QATAR UNIVERSITY

COLLEGE OF ENGINEERING

DIABETIC RETINOPATHY CLASSIFICATION USING DEEP LEARNING

BY

SARAH OBAID SHEIKH

A Thesis Submitted to

the College of Engineering

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Computing

June 2020

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Sarah Obaid Sheikh defended on 28/04/2020.

Dr Uvais Qidwai
Thesis/Dissertation Supervisor

Dr Abdelaziz Bouras
Committee Member

Approved:

Khalid Kamal Naji, Dean, College of Engineering

ABSTRACT

SHEIKH, SARAH, OBAID., Masters : June : [2020],

Masters of Science in Computing

Title: Diabetic Reinopathy Classification using Deep Learning

Supervisor of Thesis: Uvais, Qidwai.

With diabetes growing at an alarming rate, changes in the retina of diabetic patients causes a condition called diabetic retinopathy which eventually leads to blindness. Early detection of diabetic retinopathy is the best way to provide good timely treatment and thus prevent blindness. Many developed countries have put forward well-structured screening programs which screens every person diagnosed with diabetes at regular intervals. However, the cost of running these programs is increasing with ever increasing disease burden.

These screening programs require well trained opticians or ophthalmologist which are expensive especially in developing countries. A global shortage of health care professionals is putting a pressing need to develop fast and efficient screening methods. Using artificial intelligent screening tools will help process and generate a plan for the patients thus skipping the health care provider needed to just classify the disease and will lower the burden on health care professional's shortage significantly.

A plethora of research exists to classify severity of diabetic retinopathy using traditional and end to end methods. In this thesis, we first trained and compared the performance of lightweight architecture MobileNetV2 with other classifiers like DenseNet121 and VGG16 using the Retinal fundus APTOS 2019 Kaggle dataset. We experimented with different image reprocessing techniques and employed various hyperparameter tuning techniques, and found the lightweight architecture MobileNetV2 to give better results in terms of AUC score which defines the ability of the classifier to separate between the classes.

We then trained MobileNetV2 using handpicked custom dataset which was an amalgamation of 3 different publicly available datasets viz. the EyePacs Kaggle dataset, the APTOS 2019 Blindness detection dataset and the Messidor2 dataset. We enhanced the retinal features using bio-inspired retinal filters and tuned the hyper-parameters to achieve an accuracy of 91.68% and AUC score of 0.9 when tested on unseen data. The macro precision, recall, and f1-scores are 77.6%, 83.1%, and 80.1% respectively. Our results demonstrate that our computational efficient light weight model achieves promising results and can be deployed as a mobile application for clinical testing.

DEDICATION

...for my little daughter

...for my husband who helped me at every step

to make this a possibility

ACKNOWLEDGMENTS

I would like to thank my family for their utmost support and understanding. I am grateful to them for their moral support and understanding at all times. I thank my parents for their love, blessings and support.

A big word of thanks to my supervisor Dr Uvais Qidwai for his knowledge, key insights, timely support and advice, in every stage of this thesis. Without his help and guidance this thesis wouldn’t have been possible.

I would like to thank all my friends and extended family members for their continuous moral support and encouragement at all times. Additionally, I would like to thank the university for giving me the opportunity to conduct this research. I thank the university for giving me this valuable opportunity to learn and grow my skills with this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AMT	Amazon Mechanical Turk
ANN	Artificial Neural Networks
APTOS	Asia Pacific Tele-Ophthalmology Society
AUC	Area under curve
CNN	Convolution Neural Networks
DESP	Diabetic Eye Screening Programme
DM	Diabetes Mellitus
DNN	Deep Neural Network
DR	Diabetic retinopathy
IDF	International Diabetes Federation
MAC	Multiply-accumulate operations
NHS	National Health System
NPDR	Non-proliferative diabetic retinopathy
PDR	Proliferative diabetic retinopathy
POCT	Point-of-care technology
ROC	Receiver Operating Characteristics
SVM	Support Vector Machines
UK	United Kingdom
VEGF	Anti-vascular endothelial growth factor

CHAPTER 1: INTRODUCTION

Diabetes is a disease in which glucose metabolism is impaired leading to several complications. Diabetic retinopathy (DR) is one such condition which is characterized by damaged blood vessels at the back of the retina. According to the statistics provided by the International Diabetes Federation (IDF) nearly 463 million people suffer from diabetes globally and nearly one third have signs of DR [1]. Physicians have categorized DR into five different stages based on the severity viz. No DR, Mild, Moderate, Severe and Proliferative DR, characterized by symptoms shown in the retinal fundus photography images or retinal fundus images. Micro-aneurysms, Exudates, and Hemorrhages are considered indications of the presence of DR and are detected using these retinal fundus scans. In addition to that, the formation of abnormal blood vessels, called neovascularization is the characteristic for later stages of DR [2]. DR can be effectively managed in the early stages, however DR detected at later stages may cause irreversible loss of vision.

For an early diagnosis of DRs, ophthalmologists regularly advise diabetic patients to periodically undergo medical screening of their fundus. Nevertheless, retinopathies resulting from diabetes are usually undetected until considerable damage has occurred in a patient’s fundus (usually noticed by deteriorating / loss-of vision) [2]. The adequate identification / grading of DR stages can aid physicians in determining suitable intervention procedures. Diabetic patients worldwide need regular screening for the early detection which aids timely treatment to be administered quickly.

Several developed countries have already put forward a well-structured screening system to effectively manage the disease and provide quality timely treatment. The adequate identification/grading of DR aids physicians in identifying suitable intervention procedures, allowing timely treatment to be administered quickly.

The cost of running these screening programs is high and the lack of sufficient trained healthcare providers has forced the medical community to look for alternative ways to save time and resources in the grading of DR. Also, less developed nations are looking to reduce the cost of grading DR. Using automated computerized approach involving artificial intelligence is the current state-of- the-art approach to solve this issue [3][4]. Artificial Intelligence (AI) is the simulation of human intelligence with the help of complex algorithms by a software/machine where the algorithms learn to detect patterns in the data and then predict/detect patterns in unseen data.

With the rise in the users of smartphone-based technology, mobile based retinal imaging is the need of the hour providing cheap, faster and smarter point-of-care technology (POCT) for screening. Classifying the stages of DR using mobile technology screening would help generate a treatment plan for the patients, thus reducing the global disease burden and provide budget friendly, cost effective tool. Some recent studies have evaluated the performance of smartphone-based retinal imaging in the research community [5] [6] [7]. High-risk patients, will then be referred to the appropriate medical center for treatment. As the patients requiring treatment would be less than 5% of the screened patients, smartphone based automated screening tools will significantly be a stepping stone in effective management of DR.

Smartphone-based AI to detect Diabetic retinopathy severity has been studied previously in a couple of studies [5] [6]. Since mobile devices have less memory capacity and less computation efficiency, and most of the state-of-the art research work focusses on using architectures which are dense, heavy and computationally expensive, there is a need to look for alternatives. In this research we try to accomplish the task of developing a mobile based classification system for grading the severity of retinal fundus images. The system uses a much more efficient and light weight architecture of

MobileNetV2 which is known to give better performance than its predecessor MobileNetV1 without compromising on the key desired characteristics of the developed model viz. low latency and increased efficiency [8] [9]. We trained and tested the model on a custom-made dataset which is an amalgamation of 3 publicly available retinal fundus images datasets viz. the EyePACS, Kaggle APTOS dataset, and the Messidor2 dataset. We used bio-inspired retinal filters and fine-tuned the hyperparameters to achieve promising results.

The following sections are organized as follows. Section 1.1 gives an overview of Diabetic Retinopathy. The next few chapters are organized as follows: Chapter 2 highlights the related background necessary to understand the concepts presented in the thesis work and Chapter 3 presents the related work in the domain of DR grading using convolution neural network especially transfer learning giving a brief insight into MobileNetV2 architectures used in related studies. Chapter 4 presents the methodology used i.e. the dataset, pre-processing methods employed to the digital fundus images of our custom dataset, and the hyperparameter tuning of our model. Chapter 5 presents the experimental settings, the results, and a comparison of the results to previous related studies and finally Chapter 6 finally concludes the work with a list of contributions and future directions for research.

