2008, Gracas, Gama et al. 2012, Viswanathan, Khojasteh et al. 2018). These features include Jitter, Shimmer, Pitch, Harmonics to Noise Ratio, Autocorrelation, voiced and unvoiced frames (Sakar, Isenkul et al. 2013) which provide information such as the breathiness and hoarseness of the voice. However, features such as Jitter and shimmer measure the short-term fluctuations in phonation and have been found to be unsuitable to evaluate the fluctuations in more severely impaired voices (Lee, Zhou et al. 2008). To address this shortcoming, studies have used a combination of time-frequency based non-linear dynamics features such as detrended fluctuation analysis, correlation dimension and recurrence period entropy for classification of the two groups (Little, McSharry et al. 2007, Rahn, Chou et al. 2007, Tsanas, Little et al. 2010) and have obtained different classification accuracies.

In recent years, application of deep-learning in speech processing such as speech emotion recognition (Fayek, Lech et al. 2017) and voice pathology detection (Harar, Alonso-Hernandezy et al. 2017) has achieved promising outcomes. Deep-learning can extract automatic learnable features from large data arrays and model high-level abstractions. Unlike conventional machine learning methods, it supports unsupervised learning from raw data and identifies the most important features relevant to the study objectives. This has been the major motivation for this current study to use DL to distinguish between normal and pathological voices in patients with PD.

Deep convolutional neural network (DCNN) has the potential of identifying the suitable features of the data to automatically classify the signal and hence has been proposed for

speech-based diagnosis of PD. This chapter presents a framework based on DCNN for obtaining the spectral features to distinguish between speech recordings of PD and Healthy volunteers. It also reports the effect of parameters such as audio length, frame size, number of convolutional layers, feature maps and the topology of fully connected layers on the performance of the DCNN.

7.3 Material and Experimental Protocol

The recordings obtained by my colleagues, Ms Rekha Viswanathan and other members of the BioSignal Lab were used with their permission for this study. Total of 81 subjects participated in this study comprising of 41 PD (mean age ± SD=70.8 ± 8.2) in their mild to moderate stage of the disease (UPDRS ≤ 40) and 40 age-matched (mean age ± SD = 66.9 ± 6.21) control subjects (UPDRS ≤ 12). Patients were recruited from the Dandenong Neurology, Centre, Melbourne, Australia and control subjects were recruited from retirement villages. Non-English speakers, participants with Dementia, stage four PD, presence of stuttering or stammering in the voice and subjects with any other neurological or psychological illness were excluded from the study. Subjects were asked to produce sustained /a/ vowel phonation (as in "car") of length greater than 5 seconds. Voice samples were collected using Apple iPhone 6S plus® with a high quality omni directional condenser microphone. All voice samples were recorded in .wav format (i.e. no compression) with sampling rate of 8 KHz and 32-bit resolution. All the PD subjects were on Levodopa medication during the experiment and they were considered as in 'on state' condition. The experimental protocol was approved by RMIT University Human Research Ethics Committee (I was not a co-applicant in the Ethics) and in accordance with Declaration of Helsinki (revised 2004). All participants were enrolled after obtaining their oral and written consent.

7.4 Methodology

This chapter has investigated the use of DCNN to distinguish between the spectrogram of voices of PD and Healthy subjects. It has identified the suitable framework of the DCNN and the number of frames of the spectrogram that are input to the DCNN (Figure 5.15).To overcome any inter-experimental differences such as distance between the mouth and the microphone, all recordings were normalized prior to the analysis.

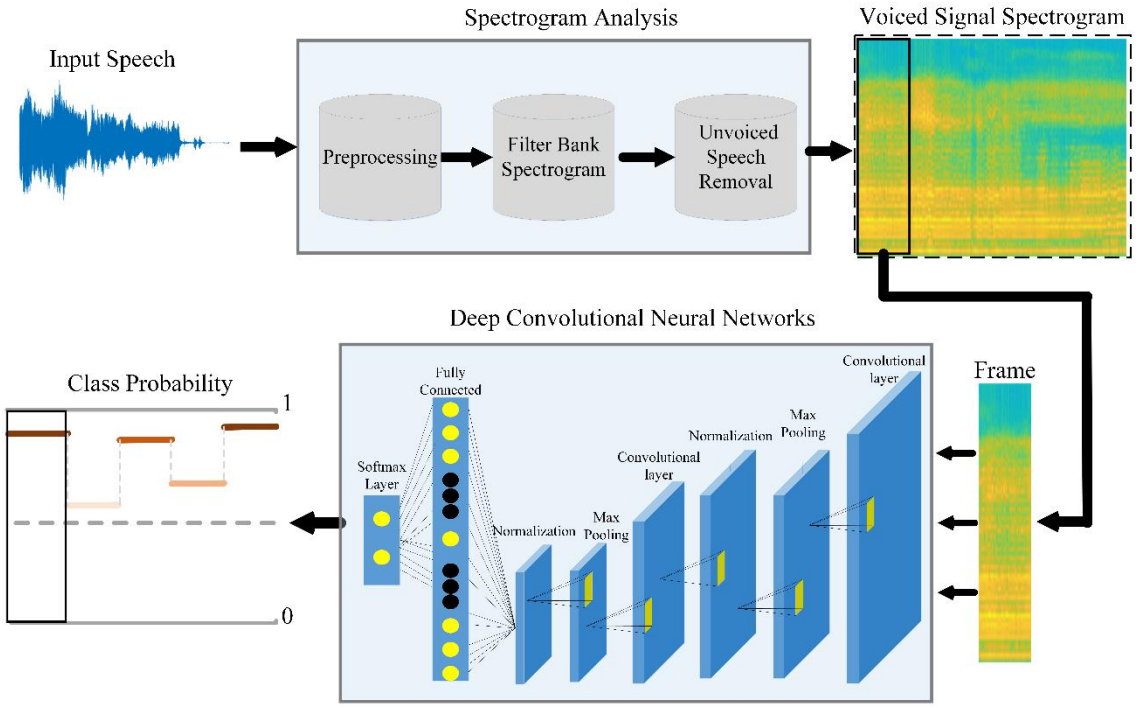


Figure 5.15. Overview of the proposed framework.

7.4.1 Deep Convolutional Neural Network

We investigated seven DCNN architectures to determine the effect of some of the parameters on the performance of the proposed algorithm. Factors such as the number of convolutional layers, feature maps as well as different topologies of the fully connected layers were considered. Table 5.7 lists seven network architectures. In all networks, rectified linear unit (ReLU) was used after each convolutional layer and a normalization layer was

applied after each max-pooling layer. The backpropagation algorithm with stochastic gradient descent (SGD) was used for the training process (Pang, Yu et al. 2017).

Table 5.7. Proposed DCNN's architectures. Conv ($a \times b \times c$): convolutional layer with a feature maps and size of $b \times c$ pixels, MP ($e \times f$) : max-pooling layer with kernel size of $e \times f$ pixels, FC (n): fully connected layer with n neurons.

