AUTONOMOUS RETINAL IMAGE ANALYSIS AND CONTENT-BASED RETRIEVAL SYSTEM FOR DIAGNOSING DIABETIC RETINOPATHY USING DEEP CONVOLUTIONAL FEATURE EXTRACTION

W.O.K.I.S. Wijesinghe

188081J

Degree of Master of Science

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

November 2019

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or another medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

Name: W.O.K.I.S. Wijesinghe

The above candidate has carried out research for the Masters thesis/dissertation under my supervision.

Signature of the supervisor: Date:

Name of the supervisor: Dr. Charith Chitraranjan

ABSTRACT

The automatic classification and content-based image retrieval (CBIR) for a given retinal image of diabetic retinopathy (DR) are very essential since this is the leading source of permanent loss of vision in the working-age individuals all over the world today. Current clinical approaches require a well-trained clinician to manually evaluate fundus photographs of retina and locate lesions associated with vascular abnormalities due to diabetes, which is time-consuming. The principal objective of this research is to classify the severity level and retrieve semantically similar retinal imageries to a given query image for effective treatment.

Recently, deep CNN-based feature extraction has been used to predict DR from fundus images with reasonable accuracy whereas effective and comprehensive deep retinal image retrieval model for DR is not available in the literature. However, techniques such as singular value decomposition (SVD), global average pooling (GAP) and ensemble learning have not been used in automatic prediction of DR.

In this research, it is suggested a combination of deep features extracted from an ensemble of pretrained-CNNs (VGG-16, ResNet-18, and DenseNet-201) as a single feature vector to accomplish the research objectives. The experimental outcomes of this research demonstrate a promising accuracy of over 98% for both tasks. A classification model was built as the first step and then it was extended it to a retrieval model by using a deep supervised hashing approach in order to perform efficient retinal image retrieval, where it implicitly learn a good image representation along with a similarity-preserving compact binary hash code for each image. This research was evaluated using prominent CNN architectures (VGG, ResNet, InceptionResNetV2, InceptionV3, Xception, and DenseNet) that can be used for transfer learning. Moreover, GAP and SVD were used as dimensional reduction techniques in order to diminish processing time and memory utilization while preserving classification accuracy and retrieval performance.

ACKNOWLEDGEMENTS

It is a great pleasure for me to acknowledge the assistance and contributions of all the people who helped me to make my research success. My research would not have been accomplished without the dedicated assistance given by those individuals.

First of all, I would like to express my genuine gratitude to my supervisor, Dr. Charith Chitraranjan for his invaluable and constructive support in providing related knowledge, encouragement, advice and kind co-operation throughout the research. This would not have been a success without his continuous and incredible assistance during the research.

Next, I would like to bestow my honor to Dr. Manel Pasquel who is an eye-consultant at the National Eye Hospital for her continuous guidance to accomplish my research. Her continuous inspiration and domain knowledge helped me during the entire research. Furthermore, my heartfelt gratitude is expressed towards Mr. Amila Chandrasekara who is an ophthalmic assistant of Vision Care (Pvt) Ltd to give his helping hands to overcome the problems on all occasions.

This entire research project was funded by the Senate Research Grant of the University of Moratuwa, and I honestly thank for its financial support.

Moreover, I kindly appreciate the feedbacks, patience, motivation, and enthusiasm provided by my colleagues to accomplish my research. Finally, I would like to give my special thanks to my lovely parents and family members who helped and supported me in various ways to successfully complete my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAO	American Academy of Ophthalmology
AUC	Area Under Curve
CNN	Convolutional Neural Network
ANN	Artificial Neural Network
DR	Diabetic Retinopathy
CBIR	Content-based Image Retrieval
GAP	Global Average Pooling
SVD	Singular Value Decomposition
SGD	Stochastic Gradient Descent
mAP	mean Average Precision
SVM	Support Vector Machine
RBF	Radial Basis Function
NPDR	Non-Proliferative Diabetic Retinopathy
PDR	Proliferative Diabetic Retinopathy
KSH	Supervised Hashing with Kernels
MLH	Minimal Loss Hashing
SH	Spectral Hashing
LSH	Locality Sensitive Hashing
ReLU	Rectified Linear Unit
LBP	Local Binary Patterns
DT-CWT	Dual-Tree Complex Wavelet Transform
LGN	Lateral Geniculate Nucleus
OCT	Optical Coherent Tomography
WHO	World Health Organization
MA	Microaneurysm

1. INTRODUCTION

1.1 Background

During the past few decades, diagnosing diseases by analyzing medical images has been a prominent research field in the biomedical area. The rapid growth of the volume of real-world data collected through medical treatments has produced incredible excitement in the healthcare domain. Digital colour fundus imagery provides a significant effect in order to develop novel bits of knowledge and disrupt the concerns of retinal diseases among this gathered information.

Today, medical imaging is broadly used for diagnosing diseases, prioritizing treatments and judging responses to treatments. One of the key reasons is that the workload significantly increases for a specialist because of an extensive number of patients taking part in the disease screening process and thus patients should stay at the hospitals for a very long period of time. For example, the number of diabetic patients is growing every year making it problematic for the health-care system to frequently diagnose complications such as diabetic retinopathy by analyzing retinal fundus images through a screening process and provide necessary instructions in order to minimize the risk of life-long conditions such as vision loss. DR is the most common vision-threatening retinal disease due to diabetes over a prolonged period of time. The following is a brief background description of the anatomy of the human vision system, retinal fundus photography and diabetic retinopathy.

1.1.1 Anatomy of the Human Vision System

The human vision system includes three key functional parts namely the eye, the LGN (lateral geniculate nucleus) and the visual cortex which is a part of the cerebral cortex in the brain that deals with visual information. The eye is a roughly spherical and sensitive organ which is responsible for all the visual information that passes to the brain. The human brain performs complex image processing whereas the eye acts as a biologically equivalent camera.

The light rays that enter the eye through the cornea, pupil, and lens, subsequently pass over the vitreous (clear gel-like substance that fills the middle of the eye) before concentrating on the surface of the retina. The retina, which is a light-sensitive tissue that detects light rays, converts them into electrochemical signals using photoreceptors (rods and cones) and sends them through the optic nerve to the visual centers in the brain [1]. The LGN, which is the principal central connection for the optic nerve to the visual cortex, collects visual information directly from the ganglion cells in the retina (see Figure 1.1).

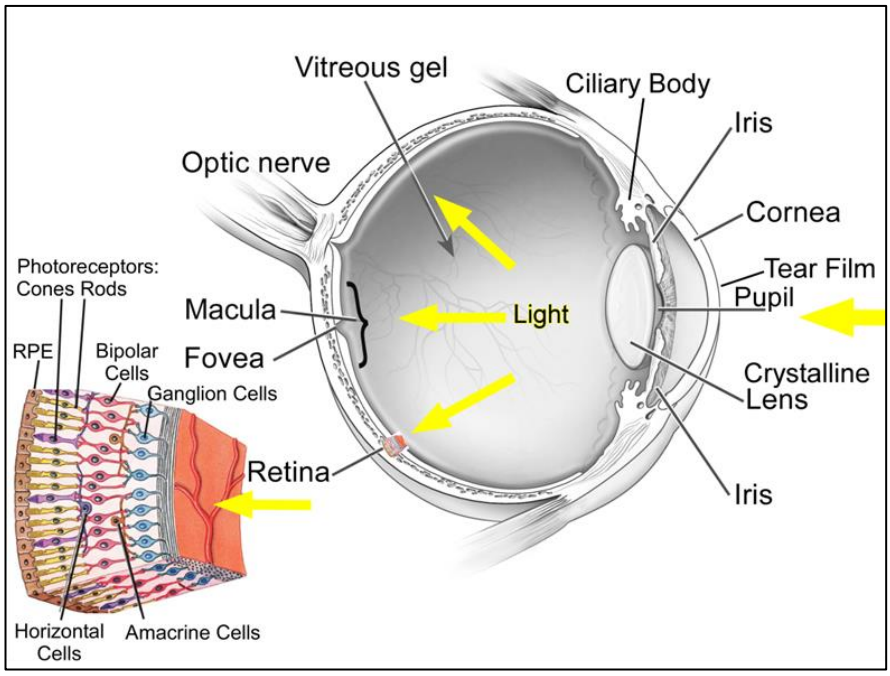


Figure 1.1: Anatomy of the human eye [1]

1.1.2 Retinal Fundus Photography

There are diverse diagnostic tools available in order to capture the interior structure of the human eye including digital colour fundus photography and OCT (Optical Coherent Tomography). These aforementioned two approaches are non-invasive imagery tests. OCT delivers cross-sectional pictures of the retina. However, retinal fundus photography affords images of the interior structure of the human eye covering the retina, optic disc, blood vessels, macula and fovea [2]. The OCT imagery is supportive if eye diseases are identified in ocular tissues. In contrast,

digital colour fundus photographs are very effective if the eye diseases are detected at the retina [2]. Moreover, retinal fundus imaging is the quicker and easier technique to observe the abnormal features of the retina.

Fundus photography is most frequently used for early disease identification and clinical educations. There are two types of retinal fundus photography namely standard and wide-field. Standard colour fundus photographs, which capture 30 degrees of the posterior pole of a patient eye including the macula and the optic nerve whereas wide-field colour fundus photographs capture the seven fundus fields of a patient eye and combined together to generate a montage image that displays a 75-degree of view. The left-side image in Figure 1.2 is an example for a standard fundus photograph and the right-side is an example for a wide-field photograph.

Figure 1.2: Standard and wide-field fundus photographs

1.1.3 Diabetic Retinopathy

Diabetic Retinopathy (DR) is a retinal disease that can affect individuals with diabetes. It occurs due to the presence of high glucose levels in the blood, bringing harm to the small veins in the human retina. There are diverse types of abnormal lesions that occur due to DR such as microaneurysms, hemorrhages, soft exudates, hard exudates, and neovascularization. These are extremely crucial in order to classify whether images show clinical signs of the disease. Microaneurysms, the minor bulges that form on the tiny blood vessels, are the earliest clinically detectable lesions through retinal fundus photographs. Neo-vascularization occurs in the PDR

(Proliferative Diabetic Retinopathy) level and formation of new fragile blood vessels, causing hemorrhages. These hemorrhages may cause severe vision difficulties. Hard exudates are protein and lipid formations leaked from damaged blood vessels and appear in yellow coloured clusters in the retinal surface. Soft exudates are due to obstruction of retinal arterioles [3].

DR will lead to blindness if untreated whereas timely treatment can stop or slow down the loss of vision. Therefore, people with diabetes should undergo regular eye screening for DR. Typically well-prepared specialists and ophthalmologists utilize a five-class severity scale, as shown in Table 1.1, to depict the severity grading of DR, to be specific diabetes without retinopathy, Mild-NPDR (Mild non-proliferative DR), Moderate-NPDR (Moderate non-proliferative DR), Severe-NPDR (Severe non-proliferative DR) and PDR [4]. Figure 1.3 shows an example retinal imagery captured from ophthalmoscope/funduscope for each severity stage indicated in Table 1.1.

Table 1.1: Diabetic Retinopathy Severity Stages

Disease Severity Level	**Findings Observable via Ophthalmoscopy**
Diabetes without Retinopathy	No visible signs of abnormalities
Mild-NPDR	Presence of MAs only
Moderate-NPDR	More than MAs and less than severe NPDR
Severe-NPDR	Any of the clinical symptoms below: • >twenty intraretinal hemorrhages • Venous bleeding • Intraretinal microvascular abnormalities • No symptoms of PDR
PDR	Any or all of the following: • Neo-vascularization • Vitreous hemorrhage

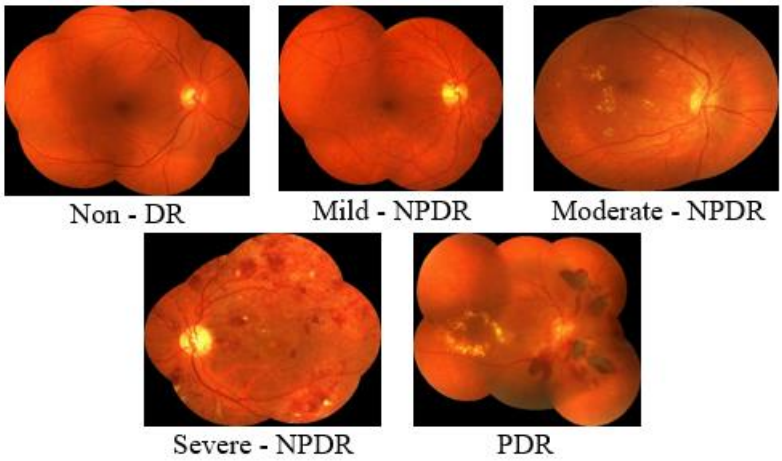


Figure 1.3: Severity stages of DR

1.2 Motivation for the Research

According to a report provided by the WHO (World Health Organization), DR had affected almost 145 million people (35% of those with diabetes) all over the world in 2015, with almost 45 million people suffering from PDR that could lead to severe vision loss [5]. Today, most of the individuals who are suffering from DR all over the world do not undergo regular documented screening according to the guidelines provided by the AAO [4].

Early diagnosis of DR by analyzing retinal images is in high demand as numerous people are left out from the healthcare centers due to restricted assets, particularly in rural areas, such as qualified clinicians or appropriate equipment. In contrast, the traditional DR diagnosing system requires a manual assessment process, which is tedious and depends heavily on the skill of ophthalmologists and well-prepared specialists. The current framework will turn out to be significantly inadequate [6] as the number of individuals with diabetes increases. Hence, automatic severity stage classification and similar case(s) retrieval from a retinal image database can be used for screening and treatment prioritization in order to assist and accelerate the clinical decision-making process for DR to diminish irreversible vision loss among diabetic patients.

1.3 Research Statement

Previous studies have explored the use of machine learning and image processing techniques for automatic classification and CBIR systems of DR [7][8][9]. When we consider the performance of previous studies proposed by numerous research groups, there is space for further improvement of classification and retrieval models of DR and it can be done by tuning hyperparameters or an ensemble of pretrained CNNs as feature extractors or using an ensemble learning approach through weak learners. In the recent past, deep feature extraction using pretrained-CNNs has been used to predict the five severity stages of DR from fundus images with reasonable accuracy whereas an effective and comprehensive deep retinal image retrieval model for DR is not available in the literature.

Hence, the principal research questions to be addressed by this study can be formulated as:

1. *Improve the performance (accuracy) of DR classification model through an ensemble Deep CNN approach*
2. *Improve the mAP (mean Average Precision) of CBIR through a novel deep supervised hashing technique*

1.4 Objective of the Research

Recently, deep convolutional neural networks have manifested superior performance in image classification and content-based image retrieval particularly in the biomedical field compared to conventional feature extraction-based image classification and retrieval methods. Hence, we utilize deep learning strategies so as to achieve the research objectives. The principal objectives of this research are as per the following:

1. Build a prediction model to classify the severity level of Diabetic Retinopathy using retinal images through an ensemble of deep CNNs. This can be used for treatment prioritization and automated screening.

2. Build a content-based image retrieval system to search the collections for retinal images that have characteristics similar to the case(s) of interest because access to clinically relevant stored data will allow for more informed and effective treatment.

1.5 Overview of Research Methodology

This section describes a brief overview of the research methodology. The extra black margins were removed in the retinal imagery as the first step of the preprocessing stage and then we transformed them in such a way that it would be feasible for any CNN to converge in a reasonable time by rescaling each image into 224px x 224px. Subsequently, we evaluate six prominent pretrained-CNN architectures to determine the best performing CNN for the severity stage classification task and aim to improve the performance compared to the current-state-of the art approaches We propose an ensemble of deep feature extraction technique by applying global average pooling (GAP) to the last pooling layer of each pretrained CNN and then apply singular value decomposition (SVD) for improved prediction of DR. We use GAP and SVD as dimensional reduction techniques in order to reduce memory consumption and processing time while preserving classification performance.

As the next step, the classification model is extended to a retrieval model by using a novel deep supervised hashing approach in order to perform efficient retinal image retrieval, where it implicitly learns a good image representation along with a similarity-preserving compact binary hash code for each image. This approach maps the image pixels to a lower-dimensional space and then generates compact binary codes to speed up the retrieval process. We use hamming distance to retrieve a group of candidate retinal images from the retinal database with similar compact binary codes for a given query image. The cosine similarity is used over the lower-dimensional feature space in order to further filter the retrieved candidate list since identical compact binary hash codes may produce for clinically similar retinal images.

1.6 Contributions and Research Articles

As the first part of this research, a novel ensemble deep CNN architecture has been proposed in order to help early diagnosis of diabetic retinopathy by classifying a given retinal image based on its severity stage. This prediction model has a significant clinical implication for early disease diagnosis since the proliferation of DR can occasionally be speedy and leading to irreversible vision loss due to blood vessel damage in the retina.

As the second part of this research, we have developed a content-based retinal image retrieval model based on a novel deep supervised hashing technique by extending the aforementioned classification model. This would allow clinicians to retrieve clinically similar images from a database of retinal images for a given image (Top k image retrieval).

These two approaches are beneficial for practitioners to diagnose the progression of the disease more accurately while prioritizing the treatment plans for the patients.

Moreover, the following research papers have been accepted and presented so far.