1.1 An Overview of Diabetic Retinopathy

Understanding Diabetic Retinopathy (DR), what causes it and the characteristic changes in the retina are imperative to building good artificial intelligent systems for screening purposes. Diabetic Retinopathy is a long-term condition seen in Diabetic patients. Diabetes Mellitus (DM), commonly known as Diabetes, is a chronic disease characterized by high blood sugar levels over a long-term period [10]. Uncontrolled diabetes may lead to many long-term conditions such as diabetic foot disease, diabetic kidney disease or diabetic retinopathy. DR is a complication which slowly damages the retina and is estimated to affect nearly 191 million people by 2030 [11]. High blood glucose levels damage or bock the blood vessels which nourish the retina and thus cause lasting harm. In response to this harm, the body then develops new blood vessels to maintain the nourishment which are weak and can easily lead to leakage and bleeding. [12]. This causes medical disorders like blurred vision and in some sever cases may lead to vision loss [13]

1.1.2 Diabetic Retinopathy Stages Classifications

The physicians have laid out the stages of DR by pointing out the symptoms. A sample retinal image with the manually annotated common DR lesions [14] is shown in Figure 1.

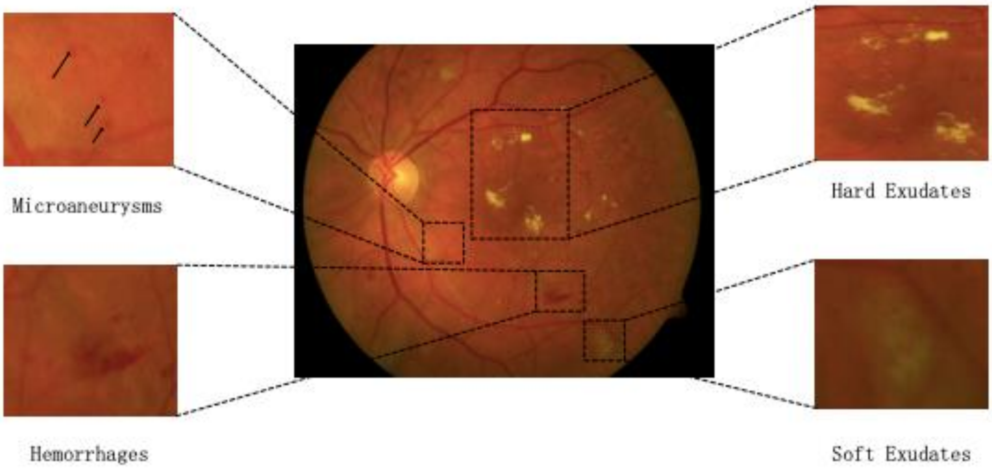

Figure 1. Annotated diabetic retinopathy image showing lesions.

The typical lesions of DR are briefly discussed below:

- Hard Exudates: Hard exudates are one of the retinal lesions that define DR. Hard exudates usually appear in the retinal images as tiny yellow-white spots with sharp edges and different sizes [15].
- Soft Exudates: Soft exudates, also known as cotton wool spots, appear as white patches with blurred and hazy edges [16]. Exudates, including soft exudates and hard exudates, are one of the most common early lesions of DR [17].
- Microaneurysms: Microaneurysms are the earliest clinically visible signs of non-proliferative diabetic retinopathy (NPDR) and are caused by dilatations of thin blood vessels. Microaneurysms usually appear as small red dots with sharp edges (20 to 200 microns) in clusters [18].
- Haemmorhges: As microaneurysms rupture in the deeper layers of the retina, they form hemorrhages.

In the United Kingdom (UK), the National Health System (NHS) is the publicly funded healthcare system and they have laid out the Diabetic Eye Screening Programme (DESP) According to them, for grading the retinal fundus images usually up to three human graders are required. These graders must meet the NHS DESP quality assurance standards and should be able to assess the images to determine a disease severity grade. They should then use these grades to produce a ‘final grade’ for each eye according to the highest level of severity observed. The stages of DR can be classified based on the presence of clinical features. The four widely used DR levels are defined in the table below [19] in Table 1.

Table 1. Stages of DR [19].

Levels	Description
R0	No DR
R1	Presence of Microaneurysms, retinal hemorrhages and any exudates
R2	Presence of Intra-retinal microvascular abnormalities, venous loop, venous beading
R3	Presence of new blood vessels (neovascularization), pre-retinal or vitreous hemorrhage

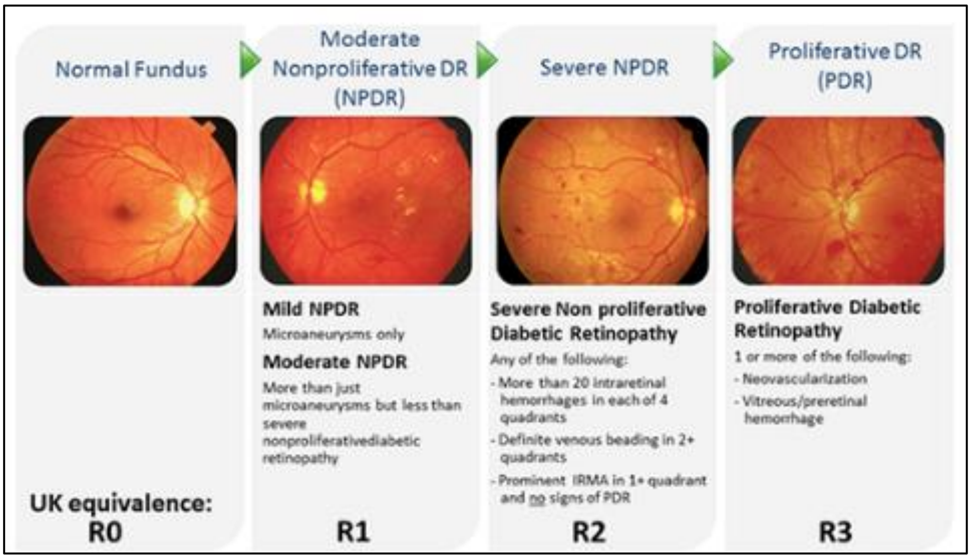
Normal Fundus
Moderate Nonproliferative DR (NPDR)
Severe NPDR
Proliferative DR (PDR)
Mild NPDR
Microaneurysms only
Moderate NPDR
More than just microaneurysms but less than severe nonproliferativediabetic retinopathy
Severe Non proliferative Diabetic Retinopathy
Any of the following:
- More than 20 intraretinal hemorrhages in each of 4 quadrants
- Definite venous beading in 2+ quadrants
- Prominent IRMA in 1+ quadrant and no signs of PDR
Proliferative Diabetic Retinopathy
1 or more of the following:
- Neovascularization
- Vitreous/preretinal hemorrhage
UK equivalence:
R0
R1
R2
R3

Figure 2. Stages of DR - UK equivalence for Diabetic retinopathy stages [20].

Figure 2 shows the various stages of Diabetic Retinopathy and the UK equivalence scale. The normal fundus equates to R0. The stage characterized by microaneurysms only is the Mild non-proliferative DR stage and the next stage where there are distortions in the blood vessels surrounding the retina causing swelling and degradation in the ability of blood transportation. Both mild and moderate non-proliferative DR equate to R1 in the UK equivalence scheme. Severe Non-proliferative DR are characterized to the blockage of more blood vessels and are depicted by more than 20 intra retinal hemorrhages in each quadrant along with venous beading. It equates to R2 in UK equivalence standards. Proliferative Diabetic retinopathy is characterized by vitreous hemorrhage and/or Neovascularization (new blood vessels formation) which are prone to bleed and leak more often. They equate to R3 stage in the UK equivalence standards [21] [22].

1.1.3 Fundus Photography

Color Fundus Retinal Photography which is widely used for screening the retina of patients uses a fundus camera to record the conditions of the eye focusing a 3D eye retina structure into a 2D plane. They record color images of the retina. The camera is basically a microscope with low power and has a camera attached to photograph the interior of the eye. The camera can clearly picturize the internal retinal features including the retinal vasculature, optic disc, macula, etc. The image is formed by the amount of light reflected by the interior of the eye. The imaging light is sent to the retina via the pupil to form the image by the reflected rays from the retina (Figure 3).

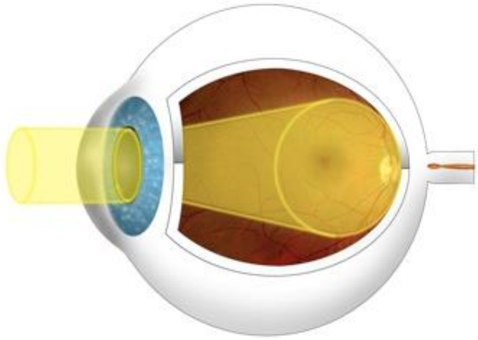

Figure 3. The scope of fundus imaging [32].