Net 1	Net 2	Net 3	Net 4	Net 5	Net 6	Net 7
Conv1 32×3×3	Conv1 16×3×3	Conv1 16×3×3	Conv1 32×3×3	Conv1 32×3×3	Conv1 32×3×3	Conv1 32×3×3
MP1 2×2	MP1 2×2	MP1 2×2	MP1 2×2	MP1 2×2	MP1 2×2	MP1 2×2
FC1 512	Conv2 16×3×3	Conv2 32×3×3	Conv2 16×3×3	Conv2 16×3×3	Conv2 16×3×3	Conv2 16×3×3
	MP2 2×2	MP2 2×2	MP2 2×2	MP2 2×2	MP2 2×2	MP2 2×2
	FC1 512	FC1 512	FC1 256	FC1 512	FC1 256	Conv3 16×3×3
					FC2 128	MP3 2×2
						FC1 512

7.4.2 Experimental Setup

Data Preparation

The dataset was randomly split into training (70%), validation (15%) and test (15%) sets. The gender ratio between male and female was considered when splitting the data in the three sets.

All data was recorded with sampling rate of 8 KHz and they were normalized to get zero mean and unit variance. The variance and mean were computed and used to normalize the data. The spectrogram was obtained using 25 ms hamming window and 15 ms overlaps between the windows.

The silent periods in each audio were removed by considering each recording individually, obtaining the mean of each window of the spectrogram and identifying the threshold (T_s) as the centroid between maximum and minimum. The windows corresponding to mean less than T_s were removed, and the new spectrogram generated by concatenating the remaining windows.

To assess the suitable frame size for analysis, the input frame to the DCNN was varied such that it consisted of multiple spectrogram windows. We investigated five different frame sizes of 20, 40, 60, 80 and 100 windows; (e.g. the frame size of 20 corresponding to 215ms). For the training set, overlaps between the frames were considered by the stride of two windows. Data augmentation technique was applied on frames to increase the number of training sets. Data augmentation consisted of flipping (vertical and horizontal) and rotation of the frames (Roth, Lee et al. 2015). For the validation and test sets, the frames did not have any overlap and data augmentation was not applied on these sets to have reliable results. Statistics of frames using different window seizes in training, validation and test sets is shown in Table 5.8.

Table 5.8. Statistics of frames using different window seizes in training, validation and test sets.

Frame Size (Number of windows)	Training		Validation		Test	
	PD	H	PD	H	PD	H
20	6,471	9,317	1,318	1,810	1,121	1,184
40	3,211	4,312	884	1,013	792	980
60	1,890	2,647	405	573	439	560
80	1,454	1,963	292	417	229	405
100	1,060	1,532	222	323	222	335

DCNN setup and Training

In training phase of DCNN, the parameters were empirically set as follow: the momentum and learning rate were set to 0.9 and 0.01, respectively. Gaussian filter's weights for all convolutional layers were set to 0.01. The Gamma value was 0.1 in each iteration and step-down policy by step size of 33 was used for SGD implementation. The batch size for training, validation and test were set to 128, 32 and 32, respectively. The maximum number of epochs was 100.

The training for the seven proposed networks were performed individually. The weights were fixed for each training and then used in the testing phase. In this study, GeForce GTX 1070 and café platform (Jia, Shelhamer et al. 2014) were used.

7.5 Results and Discussion

To distinguish between the spectrogram frame of the PD and Healthy individual, the test set was used for evaluating the performance of the seven networks using five different frame sizes. For this purpose, four evaluation parameters including accuracy, sensitivity, specificity and positive predictive value (PPV) were used (Dice 1945). To obtain reliable and repeatable results, the experiment was repeated 10 times by random selection of training, validation and test sets and the mean of the evaluation parameters were considered and shown in Figure 5.16. It can be seen that Net 5 with frame size of 80 achieved the best accuracy and sensitivity of 75.7 % and 0.66, respectively. The frame size of 80 which is equal to 815 ms (25+79×10) of audio was found as the suitable size for distinguishing between PD and Healthy. With the frame size of 100, net 5 obtained PPV and Specificity of 86.36% and 0.82, respectively.

By changing topologies of the fully connected layers from 512 neurons in one fully connected layer (Net 5) to 256 (Net 4) and two fully connected layers (Net 6), sensitivity decreased to 0.60 and 0.39. This decrease could be caused by turning the Net 5 to an over fitted and under fitted network as in Net 6 and Net 4, respectively.

Net 7 was developed with one more convolutional layer compared to Net 5 to allow for assessing the effect of increasing the depth. However, the results show that almost all

evaluation parameters significantly declined, particularly the accuracy and sensitivity to 49.0 % and 0.38, respectively. Similarly, in Net 1, which was designed by only one convolution layer, the performance was poor. Thus, changes to the number of convolution layers of the Net 5 did not provide better results. The results also show that changing the number of feature maps did not improve the performance compared with Net 5.