- "Transfer Learning with Ensemble Feature Extraction and Low-rank Matrix Factorization for Severity Stage Classification of Diabetic Retinopathy", W.O.K.I.S. Wijesinghe, H.V.L.C. Gamage, C. Chitraranjan, Accepted and presented at the 31st International Conference on Tools with Artificial Intelligence (ICTAI), USA, 2019.
- "Deep Supervised Hashing through Ensemble CNN Feature Extraction and Low-rank Matrix Factorization for Retinal Image Retrieval of Diabetic Retinopathy", W.O.K.I.S. Wijesinghe, H.V.L.C. Gamage, C. Chitraranjan, Accepted and presented at the 19th International Conference on BioInformatics and BioEngineering (BIBE), Greece, 2019.
- "A Smart Telemedicine System with Deep Learning to Manage Diabetic Retinopathy and Foot Ulcers", W.O.K.I.S. Wijesinghe, H.V.L.C. Gamage, I. Perera, C. Chitraranjan, Accepted and presented at the Moratuwa Engineering Research Conference (MERCon), 2019. (Collaborative Research Work)

1.7 Organization of the Thesis

The topics covered in this dissertation is structured as per the following. Chapter two encompasses the previous studies that have explored the use of image processing and machine learning techniques for automatic DR classification and content-based retinal image retrieval. Chapter three explains the datasets that we used during our research. Chapter four describes the methodology of the autonomous DR classification and CBIR model architectures in detail. Chapter five outlines not only the experimental analysis but also the evaluations of our methodology for each task in terms of accuracy, F1-measure, and mAP (mean Average Precision) and compares them with other state-of-the-art approaches. The final chapter, chapter six summarizes our findings and concludes the overall results by comparing it with other recently published works.

2. LITERATURE REVIEW

This chapter elaborates on the literature related to automated fundus image analysis for the assessment of diabetic retinopathy and divided into three main sections. Section 2.1 describes deep learning techniques, particularly convolutional neural networks that were used for image analysis tasks. Section 2.2 explicates the content-based image retrieval techniques and its limitations that have been studied during the past few decades including semantic compact binary hash code embedding approaches in order to retrieve similar case(s) for given query imagery. Section 2.3 explains the diabetic retinopathy analysis methods and their limitations that have been published in previous years based on severity stage classification, abnormal lesion detection methods, retinal blood vessel segmentation methods and CBIR system to retrieve a similar case(s) of DR in retinal images. Finally, section 2.4 comprises a summary of this chapter.

2.1 Deep Learning for Image Analysis

The computerized techniques for analyzing the images are challenging tasks as a result of the heterogeneity and complexity of digital colour imagery. Human intervention usually requires for diagnosing ailment by recognizing the most distinctive features through images [10]. Machine learning methods [11] demonstrate better performance through supervised and unsupervised learning techniques by addressing the aforementioned challenges. Recently, Convolution Neural Networks (CNNs) show their remarkable performance and impressive learning power in analyzing numerous types of images including medical images [12] where tasks that are heavily dependent on feature extraction, such as image classification and localization [13], video analysis [14], object detection [15] and other numerous tasks such as segmentation [16]. CNN based approaches normally beat other different methodologies in the previously mentioned fields, which proves that CNNs can capture the semantic information of the imagery by learning robust features. Hence, the most reasonable route is to utilize deep learning to learn semantic features for the

image datasets. The succeeding subsections enlighten about Convolution Neural Networks and the parameters required to implement different CNN architectures.

2.1.1 Convolutional Neural Networks

In Artificial Neural Networks (ANN), the most commonly applied type in the domain of image classification, object detection, and CBIR tasks is the Convolutional Neural Network (CNN) [17]. The effectiveness and efficiency of CNNs in image recognition tasks are one of the key reasons why the research communities in Artificial Intelligence has woken up to the usefulness of deep learning in the recent past. CNNs are notable designs in deep learning area that inspired by the natural visual perception mechanism of living animals. The history behind the Convolutional Neural Networks begins with a research experiment done by two scientists namely, Wiesel and Hubel in 1959. They discovered that the biological cells in the visual cortex, which process visual stimuli of animals, are responsible in order to detect light in the receptive fields. LeCun et al. [18] published research work in the early 90s by introducing a novel neural network architecture, LeNet-5, in order to classify the handwritten digits. During the model training process, they used the backpropagation algorithm in order to learn the weights. However, the proposed deep CNN architecture did not perform well due to limited resources such as computational power and data availability. In the past few decades, many research groups have been implemented with various techniques in order to overcome the difficulties faced during the training process of CNNs. Krizhevsky et al. [19] introduced AlexNet, a CNN architecture, which was won the ImageNet challenge in 2012 and showed significant performance and learning power compared to the previous approaches. Numerous deep CNN architectures have been developed by various research groups to obtain more accurate experimental results, to be specific VGGNet [20], InceptionV3 [21], Xception [22], ResNet [23], and DenseNet [24] with the success of AlexNet architecture.

A CNN typically consists of two basic parts namely, a feature extractor or a feature learning part and a classifier (see Figure 2.1). Feature extractor includes an input

layer, multiple stages of convolutional layers with filters (Kernels) followed by an activation function and pooling layers where classifier consists of a standard MLP (Multilayer Perceptron) Neural Network which comprises fully connected layers including the classification layer (i.e. softmax layer as shown in Figure 2.1). The layers in a CNN are organized in such a way that they distinguish much simpler patterns such as lines, curves, textures through the early layers and more complex patterns (i.e. faces, objects, etc.) from later layers. Figure 2.1 shows a complete flow of a CNN architecture.

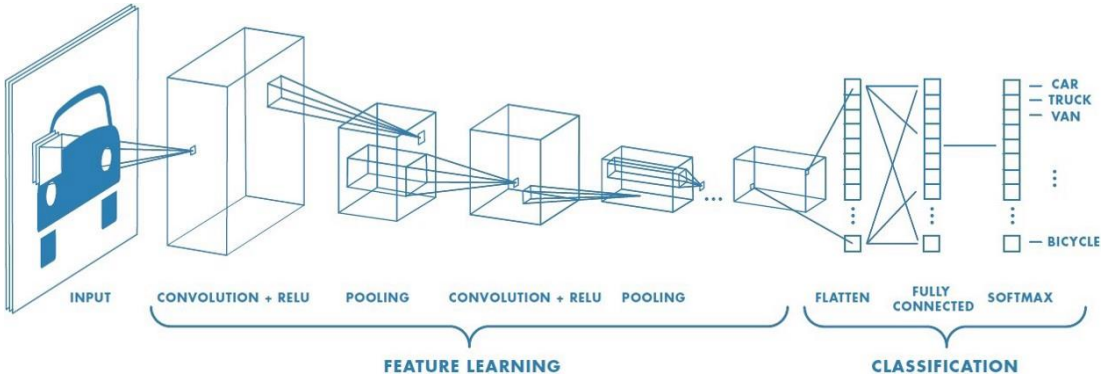


Figure 2.1: Deep Convolutional Neural Network (CNN) for classification

Input Layer

The Input layer also is known as input volume in CNN is an image represented by a three-dimensional matrix [width x height x depth/channels]. First, the input image needs to reshape into a vector format. For example, if the image dimension is 64 x 64 x 3, then convert it into 12288 x 1 before feeding into the input layer. But if the batch size is N (when we apply batch or mini-batch gradient descent) then the dimension of the input will be (12288, N).

Convolutional Layer

These are the core building blocks of any CNN architecture. A convolutional layer comprises of several filters that are learned through backpropagation in order to extract different features through the input volumes. These filters ensure that a

neuron in the next layer is connected to a small region of the previous layer (input), known as the local receptive field. Each local receptive field of the input volume is connected to each filter in the convolution layer to perform convolution operation and take the dot product between the filter and the local receptive field in order to compute a single value of the output volume (feature map). Then we move the filter over the subsequent local receptive field of the same input volume by a certain Stride value and apply the same aforementioned operation again. We then repeat the same process until we go through the whole input volume. The output volume will be the input to the succeeding layer. The number of channels of the input image is as same as the channels of a filter in the corresponding convolution layer. Here all neighborhood responsive fields of the input volume share each filter in order to create feature maps. The benefit of sharing weights concept is to decrease the complication of the model by reducing the number of learnable weights to accelerate the training process. Then add bias terms to the feature map and pass it through a non-linear activation function in order to create an activation map.

Activation Function

CNNs consist of linear and non-linear functions. The activation functions that we use in order to create activation maps are the non-linear components. These activation functions apply after the convolutional layers to introduce non-linear behaviour to distinguish non-linear features and improve the model performance in the CNN.

The ReLU (Rectified Linear Unit) activation function is a non-linear function that is widely used in CNN architectures. A previous study [25] has been revealed that we can train CNNs competently when we use the ReLU as the activation function in convolutional and fully-connected layers, except the classification layer. The mathematical formula of the ReLU function is described below.

$a^{[L]} = \max(z^{[L]}, 0)$, where $z^{[L]}$ is the input to the activation function in the L^{th} layer and $a^{[L]}$ represents the output. ReLU retains the $z^{[L]}$ if it is positive and prunes to zero if $Z^{[L]}$ is negative.

Pooling Layer

The pooling layer is used to shrink the spatial dimensions of the input image after applying the convolution operation. The main purpose of having this kind of layer is to reduces the required amount of computational time and learnable parameters by lowering the resolution of the activation maps. It takes the activation map that is generated through the convolutional layer followed by an activation function and outputs a single value per each local receptive field according to the window size. The pooling operation is performed on every channel of the input volume individually. There are four types of Pooling filters namely, Max Pooling, Min Pooling, Average Pooling, and Global Average Pooling (GAP). Max Pooling returns the maximum value from each local receptive field of the image covered by the pooling filter. Min Pooling returns the minimum value from each local receptive field of the input volume covered by the pooling kernel. In contrast, Average Pooling returns the average of all the values from each local receptive field of the input volume covered by the filter. The following figure (Figure 2.2) shows Min, Max, and Average Pooling operations relative to the given 2D input image.

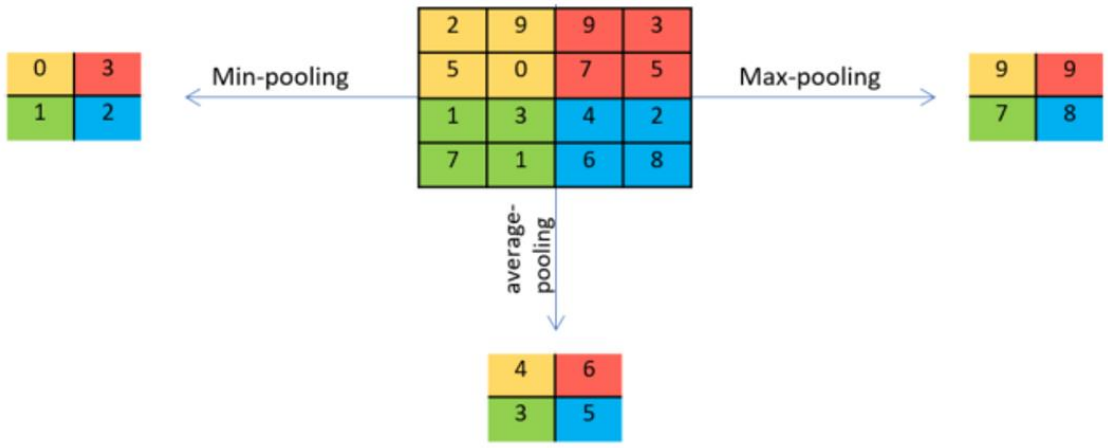


Figure 2.2: Min, Average and Max Pooling with 2 × 2 filters and stride 2

In the recent past, data scientists have used GAP layers to minimize over-fitting by reducing the total number of learnable parameters in CNN models. However, these layers perform a more extreme type of dimensional reduction, where an input feature map with dimensions h x w x d is reduced in size to have dimensions 1 x 1 x d. GAP layers reduce each feature map to a single number by obtaining the average of all cell values as shown in Figure 2.3.

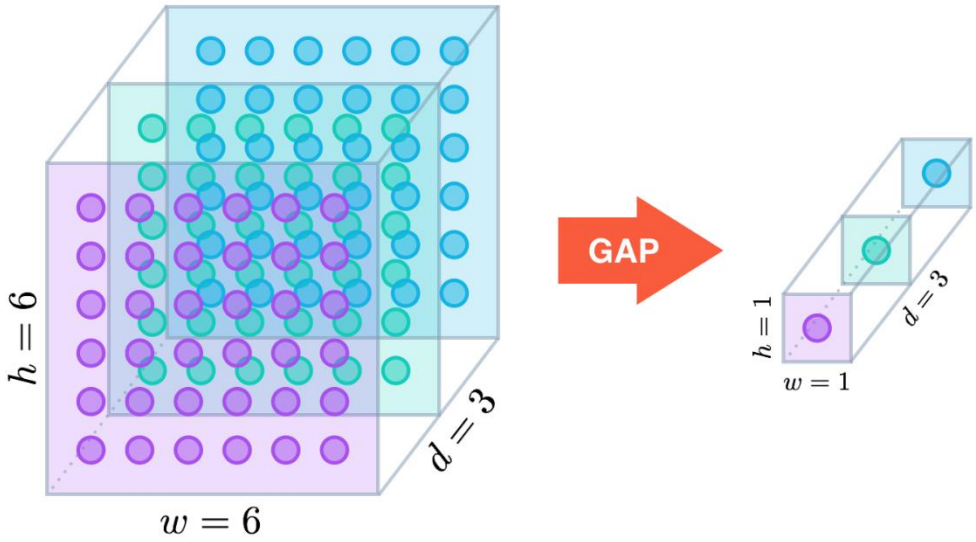


Figure 2.3: Global Average Pooling operation

Fully Connected Layers

Fully connected layers [19] are used after the convolutional and pooling layers (feature extraction part) for classification tasks. The main purpose of having these layers is to generate specific semantic representation. Each neuron in a fully connected layer connects to every neuron in the preceding layer. This structure leads to a simple matrix-vector computation when calculating the output of each layer, but it also leads to a vast set of learnable parameters. An artificial neural network (ANN) that comprises dense (fully-connected) layers can be very effective for low dimensional data, but the computational cost can turn out to be very high for high dimensional data such as images. In order to overcome the overfitting problem due to

huge trainable parameters, usually, researchers add a dropout layer after the fully connected layers.

Regularization

One of the most challenging problems that face during the training process of any CNN model is a high-variance problem also known as overfitting. Due to this, the model cannot generalize to new instances that did not appear in the training set, since it memorizes the local patterns of the training dataset. Thus, overfitting is a crucial problem that needs to handle in deep CNNs [26]. Numerous research works have been published to reduce the high-variance problem. The most commonly used techniques are dropout [27] and L2 regularization (ridge regression) [26].

Dropout

Dropout is a regularization technique that is used in the machine learning algorithm in order to prevent overfitting problems. The main purpose of a dropout layer is that it disables or ignores the neurons randomly in each iteration during the training phase. The dropping out of a neuron indicates that we temporarily eliminating it from the neural network along with its all incoming and outgoing connections during

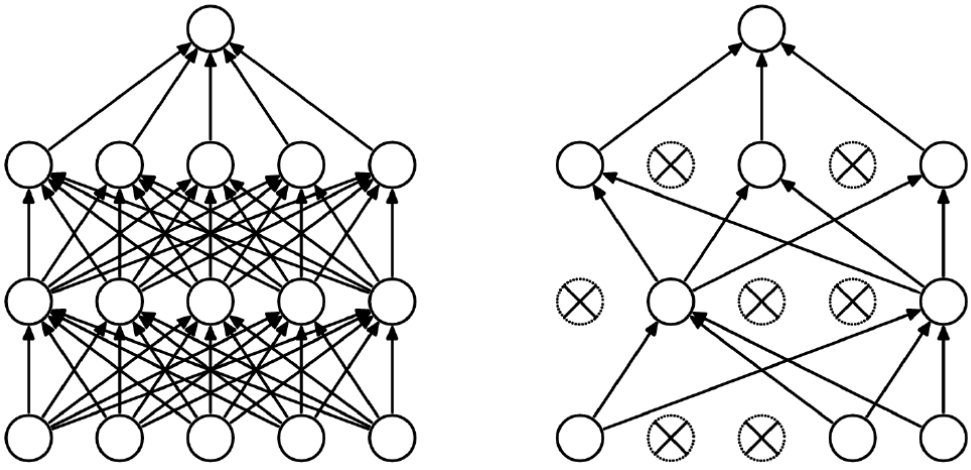


Figure 2.4: An illustration of an ANN by applying dropout for each layer

a particular forward and backward pass. Figure 2.4 represents an illustration of smaller ANN produced by applying dropout for each layer. Dropout can also be used subsequently to the convolutional layers on CNNs. However, it is not desirable to apply the dropout value in the early convolutional layers since it affects the information to vanish in the entire neural network [28], and this leads to downgrading the network performance. Moreover, the dropout layers are not actively involved during the testing time.

L2 Regularization

L2 regularization also is known as ridge regression delivers a technique to downgrading the high-variance problem (overfitting) of any deep neural network architecture on the training set while improving the performance on unseen data.

$$J_{regularized} = \underbrace{-\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)}\log\left(a^{[L](i)}\right)+(1-y^{(i)})\log\left(1-a^{[L](i)}\right)\right)}_{\text{cross-entropy cost}} + \underbrace{\frac{1}{m}\frac{\lambda}{2}\sum_{l}\sum_{k}\sum_{j}W_{k,j}^{[l]2}}_{\text{L2 regularization cost}}$$

Figure 2.5: An L2-regularized version of the cost function used for an ANN

This method adds the sum of squared magnitude of weights as penalty term to the cost function (e.g. cross-entropy) in order to penalize the weight matrices from being too large and it leads to achieving much simpler models.