Lack of appropriate intervention causes DR to progress from mild to severe stages and thus it is important to recognize the stages when referral for treatment may be most beneficial. DR is broadly classified into two categories: non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Management of NPDR involves close monitoring by an ophthalmologist or optometrist and optimization of the patient’s glycemic control by an internist. Treatment at the proliferative stages involves laser photocoagulation, intravitreal injections of anti-vascular endothelial growth factor (VEGF) agents, or surgery to repair a retinal detachment.

1.2 Motivation

DR is a major cause of blindness and manual diagnosis of DR from retinal images is a time-consuming process and cumbersome process. Further it requires expertise of trained healthcare professionals which are often scarce, thus making it a challenging process. As a result, several studies have used deep CNN models to diagnose DR automatically. However, these methods employed very deep CNN models (e.g., ResNet-based method [23], Inception Net-based method [24], InceptionNet-ResNet-based model [25]) which require vast computational resources. Research on DR screening using computationally efficient CNN models has a great practical significance. This thesis is therefore focused on applying accurate and computationally efficient CNN model of MobileNetV2 for automatic DR classification that h can be deployed in mobile environments with scare resources like memory and computation power.

1.3 Research Questions

In this research we aim to research the following questions:

- Can lightweight architectures like MobileNetV2 effectively classify the severity of DR using retinal images and give comparable results to the commonly used heavy and dense architectures like DenseNet, VGG etc. using transfer learning?

- Can transfer learning model built using MobileNetV2 architecture give a good performance using a custom dataset of handpicked retinal fundus images employing various image processing and finetuning techniques?
- Does the developed model give satisfactory results when compared to other related work in the literature and does it hold a good scope of deploying in real-time mobile applications for clinal testing?

1.4 Research Aims and Objectives

The aim of this research is to classify the severity of Diabetic Retinopathy (DR) using retinal fundus images using light weight mobile networks.

The **Objectives** are as below:

1. To compare the performance of models built using light weight mobile networks like MobileNetV2 with the commonly used heavy and dense transfer learning algorithms on retinal fundus images.
2. To train and test a lightweight mobile-friendly classifier using MobileNetV2, using a custom-made dataset which can be deployed in mobile-based screening systems to classify the severity of DR.
3. To compare the developed model with existing work in the literature and assess the scope of real time deployment of the model in a mobile environment for testing in a clinical setting.

Summary

In this chapter we have laid forth the foundations of the study, the problem, the motivations behind the study, giving an overview of Diabetic Retinopathy and the various stages and classifications used in the medical community. We further put down the objectives of the study which has been described further in the following chapters.

CHAPTER 2: BACKGROUND

This chapter is a brief overview of various necessary background concepts that are needed to understand the contents of the thesis. The chapter highlights the concept of Artificial Intelligence (AI), the basic intuition of Machine Learning (ML), foundations of Deep Learning and its relationship to ML. Further, we explain the Convolution Neural Networks (CNNs) which are popularly used in the field of Computer Vision. We further understand the architectures of the CNNs which were used in this thesis. We also understand the intuition behind transfer learning and how we have leveraged it for our experiments. We also present the structure of the light weight, computationally efficient MobileNetV2 architecture.

2.1 Artificial Intelligence

Artificial Intelligence (AI) is a field of computer science where the machines demonstrate intelligence just like the natural intelligence displayed by humans and animals. AI research has seen a resurgence, following the phase of AI winter, after being founded in 1955, as an academic discipline. This was mostly due to the availability of large amounts of computational power, the enormous amounts of data and the theoretical understanding that was necessary to simulate the human intelligence.

The field of AI includes many subfields like knowledge representation, reasoning, learning, planning, natural language processing, perception and the ability to move and manipulate objects. Automated reasoning is an area which involves knowledge representation and reasoning to help computers acquire the ability of reasoning. Natural language Processing deals with the development of computers ability to process large amounts of natural language data and analyze it. Machine perception deals with developing the ability of the computer to absorb the sensory inputs and be able to see, feel and perceive the way humans do by gathering information from the world using machine vision, hearing and touch. Fields of computer vision, machine hearing and machine touch fall under this category. Hence, learning which is formally termed as Machine Learning (ML) is a subset of AI where the machine learns from data by identifying patterns in the data and forming inferences using several algorithms and statistical techniques.

2.2 Machine Learning

It is a subset of AI which allows the machine to learn from data implicitly without being actually programmed. The machine learns from the data by identifying patterns and making predictions. The popular machine learning algorithms include, the decision tree, k-nearest neighbors, naïve bayes, neural networks and kernel methods like support vector machines (SVMs). The task of these learning algorithms is to learn from data. In machine learning the algorithms essentially learns from an experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [70]. Examples of these tasks are classification, regression, anomaly detection etc. Examples of the performance measure P is accuracy, error rate etc. Examples of the learning experience are supervised, unsupervised and reinforcement. The learning experience basically denotes the type of learning that the algorithms are allowed to have during the learning process. In supervised learning experience the algorithm has a dataset which has labels while unsupervised learning experiences are based on data which don’t have labels. Reinforcement learning experience deals with the model learning from the environment through feedback loops. A dataset is a collection of examples which is in turn a collection of features. The entire collection of steps that make up the ML pipeline include data collection, data cleaning, splitting the dataset into training, validation and testing data, feature selection and engineering, model building, hyperparameter tuning, and model optimization.

2.3 Deep Learning

Deep Learning is a subset of Machine Learning which comprises of architectures that use Deep Neural Networks (DNN). Figure 4. depicts the relationship of Artificial Intelligence, Machine Learning and Deep Learning.

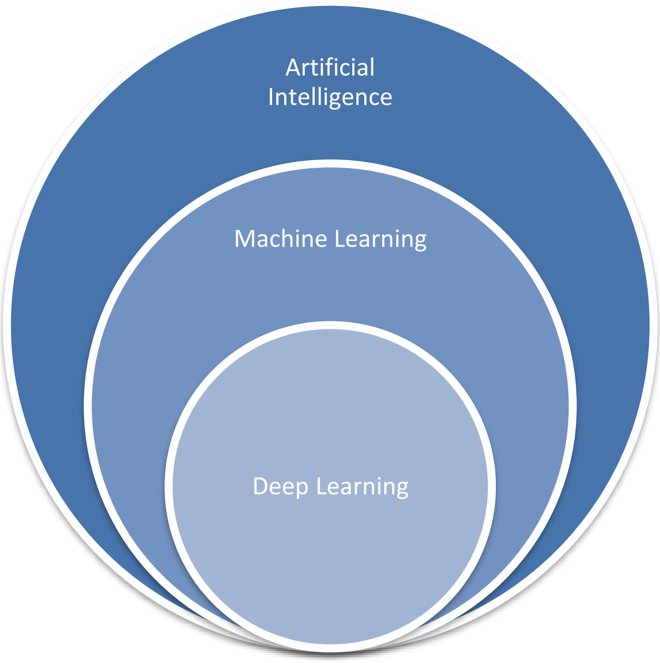

Figure 4. Relationship between AI, ML and DL.

Deep Neural Networks trace their origins back to the 1940s and 1950s and are inspired by the biological neural networks of the brain [26]. A typical neural network consists of 3 layers called input layer, the hidden layers and the output layers. The input layer takes the input from the data and passes it to the hidden layers which perform a non-linear combination of the information from the previous (input or hidden) layers. All these layers are composed of a neuron which is similar to structure of the biological neuron of the human brain. The building block of a neural network is a neuron which is depicted in the figure 5.