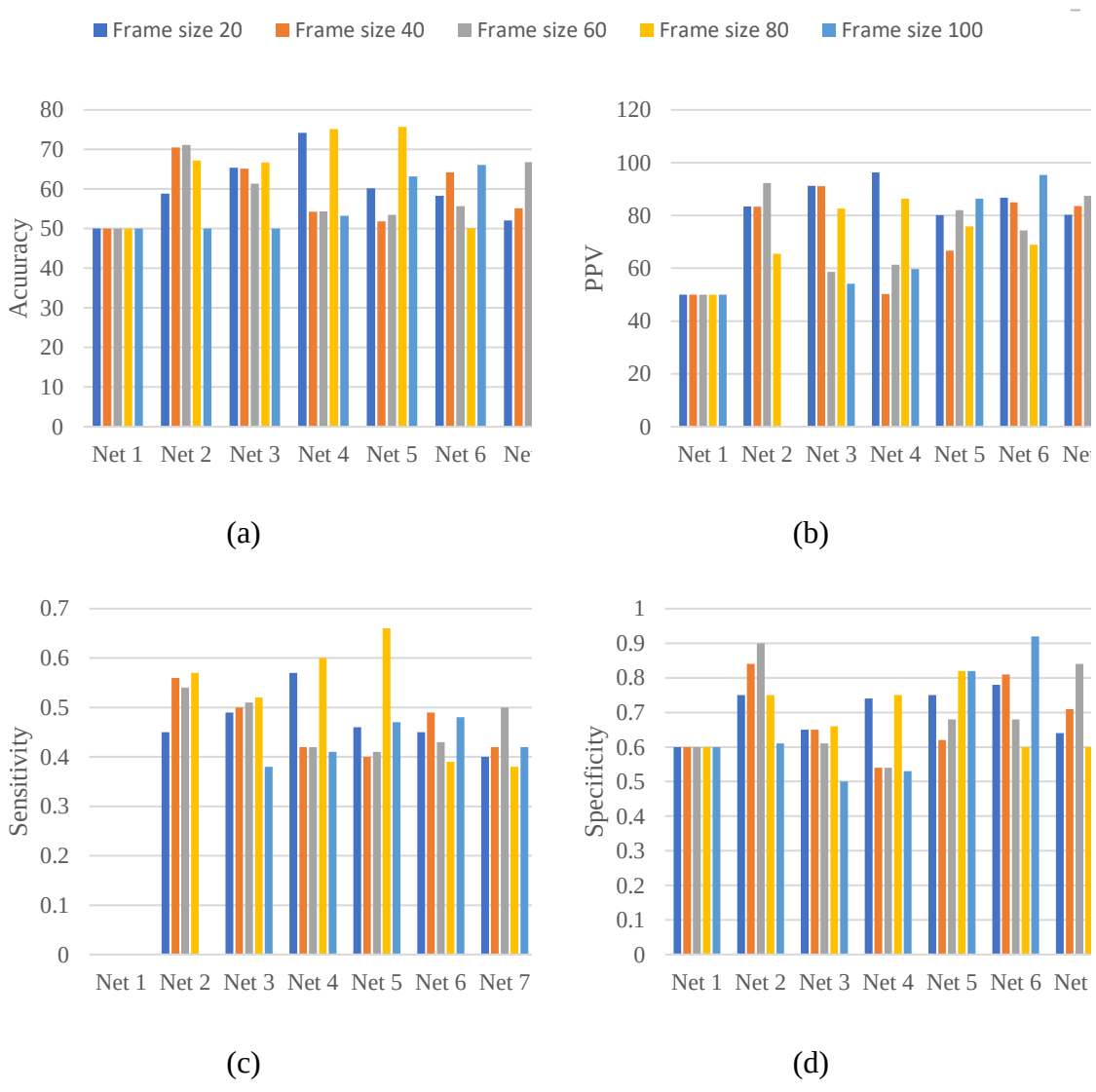


Figure 5.16. Performance of the proposed networks corresponding to different frame sizes: (a) Accuracy (b) PPV (c) Sensitivity (d) Specificity.

7.6 Conclusion

In this chapter, we have presented a DCNN-based framework to distinguish between PD and Healthy speech signal using sustained phoneme /a/. By assessing multivariate deep features, the network with two convolutional layers, 512 neurons and one fully connected layer achieved the best results when tested on two secs of voice data. The future work includes increasing the number of samples and considering other phonemes such as sustained /u/ and /o/ as pronounced in the words “cool” and “go” respectively.

Chapter 8

8 Conclusion and Future Work

8.1 Conclusion

DL and machine learning methods have been used for addressing the problems in the biomedical applications, such as DR assessment and PD diagnosis. Although some reported outcomes have shown good results in using these methods, they still suffer from the poor sensitivity and specificity as well as the limitations such as dataset size and heuristic selection of the network parameters.

This thesis has comprehensively studied various DL methods for detection of DR signs and diagnosis of PD. The finding of this thesis corresponding to the objectives can be summarised as follow:

i. The study of color space analysis of fundus images for automatic exudate detection shows the importance in the choice of color space for better representation of the fundus images. The results confirm that HSI and PCA-RGB spaces outperform other spaces. This also introduces the new color space "PHS" and the results show that this space achieves the most reliable and repeatable results compared to the conventional spaces (objective i and ii).

ii. Investigation of different DL methods for exudate detection confirms that "Resnet-50 + SVM" obtains better performance compared to the other pre-trained, CNN and DRBM models. This also shows using the pre-trained model not only achieves better results compared to other models, but also it does not require large dataset and choice of the network's parameters (objective iii).

iii. The study of novel pre-processing layer in CNN architecture shows the importance of using a suitable pre-processing method for improving the performance of system for detection of DR signs. It can be concluded that image enhancement has a

significant impact on the performance of the CNN-based systems. This also shows that CE technique is the most suitable image enhancement method for DR signs detection (objective iv).

iv. Using deep probabilistic features for detection of DR signs shows the significant impact of post-processing algorithms for improving the sensitivity and specificity of the CNN-based models. The results confirm that using the score values of the softmax layers (probability map) followed by a suitable post-processing method can increase the performance of the system by reducing the error rate in the size of the segmented areas (objective v and vi).

v. The study of Parkinson's disease analysis tests the efficiency of the CNN model for distinguishing between Parkinson's and healthy voices using spectral features. It can be concluded that 814 ms of voice is the most suitable frame size carrying enough information for Parkinson's disease diagnosis. The results also show that the network with two convolutional layers, 512 neurons and one fully connected layer achieves the best results compared to other topologies (objective vii and viii).

8.2 Novelty

The novelties of this work can be briefly summarised as follow:

i. Propose the "PHS" space as the new color space of fundus images for improving the performance of the CNN-based models (Chapter 3).

ii. Compare the performance of the different deep learning methods for the first time for detection of exudates in the fundus images (Chapter 4).

iii. Introduce the application of the pre-processing layer in the CNN architecture for the first time for detection of the DR signs (Chapter 5).

iv. Develop the CNN-based framework using probability maps for detection of DR signs (Chapter 6).

v. Propose the DL-based model for distinguishing between Parkinson's and healthy voices using spectral features (Chapter 7).

8.3 Future work

This thesis has investigated DL methods for automatic DR assessment and PD diagnosis. The findings have assured current methodologies but demonstrated the limitations and the need for further improvement. Based on the findings of this thesis, the future work is suggested as follow:

i) The proposed space ("PHS") has achieved promising results for increasing the performance of the system for detection of DR signs. However, the success of this space needs to be tested for the other applications.

ii) Compared to pre-trained CNN models, the DRBM model has achieved poor performance due to lack of suitable data presentation and use of shallow network. It is expected to improve the performance of such models by using deeper and more complex models such as Deep Belief Networks and Deep Boltzmann Machines.

iii) It has been shown that the CE method is able to increase the performance of the system for automatic DR signs detection although it has been tested on only 80 retinal images. In future, this method will be tested on a larger number of images from different datasets to assure the reliability and repeatability of the method.

iv) The proposed framework has achieved good results in detection and distinguishing between all DR signs. However, that it is not able to differentiate between hemorrhages and microaneurysms if there is an overlap between them due to improper labelling of such cases in the ground truth dataset. In future, the overlapped areas need to be labelled and classified as a new group (i.e. overlapped hemorrhages and microaneurysms)

v) The CNN-based model has shown the capability of such models for PD diagnosis using sustained phoneme /a/. To improve the performance of system, future work includes increasing the number of samples and considering other phonemes such as sustained /u/ and /o/ as pronounced in the words "cool" and "go" respectively.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng (2016) "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." arXiv e-prints.

Abbasi-Sureshjani, S., B. Dashtbozorg, B. M. ter Haar Romeny and F. Fleuret (2018). Exploratory Study on Direct Prediction of Diabetes Using Deep Residual Networks, Cham, Springer International Publishing.