In an ANN, error or cost function depends on all the learnable parameters including bias terms, $W^{[1]}$, $b^{[1]}$ through $W^{[L]}$, $b^{[L]}$, where L represents the number of layers in the given neural network. The cost function is the sum of all the losses over the m number of training examples. When we apply L2-regularization, we add an extra term to the cost function namely L2 regularization cost as shown in Figure 2.5, which indicates the sum over the squared norm of the parameters W divided by λ (i.e. regularization parameter) over 2m. Here, the norm of weight matrix $W^{[l]}$ is defined as the sum from k=1 through $k=n^{[l-1]}$ and sum from j=1 through $j=n^{[l]}$, since $W^{[l]}$ is an $n^{[l-1]}$ by $n^{[l]}$ dimensional matrix, where $n^{[l]}$ represents the number of neurons in a given layer *l*.

2.2 Content-based Image Retrieval

There are a growing number of medical images such as retinal, endoscopy, CT scan, MRI and x-rays are taken daily in numerous hospitals and health-care centers [29]. Medical image retrieval has tremendous importance, particularly in clinical decision support and in research fields such as medical image analysis and education. Medical image retrieval is in high-demand for decision-making processes because the historical imagery of different patients in hospitals and health-care centers have vital information for the forthcoming diagnosis, where an application that retrieves similar case(s) can assist in making a more precise diagnosis and deciding on prioritizing treatments.

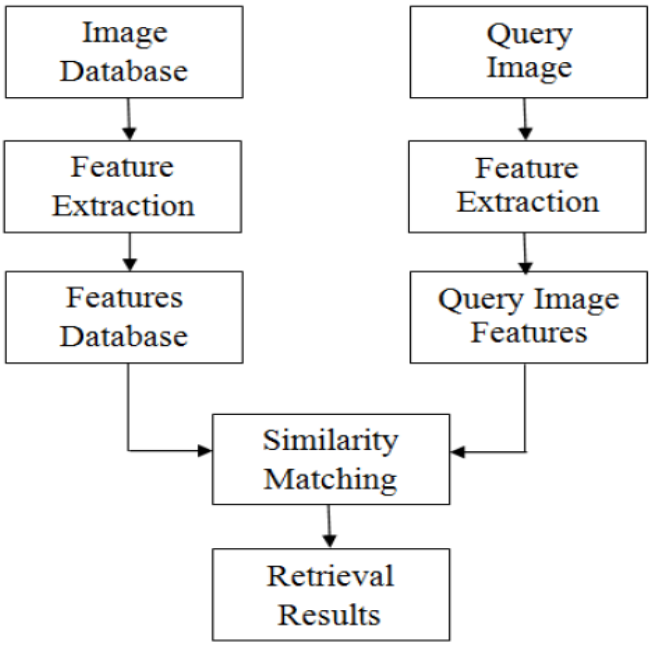


Figure 2.6: Architecture diagram of a content-based image retrieval system

In CBIR, query image and images in the database are encoded into real-valued features through an image processing or a deep learning technique. The easiest mode of searching for related imageries is to rank the images in the database by utilizing a distance metric (e.g. Euclidean distance) relative to the query image and then retrieving the neighboring ones in the feature space. Nevertheless, the memory and time consumption are very high for an image database with a considerable number of

images if we perform a linear search through the entire database. The architecture diagram of a CBIR system is shown in Figure 2.6.

2.2.1 Hashing Techniques for Content-based Image Retrieval

Hash embedding methods are proposed to map image contents to compact binary hash codes in order to address the inadequacy of real-valued features [30]. The memory and time consumption for the searching process can be significantly reduced since the images are denoted by compact binary hash codes instead of real-valued feature vectors. Numerous hashing algorithms [31]-[32] have been proposed for content-based image retrieval. These hash code embedding approaches we can be divided into two major classes namely supervised and unsupervised approaches. Locality Sensitive Hashing (LSH) [31] is the most representative approach for unsupervised methods, where it is used to generate hash codes by projecting the data points to random hyperplanes in order to map images to a new feature space. Recent studies [33] have revealed that utilizing supervised hashing can enhance the binary codes learning performance.

The most important part of the supervised hashing technique is the real-valued feature vectors that are used to derive the hash code. The quality of extracted features directly affects the accuracy of the retrieval model. Recently, convolutional neural networks have shown their remarkable learning-power and the impressive performance in tasks that are heavily dependent on feature extraction, in particular, image classification and localization [34], video analysis [35], object detection [36] and other numerous tasks such as segmentation [37]. CNN based approaches typically outflank conventional approaches in the previously mentioned fields, which proves that CNNs can learn powerful features by capturing the semantic information of the imagery. Hence, deep learning is the most suitable way to learn compact binary codes. Hashing methods using deep learning techniques [38]-[32] in content-based image retrieval demonstrate high performance over the conventional hashing techniques such as LSH [31], KSH [33], MLH [39] and SH [40], since the shallow techniques limit the retrieval performance of the learned compact binary hash codes

due to the lack of semantic information under the drastic appearance variations in data.

Although state-of-the-art deep learning-based techniques [41][42][32] have accomplished very good performance in image retrieval, they either have very high memory or training-time requirements. DSH [42] uses a matrix decomposition technique in order to learn the hash codes for images in the preprocessing stage. This method requires a pair-wise similarity matrix of the dataset as the input. But it is not favorable in the situations where the considerable amount of data since it needs both significant storage in order to store a large sparse matrix for representing the similarity between data points in the training set and computational time. DLBHC [32] uses end-to-end learning rather than using feature extraction through transfer learning. This approach required more computational time in the training process to achieve substantial performance. DNNH [41] consists of triplet-based constraints in order to describe more complex semantic relations. Hence the training process becomes more difficult because the output layer comprises of the parameterized piece-wise threshold function and sigmoid non-linearity. Thus, DNNH performs inferior to the pairwise deep-learning technique compared to the DSH technique, where the constraints based on the image triplets generation cannot provide more information than the pairwise ones since the imageries only contain category labels.

2.3 Overview of DR Analysis Methods

The early diagnosis of DR is a vital importance factor to slow down the disease progression and allow for planning the treatments. Rapid growth in computer-aided systems based on abnormal feature detection, severity stage classification and clinically similar case(s) retrieval for DR analysis and other eye-related diseases developed over the past few periods [43]. This section elaborates on four subsections as follows. The first part is based on the severity stage classification approaches for DR analysis. The next part is based on abnormal feature detection approaches for DR analysis. The following subsection is based on the blood vessel segmentation and the last subsection is based on content-based image retrieval methods.

2.3.1 DR Classification Methods

During the recent past, there has been a rapid development of autonomous disease grading systems. The main idea of using computerized systems to assist with the disease diagnosis through medical images is more practical. Previous studies have explored the use of image processing, machine learning, pattern recognition and statistical methods for automatic DR classification through digital colour fundus images according to the different grading systems including the international standard grading scale [4]. The different approaches can be categorized into two parts namely, explicit feature extraction and implicit feature extraction, i.e., deep learning. A more robust autonomous system plays a vital role in the early detection of the disease and supports ophthalmologists to recommend treatment prioritization on a timely basis. The existing approaches for automatic severity stage classification of DR are explicated below.

Explicit Feature Extraction Methods

Former studies have used shallow machine learning classifiers to diagnose diabetic retinopathy through image processing-based feature detectors by localizing the optic disc and the blood vessels and count the presence of abnormalities such as red lesions, microaneurysms, cotton wool spots, hemorrhages, and hard exudates.

Lee et al. [44] developed an automatic DR diagnosis approach for the NPDR stage through the techniques of image processing including image normalization and noise removal by detecting three lesions to be specific, microaneurysms and hemorrhages, cotton-wool spots (soft exudates) and hard exudates. They detected the aforementioned lesions on the basis of the contrast of the colour between the retinal background and lesions. The authors achieved an accuracy of 81.7% at the NPDR level. Later, Roychowdhury et al. [45] developed a two-stage hierarchical classification architecture where in the first stage, the non-lesions removed and in the second stage, the red lesions classified as microaneurysms and hemorrhages and the bright regions classified as cotton wool spots and hard-exudates. The authors have

analyzed their dataset with Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Gaussian Mixture Models (GMM) utilizing Ada Boost as for the feature ranking algorithm. This system used to classify fundus images only for the severity stages of NPDR (mild, moderate and severe) and achieved 0.904 Area Under Curve (AUC) with 100% sensitivity and 53.16% specificity.

The methods described in [44][45] are only suitable for classifying the severity stage of NPDR and recommending treatment, but their classification performance needs to be evaluated on other severity stages such as diabetes without retinopathy and PDR as well.

Acharya et al. [46] extracted four features microaneurysms, blood vessels, hemorrhages and exudates from the green channel of colour retinal images using image processing techniques, and fed them into an SVM. The authors achieved a sensitivity of 82% and specificity of 86%. They used 331 fundus images for their study. Moreover, they used a similar grading system (five severity stages of DR) according to the international guidelines, but there is room for further improvement of model accuracy and it can be done by increasing the number of features, tuning hyperparameters or combining weak learners in ensembles for classification.

The authors in [47] reported an accuracy of 93%, sensitivity of 90% and specificity of 100% on a three-severity level classification task (non-DR, NPDR, and PDR) by extracting features such as area of blood vessels, area of exudates, and texture features utilizing image processing techniques for 140 images and then fed into a small ANN. But according to the international standard, NPDR can be further divided into mild, moderate and severe and therefore, the authors should re-evaluate their model performance for these subcategories as well.

Sinthanayothin et al. [48] trained an MLP (Multilayer Perceptron) neural network by feeding the features extracted through image processing techniques in order to classify retinal images to normal, abnormal or unknown. They have taken into consideration of exudate regions detection as the prior criteria for the normal and

abnormal retinal image classification since they failed to detect microaneurysms and hemorrhages from their system. In contrast, optic disk detection used as the prior criteria for unknown images. The authors used 484 normal fundus images and 283 retinal images with DR for their study and achieved 80.21% sensitivity and 70.66% specificity. Larsen et al. [49] developed a DR prediction system to classify whether a given patient suffers from diabetic retinopathy or not. The authors used 260 retinal imagery for their study and among them 137 taken from diabetic patients. They have done their experiments through automated lesion detection by extracting microaneurysms and hemorrhages using image processing techniques. This approach demonstrated 96.7% sensitivity and 71.4% specificity.

Later, Singalavanija et al. [50] developed a computer-aided system to classify retinal images into three categories namely, normal, abnormal or unknown by recognizing the exudates, hemorrhages, MAs, blood vessels, optic disc and fovea from DR imageries through image processing techniques with a sensitivity of 74.8% and a specificity of 82.7%. The authors preprocessed the raw images by enhancing the local contrast in order to obtain more uniform images. Blood vessels, fovea and optic disk were detected by recognizing the location, intensity variation and the continuity of the vascular network. They used a recursive region growing segmentation algorithm and a colour and template matching technique to identify the exudates regions and hemorrhages respectively. They used 900 retinal images including 600 from normal patients and 300 images from diabetic patients for their research. Kahai et al. [51] proposed an automated binary classification system (normal or abnormal) for diagnosing DR. During their study, the authors used only an NPDR (moderate to severe stages) dataset. They considered these three stages in NPDR for the case of the presence of microaneurysms. The model displayed a YES decision (abnormal) related to the presence of MAs (microaneurysms) for the moderate to severe cases of NPDR and a NO decision (normal) related to the absence of MAs. They used the Bayesian optimality technique for the classification in order to recognize the pathologies (MAs). Their approach was successful in classifying DR to normal or abnormal with 100% sensitivity and 67% specificity.

Giraddi et al. [52] have been proposed an approach to classify the input retinal image into two classes namely normal or abnormal based on the colour and GLCM texture features of hard exudates. The authors used SVM and KNN classifiers in order to give a comparative analysis. Finally, they have achieved 83.4% true positive rates for SVM and 92% for KNN. Moreover, they have concluded that KNN beats SVM for both colour and texture features based on their evaluation results.

These approaches [48]-[52] can be used for analyzing the presence of pathologies in retinal images and panning treatment based on normal or abnormal, but not effective for the severity stage classification of DR through retinal images since the authors did not evaluate their model performance in mild-NPDR, moderate-NPDR, severe-NPDR and PDR stages.

Yun et al. [53] developed an early diagnosis system for severity stage classification of DR. During their study, they analyzed 124 colour fundus images, which classified into four groups, in particular, normal retina, moderate-NPDR, severe-NPDR and PDR. The authors extracted features from techniques of image processing and fed them into a three-layer ANN classifier for classification. This approach reaches a sensitivity of more than 90% and specificity of 100%. But the authors did not consider about mild-NPDR stage since it is the early stage of NPDR. Hence, they should re-evaluate their model performance for mild-NPDR as well.

Li et al. [54] developed a DR screening system and distinguished PDR from NPDR automatically utilizing digital colour fundus photographs. The authors evaluated the severity of DR by analyzing the blood vessel patterns and the occurrences of bright lesions of the retinal imagery. They extracted bright lesions through morphological reconstruction. Moreover, they used multiscale matched filters to extract retinal blood vessels and vessel net density to analyze vessel patterns. They achieved a sensitivity of 80.5%. This approach can be used for NPDR and PDR, but they must be validated to analyze their system performance in three subcategories of NPDR namely, mild-NPDR, moderate-NPDR, and severe-NPDR as well as diabetes without retinopathy.

Hasan et al. [55] proposed an approach to recognize PDR by detecting neovascularization using image processing techniques such as image normalization, morphology-based operator, Gaussian filtering, thresholding and compactness classifier. The authors tested their approach using different databases with varying quality and resolution of the images. Their approach demonstrated a sensitivity of 89.4% and specificity of 63.9%. The main drawback of this system is that it detects only one severity stage (PDR), but they must be extended their approach to recognizing other severity stages based on the abnormal feature detection such as MAs, hard and soft exudates, hemorrhages as well.

Rahim et al. [56] proposed a DR and maculopathy decision support system recently by analyzing retinal images using fuzzy image processing techniques such as fuzzy filtering and fuzzy histogram equalization along with Circular Hough Transform and other different feature extraction methods such as green channel extraction. The authors used four retinal localization approaches namely, optic disc localization, blood vessel, macula, and fovea detection during the preprocessing stage. After extracting features, they have used several classification algorithms (KNN, SVM, and Naïve-Bayes) in order to train with the dataset and evaluate the generalized performance. This approach accomplished good overall performance with 93% accuracy, a sensitivity of 86.79% and a specificity of 100%, but this can only be used to analyze the presence of the retinopathy and maculopathy in retinal images. Hence, this system is not valid to evaluate the severity stages of the DR.

Implicit Feature Extraction Methods

Most of the recent work in automatic severity stage classification of diabetic retinopathy has been used deep CNNs. In a very recent work, Alban et al. [57] trained and evaluated three dissimilar CNN models: a custom CNN architecture built as a baseline where all convolution and fully connected layers were trained, a classifier built using a pretrained AlexNet [58] where only the last fully connected layer was retrained, and GoogleNet [59] constructed similarly to AlexNet. The

authors achieved best with an AUC of 79% and an accuracy of 45% for the five-class severity stage classification using GoogleNet.

Butterworth et al. [60] used a transfer learning approach to extract features from pretrained CNNs. The authors trained a linear SVM from deep features extracted from pretrained AlexNet [58] and Resnet-34 [61] architectures and achieved 25% and 76% accuracies respectively. Models trained with the standard Categorical Cross-Entropy loss function and an MSE (Mean Squared Error) loss function.

There is a room for further improvement of model accuracy in [57] and [60], and it can be done by tuning hyperparameters or developing ensemble architectures using weak learners or using an ensemble of pretrained CNNs as feature extractors.

Thambawita et al. [62] presented a deep ensemble CNN-based approach to improve the multi-class classification of Gastrointestinal tract diseases. The authors used the combination of pretrained Resnet-152 and Densenet-161 with an additional multilayer perceptron (MLP) as the classifier and achieved 95.80% accuracy.

2.3.2 Abnormal Lesion Detection Methods

Abnormal lesions in the retina are the key indicators for recognizing the presence of a disease. These lesions can be further categorized into bright and dark regions. In the DR perspective, abnormal lesion detection approaches are vital importance for the ophthalmologists to recognize the pathology or abnormality on the retinal tissue and allow them to treat the relevant regions where symptoms appeared. We can identify several abnormal lesions such as microaneurysms, hemorrhages, cotton wool spots and hard exudates utilizing the colour retinal images. The related literature review can be divided into two major sections. The first section is associated with the automatic analysis of microaneurysms and hemorrhages and the next section is based on the analysis of exudate regions.

Automatic Analysis of Microaneurysms and Hemorrhages

Baudoin et al. [63] proposed an automatic detection of MAs in fluorescein angiograms using mathematical morphological concepts. They performed different top-hat transformations in order to extract MAs. However, they have faced several difficulties when recognizing the MAs with fuzzy boundaries due to the fluorescein leakage in the retinal tissue. The authors used 25 angiograms for this study. Later, Spencer et al. [64] used a bilinear top-hat transformation and matched filtering to segment the image and then applied a thresholding function in order to extract and count MAs. They evaluated their approach with MA counts performed manually by five clinicians.

Zhang et al. [65] came up with a new top-down approach to detect retinal hemorrhages using PCA. The authors calculated an evidence value for each pixel by utilizing SVM after applying the colour normalization in the preprocessing stage. They fed features, which were extracted from two-dimensional PCA to the SVM classifier. After finding the hemorrhage feature location, they used a post-processing step to segment the boundary if the hemorrhages fall in the ROI (region of interest). This approach expected to achieve higher accuracy for the classification task.