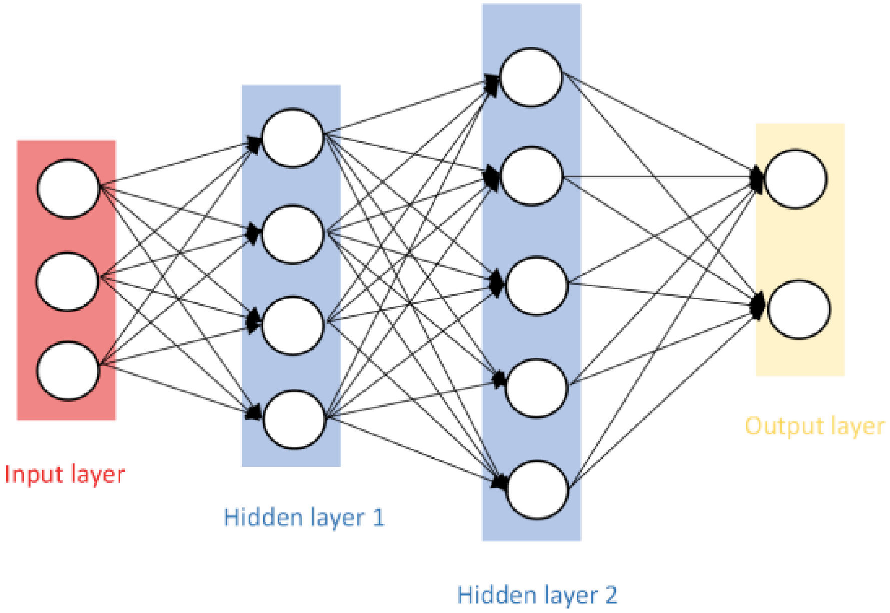

Figure 5. The basic structure of a neural network.

.

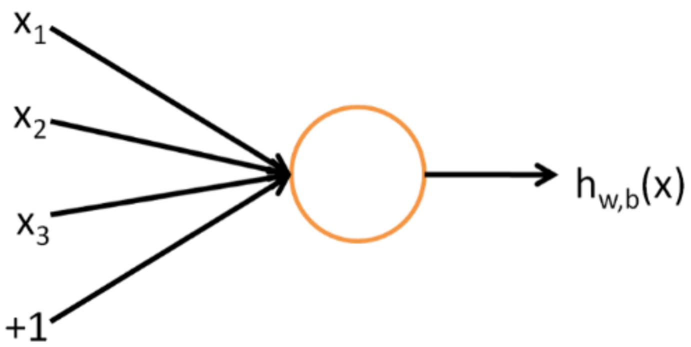

Figure 6. A neuron.

A neuron (Figure 6) is basically a placeholder for a mathematical function called as the activation function which activates this neuron by specific degrees based on the inputs and fires and output. Each layer of a neural network is a collection of neurons which take inputs and provide an output. The neural network in figure 4 has an input layer which takes the input and feeds it forward to the hidden layers which are activated based on the specific activation functions and which then produce an output based on their activations. The parameters are the weights, denoted by w (also a [nx1] vector), and a bias scalar, denoted by b. An activation, denoted by a, is a scalar defined by:

$$a = \sum_{i=1}^{n} x_i w_i + b$$

A point-wise non-linear function, denoted by σ (·), is then applied to generate the output, y = f(a) = σ($\sum_{i=1}^{n} xi\ wi + $ b). Some choices for the non-linearity function σ includes, sigmoid function, Tanh function, ReLU or the Leaky ReLU.

Other terminology used in the literature for this type of structure includes Artificial Neural Networks (ANN), Multi-layer Perceptron (MLP), and a fully-connected network. By convention, the number of layers is equal to the number of hidden layers plus the output layer (i.e., excludes the input layer).

Once a model architecture has been selected, the next step is to train the model as denoted by the steps below:

- Given the dataset of input x and output y, pick an appropriate cost function, C.
- Forward pass the input examples through the model to arrive at the predictions. Here the neurons are fired based on the activation's functions.
- Calculate the error using the cost function C to compare the predictions.
- Apply back-propagation to pass the error back through the model, adjusting the parameters to minimize the loss L.

- Once the gradients are established, use Stochastic Gradient Descent (SGD) to update the network weights.

In SGD, we start with some initial set of parameters, denoted by θ^0, and the updates are defined by $\theta^{k+1} \leftarrow \theta^{k} + \eta \Delta\theta$, where k is an iteration index, η is the learning rate, and the gradients $\Delta\theta = \partial L / \partial \theta$.

The backpropagation algorithm involves:

1. A forward pass where for each training example we compute the output for all the layers, $x_i = F_i (x_{i-1}, w_i)$;
2. A backwards pass where we compute cost derivatives iteratively from top to bottom $\partial C / \partial x_{i-1} = \partial C / \partial_{xi} \cdot \partial F_i (x_{i-1}, w_i) / \partial x_{i-1}$
3. Compute gradients and update the weights.

A constant learning rate η is typically not optimal. Techniques to optimize the learning rate are AdaGrad, RMSProp and ADAM. A momemtum term can be added to the weight update to encourage updates to follow the previous direction. This usually helps speed up convergence. The update then become $\theta\ k+1 \leftarrow \theta\ k + \alpha(\Delta\theta)\ k-1 - \eta\ \Delta\theta$, where α is typically around 0.9.

2.3.1 Convolutional Neural Networks

A popular class of Neural Networks are the Convolution Neural Networks (CNNs) which have shown exceptional performance in the field of image classification and recognition. The CNN network architecture has a number of hidden layers which help in the extraction of useful features and a fully connected layer at the end used for the classification. The CNN model is feed-forward, where input images are fed to convolution layer(s), non-linearities, pooling layers, and finally to feature maps. The CNN architecture, as shown in Figure 7, comprises of alternating convolution and pooling operation to reduce computation time and to understand the spatial details more effectively [27] [28] [29].

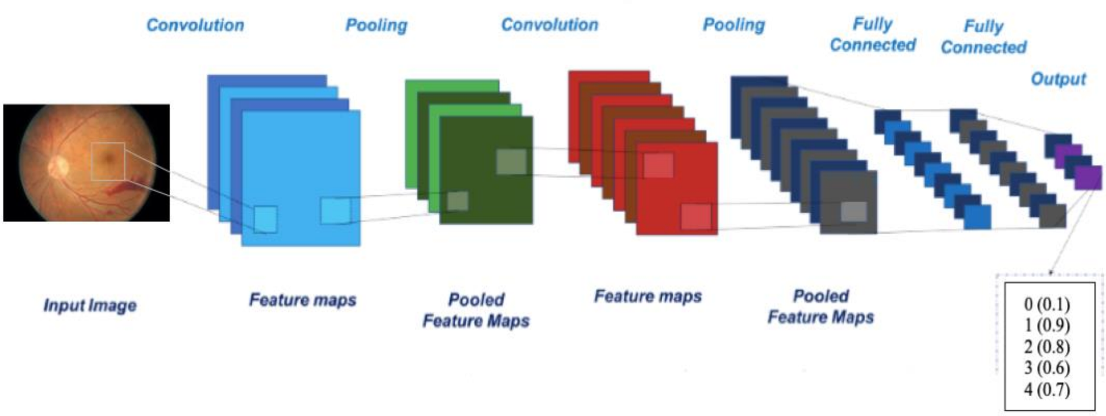

Figure 7. A convolution neural network.

The optimization process of the network is done by the backpropagation algorithm and the stochastic gradient descent algorithm. The convolutional layers apply a convolution operation with a fixed sized filter, e.g. 3x3 or 5x5, across the image. The filter is learned during training, and options such as the stride, padding, and dilation can be set. The pooling layer helps to reduce the spatial size of the representation, thus reducing the number of parameters and computation, and to prevent overfitting.

ImageNet [69] is a large-scale image database with over 14 million labeled images and over 20K classes. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual competition, where teams compete to build high performing networks. In 2012 major breakthrough was achieved via the AlexNet CNN of Krizhevsky et al. [30], achieving a top-1 and top-5 error rates of 39.7% and 18.9% which was considerably better than the previous state-of-the-art results.

2.3.1.1 DenseNet161 architecture

DenseNet (Densely Connected Convolutional Networks) is one of the latest neural networks for visual object recognition [35]. It’s considerably similar to ResNet

[34] but has some fundamental differences. It addresses the issue of “vanishing gradients”. Specifically, the DenseNet authors point out that with increasingly deep CNNs as information about the input or gradient passes through many layers, it can vanish and “wash out” by the time it reaches the end (or beginning) of the network. ResNet architecture proposed Residual connection, from previous layers to the current one. This is exactly how the DenseNet layers deal with this issue. DenseNet created an architecture to ensure maximum information flow between layers in the network, by connecting all layers (with matching feature-map sizes) directly with each other. In contrast to ResNet, DenseNet never combines features through summation before they are passed into a layer, rather it combines features by concatenating them. The “denseness” occurs because this network introduces L(L+1) 2 connections in an L-layer network, instead of just L in traditional architectures [35]. Figure 8 depicts a 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature maps as input.