Abbod, M. F., J. W. F. Catto, D. A. Linkens and F. C. Hamdy (2007). "Application of Artificial Intelligence to the Management of Urological Cancer." The Journal of Urology **178**(4): 1150-1156.

Abràmoff, M. D., J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer and G. Quellec (2010). "Automated early detection of diabetic retinopathy." Ophthalmology **117**(6): 1147-1154.

Acharya, U. R., E. Y. Ng, J. H. Tan, S. V. Sree and K. H. Ng (2012). "An integrated index for the identification of diabetic retinopathy stages using texture parameters." J Med Syst **36**(3): 2011-2020.

Agurto, C., V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russell, M. Abramoff and P. Soliz (2010). "Multiscale AM-FM methods for diabetic retinopathy lesion detection." IEEE Trans Med Imaging **29**(2): 502-512.

Akram, M. U., A. Tariq, S. A. Khan and M. Y. Javed (2014). "Automated detection of exudates and macula for grading of diabetic macular edema." Computer Methods and Programs in Biomedicine **114**(2): 141-152.

Ali, S., D. Sidibé, K. M. Adal, L. Giancardo, E. Chaum, T. P. Karnowski and F. Mériaudeau (2013). "Statistical atlas based exudate segmentation." Computerized Medical Imaging and Graphics **37**(5-6): 358-368.

Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression." The American Statistician **46**(3): 175-185.

Azizpour, H., A. S. Razavian, J. Sullivan, A. Maki and S. Carlsson (2015). From generic to specific deep representations for visual recognition. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

Bae, J. P., K. G. Kim, H. C. Kang, C. B. Jeong, K. H. Park and J. M. Hwang (2011). "A study on hemorrhage detection using hybrid method in fundus images." J Digit Imaging **24**(3): 394-404.

Balafar, M. (2014). "Gaussian mixture model based segmentation methods for brain MRI images." Artificial Intelligence Review **41**(3): 429-439.

Bastien, F., P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley and Y. Bengio (2012) "Theano: new features and speed improvements." arXiv e-prints.

Batmaz, Z., A. Yurekli, A. Bilge and C. Kaleli (2018). "A review on deep learning for recommender systems: challenges and remedies." Artificial Intelligence Review.

Bengio, Y., P. Lamblin, D. Popovici and H. Larochelle (2006). Greedy layer-wise training of deep networks. Proceedings of the 19th International Conference on Neural Information Processing Systems. Canada, MIT Press**:** 153-160.

Cao, W., J. Shan, N. Czarnek and L. Li (2017). Microaneurysm detection in fundus images using small image patches and machine learning methods. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE.
Charron, O., A. Lallement, D. Jarnet, V. Noblet, J.-B. Clavier and P. Meyer (2018). "Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network." Computers in biology and medicine **95**: 43-54.
Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille (2018). "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence **40**(4): 834-848.
Cho, K., A. Ilin and T. Raiko (2011). Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines, Berlin, Heidelberg, Springer Berlin Heidelberg.
Ciresan, D. C., A. Giusti, L. M. Gambardella and J. Schmidhuber (2013). "Mitosis detection in breast cancer histology images with deep neural networks." Med Image Comput Comput Assist Interv **16**(Pt 2): 411-418.
Collobert, R., K. Kavukcuoglu and C. Farabet (2011). Torch7: A Matlab-like Environment for Machine Learning.
Collobert, R. and J. Weston (2008). A unified architecture for natural language processing: deep neural networks with multitask learning. Proceedings of the 25th international conference on Machine learning. Helsinki, Finland, ACM**:** 160-167.
Courville, A., J. Bergstra and Y. Bengio (2011). "A spike and slab Restricted Boltzmann Machine." Journal of Machine Learning Research **15**: 233-241.
Courville, A., J. Bergstra and Y. Bengio (2011). Unsupervised models of images by spike-and-slab RBMs. Proceedings of the 28th International Conference on Machine Learning, ICML 2011.
Courville, A., G. Desjardins, J. Bergstra and Y. Bengio (2014). "The spike-and-slab RBM and extensions to discrete and sparse data distributions." IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(9): 1874-1887.
Cun, Y. L., B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel and D. Henderson (1990). Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems 2. S. T. David, Morgan Kaufmann Publishers Inc.**:** 396-404.
Das, V. and N. B. Puhan (2017). "Tsallis entropy and sparse reconstructive dictionary learning for exudate detection in diabetic retinopathy." Journal of Medical Imaging **4**(2): 024002.
Decencière, E., G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard and R. Danno (2013). "TeleOphta: Machine learning and image processing methods for teleophthalmology." Irbm **34**(2): 196-203.
Decencière, E., G. Cazuguel, X. Zhang, G. Thibault, J. C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Laÿ and A. Chabouis (2013). "TeleOphta: Machine learning and image processing methods for teleophthalmology." IRBM **34**(2): 196-203.
DeMaagd, G. and A. Philip (2015). "Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis." P T **40**(8): 504-532.
Dice, L. R. (1945). "Measures of the Amount of Ecologic Association Between Species." Ecology **26**(3): 297-302.
Ege, B. M., O. K. Hejlesen, O. V. Larsen, K. Møller, B. Jennings, D. Kerr and D. A. Cavan (2000). "Screening for diabetic retinopathy using computer based image analysis and statistical classification." Computer methods and programs in biomedicine **62**(3): 165-175.

Esteva, A., B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature **542**: 115.
Farabet, C., C. Couprie, L. Najman and Y. Lecun (2013). "Learning hierarchical features for scene labeling." IEEE Trans Pattern Anal Mach Intell **35**(8): 1915-1929.
Fayek, H. M., M. Lech and L. Cavedon (2017). "Evaluating deep learning architectures for Speech Emotion Recognition." Neural Networks **92**: 60-68.
Fleming, A. D., S. Philip, K. A. Goatman, J. A. Olson and P. F. Sharp (2006). "Automated microaneurysm detection using local contrast normalization and local vessel detection." IEEE Trans Med Imaging **25**(9): 1223-1232.
Fleming, A. D., S. Philip, K. A. Goatman, G. J. Williams, J. A. Olson and P. F. Sharp (2007). "Automated detection of exudates for diabetic retinopathy screening." Physics in Medicine & Biology **52**(24): 7385.
Fraz, M. M., W. Jahangir, S. Zahid, M. M. Hamayun and S. A. Barman (2017). "Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification." Biomedical Signal Processing and Control **35**: 50-62.
Fu, H., Y. Xu, S. Lin, D. W. Kee Wong and J. Liu (2016). DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field, Cham, Springer International Publishing.
Fu, H., Y. Xu, D. W. K. Wong and J. Liu (2016). Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. 2016 IEEE 13th international symposium on biomedical imaging (ISBI), IEEE.
Fu, X., T. Liu, Z. Xiong, B. H. Smaill, M. K. Stiles and J. Zhao (2018). "Segmentation of histological images and fibrosis identification with a convolutional neural network." Computers in biology and medicine **98**: 147-158.
Garcia, M., M. I. Lopez, D. Alvarez and R. Hornero (2010). "Assessment of four neural network based classifiers to automatically detect red lesions in retinal images." Med Eng Phys **32**(10): 1085-1093.
García, M., C. I. Sánchez, M. I. López, D. Abásolo and R. Hornero (2009). "Neural network based detection of hard exudates in retinal images." Computer Methods and programs in biomedicine **93**(1): 9-19.
Gardner, G. G., D. Keating, T. H. Williamson and A. T. Elliott (1996). "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool." The British Journal of Ophthalmology **80**(11): 940-944.
Gargeya, R. and T. Leng (2017). "Automated Identification of Diabetic Retinopathy Using Deep Learning." Ophthalmology **124**(7): 962-969.
Giancardo, L., F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin Jr and E. Chaum (2012). "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets." Medical Image Analysis **16**(1): 216-226.
Glorot, X. and Y. Bengio (2010). Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. T. Yee Whye and T. Mike. Proceedings of Machine Learning Research, PMLR. **9:** 249--256.
Goodfellow, I., A. Courville and Y. Bengio (2012) "Large-Scale Feature Learning With Spike-and-Slab Sparse Coding." arXiv e-prints.
Goodfellow, I. J., A. Courville and Y. Bengio (2012) "Spike-and-Slab Sparse Coding for Unsupervised Feature Discovery." arXiv e-prints.
Goodfellow, I. J., D. Warde-Farley, M. Mirza, A. Courville and Y. Bengio (2013). "Maxout Networks."