Quellec et al. [66] proposed an approach to detect MAs based on local template lesion matching with optimal wavelet transform technique in retinal images. The results of their approach as evaluated using 120 retinal images, which further categorized into three different modalities namely, colour photographs, colour filtered photographs, and angiographs. The authors achieved a sensitivity of 89.62%, 90.24% and 93.74% for the aforementioned three modalities respectively.

Kande et al. [67] presented an automatic red lesions detection method using ocular fundus images. This approach utilized intensities of green and red channels for a given fundus image in order to correct non-uniform illumination. They enhanced the contrast of the red lesions using matched filtering and then segmented by relative entropy-based threshold function. In order to reduce the enhanced vasculature, the

authors used morphological top-hat transformation. Moreover, the authors classify red lesions from the dark regions utilizing SVM and accomplished 96.22% sensitivity and 99.53% specificity.

Frame et al. [68] compared three classification approaches for the detection of MAs using fluorescein angiograms. First, they segmented MAs through image processing techniques and extracted a set of features for each candidate in order to train three classifiers. They used linear discriminant analysis and an ANN together with a rule-based system in order to perform a classification task. The authors achieved a higher accuracy for the rule-based approach.

Niemeijer et al. [69] published novel research based on red lesion detection in colour retinal images using a hybrid approach by combining two prior works done in [64] and [68]. They first separate the red lesions and vascular network from the background through a pixel classification technique. Next, they removed the vascular network from the segmented image in order to further filter the possible red lesions. Finally, they classified these lesions using KNN and accomplished 100% sensitivity and 87% specificity.

Automatic Analysis of Exudate Regions

Wang et al. [70] published research work to detect exudate lesions in the fundus photographs. They used several image processing techniques such as statistical classification with brightness adjustment, a local window-based verification, and a thresholding strategy. They showed 100% accuracy in terms of exudates detection through their experimental results.

Phillips et al. [71] proposed a technique for detecting and measuring the exudate regions from digital colour retinal photographs of individuals with DR. They used a global thresholding strategy to detect large high-intensity regions. In contrast, in order to segment smaller exudate regions, the authors used a block-wise local

thresholding strategy. This is a better approach to detect pixels belong to the exudate areas but bounces several false positives.

Walter et al. [72] developed a system to detect exudates in colour fundus photographs to diagnose of DR. First, they localized optic disc in the given retinal image by applying watershed transformation and the morphological filtering techniques. The authors detected exudates using its high grey level variation. Additionally, they used morphological reconstruction techniques to identify the contours of the region. They compared their experimental results with a clinician. This approach reached 92.8% of sensitivity.

Sánchez et al. [73] proposed an automatic detection approach for hard-exudate regions in fundus imagery. They segmented exudates by applying threshold dynamically from the background using mixture models. After this step, the authors used an edge detection technique in order to separate hard exudate regions from soft-exudates and other lesions. Finally, they evaluated this approach using 80 retinal images and obtained 90.2% of sensitivity and 96.8% of positive predictive value.

Xu et al. [74] proposed an approach to detect hard-exudates from retinal imagery by extracting feature vectors using gray level co-occurrence matrix (GLCM) and stationary wavelet transform (SWT). The authors used a radial basis kernel function with SVM to classify the 50 data points and achieved 84% accuracy, 88% sensitivity, and 80% specificity.

Section 2.1.2 mainly focuses on diverse techniques for the automatic abnormal lesion detection for the severity stages of DR diagnosis. It is a very challenging task to automatically detect MAs, hemorrhages and exudate regions since they appear as very tiny spots on the retinal surface.

2.3.3 Retinal Blood Vessel Segmentation Methods

This section reveals numerous techniques which are related to the retinal blood vessel segmentation in fundus images of healthy and diseased individuals. In this

section, we primarily focus on the previous studies that had been done in the areas of supervised and unsupervised learning. Most of the algorithms are constructed on image processing such as morphological methods, matched filter techniques, the grouping of edge pixels, intensity profile techniques, and vessel tracking techniques, and conventional machine learning techniques, but few types of research are based on deep learning.

Supervised Learning Approaches

Nekovei et al. [75] used an ANN as a classifier to detect vascular network in angiograms by classifying retinal blood vessel and non-blood vessel pixels. The authors classified the center pixel of a given window using gray-scale information. They used 75 angiogram images during their experimental evaluation. Later, Sinthanayothin et al. [76] developed an approach to localize the vessel network of the retinal images. They first preprocessed 112 fundus images by applying adaptive local contrast enhancement. After that, the authors applied PCA on the image and edge detection of the first principal component in order to derive the inputs. Next, they fed these inputs to train an MLP neural network and achieved a sensitivity of 83.3% and a specificity of 91.0% for vasculature localization.

Staal et al. [77] developed an approach to segment the blood vessels in 2D colour retinal images. First, the authors extract the ridges of the images and then used them to construct line elements. After this step, they partitioned the image into patches by assigning each pixel value to the closest line segment. Next, they extracted feature vectors for every pixel and fed them into a KNN classifier. The dataset consisted of 40 manually labeled retinal images. This system achieved an AUC of 0.952.

Soares et al. [78] published a research work based on automatic vasculature segmentation in retinal imagery. The authors performed segmentation by classifying every pixel in the image into two categories namely. vessel and non-vessel after extracting the feature vector of each pixel. The intensity of the pixel and the responses of the Gabor wavelet transform taken into consideration when preparing

these feature vectors. They used a Gaussian mixture model as the classifier and the performance was evaluated based on two publicly available datasets, DRIVE [77] and STARE [79]. For the DRIVE dataset, they achieved an AUC of 0.9614.

Ricci et al. [60] implemented a supervised classification approach to segment retinal blood vessels in fundus images by extracting feature vectors using line operators. The authors computed line detectors based on fixed-length average gray lines passing through the selected pixel at different orientations. They used two segmentation approaches. At the first stage, an unsupervised pixel classification performed by applying a threshold function. In the next stage, they constructed two orthogonal line detectors composed of the gray lines in order to build feature vectors to train a classifier. They performed their classification task to identify vessel pixels using SVM. They evaluated the model performance on two datasets, DRIVE [77] and STARE [79] and accomplished 95.6% and 95.8% accuracies respectively.

Lupascu et al. [80] developed a system to segment the vascular network in retinal images by extracting 41-dimensional feature vectors at a diverse spatial scale and then fed them into an AdaBoost classifier. The authors used various filters such as Gaussian filters, matched filters, and a two-dimensional Gabor wavelet in order to extract these features. They trained the AdaBoost classifier with their dataset in order to classify vessel pixels and non-vessel pixels. This approach was tested on the 20 retinal imagery from the DRIVE dataset [77] and reached AUC of 0.9561.

You et al. [81] published a research work based on blood vessel segmentation in fundus photographs using the radial projection and semi-supervised learning. In order to capture the vessel centerlines of narrow and low-contrast blood vessels, they apply the radial projection. The authors used a steerable wavelet technique to enhance vessels. Next, they generated a feature vector by calculating the strength of the line and applied it to the aforementioned enhanced vessel imagery. For the vessel structures extraction, they used the SVM classifier. This approach was evaluated based on two publicly available datasets namely, DRIVE [77] and STARE [79] with 94.3% and 94.9% mean accuracies correspondingly.

Recently, Roychowdhury et al. [82] published a research work based on blood vessel segmentation using fundus imagery. They used a three-stage process in order to develop this novel algorithm. They preprocessed the green channel of each image in order to extract a binary mask by applying a high-pass filter in the first stage. At the same time, they extracted another binary mask from a morphological reconstruction technique. The common areas of both binary masks were extracted to identify the major vessel areas. During the second stage, all residual pixels in the aforementioned two binary masks were classified utilizing a GMM (Gaussian mixture model). In the third stage, the pixels in the major vessels were combined together with classified blood-vessel pixels to obtain the last vessel imagery. The proposed approach was evaluated based on CHASE-DB1, DRIVE and STARE retinal imagery databases and reached a mean accuracy of 95.3%, 95.2%, and 95.1% respectively.

Wang et al. [83] used pretrained LeNet-5 CNN architecture as a feature extractor for addressing tiny blood vessel segmentation. The model consists of three heads at different layers of the CNN which then fed into three random forest classifiers. The ensemble of the random forest classifiers achieved 0.97 and 0.94 for model accuracy and AUC respectively on the DRIVE [77] dataset.

Unsupervised Learning Approaches

Salem et al. [84] developed an unsupervised learning algorithm in order to segment the blood vessels from colour retinal imagery. The authors used a Radius-based Clustering Algorithm (RACAL), which relies on the distance-based principle by mapping the distributions of image pixels. They further enhanced this approach to detect low contrast blood vessels with small diameters using semi-supervised learning. The performance evaluated based on the STARE [79] database and achieved 82.1% of sensitivity.

Kande et al. [85] used an unsupervised approach for blood-vessel segmentation, which is based on fuzzy c-means clustering. The authors extracted intensity values from red and green channels from the colour retinal imagery in order to address the

uneven illumination problem. They used Matched filtering for contrast enhancement of the retinal vascular network relative to the background. As the next step, the fuzzy c-means algorithm is used to recognize the pixels of the blood vessel in order to obtain a clear segmented vascular network. This approach is evaluated based on the two datasets, STARE, and DRIVE. Finally, the authors accomplish an AUC of 96.02% and 95.18% respectively.

Zhao et al. [86] developed an unsupervised learning technique that depends on an infinite active contour technique for the blood vessel segmentation from retinal images. The authors extracted pixels of the blood vessels using hybrid region information of retinal imagery. Their concept is based on an infinite perimeter regularizer, which is used to detect tiny blood vessel branching structures. Moreover, they used diverse forms of region information in order to obtain good segmentation performance. This technique was validated based on three publicly available datasets namely, STARE, VAMPIRE and DRIVE and accomplish an accuracy of 95.6%, 97.7%, and 95.4% respectively.

Section 2.1.3 illustrated various methods for the automatic retinal blood vessel segmentation in order to identify the severity stages of DR diagnosis. It is a very challenging task to automatically extract tiny blood vessels in the retinal surface since it is difficult to recognize the bifurcations (branching points) and the connectivity of blood vessels.

2.3.4 Retinal Image Retrieval Methods for DR

During the past few decades, content-based image retrieval (CBIR) has been a prominent research area in medical image analysis. It enables retrieving images from an image database that are similar to a given query image. Different research groups have been proposed numerous types of medical image retrieval approaches in the recent past. However, a comprehensive and effective deep neural network-based retinal image retrieval architecture for diabetic retinopathy (DR) is not available in the literature. The principal objective of CBIR for DR is to efficiently retrieve retinal

images that are semantically similar to a given query for effective treatment based on the severity stage of the decease. Most of the previous study in Diabetic retinopathy has been explored for classification and segmentation. Inadequate efforts have been made in CBIR as described below.

Galshetwar et al. [7] developed a CBIR system using the concept of salient point selection of edgy images and local binary patterns (LBP) extracted through inter-plane relationship. The authors enhanced their results by using color features of the original image in combination with LBP features. They conducted experiments on 1200 retinal images which are categorized in four severity stage groups. They achieved a 57.82% mAP.

C. Baby and D. Chandy [8] proposed a CBIR technique through dual-tree complex wavelet transform (DT-CWT). The authors extracted features by using a combination of two-dimensional DT-CWT and generalized Gaussian density. They used KL divergence (Kullback-Leibler Divergence) in order to calculate the similarity measure between two feature sets. They conducted experiments on 1200 retinal images which are categorized in four severity stage groups. The mAP at the top five retrieved images was obtained as 53.70% and 78.23% for severity stages of DR and Macular Edema correspondingly.

J. Sivakamasundari et al. [9] proposed a CBIR framework that relies on the edge detection technique for DR diagnosis. The authors enhanced edge information by extracting green channel and morphological operation of normal and abnormal retinal images. They used Canny edge-based detection and the Kirsch template techniques for segmenting the blood vessels. The extracted features from segmented images were used for further analysis. They applied Euclidean distance in order to measure the similarity. The authors estimated the performance for the Kirsch template-based method using precision and recall and they achieved 90% and 82% respectively. Similarly, for the Canny edge method, they achieved 80% and 38% correspondingly.

The first two literature are based on four severity stage of DR and the last one is built on normal and abnormal retinal images, but according to the international standard there exists five severity stages, namely diabetes without retinopathy (Non-DR), Mild non-proliferative diabetic retinopathy (Mild-NPDR), Moderate non-proliferative diabetic retinopathy (Moderate NPDR), Severe non-proliferative diabetic retinopathy (Severe-NPDR) and Proliferative diabetic retinopathy (PDR) [4]. In contrast, there is room for further improvement of the mAP of the aforementioned retrieval models.

2.4 Summary

This chapter primarily based on related literature which has been conducted during the last few decades. Initially, it describes the deep learning approaches especially general CNN architecture and its components that have been used for image analysis tasks. Secondly, includes a detailed annotation of content-based image retrieval techniques and their limitations. Finally describes various image processing and machine learning-based approaches that were used to classify, retrieve and segment retinal images in order to detect diabetic retinopathy.

3. DATASETS

3.1 Retinal Datasets

Two retinal imagery datasets were used to train automated classifiers during this research. One dataset was drawn from a recent Kaggle competition [87] and the other one (DIABRET) was collected from an eye hospital in Sri Lanka. These are a set of high-resolution images taken through a funduscope in a variety of conditions including colours, lighting and different orientation. Each retinal image is assigned a class based on the severity stage of DR, where each image collected from the eye hospital was labeled by a well-trained clinician and validated by an eye-consultant. Each image is labeled as 0, 1, 2, 3, 4 and the number represents the severity level of DR namely Diabetes without Retinopathy, Mild NPDR, Moderate NPDR, Severe NPDR, and Proliferative DR respectively.

The main difference of these two datasets is that Kaggle dataset consists of standard colour fundus photographs which captures 30 degree of the posterior pole of a patient eye including the macula and the optic nerve whereas DIABRET dataset consists of wide-field colour fundus photographs which capture the seven fundus fields of a patient eye and combined together to generate a montage image that displays a 75 degree field of view.

Moreover, each dataset was divided into training and testing where the training dataset represents 80% and the test set represents 20% from the entire dataset. For each dataset, we used stratified five-fold cross-validation on its training data in order to select the best hypothesis by tuning the hyperparameters, while its test data is used to evaluate the performance of this best hypothesis.

These two datasets consist of imbalanced class labels. Table 3.1 displays the class proportion statistics for both of these datasets.

Table 3.1: Class distribution of two datasets

Class	Images in Kaggle	Images in DIABRET
Diabetes without Retinopathy	1170	440
Mild-NPDR	558	217
Moderate-NPDR	710	191
Severe-NPDR	200	160
PDR	162	123

3.2 Gastrointestinal-tract Endoscopy Dataset

In order to measure the effectiveness and efficiency of our two approaches (classification and content-based image retrieval tasks), we further used another medical image dataset called KVASIR [88] which consists of 8000 Gastrointestinal tract images captured through endoscopic process which further categorized into eight different categories based on the anomaly type where each class holds 1000 images.

This dataset comprises three anatomical landmarks namely pylorus, cecum, and z-lines and three pathological findings; polyps, ulcerative colitis, and esophagitis. In contrast, the dataset contains another two types identified with the expulsion of polyps called dyed resection margins and lifted and dyed polyp. The dataset was divided into 90% and 10% to represent the training and test sets respectively.

3.2.1 Anatomical Landmarks in GI-tract

An anatomical landmark is a discernible component inside the GI tract which can distinguish effectively through the endoscope. Recognizable region of interest is exceptionally vital since the region of interest can be considered as a reference point to portray the area for the discoveries and for exploring along the GI tract.

3.2.1.1 Z-line

The z-line indicates the esophagogastric intersection between the squamous mucosa of the throat and columnar mucosa of the stomach. Figure 3.1 demonstrates a case of

a z-line. It is noticeable through an endoscope as an unmistakable limit where the white mucosa in the throat meets the red gastric mucosa as should be obvious in Figure 3.1. The significance of detecting z-line is to choose whether an ailment is accessible or not.

3.2.1.2 Pylorus

The pylorus links the stomach into the initial segment of the little bowel called duodenum. The significance of distinguishing the pylorus is for endoscopic instrumentation to the duodenum which considered a difficult move in endoscopy. The subsequent picture in Figure 3.1 demonstrates a case of the pylorus.

3.2.1.3 Cecum

The Cecum has a huge cylinder-like structure in the lower stomach hole. Regularly it gets undigested nourishments. The significance of perceiving the cecum is the verification of complete colonoscopy. The last picture in Figure 3.1 demonstrates a case of Cecum.

(a) Z-line (b) Pylorus (c) Cecum

Figure 3.1: Anatomical Landmarks of Endoscopic imagery

3.2.2 Pathological Findings

An obsessive finding is an unusual element inside the GI tract. It is considered as a harm in the ordinary mucosa through an endoscope. This harm might be the side effects of a progressing or originator of disease.

3.2.2.1 Esophagitis

This is an aggravation or irritation of the throat. They are observable as a break in the esophageal mucosa. The first picture in Figure 3.2 demonstrates a case of Esophagitis.

3.2.2.2 Polyps

Polyps are masses of injuries that structure inside the entrail. Despite the fact that a large portion of the polyps is kind, some of them may prompt colorectal malignancy. Consequently, the recognition of polyps is significant. The subsequent picture in Figure 3.2 demonstrates a case of polyps.