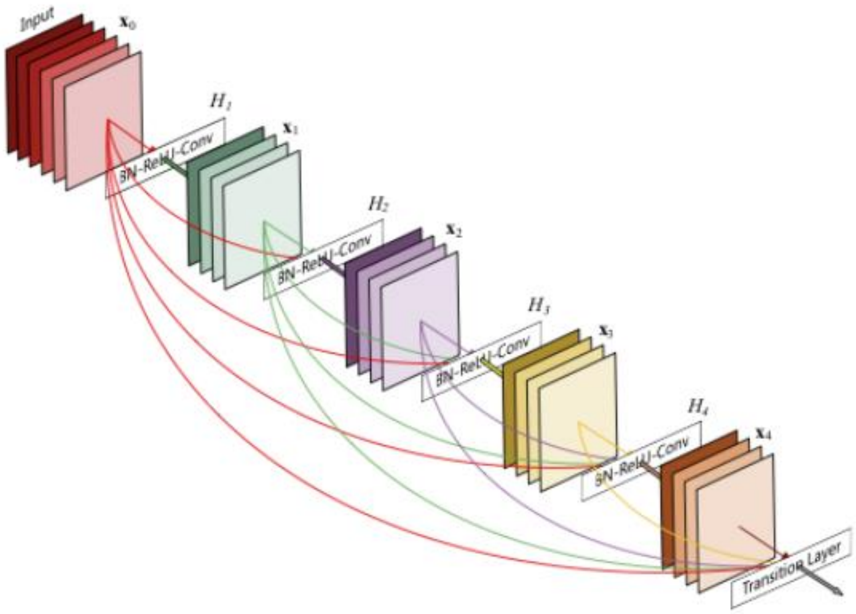

Figure 8. Schematic diagram of DenseNet architecture, reproduced from [35].

The biggest advantage of DenseNets is the improved flow of information and gradients throughout the network as each layer has direct access to the gradients from the loss function and the original input signal, leading to an implicit deep supervision.

2.3.1.2 VGG16 architecture

VGG16 is a deep CNN used for object recognition which was developed and trained by Oxford's renowned Visual Geometry Group (VGG) and gave 92.7% top-5 test accuracy in ImageNet in the Large-Scale Visual Recognition Challenge 2014 (ILSVRC2014) [33]. ImageNet is a *huge* database of images for academic researchers with 14 million hand-annotated images of what is in the picture. The main contribution of VGG is to show that classification/localization accuracy can be improved by increasing the depth of CNN in spite of using small receptive fields in the layers. (especially earlier layers). Neural networks prior to VGG used bigger receptive fields (7*7 and 11*11) as compared to 3*3 in VGG, but they were not as deep as VGG. There are few variants of VGG, the deepest one is with 19 weight layers. In this research, the model with 16 layers has been used which is popularly known as VGG-16.

VGG-16 is named so as it has 16 layers as shown in figure 9. It consists of Convolutional layers, Max Pooling layers, Activation layers, Fully connected layers arranged together as shown in figure 9 below which gives an overview of the layers in a VGG-16 network.

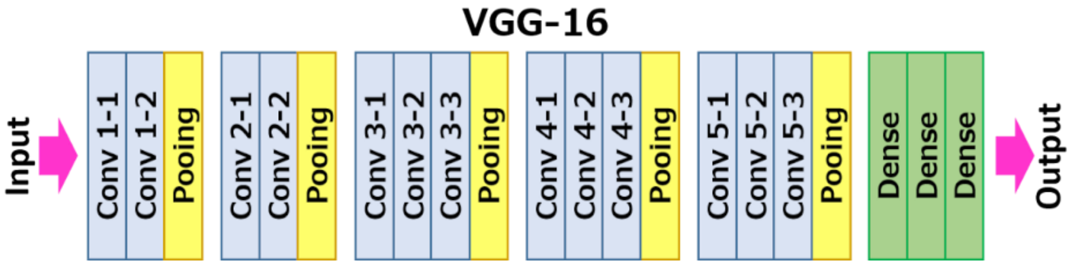

Figure 9. VGG-16 Layer Architecture.

There are 13 convolutional layers in the VGG-16 network, 5 Max Pooling layers to down sample the size of the network and 3 Dense layers which sums up to 21 layers but since only 16 weight layers, are there the network is named as VGG-16.

2.3.1.3 MobileNetV2 architecture

Transfer learning using Image net weights on MobileNetV2 has been taken into consideration for this thesis as this network is the state-of-the-art approach in the most mobile compatible networks. The biggest drawback of the commonly used CNNs such as DenseNets and VGG etc. is that they are computationally expensive. MobileNets are neural networks which are very efficient for mobile devices. MobileNetV2 is an enhancement of MobileNetV1 and is much more efficient and powerful than its predecessor. The original MobileNetV1 is a CNN which uses depthwise separable convolutions and basically splits the convolution layer into two sub tasks i.e. The input is filtered by a *depthwise convolution layer* and then a *pointwise convolution* of size 1x1 combines these filtered values to create new features. These two layers are together termed as *'depthwise separable convolution block'* which performs the tasks of a normal CNN but much faster, almost about 9 times as fast as other neural networks giving about the same accuracy [11]. The structure of the MobileNetV1 has these layers followed batch normalization and the ReLU6 activation function is used which is known to give better performance than the regular ReLU. At the end, there is a global average pooling layer, followed by a fully connected layer or a 1x1 convolution, and a softmax. The depth multiplier which is also known as the width multiplier is a hyperparameter which can be tuned and controls the number of channels in each layer [8].

Table 2. Depthwise Separable vs Full Convolution MobileNet reproduced from [8] on ImageNet dataset.

Network	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Conv MobileNet	71.7%	4866	29.3
MobileNet	70.6%	569	4.2

Table 2 clearly demonstrates the effect of depth wise separable convolutional layer. Conv MobileNet, which uses the standard convolutional layer gives an accuracy of 71.7% while MobileNet which is based on the depth wise separable convolutional layers gives an accuracy of 70.6%, when trained on the ImageNet dataset. Additionally, the Mult-Adds operations and the learnable parameters are reduced significantly due to the depth wise separable layer. Hence significant improvement in the efficiency is observed by reducing the parameters and Mult-Adds significantly while trading off the accuracy only by approximately 1%. This explains the contribution of depthwise separable layer which make up the MobileNet architecture, thus making them highly efficient for mobile devices.

According to Sandler et.al. [9] MobileNetV2 s similar to MobileNetV1, with differences in the architecture which contribute to its effectiveness. It also uses depthwise separable convolutions but the structure of the building block has residual skip connections and the expansion and projection layers, in addition. The building block consists of three convolution layers i.e. an expansion layer in which a 1x1 convolution layer expands the number of channels, a second layer which called the depthwise convolution layer and filters the inputs, followed by a third layer which is called the projection layer (and is a 1x1 pointwise convolution) which makes the number of channels smaller [9]. The expansion factor gives the factor by which the data gets expanded in the expansion layer and is a hyperparameter with a default value of 6. The network also has residual connections which helps with the flow of gradients through the network. Similar to MobileNetV1, every layer has batch normalization and the activation function is ReLU6 but the output of the projection layer in MobileNetV2 does not have an activation function applied to it. The complete MobileNetV2 architecture consists of 17 such building blocks followed by a regular 1×1 convolution, and a global average pooling layer, and a classification layer which is a 1x1x N layer where N is the number of classes in our problem [9].

Sandler et.al. mention that MobileNetV2 performs 300 million MACs which are the multiply-accumulate operations for an RGB image of 224x224, while MobileNetV1 performs 569 million MACs [12]. Additionally, V2 has nearly 20% less parameter counts than V1 has and this explains why V2's is more computationally efficient that V1 for mobile devices which have low memory access and less computation power [9].

Sandler et al. [9] termed the 2 building blocks as the inverted residual building block with linear bottlenecks which is fact the expansion layer and the reduction building block with linear bottlenecks which is the projection layer. Table 3 shows the architecture of the MobileNetV2 network and figure 10 shows the building blocks of the network. Similar to MobileNetV1 architecture, a hyperparameter α is introduced to trade off performance and computational cost. It is found that MobileNetV2 achieves an Accuracy of 0.720 compared to an Accuracy of 0.706 achieved by MobileNetV1 while using 19% fewer parameters and 48% fewer MAdds on ImageNet dataset [9].