Gracas, R. R., A. C. Gama, F. E. Cardoso, B. P. Lopes and I. B. Bassi (2012). "Objective and subjective analysis of women's voice with idiopathic Parkinson's disease." Arq Neuropsiquiatr **70**(7): 492-496.
Grassmann, F., J. Mengelkamp, C. Brandl, S. Harsch, M. E. Zimmermann, B. Linkohr, A. Peters, I. M. Heid, C. Palm and B. H. Weber (2018). "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography." Ophthalmology **125**(9): 1410-1420.
Grinsven, M. J. J. P. v., B. v. Ginneken, C. B. Hoyng, T. Theelen and C. I. Sánchez (2016). "Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images." IEEE Transactions on Medical Imaging **35**(5): 1273-1284.
Gulshan, V., L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega and D. R. Webster (2016). "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus PhotographsAccuracy of a Deep Learning Algorithm for Detection of Diabetic RetinopathyAccuracy of a Deep Learning Algorithm for Detection of Diabetic Retinopathy." JAMA **316**(22): 2402-2410.
Hajeb Mohammad Alipour, S., H. Rabbani and M. R. Akhlaghi (2012). "Diabetic Retinopathy Grading by Digital Curvelet Transform." Computational and Mathematical Methods in Medicine **2012**.
Hansen, M. B., M. D. Abramoff, J. C. Folk, W. Mathenge, A. Bastawrous and T. Peto (2015). "Results of Automated Retinal Image Analysis for Detection of Diabetic Retinopathy from the Nakuru Study, Kenya." PLoS One **10**(10): e0139148.
Harangi, B. and A. Hajdu (2014). "Automatic exudate detection by fusing multiple active contours and regionwise classification." Computers in biology and medicine **54**: 156-171.
Harangi, B. and A. Hajdu (2014). Detection of exudates in fundus images using a Markovian segmentation model. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE.
Harar, P., J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, R. Burget and Z. Smekal (2017). Voice Pathology Detection Using Deep Learning: a Preliminary Study. 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI).
He, K., X. Zhang, S. Ren and J. Sun (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.
He, K., X. Zhang, S. Ren and J. Sun (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Hinton, G., L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath and B. Kingsbury (2012). "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." IEEE Signal Processing Magazine **29**(6): 82-97.
Hinton, G. E. (2002). "Training products of experts by minimizing contrastive divergence." Neural Comput. **14**(8): 1771-1800.
Hinton, G. E. (2012). A Practical Guide to Training Restricted Boltzmann Machines. Neural Networks: Tricks of the Trade: Second Edition. G. Montavon, G. B. Orr and K.-R. Müller. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 599-619.
Hinton, G. E., S. Osindero and Y.-W. Teh (2006). "A Fast Learning Algorithm for Deep Belief Nets." Neural Computation **18**(7): 1527-1554.
Hinton, G. E. and R. R. Salakhutdinov (2006). "Reducing the dimensionality of data with neural networks." Science **313**(5786): 504-507.

Hinton, G. E. and R. R. Salakhutdinov (2006). "Reducing the Dimensionality of Data with Neural Networks." Science **313**(5786): 504-507.
Hove, M. N., J. K. Kristensen, T. Lauritzen and T. Bek (2004). "Quantitative analysis of retinopathy in type 2 diabetes: identification of prognostic parameters for developing visual loss secondary to diabetic maculopathy." Acta Ophthalmologica Scandinavica **82**(6): 679-685.
Hsu, W., P. Pallawala, M. L. Lee and K.-G. A. Eong (2001). The role of domain knowledge in the detection of retinal hard exudates. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, IEEE.
Imani, E. and H.-R. Pourreza (2016). "A novel method for retinal exudate segmentation using signal separation algorithm." Computer Methods and Programs in Biomedicine **133**: 195-205.
Imani, E., H.-R. Pourreza and T. Banaee (2015). "Fully automated diabetic retinopathy screening using morphological component analysis." Computerized Medical Imaging and Graphics **43**: 78-88.
Ioffe, S. and C. Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift."
Jaafar, H. F., A. K. Nandi and W. Al-Nuaimy (2010). Detection of exudates in retinal images using a pure splitting technique. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE.
Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell (2014). Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM international conference on Multimedia, ACM.
Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. Proceedings of the 22nd ACM international conference on Multimedia. Orlando, Florida, USA, ACM**:** 675-678.
Kahai, P., K. R. Namuduri and H. Thompson (2006). "A decision support framework for automated screening of diabetic retinopathy." Int J Biomed Imaging **2006**: 45806.
Kamal Maried, E., M. Abdalla Eldali, O. Omar Ziada and A. Baba (2017). A Literature Study of Deep learning and its application in Digital Image Processing.
Kauppi, T., V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen and J. Pietilä (2007). The diaretdb1 diabetic retinopathy database and evaluation protocol. BMVC.
Kauppi, T., V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen and J. Pietilä (2007). The DIARETDB1 Diabetic Retinopathy Database and Evaluation Protocol**:** 1-10.
Kaur, J. and D. Mittal (2018). "A generalized method for the segmentation of exudates from pathological retinal fundus images." Biocybernetics and Biomedical Engineering **38**(1): 27-53.
Khan, S. and T. Yairi (2018). "A review on the application of deep learning in system health management." Mechanical Systems and Signal Processing **107**: 241-265.
Khojasteh, P., B. Aliahmad, S. P. Arjunan and D. Kumar (2018). Introducing a Novel Layer in Convolutional Neural Network for Automatic Identification of Diabetic Retinopathy. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.
Khojasteh, P., B. Aliahmad and D. K. Kumar (2018). "Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms." BMC Ophthalmology **18**(1): 288.