3.2.2.3 Ulcerative Colitis

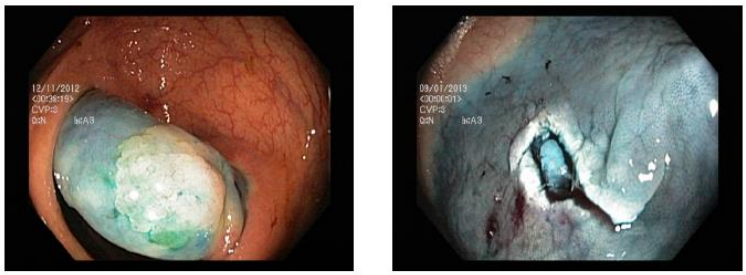
Ulcerative colitis (UC) is an incendiary entrail disease and it impacts the whole entrail. This can cause enduring irritation or wounds in the entrail. The last picture in Figure 3.2 demonstrates a case of ulcerative colitis.

(a) Esophagitis (b) Polyps (c) Ulcerative Colitis

Figure 3.2: Pathological Findings of Endoscopic imagery

3.2.3 Polyps Removal

At times, polyps expulsion is operated during the endoscopy. One of the techniques called Endoscopic Mucosal Resection (EMR) which lifts the polyp from the hidden tissue. At that point, it turns into a Dyed and Lifted Polyp as appeared in the left-side picture in Figure 3.3. The Dyed Resection Margins appeared in the right-side in Figure 3.3 which is essential to guarantee whether the polyp is totally evacuated or not.



(a) Dyed & lifted polyps (b) Dyed resection margins

Figure 3.3: Polyp evacuation of Endoscopic imagery

3.3 Summary

This chapter describes three datasets that were used for our experimental analysis during the research. The two of them consists of retinal images of diabetic patients that were collected from a recently published kaggle dataset and another dataset provided to us by an ophthalmic clinic in Sri Lanka. The GI-tract dataset is a publicly available endoscopic dataset that was used to further evaluate our CNN-based model architectures.

4. METHODOLOGY

This chapter describes the research methodology that we performed in order to develop our classification and content-based retinal image retrieval tasks. Section 4.1 explicates the data preprocessing techniques that were used before feeding images directly to the CNN-based model architectures. Section 4.2 illustrates how we handled the class imbalance problem for our two retinal datasets. In section 4.3, we discuss the overview of the methodology of our classification model and the last section demonstrates the detailed description of the methodology of our retrieval model.

4.1 Preprocessing

The extra black margins (unwanted background) were removed in the retinal imagery as the first step of the preprocessing stage. In contrast, we required to transform the images in such a way that it would be feasible for any CNN to converge in a reasonable time. Retinal images were standardized by resizing all images into 224px x 224px since they were in different dimensions and aspect ratios. Moreover, poor quality low-resolution (bad-lighting conditions) and occluded images were removed from the datasets before feeding them into the models.

4.2 Addressing Class Imbalance

The retinal dataset contains unbalanced class distributions. This data imbalance problem creates additional overheads for the classification model. The imbalanced problem was handled by incorporating the weights of the classes into the cost function. The class weights were adjusted inversely proportional to class frequencies in the input data and then passed into the fit function of models as a parameter when training. The stratified cross-validation was used during the training process. Hence, each fold contains roughly the same proportion of observations as in the training dataset.

4.3 Overview of Classification Model Architecture

In this research task, we attempted different approaches of feature extraction using pretrained CNNs on two retinal imagery datasets. In order to obtain the predicted class label through an independent classifier, we used diverse combinations of extracted feature vectors. The steps of our proposed methodology for the severity stage classification are described below. Each dataset was preprocessed as described in the preprocessing section 4.1 as the first step and addressed the class imbalance problem for the training retinal datasets as described in section 4.2. Secondly, six different CNN architectures followed by a GAP layer were used to produce feature vectors. Next, the last feature vector was acquired by joining vectors from the past advance for classification. This method can be referred to as a CNN Transfer Learning Ensemble feature extraction approach.

The set of experiments described in chapter 5 revealed that ResNet-18, VGG-16, and DenseNet-201 pretrained CNNs as feature extractors produced the most accurate results both in cross-validation and predictions on the test data. Therefore, we decided to use a combination of features extracted from VGG-16, ResNet-18, and DenseNet-201 pretrained CNNs in order to build an ensemble model. Each image from each dataset is used as input for the DenseNet-201 CNN and 1920 features extracted from the feature extractor by applying the global average pooling layer. Similarly, each dataset was processed through both VGG-16 and ResNet-18 similar to the DenseNet-201 and we received two sets of 512 features each. Subsequently, for each image we got a vector with 2944 (1920 + 2x512) features that represent the image. As the next step, in order to eliminate redundant and noisy features, we normalized and applied SVD on this concatenated feature vector. The optimal number of features was selected by using truncated SVD with a variance threshold of 95%.

Finally, we fed this into a 128 units single hidden layer (with the ReLU activation) ANN (Artificial Neural Network) model with a softmax activation layer to obtain the best classification accuracy. For the model training process, we used a stratified five-

fold cross-validation. Each fold has the same proportion of observations as in the training dataset. The overall architecture of our novel classification model is shown in Figure 4.1. We tried numerous model configurations for our classifiers before we

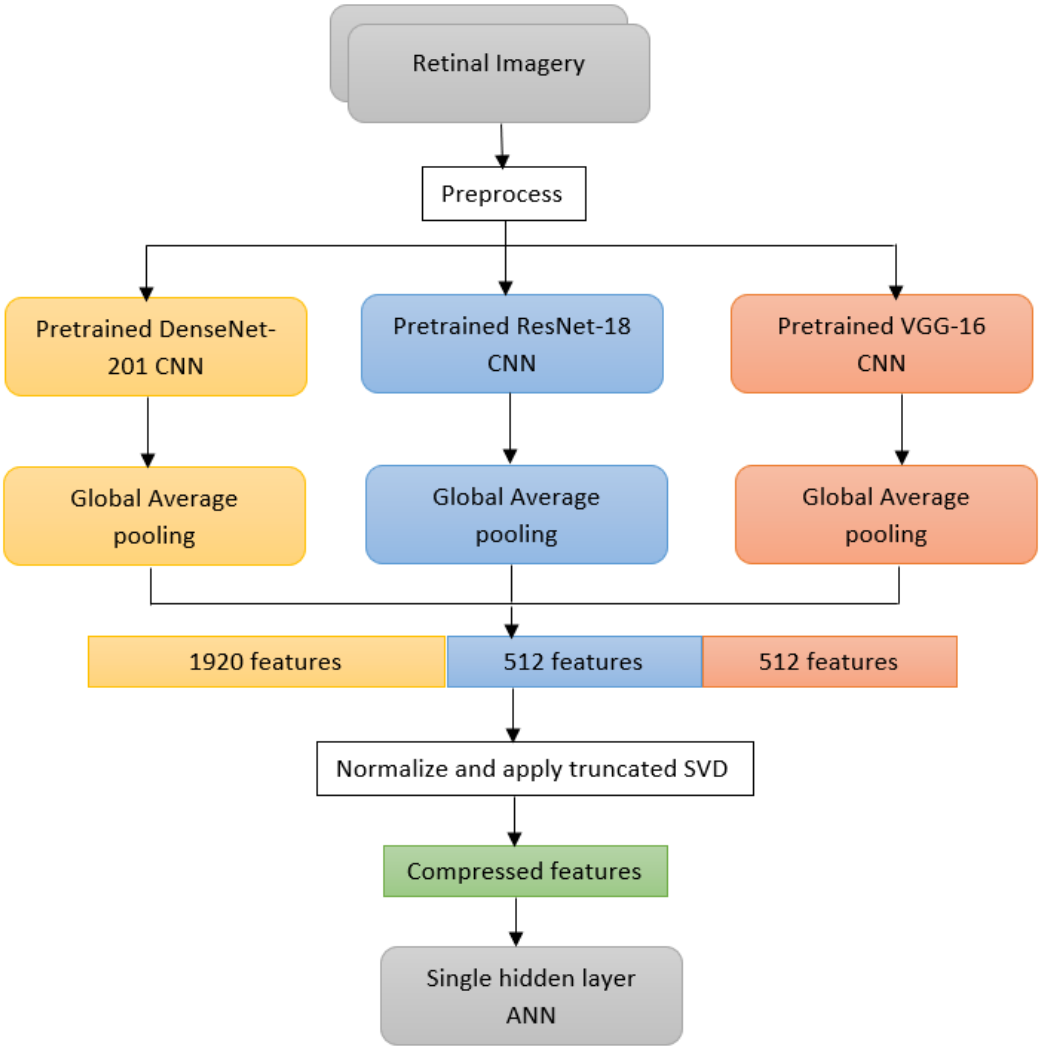


Figure 4.1: Ensemble Method

found the best solution (single hidden layer ANN) that worked for our specific problem. The mini-batch stochastic gradient descent (SGD) algorithm was used in order to train the ANN classifier. The batch size was set to 64 for training sets as well as validation datasets and momentum was set to 0.9. We penalized large weights by a factor of 1×10^{-3} by using ridge regression (L2-regularization). The dropout was set to 0.8 probability to stop the activations for the hidden layer in the ANN

classifier. In order to normalize the activation values of the hidden layer, we used Batch Normalization [89] and it is focused on faster optimization by reducing the internal covariate shift, which constantly changes the distribution of the activations during model training. The cost function was categorical cross-entropy and the learning-rate we set to 1×10^{-3}. After 400 epochs, the gradient descent converged to the optimal solution and the training took approximately seven hours. All weights in the ANN were initialized using the He [90] initialization scheme.

4.4 Overview of Retrieval Model Architecture

The main objective of this study is to find a hash function which solves the CBIR task for diabetic retinopathy. Given N number of retinal images $X=\{x_1, x_2, \ldots\ldots, x_N\}$ belonging to five categories as described in section 3.1. The class label is defined as $Y=\{y_1, y_2, \ldots\ldots.., y_N\}$ where each $y_i \in\{0, 1, 2, 3, 4\}$. Our goal is to learn a hash function H(x) which maps retinal images to compact binary hash codes $b_i=H(x_i)$ and $b_i \in \{0, 1\}^D$ where D represents the length of the hash code. This hash function satisfies the two properties as per the following. b_i and b_j are close to each other in the hamming space when $y_i=y_j$ and far away when $y_i \neq y_j$.

There are three main components of this approach. The first component is to train the ensemble CNN model as described in section 4.3 using the retinal image dataset to learn rich mid-level signatures of the images. The second component is used to train another ANN which comprises a single hidden layer with sigmoid activation as shown in Module 2 of Figure 4.2 by feeding the extracted features from the feature extractor of our classification model to learn binary hash codes. The third component retrieves retinal images similar to the query image through the hierarchical deep search as described in section 4.4.2. This step is used as a coarse-to-fine strategy to retrieve similar clinically relevant retinal images by utilizing the learned compact binary codes and mid-level signatures of the images. Our approach for learning compact binary hash embeddings is explained in section 4.4.1. Moreover, the combination of two retinal imagery datasets was used for this study. The proposed image retrieval architecture through hierarchical deep search is shown in Figure 4.2.

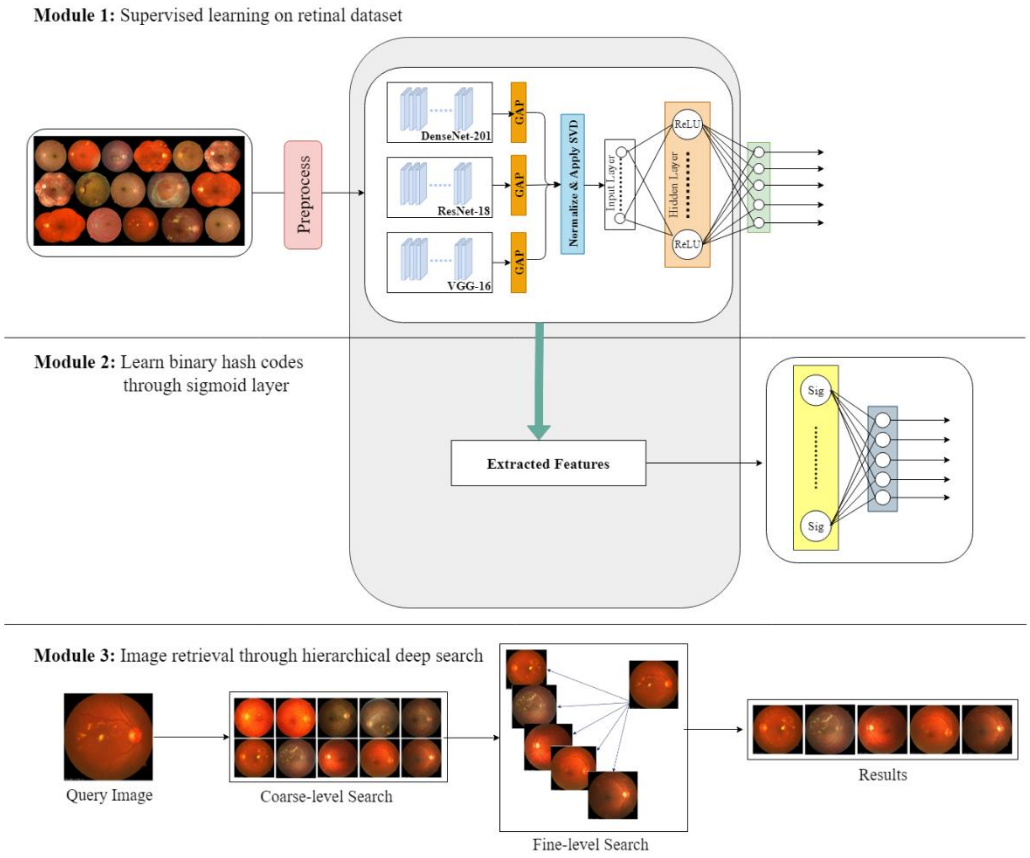


Figure 4.2: Image Retrieval CNN-based Architecture

4.4.1 Learning Binary Hash Codes

We can use rich mid-level signatures of the images that are extracted from our feature extractor (by removing the softmax layer) as shown in Module 2 of Figure 4.2 to perform the similarity measurement with the given query image. However, these image representations are high-dimensional feature vectors that are inefficient for content-based image retrieval in a corpus which consists of a considerable number of images. A feasible approach in order to perform efficient image retrieval is to transform the feature vectors into binary hash codes to reduce memory and time consumption. Then such compact binary codes can be rapidly compared using Hamming distance.

We attempted to learn good image signatures and a hash function through our ensemble CNN model architecture. In order to learn a hash function, we assumed that the final predictions of the classification layer with softmax activation rely on a

set of hidden neurons with each neuron being on or off according to a given threshold. That is to say, retinal images inducing similar binary activations would have the same severity stage of DR. In order to accomplish this approach, we fed extracted features of our dataset from the feature extractor as shown in module 2 of Figure 4.2 to another single hidden layer ANN (ANN-2) classifier where each neuron in the hidden layer contains sigmoid activation. The sigmoid layer is a dense layer (fully connected layer), which is connected to the succeeding softmax layer that accomplishes classification. All weights were initialized using the He [90] initialization scheme before we train the ANN (ANN-2). We used the sigmoid layer as the feature extractor to retrieve the semantic binary codes for each retinal image by removing the softmax layer from the ANN (ANN-2). We have done experiments by changing the number of neurons into 28, 64, 128 and 256 of the sigmoid layer at each time when the model is training in order to identify the best suitable length of the binary code to retrieve similar retinal images with higher accuracy.

4.4.2 Hierarchical Deep Search for Image Retrieval

In order to retrieve the retinal images with higher accuracy, we implemented a coarse-to-fine search strategy. First, we computed the hamming distance using the generated binary codes in order to retrieve a group of candidates that are similar to the query image. This candidate list then sorted using the cosine similarity with the query image based on the rich mid-level signatures of the images that are extracted from the feature extraction part of the classification model since clinically similar retinal images may have identical compact binary hash codes.

In the coarse level search strategy, first, we extracted features through the sigmoid layer as described in section 4.4.1 for a given image as the image signature. The binary codes were then obtained by binarizing each activation value of the sigmoid layer by 0.5 thresholds. If an activation value greater than or equal to the 0.5 then it outputs one else zero.

Let $X=\{x_1, x_2, x_3, \ldots\ldots, x_N\}$ represent the retinal dataset of N images for the retrieval task. The corresponding binary codes of all the retinal images are denoted as $H_b=\{b_1, b_2, b_3, \ldots\ldots., b_N\}$ with $b_i \in \{0, 1\}^D$. Given a query image I_q and its binary codes b_q, we identify an m candidate pool, $L=\{x_1, x_2, x_3, \ldots\ldots, x_m\}$, if the Hamming distance between b_q and $b_i \in H_b$ is lower than or equal to 0.5 thresholds.

During the fine level search, we used rich mid-level features extracted from the feature extractor (see Module 2 of Figure 4.2) of our classification model for all the retrieved m retinal images in the candidate list L in order to rank them according to the distance with respect to the query image. Here we took the cosine similarity between each rich mid-level feature vector of our candidate list L with the real-valued feature vector of the query image I_q. The similarity of the two images is in higher value if they have larger cosine similarity. We retrieve top k (e.g. k=20) ranked images by ranking m ($m \geq k$) candidates in the list L in descending order according to the similarity score.

4.5 Summary

This chapter explicates the methodology for our two tasks namely classification model construction to predict severity stages of the diabetic retinopathy and similar case(s) retrieval for a given query image according to the proposed research objectives. First, we built a classification model by extracting deep features through an ensemble of pretrained-CNNs (VGG-16, DenseNet-201, and ResNet-18) followed by a GAP layer as a single feature vector and then extend it to a retrieval model by using a deep supervised hashing approach in order to perform efficient retinal image retrieval, where we implicitly learn a good image representation along with a similarity-preserving compact binary hash code for each image. Moreover, we used a technique of reducing memory consumption and processing time while preserving classification and retrieval performance by using dimensional reduction based on SVD.