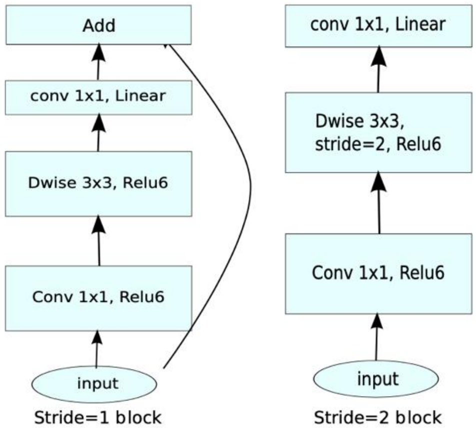

Figure 10. Inverted Residual (left) and Linear Bottleneck (right), Convolution blocks of MobileNetV2 reproduced from [9].

Table 3. MobileNetV2 vs MobileNetV1 reproduced from [9] on ImageNet dataset.

Network	ImageNet Accuracy	Million Mult-Adds	Million Parameters
MobileNetV1	70.6%	575	4.2
MobileNetV2	72.0%	300	3.4

2.4 Transfer Learning using Convolution Neural Networks

Transfer learning is a technique developed to address the issue of applying deep learning for domains with limited data [31]. The basic idea is to leverage the fundamental learning blocks built with a particular Deep Neural Network (DNN), such as ResNet, and to “re-train” the DNN for a particular domain of interest. Thus, one can utilize the strong “fixed feature extractor” capabilities of a DNN built on the millions of training examples from ImageNet to detect features common to all domains, such as object edges, and then re-train just the “top layer” for classification with the limited training data from a particular target domain.

Transfer Learning reduces the training time significantly. It also reduces the size of the dataset used for learning and this makes it very useful in domains which have little data. Training the final layers means that we transfer the knowledge that the classifiers learned from the its original training and can now apply it the smaller dataset from a different domain without the need for extensive training, thus saving training time and computational resources, and building models which give accurate results.

In this study, we have tried experimenting with the famous *DenseNet121, VGG16, and MobileNetV2.* We have tried to classify retinal images into respective categories using the original pretrained network which was trained on image net weights and we modified those networks by removing the top classification layers and by replacing them with pooling layers, dense layers and the final classification layer of

5 x1 nodes for the 5 output classes. This section of the overview of deep learning will briefly review 3 DNNs that were evaluated in this report. These networks are DenseNet121, VGG16, and the lightweight MobileNetV2.

Summary

In this chapter we have discussed in detail the basics behind the deep neural networks, the feed-forward process and the backpropagation algorithm which is used to update the weights learned from the input. We have also looked at Convolution Neural Networks and have discussed the various CNN architectures used in this thesis. The following chapter presents the related work done in the field of DR severity grading using artificial intelligence.

CHAPTER 3: RELATED WORK

In this chapter we present the related work highlighting key studies which have worked on the problem of diabetic retinopathy grading using computerized approach. We will study the various techniques and approaches used by the previous studies and how our study stands out from them.

3.1 Related work on DR screening

DR screening is based on retinal examinations by trained ophthalmologist or optometrist, or by digital retinal images which can detect early stages of DR. [36]-[44]. Manual screening for DR using digital imaging is known to give satisfactory results, however with the rise in DR prevalence coupled with the limited number of trained eye care professionals or certified readers, current DR screening programs that rely on expensive labor-intensive manual assessment may fail to address the rising screening demand, especially in the rural or non-developed nations which lack access to qualified screeners. To address such issues two alternative methods, exist in the literature i.e. crowdsourcing and automated computer aided retinal image analysis. The following sections gives some insights into the methods.

3.1.1 Crowdsourcing

Amazon's Mechanical Turk (AMT) is a crowdsourcing platform which lets us hire remote workers to do specific jobs which computers may be unable to do. One study [45] used Amazon Mechanical Turk to find an accuracy of 81.3% for distinguishing between normal and abnormal images while another study [46] found an accuracy of greater than 90% in distinguishing normal and severely abnormal images. Both the studies indicated a sensitivity of greater than 93% indicating that this screening modality has a good potential to address the increasing volume of images that will need to be screened with increasing diabetes prevalence rates. However, this technique has its obvious limitations i.e. there is no control over who is the “the crowd” in such platforms as it is counter to the spirit of crowdsourcing, but in reality, it is indeed necessary to know and have control over who is performing the analysis to validate the quality of image analysis [47].

3.1.2 Computer-aided retinal image Analysis

There is a plethora of studies which are focused on DR grading via automated mechanisms based on machine learning algorithms. The studies can be grouped into two very broad categories i.e. one which uses traditional approaches and the other which follow end to end methods.

3.1.2.1 DR Screening Using Traditional Methods

The traditional methods comprise of different phases which are distinct and separate. The phases usually are (1) feature extraction, (2) training and (3) classification/grading. The feature extraction phase depends on one or more of the DR symptoms. Usually the target is to extract the features of DR such as changes in the blood vessels which usually requires blood vessel segmentation and/or detection of lesion and classification which includes exudates, microaneurysms, hemorrhages, etc. For example, Nijalingappa and Sandeep [2] used traditional techniques to detect Microaneurysm in digital fundus images on a local dataset and achieved an accuracy of 87%. Microaneurysms are localized capillary dilations in circular shape and appear like small red dots often in clusters. Similarly, Seoud et al. [48] conducted experiments for detecting red lesions (i.e., Microaneurysm and Hemorrhages) in 2016 and achieved a sensitivity of 96%. They built a traditional model based on both Random Forests and Decision Trees algorithms.

In [49], Support Vector Machines (SVMs) have been used to detect and classify the exudates and predict the severity of DR. In the research conducted by Acharya et al, an automated system was developed for identifying the five classes using the features extracted. Some studies [50], [51] extracted exudates as features to classify the retinal images while some others used number of microaneurysms [52] while some other studies focused on retinal blood vessels [53] to identify diabetic retinopathy. Hence within the literature there are various trends and techniques to classify the severity of diabetic retinopathy and these techniques are heavily varied.

There is variation in the type of image processing techniques used as well. Rakshitha et al. [54] used several image transformation methods to enhance the images. She used contourlet transform, curvelet transform and wavelet transform and compared the performances of the classifiers. Another study [55] experimented with discriminant texture features to detect diabetic retinopathy. Most of these image processing techniques aim at identifying the lesions (exudates, hemorrhages, microaneurysms etc.) and the blood vessels to detect diabetic retinopathy.

Medical doctors and/or domain experts are required to put in their expert knowledge in the case of feature-based classification methods. This process of feature selection and extraction manually is often time consuming and varies vastly. Current trends in the literature point towards the DL based systems which use CNN'S. these models have been able to outperform the traditional approaches which use extraction of features. Hence end to end methods which primarily use the state-of-the-art deep learning models depend on pixel locations, correlations between the pixels, the color intensities and/or the combination of these features to create more sophisticated features. Over the last few years this step led to the automation of the feature extraction phase without any manual investigation or selection of the features. CNNs have demonstrated encouraging results in both the identification and grading of DR symptoms. Deep Learning algorithms are mainly used for classification based either from scratch or for transfer learning.

3.1.2.2 DR Screening Using Deep Learning & Transfer Learning technique

Transfer learning uses the technique of transferring the weights from a pre-trained model to adjust to our learning dataset. Maninis et al. [56] used VGG-16 for transfer learning for optic disk and blood vessel segmentation and achieved good results. Mohammadian et al. [57] used the Inception-V3 algorithm and fine-tuned the parameters and also used the Xception pre-trained models and achieved promising results. In addition to balancing the dataset they used data augmentation techniques and achieved an accuracy of 87.12% using Inception-v3 algorithm.

The study published in [58] demonstrated how transfer learning could solve the issue of insufficient training data. The authors in this paper use transfer learning for retinal vessel segmentation. Along similar lines studies like [59] and [60] used transfer learning and in [60] the best model was developed using VggNet-16 which achieved a 78.3% accuracy. Hence, we see a number of recent studies are targeting transfer learning to solve the inadequacy of data in the domain and also reducing the training time.

With the increase in smartphone usage, smartphone-based Point-Of-Care-Technology (POCT) is becoming popular. Rajalakshmi et. al. [5] and Xu et. al. [61] in two separate studies have used smartphone-based applications coupled with a fundus camera hardware for diabetic retinopathy grading. We lay forth the foundations of our study based on the related literature existing in the domain.