Khojasteh, P., L. A. P. Júnior, T. Carvalho, E. Rezende, B. Aliahmad, J. P. Papa and D. K. Kumar (2019). "Exudate detection in fundus images using deeply-learnable features." Computers in biology and medicine **104**: 62-69.
Kim, E., M. Corte-Real and Z. Baloch (2016). A deep semantic mobile application for thyroid cytopathology, SPIE.
Kleesiek, J., G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus and A. Biller (2016). "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping." NeuroImage **129**: 460-469.
Krizhevsky, A., I. Sutskever and G. E. Hinton (2012). ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Lake Tahoe, Nevada, Curran Associates Inc.**:** 1097-1105.
Krizhevsky, A., I. Sutskever and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems.
Larochelle, H. and Y. Bengio (2008). Classification using discriminative restricted Boltzmann machines. Proceedings of the 25th international conference on Machine learning, ACM.
Lazar, I. and A. Hajdu (2013). "Retinal microaneurysm detection through local rotating cross-section profile analysis." IEEE Trans Med Imaging **32**(2): 400-407.
LeCun, Y., Y. Bengio and G. Hinton (2015). "Deep learning." Nature **521**: 436.
Lee, H., R. Grosse, R. Ranganath and A. Y. Ng (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th International Conference On Machine Learning, ICML 2009.
Lee, H., L. Yan, P. Pham and A. Y. Ng (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference.
Lee, H. J., S. H. Hong, M. Kim, S. Y. Ha, S. M. An and W. J. Shim (2011). "Tracing origins of sewage and organic matter using dissolved sterols in Masan and Haengam Bay, Korea." Ocean Science Journal **46**(2): 95-103.
Lee, V. S., X. P. Zhou, D. A. Rahn, 3rd, E. Q. Wang and J. J. Jiang (2008). "Perturbation and nonlinear dynamic analysis of acoustic phonatory signal in Parkinsonian patients receiving deep brain stimulation." J Commun Disord **41**(6): 485-500.
Leontidis, G., B. Al-Diri and A. Hunter (2017). "A new unified framework for the early detection of the progression to diabetic retinopathy from fundus images." Computers in biology and medicine **90**: 98-115.
Li, Q., W. Cai, X. Wang, Y. Zhou, D. D. Feng and M. Chen (2014). Medical image classification with convolutional neural network. 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV).
Li, Q., B. Feng, L. Xie, P. Liang, H. Zhang and T. Wang (2016). "A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images." IEEE Transactions on Medical Imaging **35**(1): 109-118.
Li, X., T. Pang, B. Xiong, W. Liu, P. Liang and T. Wang (2017). Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI).
Lin, M., Q. Chen and S. Yan (2013). "Network In Network."
Liskowski, P. and K. Krawiec (2016). "Segmenting Retinal Blood Vessels With Deep Neural Networks." IEEE Transactions on Medical Imaging **35**(11): 2369-2380.

Little, M. A., P. E. McSharry, S. J. Roberts, D. A. Costello and I. M. Moroz (2007). "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." Biomed Eng Online **6**: 23.
Liu, Q., B. Zou, J. Chen, W. Ke, K. Yue, Z. Chen and G. Zhao (2017). "A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images." Computerized Medical Imaging and Graphics **55**: 78-86.
Long, J., E. Shelhamer and T. Darrell (2015). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Long, J., E. Shelhamer and T. Darrell (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition.
Loong, T. W. "Understanding sensitivity and specificity with the right side of the brain." BMJ **327**(7417).
Mazo, C., J. Bernal, M. Trujillo and E. Alegre (2018). "Transfer Learning for Classification of Cardiovascular Tissues in Histological Images." Computer Methods and Programs in Biomedicine.
Mikolov, T., A. Deoras, D. Povey, L. Burget and J. Černocký (2011). Strategies for training large scale neural network language models. 2011 IEEE Workshop on Automatic Speech Recognition & Understanding.
Milletari, F., N. Navab and S. Ahmadi (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV).
Mittelman, R., B. Kuipers, S. Savarese and H. Lee (2014). "Structured Recurrent Temporal Restricted Boltzmann Machines." Proceedings of the 31St International Conference on Machine Learning **32**: 1647-1655.
Mo, J., L. Zhang and Y. Feng (2018). "Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks." Neurocomputing **290**: 161-171.
Mohamed, Q., M. C. Gillies and T. Y. Wong (2007). "Management of diabetic retinopathy: a systematic review." Jama **298**(8): 902-916.
Montavon, G., G. B. Orr and K.-R. Müller (2012). Neural Networks: Tricks of the Trade Second Edition. Berlin, Heidelberg, Springer Berlin Heidelberg : Imprint: Springer.
Montúfar, G., N. Ay and K. Ghazi-Zahedi (2015). "Geometry and expressive power of conditional restricted Boltzmann machines." Journal of Machine Learning Research **16**: 2405-2436.
Montúfar, G. and J. Morton (2015). "Discrete restricted Boltzmann machines." Journal of Machine Learning Research **16**: 653-672.
Mookiah, M. R. K., U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng and A. Laude (2013). "Computer-aided diagnosis of diabetic retinopathy: A review." Computers in biology and medicine **43**(12): 2136-2155.
Murphy, K., A. Torralba, D. Eaton and W. Freeman (2006). Object Detection and Localization Using Local and Global Features. Toward Category-Level Object Recognition. J. Ponce, M. Hebert, C. Schmid and A. Zisserman. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 382-400.
Nair, V. and G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10).
Naqvi, S. A. G., M. F. Zafar and I. u. Haq (2015). "Referral system for hard exudates in eye fundus." Computers in Biology and Medicine **64**: 217-235.
Naqvi, S. A. G., M. F. Zafar and I. ul Haq (2015). "Referral system for hard exudates in eye fundus." Computers in biology and medicine **64**: 217-235.