5. EXPERIMENTAL ANALYSIS & MODEL EVALUATION

This chapter describes the experimental setup and the model evaluation that we conducted in order to properly evaluate our classification and content-based retinal image retrieval tasks. Section 5.1 elucidates the experimental analysis for classification task including fine-tuning CNN models and feature extraction through pretrained-CNN models. Section 5.2 explains the current state-of-the-art deep learning-based ensemble approaches for the comparison with our approach. In sections 5.3 and 5.4, we discuss the evaluation and results of the ensemble classifier and retrieval model respectively.

5.1 Experimental Analysis for Classification Model

We experimented with six different prominent CNNs using two methods as described in sections 5.2 and 5.3. The subsequent set of experiments measures the accuracy and F1-measure of our two retinal test datasets separately using CNNs and classifiers. We attempt to improve the accuracy of validation sets by fine-tuning different hyperparameter values. Moreover, we have done experimental analysis for the GI-tract dataset as well. In order to prove the effectiveness and efficiency of our classification approach, we further test it on another medical image dataset called KVASIR which is described in section 3.2.

5.2 Fine-tuning CNN Models

In our first experiment, we attempt to perform DR classification by fine-tuning pretrained-CNNs, which included DenseNet-201 [24], ResNet-18 [23], InceptionV3 [21], InceptionResNetV2 [91], VGG-16 [20] and Xception [22]. First, we load a domain transferred standard CNN architecture and replace the last fully connected layer (output layer) with a custom softmax layer which comprises five neurons to perform the classification task. We initialize the weights of layers from the input layer to the last pooling layer using ImageNet weights and weights of the custom softmax layer using the He [90] initialization scheme.

We first froze up to the last pooling layer and warm-up the newly added fully connected head by fine-tuning the randomly initialized weights, because if we allow the gradient to backpropagate from these random weights through the entire network, we risk vanishing the powerful low-level features from the pretrained early convolutional layers. Next, we unfroze the rest of the network (allow all layers to train including softmax layer) and continue the training process.

Table 5.1: Results of Fine-tuning Pretrained CNN Models

Dataset	**Pretrained CNN**	**Accuracy**	**F1-Measure**
DIABRET	DenseNet-201	68.24%	0.6581
	ResNet-18	65.33%	0.6332
	VGG-16	66.41%	0.6339
	InceptionV3	63.87%	0.6063
	Xception	65.93%	0.6395
	InceptionResnetV2	59.36%	0.5721
Kaggle	DenseNet-201	77.87%	0.7551
	ResNet-18	77.43%	0.7479
	VGG-16	74.17%	0.7231
	InceptionV3	71.56%	0.6902
	Xception	72.82%	0.6779
	InceptionResnetV2	71.58%	0.6586
KVASIR	DenseNet-201	74.13%	0.7257
	ResNet-18	74.07%	0.7203
	VGG-16	73.24%	0.6828
	InceptionV3	67.77%	0.6329
	Xception	70.43%	0.6778
	InceptionResnetV2	65.71%	0.6303

We fine-tuned each CNN up to 400 epochs using the mini-batch Stochastic Gradient Descent (SGD) algorithm and set the momentum to 0.9. The learning rate was set to 0.001 and we used a minibatch size of 64. We set the L2- regularize parameter to

0.001 to penalize large weights. The best result for both datasets is achieved using the DenseNet-201 architecture. Moreover, we have done the same experimental setup for the KAVISIR dataset as well. The experimental result for this approach is shown in Table 5.1. This approach presented low accuracy and required more computational time. Hence, we discarded this method.

5.3 Feature Extraction Based on CNN Models

In our second experimental setup, we first extracted the features of the retinal image using a deep CNN. Then, we fed the extracted features to a classification model in one of three different approaches. The simplest is that we fed the features directly to the classification model (see Table 5.2, Table 5.5 and Table 5.8). In the second approach, we applied the Global Average Pooling (GAP) before feeding the extracted features to the classification model (see Table 5.3, Table 5.6 and Table 5.9). In the third approach, we applied truncated Singular Value Decomposition (SVD) in addition to Global Average Pooling (GAP) before feeding the extracted features to the classification model (see Table 5.4, Table 5.7 and Table 5.10).

We experimented with each pretrained CNN described in section 5.2 for feature extraction. As for the classification model, we experimented with SVMs, ANNs and Random Forest. Altogether, this leads to 6x3x3 experimental combinations corresponding to the 6 different feature extractors, 3 different approaches of feeding the features to the classifier and the 3 different types of classification models. The summary of the results is in Table 5.2 to Table 5.10 for the ANN, Random Forest classifiers and SVM, for each dataset.

We have achieved the best result with the DenseNet-201 feature extractor for both datasets along with a GAP layer and SVD, and a single hidden layer ANN as the classifier. The accuracy values reported in Table 5.2 to Table 5.10 for each experiment are based on predictions for the test datasets. Cross-validation on the training sets too confirm a similar ranking of configurations in terms of accuracy/F1 measure. According to these results, we proposed a combination of feature vectors

extracted through VGG-16, DenseNet-201, and ResNet-18 CNN architectures as shown in Figure 4.1 to increase our classification accuracy.

Table 5.2: Results for different CNN feature extractors with ANN

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	59.43%	0.4915
	ResNet-18	61.56%	0.5078
	VGG-16	55.36%	0.4703
	InceptionV3	57.89%	0.4851
	Xception	61.37%	0.4923
	InceptionResnetV2	62.15%	0.5310
Kaggle	DenseNet-201	80.86%	0.7163
	ResNet-18	79.01%	0.6664
	VGG-16	77.43%	0.6435
	InceptionV3	71.59%	0.6033
	Xception	71.85%	0.6052
	InceptionResnetV2	70.48%	0.5871
KVASIR	DenseNet-201	67.52%	0.6683
	ResNet-18	68.27%	0.6781
	VGG-16	65.39%	0.5907
	InceptionV3	67.11%	0.5892
	Xception	63.12%	0.5839
	InceptionResnetV2	64.89%	0.5767

Table 5.3: Results for different CNN feature extractors followed by a GAP layer with ANN

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	92.94%	0.9212
	ResNet-18	91.60%	0.9028
	VGG-16	90.28%	0.9001
	InceptionV3	84.64%	0.8144
	Xception	84.62%	0.8224
	InceptionResnetV2	83.25%	0.8123
Kaggle	DenseNet-201	87.89%	0.8234
	ResNet-18	87.68%	0.8199
	VGG-16	86.70%	0.8152
	InceptionV3	75.94%	0.6824
	Xception	79.17%	0.6957
	InceptionResnetV2	77.28%	0.6831
KVASIR	DenseNet-201	90.74%	0.9003
	ResNet-18	88.43%	0.8731
	VGG-16	84.37%	0.8411
	InceptionV3	80.72%	0.7947
	Xception	79.93%	0.7792
	InceptionResnetV2	77.57%	0.7591

Table 5.4: Results for different CNN feature extractors followed by a GAP layer and SVD with ANN

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	97.20%	0.9691
	ResNet-18	96.71%	0.9627
	VGG-16	96.58%	0.9624
	InceptionV3	96.47%	0.9613
	Xception	96.14%	0.9587
	InceptionResnetV2	95.71%	0.9559
Kaggle	DenseNet-201	92.88%	0.9146
	ResNet-18	90.09%	0.8740
	VGG-16	88.58%	0.8567
	InceptionV3	86.53%	0.8395
	Xception	87.16%	0.8462
	InceptionResnetV2	87.42%	0.8314
KVASIR	DenseNet-201	95.28%	0.9496
	ResNet-18	93.26%	0.9319
	VGG-16	92.87%	0.9228
	InceptionV3	90.71%	0.9006
	Xception	90.96%	0.8933
	InceptionResnetV2	91.91%	0.9172

Table 5.5: Results for different CNN feature extractors with SVM

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	65.22%	0.5420
	ResNet-18	64.07%	0.5236
	VGG-16	54.22%	0.4475
	InceptionV3	60.70%	0.5027
	Xception	62.21%	0.4893
	InceptionResnetV2	64.76%	0.5506
Kaggle	DenseNet-201	84.95%	0.7402
	ResNet-18	76.82%	0.6550
	VGG-16	74.09%	0.6113
	InceptionV3	71.83%	0.5929
	Xception	74.02%	0.6175
	InceptionResnetV2	71.24%	0.5880
KVASIR	DenseNet-201	70.22%	0.6991
	ResNet-18	69.12%	0.6243
	VGG-16	67.15%	0.5325
	InceptionV3	61.50%	0.5045
	Xception	60.19%	0.5864
	InceptionResnetV2	59.76%	0.5436

Table 5.6: Results for different CNN feature extractors followed by a GAP layer with SVM

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	68.02%	0.6283
	ResNet-18	66.76%	0.5941
	VGG-16	64.57%	0.5747
	InceptionV3	58.38%	0.5254
	Xception	60.55%	0.5342
	InceptionResnetV2	59.55%	0.5419
Kaggle	DenseNet-201	84.06%	0.7705
	ResNet-18	74.16%	0.6369
	VGG-16	71.67%	0.6041
	InceptionV3	69.46%	0.5823
	Xception	70.48%	0.6010
	InceptionResnetV2	67.47%	0.5612
KVASIR	DenseNet-201	72.42%	0.7204
	ResNet-18	70.53%	0.6971
	VGG-16	69.41%	0.6738
	InceptionV3	64.83%	0.6499
	Xception	61.25%	0.5942
	InceptionResnetV2	60.78%	0.5812

Table 5.7: Results for different CNN feature extractor followed by a GAP layer and SVD with SVM

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	71.34%	0.6520
	ResNet-18	69.43%	0.6213
	VGG-16	68.80%	0.5997
	InceptionV3	61.70%	0.5407
	Xception	65.64%	0.5754
	InceptionResnetV2	64.78%	0.5772
Kaggle	DenseNet-201	85.27%	0.7790
	ResNet-18	80.37%	0.6621
	VGG-16	78.63%	0.6147
	InceptionV3	75.49%	0.5971
	Xception	76.98%	0.6266
	InceptionResnetV2	75.00%	0.5943
KVASIR	DenseNet-201	79.52%	0.7830
	ResNet-18	77.82%	0.7431
	VGG-16	73.51%	0.6998
	InceptionV3	69.50%	0.6569
	Xception	64.94%	0.6033
	InceptionResnetV2	62.51%	0.5976

Table 5.8: Results for different CNN feature extractors with Random Forest

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	60.12%	0.4989
	ResNet-18	58.63%	0.4767
	VGG-16	51.64%	0.4663
	InceptionV3	54.93%	0.4819
	Xception	55.21%	0.4338
	InceptionResnetV2	60.36%	0.5045
Kaggle	DenseNet-201	70.59%	0.5459
	ResNet-18	68.03%	0.5364
	VGG-16	67.77%	0.5127
	InceptionV3	61.96%	0.4863
	Xception	63.86%	0.5435
	InceptionResnetV2	60.06%	0.5302
KVASIR	DenseNet-201	79.53%	0.7734
	ResNet-18	76.32%	0.7561
	VGG-16	74.65%	0.7088
	InceptionV3	70.36%	0.6743
	Xception	70.06%	0.6610
	InceptionResnetV2	67.56%	0.6286

Table 5.9: Results for different CNN feature extractors followed by a GAP layer with Random Forest

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	68.09%	0.5562
	ResNet-18	67.21%	0.6214
	VGG-16	66.34%	0.5700
	InceptionV3	58.98%	0.5199
	Xception	60.19%	0.5166
	InceptionResnetV2	60.10%	0.5137
Kaggle	DenseNet-201	74.16%	0.6516
	ResNet-18	70.22%	0.6372
	VGG-16	69.11%	0.6137
	InceptionV3	64.87%	0.5381
	Xception	60.32%	0.4838
	InceptionResnetV2	61.60%	0.5287
KVASIR	DenseNet-201	80.05%	0.7964
	ResNet-18	78.91%	0.7652
	VGG-16	75.31%	0.7261
	InceptionV3	72.30%	0.6813
	Xception	70.15%	0.6714
	InceptionResnetV2	68.11%	0.6402

Table 5.10: Results for different CNN feature extractor followed by a GAP layer and SVD with Random Forest

Dataset	Pretrained CNN	Accuracy	F1-Measure
DIABRET	DenseNet-201	69.20%	0.6625
	ResNet-18	68.04%	0.6291
	VGG-16	67.62%	0.5815
	InceptionV3	62.98%	0.5546
	Xception	57.58%	0.4983
	InceptionResnetV2	60.78%	0.5377
Kaggle	DenseNet-201	77.18%	0.6525
	ResNet-18	71.72%	0.6014
	VGG-16	69.92%	0.5929
	InceptionV3	66.85%	0.5844
	Xception	65.34%	0.5555
	InceptionResnetV2	65.94%	0.5652
KVASIR	DenseNet-201	82.31%	0.8045
	ResNet-18	80.89%	0.7991
	VGG-16	78.52%	0.7601
	InceptionV3	74.93%	0.7092
	Xception	71.63%	0.6824
	InceptionResnetV2	69.05%	0.6541

5.3.1 Hyperparameter-tuning

For the ANN model, hyperparameters were selected from the parameter space using the orthogonalization concept (a randomized search strategy). In this approach, we examined our five-fold cross-validation error and the training error results. If the training error is low (i.e. does not under-fit the data) but the gap between the training error and cross-validation error is high, we observed that our model failed to generalize to new examples. Hence, in order to deal with such an overfitting problem, we used regularization techniques such as dropout, L2 regularization, batch-normalization and normalize extracted features by removing the mean and scaling to unit variance. In contrast, if the training set is not performed well on the cost function, we first increased the number of epochs to run gradient descent longer. As the next step, we made an effort to use Adam optimizer for all the cases, because it leads to a much faster convergence time. However, we achieved the best results for the two datasets through SGD with momentum. We settled on a learning rate of 0.001 for all of these situations because changing the learning rate below or beyond did not lead to noticeable improvements for our results. In addition to the aforementioned steps, we changed the number of hidden layers (1 to 3 layers) and neurons per layer (64, 128, 256 and 512) to improve our model performance. We got the highest prediction percentage over the five-fold cross-validation for a single hidden layer with ReLU activations (128 hidden units) ANN for our two datasets.

In random Forest, we fine-tuned n-estimators, max features, min samples leaf, and max depth parameters through GridSearchCV, which is an exhaustive search strategy over specified parameter values for a given estimator. N-estimators denotes the number of trees in the random forest. Typically, the higher the number of trees the better to learn the model. Though adding a lot of decision trees slow down the training process and accuracy considerably, thus we did a hyper-parameter search to find the best number of trees. We used an array of values ranging from 100 to 1000 by steps of size 100 for N-estimators. The best number of trees we have got was 500 for both cases. Max depth denotes the depth of each tree in the forest. The deeper the decision tree, the more splits it has and it captures further information about the

dataset. We fitted each decision tree with depths (max depths) ranging from 5 to 35 by steps of size 5. Min samples leaf is the minimum number of samples that required to be at a leaf node and we used the values ranging from 10 to 50 by steps of size 10. Max features denote the number of features to consider when looking for the best split and the values ranging from 10 to 150 by steps of size 10. The best values for max depth, min samples leaf, and max features were 5, 40 and 20 respectively for the best fit random forest model for the Vision Care dataset. For the Kaggle dataset, we achieved maximum accuracy with 10, 40 and 30 for the aforementioned hyper-parameters respectively.

Similarly, for SVM we used GridSearchCV to find the optimum values for the kernel (linear or RBF) and soft-margin parameter (C). In SVM, kernel supports to find a hyperplane in the higher dimensional space without increasing the computational cost much. We used linear and radial basis function as the kernels. For large values of soft-margin parameter C, the learning algorithm will select a smaller margin hyperplane if that hyperplane does a better job of getting all the training data classified properly. Conversely, a very small value of C will cause the algorithm to look for a larger margin separating hyperplane, even if that hyperplane misclassifies more data points. In our case, we used different values in the range of 0 to 1 (0.25, 0.5, 0.75, 1). Best classification accuracy achieved with the linear kernel when C = 0.5 and C = 0.75 for Vision Care and Kaggle datasets respectively.

We used different batch sizes such as 8, 16, 32, 64 and 128. But we achieved maximum performance when we use mini-batch size as 64.

5.4 Comparison Models for the Classification Task

In order to perform a proper comparison of our proposed approach relative to the current state-of-the-art classification approaches, we used three deep learning models from the literature and we evaluated them using our two retinal datasets. This section describes the models that we used for the comparisons and Table 5.11 summarizes the obtained results.

5.4.1 Method 1 : ResNet-152 + DenseNet-161 + ANN

For the first model, we used the ensemble CNN-based approach based on pretrained ResNet-152 and DenseNet-161 described in [92] to extract the features and feed into a single hidden layer ANN. We achieved the best performance for the SGD optimizer with a 0.001 learning rate and 32 hidden neurons with ReLU activation. We used 200 epochs in order to train the model. All weights in the ANN were initialized using the Xavier initialization scheme.