3.1.2.3 DR screening using MobileNetV2

MobileNetV2 is similar to MobileNetV1, with differences in the architecture which contribute to its effectiveness. It also uses depth wise separable convolutions, like its predecessor but the structure of the building block has residual connections and the expansion and projection layers, in addition. MobileNetV1 are a neural network architecture which are very efficient for mobile devices. MobileNetV2 is an enhancement of MobileNetV1 and is extremely powerful for mobile devices which are targeting models which give low latency and high computational efficiency.

In [62] the authors experimented with MobileNet and MobileNetV2 among other transfer learning approaches on the problem of diabetic retinopathy grading. They used MobileNetV2 which is an improvement of the MobileNetV1 architecture in terms of speed and computational efficiency and also gave good performance. They achieved an accuracy of 78.1% using MobileNetV2 and an accuracy of 58.3% using MobileNetV1.

Further, we see that Gao et.al. in [63] divided the problem into a 2-class problem of DR and RDR which is referable DR. According to them, grade 0 and 1 of the Messidor dataset form the DR category while grade 2 and 3 are considered the referable DR category which needs urgent attention. Their MobileNetV2 model has achieved an accuracy of 90.8% for class DR and 92.3% for class RDR [63]. Due to a difference in the annotation scales used in the datasets Gao et.al. [63] transformed the problem into a binary classification task of referable vs non- referable.

To the best of our knowledge no other study has first validated the performance of pre-trained DNNs on retinal fundus image and then built a custom dataset of handpicked images to build a light weight mobile friendly model using MobileNetV2 to classify retinal images for DR retinal screening. In this study, we used a custom-made dataset of handpicked retinal fundus images from 3 different publicly available datasets and then trained and tested our classifier using MobileNetV2 on these images after using several preprocessing techniques. In this study we present the methodology and the results of applying transfer learning algorithms to retinal fundus datasets and then using MobileNetV2 on a custom dataset to fine tune the model and build models specifically for the needs of mobile devices which has the potential to be used for screening DR using mobile devices.

Summary

In this chapter we have analyzed the various studies in the literature which have done similar work and laid forth the methodology used in our study in the next chapter. We have tried to cover the broad areas where automated diabetic retinopathy severity classification was being carried out. The nest chapter will explain the methodology followed in this research.

CHAPTER 4: METHODOLOGY

In this chapter we lay forth the foundations of the research by explaining the dataset used in the study, the architecture of our model, the preprocessing techniques and the hyperparameter tuning that was done to achieve the optimal performance. Figure 11 and Figure 12 explain the methodology diagrammatically.

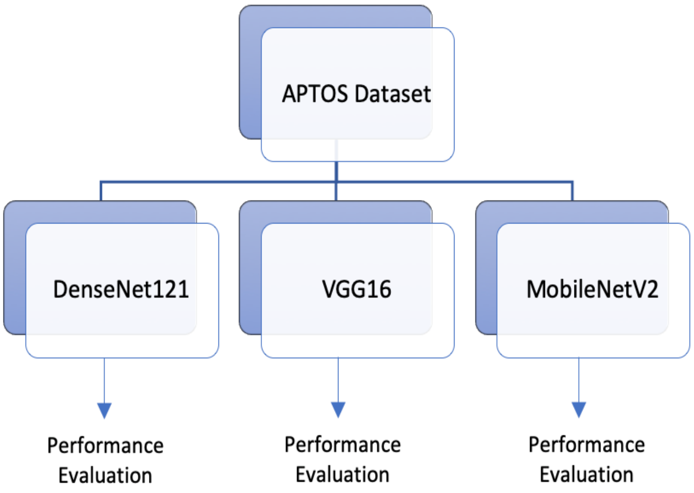

Figure 11. Diagrammatic overview of the methodology – Part 1.

4.1 Datasets

For the first part of our study we trained our models using pretrained networks on a publicly available Diabetic retinopathy dataset from Kaggle i.e. the APTOS 2019 Blindness Detection dataset from Kaggle [64]. The dataset comprised of retinal fundus images which were split first into train and validation where 10% of the data was randomly chosen for validation. The train data was split into training and testing data using train_test_split from sklearn with the test_size being 15% of the remaining data. The dataset has retinal fundus images which are rated by a clinician for the presence of diabetic retinopathy in each image on a scale of 0 to 4. The five classes are as follows:

0 - Negative or No DR: Patient has no disease.

1 - Mild DR (Stage 1): Patient has mild level of disease.

2 - Moderate DR (Stage 2): Patient has moderate level of disease.

3 - Severe DR (Stage 3): Patient has severe level of disease, the most part of the retina is damaged, can lead to complete blindness.

4 - Proliferative DR (Stage 4): Patient has proliferative level of disease. The patient's eye is damaged to an extent where treatment is elusive, about 80 percent of blindness exists.

Further in order to build an accurate, computationally efficient and robust system, we prepared a custom dataset using 3 different publicly available datasets to train and test our model using the light-weight MobileNetV2 architecture. The final dataset that we used in our experiments is an amalgamation of retinal fundus images from the three datasets mentioned below:

1. **EyePacs dataset**: EyePacs has provided a large dataset of retinal images from diabetic screening programmes [65]. The dataset is sponsored by the California Health care Foundation and was used in the Kaggle DR Detection Challenge. This dataset comprises a large set of high-resolution retina images taken under a variety of imaging conditions. It has a total of 88,702 JPEG images. The competition sponsors pre-allocated 31,615 (35.6%) of these images for training, 3,511 (4.0%) for validation, and 53,576 (60.4%) for testing. Class labels were provided by the competition for the training, validation, and test images. The dataset is labelled with a severity grade of DR from 0 to 4. All images were rated by a certified reader according to a standard diabetic retinopathy grading scale.
2. **APTOS 2019 dataset**: APTOS stands for Asia Pacific Tele-Ophthalmology Society [64]. It is a subset of the EyePacs dataset obtained after performing some

preliminary operations and available in different image formats. It consists of 3000+ images.

3. **Messidor 2 dataset**: The Messidor 2 is an extension of the original Messidor dataset for diabetic retinopathy [66]. It contains 1500+ retinal fundus images which are labelled with 4 classes from 0 to 4.

We performed data normalization on images from the 3 datasets in a way that each of the 5 classes has relatively the same number of images. This was done to ensure that the dataset is not biased towards any one particular class. We used partial datasets from EyePacs and APTOS 2019 dataset and used 50% of the Messidor2 dataset to create this custom dataset for training our classifier in our experiments and we only choose images which were of a good quality and would contribute good features to the classifiers. The other 50% of Messidor2 dataset was kept aside and was used for testing the performance of our classifier. The testing dataset has 961 images belonging to the 5 classes, explained above.

4.2 Image Pre-processing and Augmentations

For testing the performance of the 3 classifiers we experimented with a few preprocessing techniques. For each image we have applied the following transformations i.e. horizontally flip 50% of all images, vertically flip 20% of all images, scaled images to 80-120% of their size, individually per axis, translated by -20 to +20 percent (per axis), rotated by -45 to +45 degrees, shear by -16 to +16 degrees, used nearest neighbor or bilinear interpolation (fast). We further converted images into their super pixel representation and blurred the images with a sigma between 0 and 3.0. We add further processing by blurring the image using local means and medians with kernel sizes between 2 and 7. We added gaussian noise to the images and randomly remove up to 10% of the pixels. We experimented with using inverted color channels and changed the brightness of the images (by -10 to 10 of original value) and also with changing the hue and saturation. We have also removed any ovals from the dataset and kept only the circular ones so that we can remove any bias and train the algorithms with images where every image contributes equally to the learning process. We used these transformations on the images based on our literature survey that we conducted and chose the techniques which gave promising results.