Nash, W., T. Drummond and N. Birbilis (2018). "A review of deep learning in the study of materials degradation." npj Materials Degradation **2**(1): 37.
Niemeijer, M., B. van Ginneken, S. R. Russell, M. S. Suttorp-Schulten and M. D. Abramoff (2007). "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.
Niemeijer, M., B. van Ginneken, J. Staal, M. S. Suttorp-Schulten and M. D. Abramoff (2005). "Automatic detection of red lesions in digital color fundus photographs." IEEE Trans Med Imaging **24**(5): 584-592.
Norouzi, M., M. Ranjbar and G. Mori (2009). Stacks of convolutional Restricted Boltzmann Machines for shift-invariant feature learning. 2009 IEEE Conference on Computer Vision and Pattern Recognition.
Oliveira, A., S. Pereira and C. A. Silva (2018). "Retinal vessel segmentation based on Fully Convolutional Neural Networks." Expert Systems with Applications **112**: 229-242.
Osareh, A., M. Mirmehdi, B. Thomas and R. Markham (2003). "Automated identification of diabetic retinal exudates in digital colour images." British Journal of Ophthalmology **87**(10): 1220.
Pachiyappan, A., U. N. Das, T. V. Murthy and R. Tatavarti (2012). "Automated diagnosis of diabetic retinopathy and glaucoma using fundus and OCT images." Lipids in health and disease **11**(1): 73.
Pak, M. and S. Kim (2017). A review of deep learning in image recognition. 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT).
Pang, S., Z. Yu and M. A. Orgun (2017). "A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images." Computer Methods and Programs in Biomedicine **140**: 283-293.
Papa, J. P., A. X. FalcãO, V. H. C. De Albuquerque and J. M. R. Tavares (2012). "Efficient supervised optimum-path forest classification for large datasets." Pattern Recognition **45**(1): 512-520.
Papa, J. P., A. X. Falcao and C. T. Suzuki (2009). "Supervised pattern classification based on optimum-path forest." International Journal of Imaging Systems and Technology **19**(2): 120-131.
Papa, J. P., S. E. N. Fernandes and A. X. Falcao (2017). "Optimum-path forest based on k-connectivity: Theory and applications." Pattern Recognition Letters **87**: 117-126.
Perdomo, O., S. Otalora, F. Rodríguez, J. Arevalo and F. A. González (2016). "A novel machine learning model based on exudate localization to detect diabetic macular edema."
Pereira, C., L. Gonçalves and M. Ferreira (2015). "Exudate segmentation in fundus images using an ant colony optimization approach." Information Sciences **296**: 14-24.
Pompili, A., A. Abad, P. Romano, I. P. Martins, R. Cardoso, H. Santos, J. Carvalho, I. Guimarães and J. J. Ferreira (2017). Automatic Detection of Parkinson's Disease: An Experimental Analysis of Common Speech Production Tasks Used for Diagnosis, Cham, Springer International Publishing.
Prentašić, P. and S. Lončarić (2015). Detection of exudates in fundus photographs using convolutional neural networks. 2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA).
Prentašić, P. and S. Lončarić (2016). "Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion." Computer methods and programs in biomedicine **137**: 281-292.
Quellec, G., K. Charrière, Y. Boudi, B. Cochener and M. Lamard (2017). "Deep image mining for diabetic retinopathy screening." Medical image analysis **39**: 178-193.

Quellec, G., M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener and C. Roux (2008). "Optimal wavelet transform for the detection of microaneurysms in retina photographs." IEEE Trans Med Imaging **27**(9): 1230-1241.
Rahn, D. A., 3rd, M. Chou, J. J. Jiang and Y. Zhang (2007). "Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis." J Voice **21**(1): 64-71.
Ranzato, M. and G. E. Hinton (2010). Modeling pixel means and covariances using factorized third-order boltzmann machines. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
Ranzato, M., A. Krizhevsky and G. E. Hinton (2010). "Factored 3-way restricted Boltzmann machines for modeling natural images." Journal of Machine Learning Research **9**: 621-628.
Reed, S., K. Sohn, Y. Zhang and H. Lee (2014). "Learning to disentangle factors of variation with manifold interaction." Learning to Disentangle Factors of Variation with Manifold Interaction: 1431-1439.
Ronneberger, O., P. Fischer and T. Brox (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, Cham, Springer International Publishing.
Roth, H. R., C. T. Lee, H. C. Shin, A. Seff, L. Kim, J. Yao, L. Lu and R. M. Summers (2015). Anatomy-specific classification of medical images using deep convolutional nets. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI).
Roychowdhury, S., D. D. Koozekanani and K. K. Parhi (2014). "DREAM: Diabetic Retinopathy Analysis Using Machine Learning." IEEE Journal of Biomedical and Health Informatics **18**(5): 1717-1728.
Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei (2015). "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision **115**(3): 211-252.
Sainath, T. N., A. Mohamed, B. Kingsbury and B. Ramabhadran (2013). Deep convolutional neural networks for LVCSR. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.
Sakar, B. E., M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin and O. Kursun (2013). "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings." IEEE Journal of Biomedical and Health Informatics **17**(4): 828-834.
Salakhutdinov, R. and H. Larochelle (2010). Efficient Learning of Deep Boltzmann Machines. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. T. Yee Whye and T. Mike. Proceedings of Machine Learning Research, PMLR. **9:** 693--700.
Sánchez, C. I., M. García, A. Mayo, M. I. López and R. Hornero (2009). "Retinal image analysis based on mixture models to detect hard exudates." Medical Image Analysis **13**(4): 650-658.
Sarkar, J., S. Saha and S. Agrawal (2014). "An efficient use of principal component analysis in workload characterization-A study." AASRI Procedia **8**: 68-74.
Schmidhuber, J. (1992). "Learning Complex, Extended Sequences Using the Principle of History Compression." Neural Computation **4**(2): 234-242.
Schmidhuber, J. (2015). "Deep learning in neural networks: An overview." Neural Networks **61**: 85-117.
Shan, J. and L. Li (2016). A Deep Learning Method for Microaneurysm Detection in Fundus Images. 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE).

Shaw, J. E., R. A. Sicree and P. Z. Zimmet (2010). "Global estimates of the prevalence of diabetes for 2010 and 2030." Diabetes Res Clin Pract **87**(1): 4-14.
Shelhamer, E., J. Long and T. Darrell (2017). "Fully Convolutional Networks for Semantic Segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4): 640-651.
Shen, D., G. Wu and H.-I. Suk (2017). "Deep Learning in Medical Image Analysis." Annual Review of Biomedical Engineering **19**(1): 221-248.
Shuang Yu, Y., Y. Di Xiao and Y. Kanagasingam (2017). "Exudate detection for diabetic retinopathy with convolutional neural networks." Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference **2017**: 1744.
Shyu, C. R., C. E. Brodley, A. C. Kak, A. Kosaka, A. Aisen and L. Broderick (1998). Local versus global features for content-based image retrieval. Proceedings. IEEE Workshop on Content-Based Access of Image and Video Libraries (Cat. No.98EX173).
Simonyan, K. and A. Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition."
Sinthanayothin, C., J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal and D. Usher (2002). "Automated detection of diabetic retinopathy on digital fundus images." Diabet Med **19**(2): 105-112.
Smolensky, P. (1986). Information processing in dynamical systems: foundations of harmony theory. Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. E. R. David, L. M. James and C. P. R. Group, MIT Press**:** 194-281.
Soille, P. (2003). Morphological Image Analysis: Principles and Applications, Springer-Verlag New York, Inc.
Sopharak, A., B. Uyyanonvara, S. Barman and T. H. Williamson (2008). "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods." Computerized Medical Imaging and Graphics **32**(8): 720-727.
Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov (2014). "Dropout: a simple way to prevent neural networks from overfitting." J. Mach. Learn. Res. **15**(1): 1929-1958.
Suk, H.-I. and D. Shen (2015). Deep Learning in Diagnosis of Brain Disorders. Recent Progress in Brain and Cognitive Engineering. S.-W. Lee, H. H. Bülthoff and K.-R. Müller. Dordrecht, Springer Netherlands**:** 203-213.
Sutskever, I., G. Hinton and G. Taylor (2009). The recurrent temporal restricted boltzmann machine. Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference.
Sutskever, I. and G. E. Hinton (2007). "Learning multilevel distributed representations for high-dimensional sequences." AISTATS: 544-551.
Sutskever, I., J. Martens and G. Hinton (2011). Generating text with recurrent neural networks. Proceedings of the 28th International Conference on International Conference on Machine Learning. Bellevue, Washington, USA, Omnipress**:** 1017-1024.
Swersky, K., D. Tarlow, I. Sutskever, R. Salakhutdinov, R. S. Zemel and R. P. Adams (2012). Cardinality restricted boltzmann machines. Advances in Neural Information Processing Systems.
Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich (2014). "Going Deeper with Convolutions."