5.4.2 Method 2: Ensemble of ResNet-50 and Inception V3

Shahin et.al [93] have proposed a deep ensemble CNN based architecture to classify the seven different types of skin lesions. They have used two pre-trained CNN architectures namely Inception V3 and ResNet-50 in their architecture. They calculated the average of the output probabilities from the previously mentioned CNNs and choose the class with the highest probability. We implemented this architecture and trained on our two DR datasets. We could achieve the highest accuracies with the learning rate of 0.001 along with Adam optimizer and in 100 epochs. We used ImageNet pre-trained weights in order to initialize the network parameters for the convolutional blocks. The mini-batch size was set to 32 for both models.

5.4.3 Method 3 : Ensemble of AlexNet and GoogLeNet + PCA + one-vs-one multi-class SVM

Kumar et.al [94] have developed a deep ensemble technique to classify the medical images. Their high-performance model architecture consists of AlexNet and GoogLeNet feature extractors, Principle Component Analysis (PCA) as a feature selector and a one-vs-one multi-class SVM as the classifier. They extracted features from the fine-tuned GoogLeNet and AlexNet and then fed them to a one-vs-one multi-class SVM classifier for the training. The feature vectors extracted from each CNN have concatenated to form a single-dimensional vector. In order to reduce the dimensionality, they used PCA. We used this architecture with our DR datasets and

we could achieve the best accuracy with Adam optimizer, 0.005 learning rate, and dropout 0.5 in 400 epochs.

We use the same strategy with stratified cross-validation as described in section 4.2 to handle class imbalance problems when we train the aforementioned three classifiers.

5.5 Ensemble Classifier Evaluation and Results

In order to train our proposed ensemble classifier, we used each training set of our datasets. The corresponding test sets used to estimate the unbiased performance of the generalized models. We used two main metrics namely accuracy and F1-score for the performance evaluation of the proposed CNN ensemble approach on the three datasets. Moreover, we used three comparison models described in section 5.4 in order to compare the performance of our proposed approach.

Table 5.11: Our approach with comparison models

Approach	Dataset	Accuracy	F1-Measure
Our ensemble method	DIABRET	98.69%	0.9867
	Kaggle	98.63%	0.9834
	KVASIR	97.38%	0.9721
Method 1 (see section 5.4.1)	DIABRET	83.28%	0.8176
	Kaggle	89.60%	0.8736
	KVASIR	86.09%	0.8132
Method 2 (see section 5.4.2)	DIABRET	77.81%	0.7377
	Kaggle	81.33%	0.7651
	KVASIR	79.55%	0.7493
Method 3 (see section 5.4.3)	DIABRET	85.47%	0.8510
	Kaggle	90.48%	0.8926
	KVASIR	88.61%	0.8508

The experimental results in terms of the aforementioned evaluation metrics are shown in Table 5.11. As indicated by the outcomes that appeared in Table 5.1 to

Table 5.11, our proposed novel CNN architecture achieved a higher classification accuracy and F1-measure of over 98% compared to the comparison models and single CNN model architectures depicted in sections 5.2 and 5.3. In order to further understand how the classification is performing with respect to individual classes for each test set, we have provided F1-measure as shown in Table 5.12.

Table 5.12: F1-measure for individual classes

Dataset	Class Label	F1-Measure
DIABRET	Diabetes without Retinopathy	0.9971
	Mild-NPDR	0.9922
	Moderate-NPDR	0.9881
	Severe-NPDR	0.9598
	PDR	0.9963
Kaggle	Diabetes without Retinopathy	0.9957
	Mild-NPDR	0.9905
	Moderate-NPDR	0.9876
	Severe-NPDR	0.9483
	PDR	0.9949
KVASIR	Dyed-lifted-polyps	0.9701
	Dyed-resection-margins	0.9636
	Esophagitis	0.9954
	Normal-cecum	0.9568
	Normal-pylorus	0.9342
	Normal-z-line	0.9869
	Polyps	0.9981
	Ulcerative-colitis	0.9717

5.6 Retrieval Model Evaluation and Results

We demonstrate the experimental results of the proposed image retrieval method throughout this section. In order to analyze the quality of our proposed approach, we used three evaluation metrics namely, mean Average Precision (mAP) with different

hash code length by setting the number of neurons in the sigmoid layer, precision-recall curves, and Precision curves with respect to different numbers of top returned images. Moreover, We compared the performance of our proposed approach against seven state-of-the-art hashing techniques, including two unsupervised approaches namely LSH [31] and SH [40], two shallow supervised approaches KSH [33] and MLH [39], and three deep supervised approaches DNNH [41], DHN [42], and DLBHC [32]. Even though the deep learning-based state-of-the-art approaches demonstrate significant improvements, they were still inferior to our proposed retrieval approach (see results below). This indicates that our proposed deep supervised approach is much more beneficial than the comparison models. The performance gaps among state-of-the-art-approaches and the proposed approach are based on the model architecture. All techniques use similar training and test sets for a fair comparison.

5.6.1 Results on the Retinal Dataset

In order to evaluate the accuracy and the quality of our retrieval approach, we use our test dataset as the query images and extracted top-ranked similar images for quantitative evaluation. In Figure 5.2, we illustrate three sample queries with different DR severity stages and the top five corresponding retrieved examples using our deep CNN ensemble approach.

We directly used the image pixels as input for the deep learning-based techniques including DHN, DNNH, and DLBHC. For the shallow learning-based techniques, we use the same approach as mentioned in [33] and [95] in order to represent each image by a 512-dimensional GIST vector.

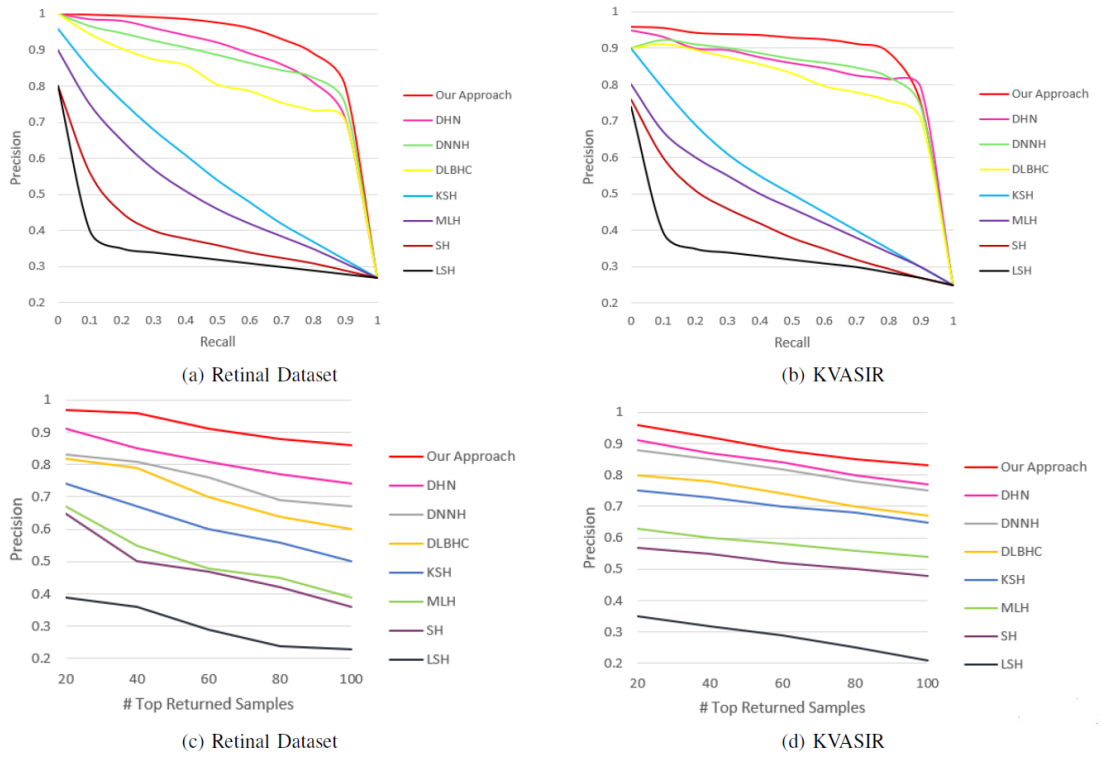


Figure 5.1: The results of comparison methods on the retinal and KVASIR datasets: (a)-(b) precision-recall curves @ 28-bits; (c)-(d) precision w.r.t. top returned samples curves @ 28-bits

The results of LSH are gained through our own implementation. The mAPs of the other baseline approaches are achieved through the open-source implementations provided by the corresponding authors. Moreover, we fine-tune each CNN architecture DNNH, DHN, and DLBHC up to 600 epochs with mini-batch Stochastic Gradient Descent (SGD) algorithm. We set the momentum of 0.9 and continue the training process with stratified cross-validation in order to give a fair comparison with our approach. We used a 0.0001 learning rate and a mini-batch size of 64. We set the L2- regularize parameter to 0.005 for penalizing the large weights.

We set the number of neurons with sigmoid activation in that layer to 28, 64, 128 and 256 to measure the effectiveness of the sigmoid layer in the ANN. Then, we apply the same configuration as described in the latter part of section 4.3 to train our model on the retinal dataset. Our approach with 28 and 64 sigmoid neurons achieved 99.30% mAP as shown in Table 5.13 and performs well against most of the test

images. We achieved 97.26% and 95.71% mAPs for 128 and 256 hash code lengths respectively for the top twenty returned images. This may be due to the over-fitting of our model when adding more neurons (such as 128 and 256 neurons) to the hidden layer. The experimental results are shown in Table 5.13 realize that the Hamming space becomes increasingly sparse and very few data points fall within the given Hamming ball when using longer binary hash codes. This explicates why our approach reaches the best performance for the relatively compact hash codes. Such compact binary representations are beneficial to save space whereas preserving retrieval accuracy.

Moreover, compared to the LSH, SH, KSH, MLH, DNNH, DLBHC and DHN our approach produces similarity-preserving binary hash codes using the retinal dataset with higher accuracy.

Table 5.13: Comparison of mAP of our approach with different hashing methods for the retinal dataset for top 20 returned images

Hashing method	**28-bits**	**64-bits**	**128-bits**	**256-bits**
Our approach	99.30%	99.30%	97.26%	95.71%
DHN	90.71%	86.75%	86.91%	87.84%
DNNH	83.37%	86.03%	85.78%	86.85%
DLBHC	82.45%	84.07%	84.40%	83.87%
KSH	73.33%	71.71%	71.71%	77.16%
MLH	67.24%	69.88%	70.13%	70.47%
SH	65.31%	65.67%	62.11%	61.24%
LSH	39.19%	43.38%	45.12%	51.07%

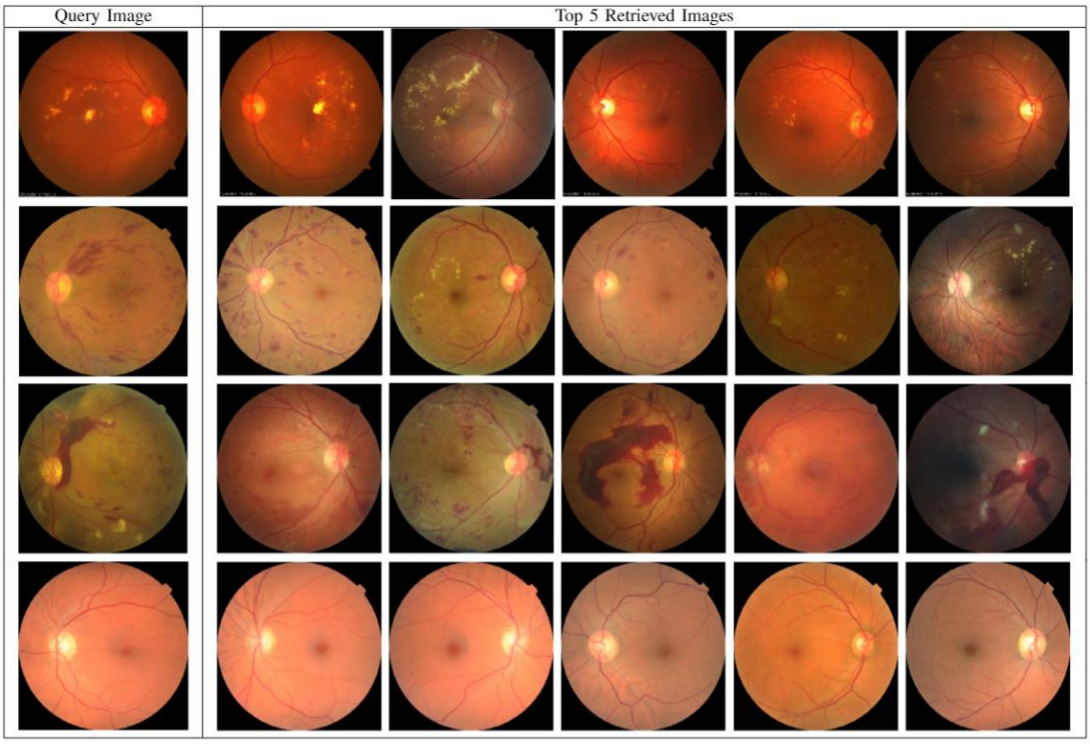


Figure 5.2: Top five returned results from the retinal image dataset

5.6.2 Results on Another Medical Dataset

We further test it on another medical image dataset called KVASIR [88] in order to demonstrate the effectiveness and efficiency of our approach. This dataset comprises 8000 Gastrointestinal tract images through the endoscopic process and these imageries are further categorized into eight different classes based on the anomaly type.

First, we have done our experiment with the classification task by setting 128 hidden layer neurons with ReLU activation and eight neurons with softmax activation for the classification layer of the ANN classifier. We then fine-tuned our classifier with stratified cross-validation using the entire training dataset. After 400 epochs, our proposed approach achieved 95.03% testing accuracy for the classification task.

During the next stage, we extracted features from the above-trained classifier and fed it into another single hidden layer ANN as described in section 4.4.1 in order to learn similarity preserving compact binary hash codes. Finally, we evaluated the retrieval

performance with respect to the KVASIR dataset and achieved 97.87% maximum mAP for the 28-bit hash code length as shown in Table 5.14. In Figure 5.3, we demonstrate three sample queries with different Gastrointestinal tract anomaly types and the top five corresponding retrieved examples using our deep CNN ensemble approach.

Table 5.14: Comparison of mAP of our approach with different hashing methods for KVASIR dataset for top 20 returned images

Hashing method	**28-bits**	**64-bits**	**128-bits**	**256-bits**
Our approach	97.87%	96.41%	95.65%	94.01%
DHN	91.11%	89.59%	88.19%	88.85%
DNNH	88.66%	88.15%	87.04%	85.93%
DLBHC	80.27%	84.83%	84.41%	82.95%
KSH	75.69%	69.15%	73.54%	74.81%
MLH	63.16%	67.58%	65.23%	69.91%
SH	57.18%	56.42%	57.35%	55.77%
LSH	35.71%	39.95%	41.29%	44.33%

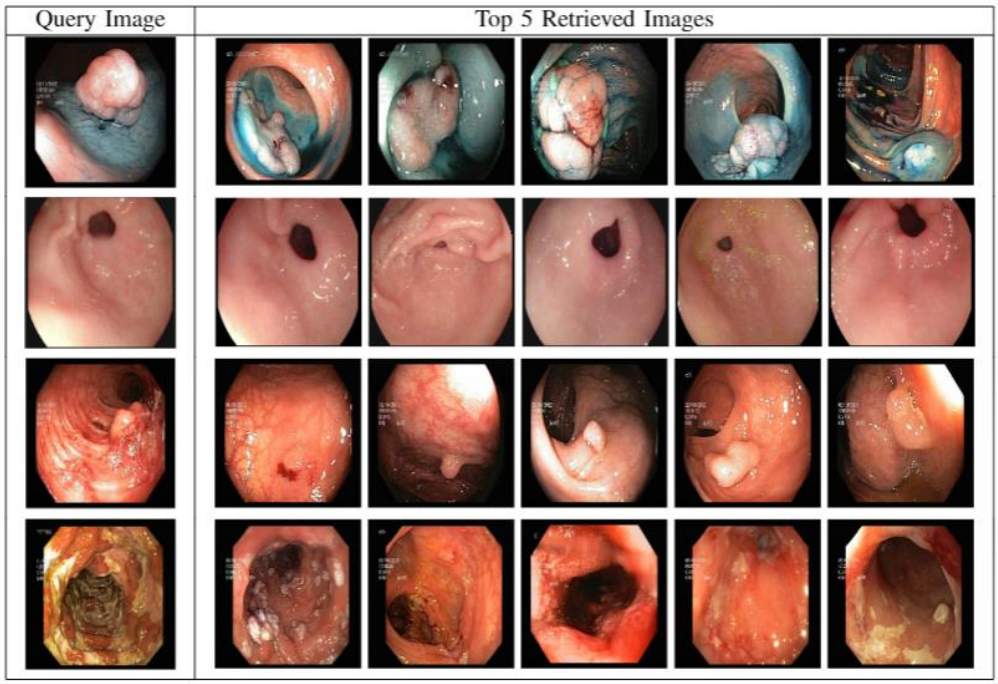


Figure 5.3: Top five returned results from KVASIR image dataset

Figure 5.1 shows the experimental results of precision-recall curves (see (a) and (b) in Figure 5.1) and precision curves with 28-bits relative to the different numbers of top returned images (see (c) and (d) in Figure 5.1) on the retinal and KVASIR datasets respectively. Moreover, these curves demonstrate that our proposed novel technique generally beats CNN-based comparison hashing techniques by a considerable margin and the conventional hash learning methods by a large margin for both datasets. Our approach is desirable for precision-oriented content-based image retrieval systems since it accomplishes particularly decent results at lower recall levels for both datasets.