Szegedy, C., L. Wei, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Tajbakhsh, N., J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway and J. Liang (2016). "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" IEEE Transactions on Medical Imaging **35**(5): 1299-1312.
Tan, C., F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu (2018) "A Survey on Deep Transfer Learning." arXiv e-prints.
Tan, J. H., H. Fujita, S. Sivaprasad, S. V. Bhandary, A. K. Rao, K. C. Chua and U. R. Acharya (2017). "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network." Information sciences **420**: 66-76.
Tang, L., M. Niemeijer, J. M. Reinhardt, M. K. Garvin and M. D. Abramoff (2013). "Splat feature classification with application to retinal hemorrhage detection in fundus images." IEEE Trans Med Imaging **32**(2): 364-375.
Tiulpin, A., J. Thevenot, E. Rahtu, P. Lehenkari and S. Saarakkala (2018). "Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach." Scientific reports **8**(1): 1727-1727.
Tompson, J., A. Jain, Y. LeCun and C. Bregler (2014). Joint training of a convolutional network and a graphical model for human pose estimation. Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1. Montreal, Canada, MIT Press**:** 1799-1807.
Tsanas, A., M. A. Little, P. E. McSharry and L. O. Ramig (2010). "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity." Journal of The Royal Society Interface.
Unnikrishnan, P., B. Aliahmad, R. Kawasaki and D. Kumar (2013). Adaptive colour transformation of retinal images for stroke prediction. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.
van Gerven, M. and S. Bohte (2017). "Editorial: Artificial Neural Networks as Models of Neural Information Processing." Frontiers in computational neuroscience **11**: 114-114.
van Grinsven, M. J., B. van Ginneken, C. B. Hoyng, T. Theelen and C. I. Sanchez (2016). "Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images." IEEE Trans Med Imaging **35**(5): 1273-1284.
van Grinsven, M. J., F. Venhuizen, B. van Ginneken, C. C. Hoyng, T. Theelen and C. I. Sanchez (2016). "Automatic detection of hemorrhages on color fundus images using deep learning." Investigative Ophthalmology & Visual Science **57**(12): 5966-5966.
van Grinsven, M. J. J. P., F. Venhuizen, B. van Ginneken, C. C. B. Hoyng, T. Theelen and C. I. Sanchez (2016). "Automatic detection of hemorrhages on color fundus images using deep learning." Investigative Ophthalmology & Visual Science **57**(12): 5966-5966.
Vincent, P., H. Larochelle, I. Lajoie, Y. Bengio and P.-A. Manzagol (2010). "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion." J. Mach. Learn. Res. **11**: 3371-3408.
Viswanathan, R., P. Khojasteh, B. Aliahmad, S. P. Arjunan, S. Ragnav, P. Kempster, K. Wong, J. Nagao and D. K. Kumar (2018). Efficiency of Voice Features Based on Consonant for Detection of Parkinson's Disease. 2018 IEEE Life Sciences Conference (LSC).

Walter, T., J.-C. Klein, P. Massin and A. Erginay (2002). "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina." IEEE transactions on medical imaging **21**(10): 1236-1243.
Walter, T., J. C. Klein, P. Massin and A. Erginay (2002). "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina." IEEE Transactions on Medical Imaging **21**(10): 1236-1243.
Walter, T., P. Massin, A. Erginay, R. Ordonez, C. Jeulin and J.-C. Klein (2007). "Automatic detection of microaneurysms in color fundus images." Medical Image Analysis **11**(6): 555-566.
Wang, Z., A. C. Bovik, H. R. Sheikh and E. P. Simoncelli (2004). "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing **13**(4): 600-612.
Watson, G. S. and J. B. Leverenz (2010). "Profile of cognitive impairment in Parkinson's disease." Brain Pathol **20**(3): 640-645.
Welfer, D., J. Scharcanski and D. R. Marinho (2010). "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images." Computerized Medical Imaging and Graphics **34**(3): 228-235.
Wu, G., M. Kim, Q. Wang, B. C. Munsell and D. Shen (2016). "Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning." IEEE Transactions on Biomedical Engineering **63**(7): 1505-1516.
Yu, S., D. Xiao and Y. Kanagasingam (2017). Exudate detection for diabetic retinopathy with convolutional neural networks. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.
Zaki, W. M. D. W., M. A. Zulkifley, A. Hussain, W. H. W. Halim, N. B. A. Mustafa and L. S. Ting (2016). "Diabetic retinopathy assessment: Towards an automated system." Biomedical Signal Processing and Control **24**: 72-82.
Zaki, W. M. D. W., M. A. Zulkifley, A. Hussain, W. H. W. A. Halim, N. B. A. Mustafa and L. S. Ting (2016). "Diabetic retinopathy assessment: Towards an automated system." Biomedical Signal Processing and Control **24**: 72-82.
Zana, F. and J. C. Klein (1997). Robust segmentation of vessels from retinal angiography. Proceedings of 13th International Conference on Digital Signal Processing.
Zhang, H. H., L. Yang, Y. Liu, P. Wang, J. Yin, Y. Li, M. Qiu, X. Zhu and F. Yan (2016). "Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples." Biomed Eng Online **15**(1): 122.
Zhang, W., R. Li, H. Deng, L. Wang, W. Lin, S. Ji and D. Shen (2015). "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation." NeuroImage **108**: 214-224.
Zhang, X., G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard and P. Massin (2014). "Exudate detection in color retinal images for mass screening of diabetic retinopathy." Medical image analysis **18**(7): 1026-1043.
Zhao, L. and K. Jia (2016). "Multiscale CNNs for Brain Tumor Segmentation and Diagnosis." Computational and Mathematical Methods in Medicine **2016**.
Zhou, S., H. Greenspan and D. Shen (2017). Deep Learning for Medical Image Analysis, Academic Press.