5.7 Summary

This chapter describes the experimental results and model evaluation for our two tasks defined according to the objectives of this research to accomplish the classification model to predict the severity stages of the diabetic retinopathy and similar case(s) retrieval for a given query image. We used two main metrics namely accuracy and F1-score for the performance evaluation of the proposed CNN-based classification architecture and used three evaluation metrics to evaluate our retrieval model namely, mean Average Precision (mAP) with different hash code length by setting the number of neurons in the sigmoid layer, precision-recall curves, and Precision curves with respect to different numbers of top returned images. Moreover, this chapter gives a detailed experimental analysis and evaluation relative to the recently published studies.

6. CONCLUSION

This chapter elaborates on the contribution related to this dissertation and future directions. The contribution of this dissertation depends on the two tasks, in particular, the classification and retrieval approach based on the objectives of the thesis. The last section describes the future work of this research study.

6.1 Contribution

Early detection of DR by examining retinal images is in high demand as various individuals are left out from the medicinal services because of limited resources, especially in provincial territories, for example, qualified clinicians or suitable equipment. Conversely, the conventional DR diagnosing framework requires a manual evaluation process, which is monotonous and depends overwhelmingly on the aptitude of ophthalmologists and well-trained practitioners. In addition, the present structure will end up being altogether deficient because of the number of people with diabetes increases. Hence, automatic severity stage classification and similar case(s) retrieval from a retinal image database can be used for screening and treatment prioritization in order to assist and accelerate the clinical decision-making process for DR to diminish irreversible vision loss among diabetic patients.

There is space for further improvement of classification and retrieval models of DR compared to the previous studies that have been proposed by numerous research groups and it can be done by tuning hyperparameters or an ensemble of pretrained CNNs as feature extractors or using an ensemble learning approach through weak learners.

In order to overcome the limitations mentioned in 2.2 and 2.3, this research introduced an ensemble model based on transfer learning with CNNs for the classification and content-based image retrieval tasks. A concatenated deep feature vector was produced by an ensemble of pretrained CNNs (ResNet-18, VGG-16, and DenseNet-201) in order to predict five-class severity levels of diabetic retinopathy. Moreover, we describe a dimensionality reduction technique with the combination of

GAP and SVD in order to reduce processing time while preserving classification accuracy. A global average pooling layer was applied after the last pooling layer of each pretrained CNN before performing the feature extraction. Next, we concatenated extracted features of each pretrained CNN, normalized and then applied truncated SVD. Due to the usage of this combination, we could overcome the over-fitting problem while maximizing the F1-measure compared to the current state of the art techniques. The proposed method with pretrained CNNs shows a promising F1-measure of over 98%.

Moreover, this classification approach was extended to a retrieval model, which learned good image signatures in order to represent the retinal images as well as a hash function through an ensemble CNN model for retinal image retrieval of diabetic retinopathy. The proposed retrieval model architecture with pretrained CNNs shows a considerable improvement compared to the other several recently published hashing techniques on the retinal and KVASIR datasets.

6.2 Future Works

This dissertation presents considerable improvements to the retinal image classification and content-based image retrieval tasks for diabetic retinopathy when compared to the previous studies described in section 2. Hence, there are numerous directions that can be engaged in future research studies. These future directions are described as follows.

A large clinical evaluation of the proposed classification and retrieval techniques can be undertaken for further validation to make commercialized software to analyze the diabetic retinopathy. This computerized analysis of fundus images allows better identification of the progress of the disease, enabling early treatment for the individuals. Moreover, a complete automated DR system can be developed in order to segment the normal (such as optic disc) and abnormal features through deep learning techniques, to analyze the quality of images.

References

[1] N. H. Lents, “The Poor Design of the Human Eye,” 2015. [Online]. Available: https://thehumanevolutionblog.com/2015/01/12/the-poor-design-of-the-human-eye/. [Accessed: 10-May-2019].

[2] H. Li and O. Chutatape, “Fundus image features extraction,” *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 4, pp. 3071–3073, 2000.

[3] M. Garcia, R. Hornero, C. I. Sanchez, M. I. Lopez, and A. Diez, “Feature Extraction and Selection for the Automatic Detection of Hard Exudates in Retinal Images,” in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 4969–4972.

[4] T. R. C. of Ophthalmologists, “The Royal College of Ophthalmologists Diabetic Retinopathy Guidelines,” *Diabet. Retin. Guidel.*, 2013.

[5] World Health Organization, *Global Report on Diabetes*. 2016.

[6] A. Tamkin, I. Usiri, and C. Fufa, “Deep CNNs for Diabetic Retinopathy Detection,” pp. 1–6, 2013.

[7] G. M. Galshetwar, L. M. Waghmare, A. B. Gonde, and S. Murala, “Edgy salient local binary patterns in inter-plane relationship for image retrieval in Diabetic Retinopathy,” *Procedia Comput. Sci.*, vol. 115, pp. 440–447, Jan. 2017.

[8] C. G. Baby and D. A. Chandy, “Content-based retinal image retrieval using dual-tree complex wavelet transform,” in *2013 International Conference on Signal Processing , Image Processing & Pattern Recognition*, 2013, pp. 195–199.

[9] Sivakamasundari J., Kavitha G., Natarajan V., and Ramakrishnan S., “Proposal of a Content Based retinal Image Retrieval system using Kirsch template based edge detection,” in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, 2014, pp. 1–5.

[10] U. T. V. Nguyen, A. Bhuiyan, L. A. F. Park, and K. Ramamohanarao, “An effective retinal blood vessel segmentation method using multi-scale line detection,” *Pattern Recognit.*, vol. 46, no. 3, pp. 703–715, Mar. 2013.

[11] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[12] M. Shakeri *et al.*, “Sub-cortical brain structure segmentation using F-CNN’S,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 269–272.

[13] D. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.

[14] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-Scale Video Classification with Convolutional Neural Networks,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[15] C. Szegedy, A. Toshev, and D. Erhan, “Deep Neural Networks for Object Detection,” in *NIPS*, 2013, pp. 2553–2561.

[16] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[17] “All About Vision - Complete Guide To Vision and Eye Care.” [Online]. Available: https://www.allaboutvision.com/. [Accessed: 02-Dec-2018].

[18] Y. LeCun *et al.*, “Handwritten Digit Recognition with a Back-Propagation Network,” in *NIPS*, 1990, pp. 396–404.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *NIPS*, 2012, pp. 1097–1105.

[20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *ICLR*, 2015.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[22] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[24] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[25] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *In the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.

[26] J. Gu *et al.*, “Recent Advances in Convolutional Neural Networks,” Dec.

2015.

[27] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” Jul. 2012.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[29] M. Esnaashari, S. Amirhassan Monadjemi, and G. Naderian, “A Content-based Retinal Image Retrieval Method for Diabetes-Related Eye Diseases Diagnosis,” *Int. J. Res. Rev. Comput. Sci.*, vol. 2, no. 6, 2011.

[30] J. Jun Wang, S. Kumar, and S.-F. Shih-Fu Chang, “Semi-Supervised Hashing for Large-Scale Search,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2393–2406, Dec. 2012.

[31] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, “Locality-sensitive hashing scheme based on p-stable distributions,” in *Proceedings of the twentieth annual symposium on Computational geometry - SCG ’04*, 2004, p. 253.

[32] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, “Deep learning of binary hash codes for fast image retrieval,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 27–35.

[33] Wei Liu, Jun Wang, Rongrong Ji, Yu-Gang Jiang, and Shih-Fu Chang, “Supervised hashing with kernels,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2074–2081.

[34] D. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.

[35] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-Scale Video Classification with Convolutional Neural Networks,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[36] C. Szegedy, A. Toshev, and D. Erhan, “Deep Neural Networks for Object Detection,” in *NIPS*, 2013, pp. 2553–2561.

[37] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[38] A. Krizhevsky and G. E. Hinton, “Using Very Deep Autoencoders for Content-Based Image Retrieval,” in *19th European Symposium on Artificial*

Neural Networks, 2011.

[39] L. Getoor, T. Scheffer, and International Machine Learning Society., "Minimal loss hashing for compact binary codes," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, p. 1216.

[40] Y. Weiss, A. Torralba, and R. Fergus, "Spectral Hashing," in *Advances in neural information processing systems*, 2009, pp. 1753–1760.

[41] H. Lai, Y. Pan, Y. Liu, and S. Yan, "Simultaneous Feature Learning and Hash Coding with Deep Neural Networks," Apr. 2015.

[42] H. Zhu, M. Long, J. Wang, and Y. Cao, "Deep Hashing Network for Efficient Similarity Retrieval *," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[43] H. Fujita *et al.*, "Computer-aided diagnosis: The emerging of three CAD systems induced by Japanese health care needs," *Comput. Methods Programs Biomed.*, 2008.

[44] S. C. Lee, E. T. Lee, Y. Wang, R. Klein, R. M. Kingsley, and A. Warn, "Computer Classification of Nonproliferative Diabetic Retinopathy," *Arch. Ophthalmol.*, vol. 123, no. 6, p. 759, Jun. 2005.

[45] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "DREAM: Diabetic Retinopathy Analysis Using Machine Learning," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 5, pp. 1717–1728, Sep. 2014.

[46] U. R. Acharya, C. M. Lim, E. Y. K. Ng, C. Chee, and T. Tamura, "Computer-based detection of diabetes retinopathy stages using digital fundus images," *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, vol. 223, no. 5, pp. 545–553, Jul. 2009.

[47] J. Nayak, P. S. Bhat, R. Acharya, C. M. Lim, and M. Kagathi, "Automated identification of diabetic retinopathy stages using digital fundus images.," *J. Med. Syst.*, vol. 32, no. 2, pp. 107–15, Apr. 2008.

[48] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija, "Automated screening system for diabetic retinopathy," in *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the*, vol. 2, pp. 915–920.

[49] N. Larsen, J. Godt, M. Grunkin, H. Lund-Andersen, and M. Larsen, "Automated Detection of Diabetic Retinopathy in a Fundus Photographic Screening Population," *Investig. Opthalmology Vis. Sci.*, vol. 44, no. 2, p. 767, Feb. 2003.

[50] A. Singalavanija, J. Supokavej, P. Bamroongsuk, C. Sinthanayothin, S.

Phoojaruenchanachai, and V. Kongbunkiat, “Feasibility Study on Computer-Aided Screening for Diabetic Retinopathy,” *Jpn. J. Ophthalmol.*, vol. 50, no. 4, pp. 361–366, Aug. 2006.

[51] P. Kahai, K. R. Namuduri, and H. Thompson, “A Decision Support Framework for Automated Screening of Diabetic Retinopathy,” *Int. J. Biomed. Imaging*, vol. 2006, pp. 1–8, Feb. 2006.

[52] S. Giraddi, J. Pujari, and S. Seeri, “Identifying Abnormalities in the Retinal Images using SVM Classifiers,” *Int. J. Comput. Appl.*, vol. 111, no. 6, pp. 5–8, Feb. 2015.

[53] W. L. Yun, U. Rajendra Acharya, Y. V. Venkatesh, C. Chee, L. C. Min, and E. Y. K. Ng, “Identification of different stages of diabetic retinopathy using retinal optical images,” *Inf. Sci. (Ny).*, vol. 178, no. 1, pp. 106–121, Jan. 2008.

[54] Q. Li, X.-M. Jin, Q. Gao, J. You, and P. Bhattacharya, “Screening Diabetic Retinopathy Through Color Retinal Images,” in *1st International Conference on Medical Biometrics*, 2008, pp. 176–183.

[55] S. S. A. Hassan, D. B. L. Bong, and M. Premsenthil, “Detection of Neovascularization in Diabetic Retinopathy,” *J. Digit. Imaging*, vol. 25, no. 3, pp. 437–444, Jun. 2012.

[56] S. S. Rahim, V. Palade, J. Shuttleworth, and C. Jayne, “Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing,” *Brain Informatics*, vol. 3, no. 4, pp. 249–267, Dec. 2016.

[57] M. Alban and T. Gilligan, “Automated Detection of Diabetic Retinopathy using Fluorescein Angiography Photographs,” 2016.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, pp. 1097--1105, 2012.

[59] C. Szegedy *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[60] D. T. Butterworth, S. Mukherjee, and M. Sharma, “Ensemble Learning for Detection of Diabetic Retinopathy,” in *NIPS*, 2016.

[61] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[62] V. Thambawita *et al.*, “The Medico-Task 2018: Disease Detection in the Gastrointestinal Tract using Global Features and Deep Learning,” Oct. 2018.

[63] C. E. Baudoin, B. J. Lay, and J. C. Klein, “Automatic detection of microaneurysms in diabetic fluorescein angiography.,” *Rev. Epidemiol. Sante Publique*, vol. 32, no. 3–4, pp. 254–61, 1984.

[64] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V Forrester, “An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus.,” *Comput. Biomed. Res.*, vol. 29, no. 4, pp. 284–302, Aug. 1996.

[65] Xiaohui Zhang and O. Chutatape, “A SVM approach for detection of hemorrhages in background diabetic retinopathy,” in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 4, pp. 2435–2440.

[66] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, “Optimal Wavelet Transform for the Detection of Microaneurysms in Retina Photographs,” *IEEE Trans. Med. Imaging*, vol. 27, no. 9, pp. 1230–1241, Sep. 2008.

[67] G. B. Kande, T. S. Savithri, P. V. Subbaiah, and M. R. M. Tagore, “Detection of red lesions in digital fundus images,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 558–561.

[68] A. J. Frame *et al.*, “A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms.,” *Comput. Biol. Med.*, vol. 28, no. 3, pp. 225–38, May 1998.

[69] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abramoff, “Automatic detection of red lesions in digital color fundus photographs,” *IEEE Trans. Med. Imaging*, vol. 24, no. 5, pp. 584–592, May 2005.

[70] Huan Wang, Wynne Hsu, Kheng Guan Goh, and Mong Li Lee, “An effective approach to detect lesions in color retinal images,” in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 2, pp. 181–186.

[71] R. Phillips, J. Forrester, and P. Sharp, “Automated detection and quantification of retinal exudates.,” *Graefes Arch. Clin. Exp. Ophthalmol.*, vol. 231, no. 2, pp. 90–4, Feb. 1993.

[72] T. Walter, J. Klein, P. Massin, and A. Erginay, “A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina,” *IEEE Trans. Med. Imaging*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002.

[73] C. I. Sánchez, M. García, A. Mayo, M. I. López, and R. Hornero, “Retinal

image analysis based on mixture models to detect hard exudates," *Med. Image Anal.*, vol. 13, no. 4, pp. 650–658, Aug. 2009.

[74] Lili Xu and Shuqian Luo, "Support vector machine based method for identifying hard exudates in retinal images," in *2009 IEEE Youth Conference on Information, Computing and Telecommunication*, 2009, pp. 138–141.

[75] R. Nekovei and Ying Sun, "Back-propagation network and its configuration for blood vessel detection in angiograms," *IEEE Trans. Neural Networks*, vol. 6, no. 1, pp. 64–72, 1995.

[76] C. Sinthanayothin, J. Boyce, H. Cook, and T. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Br. J. Ophthalmol.*, vol. 83, no. 8, p. 902, Aug. 1999.

[77] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-Based Vessel Segmentation in Color Images of the Retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[78] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[79] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[80] C. A. Lupascu, D. Tegolo, and E. Trucco, "FABC: Retinal Vessel Segmentation Using AdaBoost," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1267–1274, Sep. 2010.

[81] X. You, Q. Peng, Y. Yuan, Y. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *Pattern Recognit.*, vol. 44, no. 10–11, pp. 2314–2324, Oct. 2011.

[82] S. Roychowdhury, D. Koozekanani, and K. Parhi, "Blood Vessel Segmentation of Fundus Images by Major Vessel Extraction and Sub-Image Classification," *IEEE J. Biomed. Heal. Informatics*, pp. 1–1, 2014.

[83] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, Feb. 2015.

[84] S. A. Salem, N. M. Salem, and A. K. Nandi, "Segmentation of retinal blood vessels using a novel clustering algorithm (RACAL) with a partial supervision strategy," *Med. Biol. Eng. Comput.*, vol. 45, no. 3, pp. 261–273, Feb. 2007.

[85] G. B. Kande, P. V. Subbaiah, and T. S. Savithri, "Unsupervised Fuzzy Based

Vessel Segmentation In Pathological Digital Fundus Images," *J. Med. Syst.*, vol. 34, no. 5, pp. 849–858, Oct. 2010.

[86] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated Vessel Segmentation Using Infinite Perimeter Active Contour Model with Hybrid Region Information with Application to Retinal Images," *IEEE Trans. Med. Imaging*, vol. 34, no. 9, pp. 1797–1807, Sep. 2015.

[87] "Diabetic Retinopathy Detection | Kaggle." [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection. [Accessed: 03-Oct-2018].

[88] K. Pogorelov *et al.*, "KVASIR," in *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17*, 2017, pp. 164–169.

[89] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *ICLR*, 2015, pp. 448–456.

[90] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," Feb. 2015.

[91] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," Feb. 2016.

[92] V. Thambawita *et al.*, "The Medico-Task 2018: Disease Detection in the Gastrointestinal Tract using Global Features and Deep Learning," Oct. 2018.

[93] A. H. Shahin, A. Kamal, and M. A. Elattar, "Deep Ensemble Learning for Skin Lesion Classification from Dermoscopic Images," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, 2018, pp. 150–153.

[94] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 31–40, Jan. 2017.

[95] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, "Supervised hashing for image retrieval via image representation learning," in *Proceedings of the National Conference on Artificial Intelligence*, 2014.