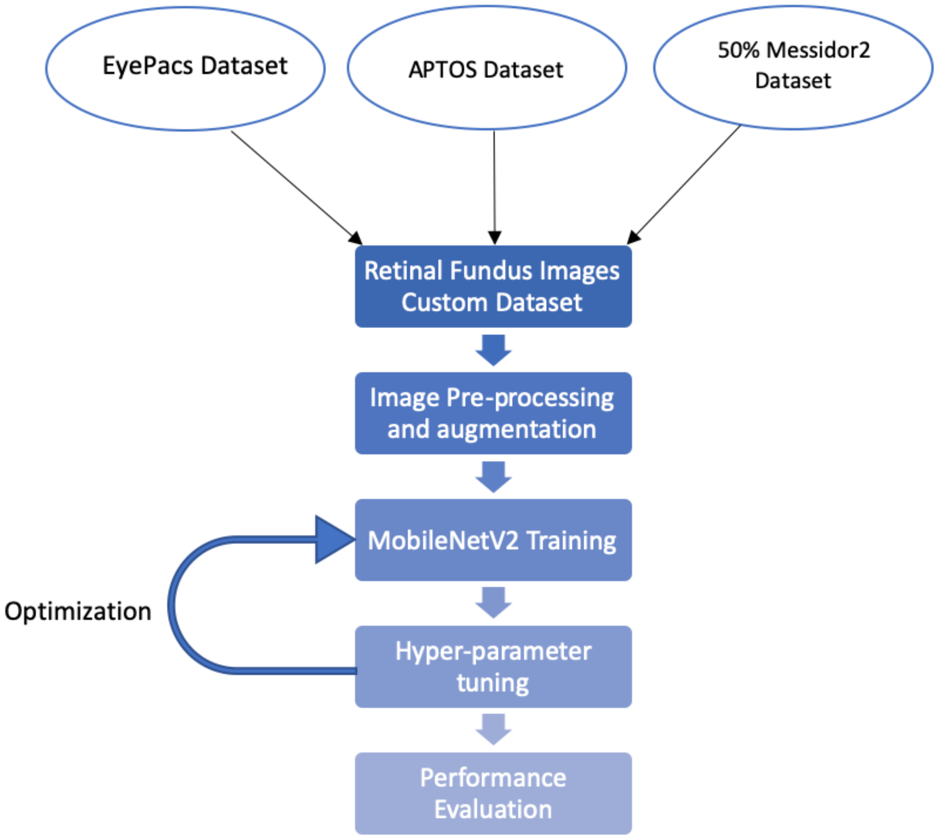

Figure 12. Diagrammatic overview of the methodology – Part 2.

Further when we worked on building a lightweight mobile friendly MobileNetV2 model which gave promising results. We employed certain image preprocessing techniques to improve the quality of the images. We changed the luminous intensity of the image which altered the brightness and made the details of the image more visible. We considered changing alpha, beta and gamma channels which are the important channels which control the amount of the light. We checked for images which were too dark or over-bright and could be augmented by altering the alpha, beta and gamma channels and controlled the light by using *alpha = 2.5, beta = 40 and gamma = 1.44.* These values were obtained using trial and error method on several dim images using OpenCV's ConvertscaleAbs function [67].

In order to transform the image to perform texture analysis with an enhanced signal to noise ratio and provide better luminance ranges to the input images we used OpenCV's Bioinspired_retina function on some of the retrieved dim images and then augmented some particular parameters [68]. This function helps in performing texture analysis and improves the signal to noise ratio. We particularly altered the default values of the following parameters in the function using retina configurations:

- photoreceptorsLocalAdaptationSensitivity: 0.69
- photoreceptorsTemporalConstant: 8.9999997615814209e-01
- photoreceptorsSpatialConstant: 5.2999997138977051e-01
- horizontalCellsGain: 0.75
- hcellsSpatialConstant: 7.0
- ganglionCellsSensitivity: 0.75

These preprocessing steps helped us in bringing up enriched features from the images, following which, we applied data augmentation using horizontal and vertical flips and rotation using -20 to +20-degree rotation. We were able to bring 3400+ images in each class with a total of 17,121 images in the final dataset. 80 % of the images were used for training, and 20 % were used for validation. We tested the performance of the dataset on unseen data, which was 50% of the messidor2 dataset kept aside specifically for this purpose.

4.3 Model Architecture and Training

We compared 2-different models for our experimentation which were heavily used in the literature i.e. *DenseNet121, VGG16 and a* light weight *MobileNetV2*. We used pretrained weights on ImageNet dataset. The last layer was modified and we introduced Global Average Pooling (GAP layers). This was done to reduce the total number of parameters and thus reduce overfitting. The GAP layers are used to reduce the spatial dimensions of a 3-d tensor just like the Max pooling layers. And then Dropout with 0.5 as regularizer. During training we freeze the first layers and trained the last 4 layers including classifier for transfer learning. Feature extractor was trained on ImageNet. These experiments were conducted with an intuition to compare the performance of the lightweight architecture of MobilenetV2 with the other heavy and dense architectures and record their results. Architecture of one of the transfer learning algorithms - the DenseNet121 is depicted in Figure 13 below.

Since the aim of the study was to build a light weight mobile friendly architecture using MobileNetV2, transfer learning using Image net weights on MobileNetV2 has be taken into consideration in this work as this is the state-of-the-art approach in the most mobile compatible networks. MobileNetsV1 are a neural network architecture which are very efficient for mobile devices. MobileNetV2 is an enhancement of MobileNetV1 and is much more efficient and powerful than its predecessor. According to Sandler et.al. [9] MobileNetV2 is similar to MobileNetV1, with differences in the architecture which contribute to its effectiveness which have been explained in Chapter 2 (section 2.2.3). The structure of our final model tested on the custom dataset is given in Table 4.

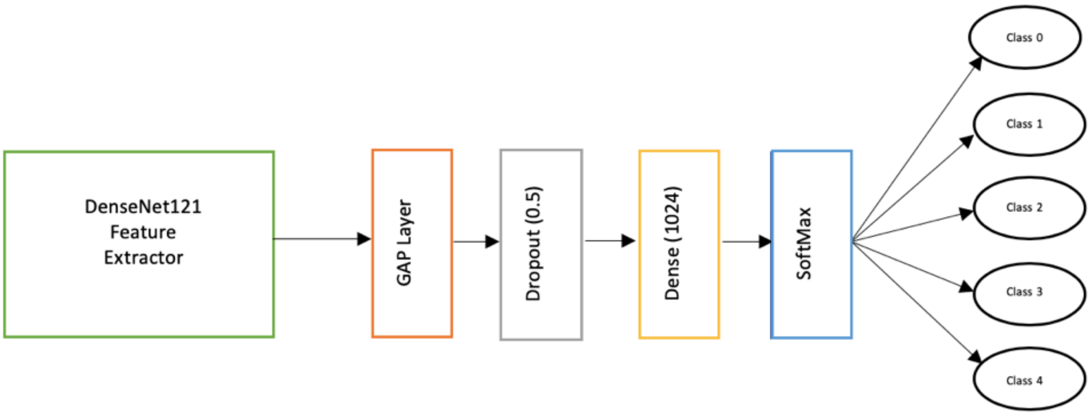

Figure 13. Schematic representation of Classification with DenseNet121.

Table 4. The structure of MobileNetV2 architecture used for DR Classification.

Input Dimension	Operator	t	c	n	S
224^2 x 3	Conv2D	-	48	1	2
112^2 x 48	Residual Module	1	24	1	1
112^2 x 24	Residual Module	6	32	2	2
56^2 x 32	Residual Module	6	48	3	2
28^2 x 48	Residual Module	6	88	4	2
14^2 x 88	Residual Module	6	136	3	1
14^2 x 136	Residual Module	6	224	3	2
7^2 x 224	Residual Module	6	448	1	1
7^2 x 448	Conv2D 1x1	-	1792	1	1
7^2 x 1792	AvgPool 7x7	-	-	1	-
1^2 x 1024	Dense	1	1024	2	1
1^2 x 512	Dense	1	512	1	1
1^2 x 5	Dense - Final	1	5	1	-

4.4 Training and Hyperparameter tuning

We trained our final model on the custom dataset and validated its performance using the validation set. The 80% of our custom dataset was used as training set and the other 20% was used for validation.

We chose a learning rate of 0.00015 where the network converged faster. We used Adam with AMSGrad as our optimizer after trying several other optimizers and used a batch size of 32 for CPU training. We used a Dropout of 0.1. We ran a grid search on number of unfreezing layers and found that when we unfreeze half of the MobileNetV2 layers the model accuracy is more and the learning is very fast and the model converges very fast.

We freezed the first 80 layers and unfreezed the rest of the layers for better training. We tested unfreezing of different layers and found that unfreezing from the 80th layer learned better features from the dataset.

4.5 Performance Evaluation

We used a number of metrics to measure the performance of our model. Evaluation metrics are tied to machine learning tasks. There are different metrics for different tasks. This experimentation is based on classification, so the focus will be on metrics that can evaluate a classification task.

A confusion matrix is an N x N matrix, where N is the number of classes being predicted. For the problem in hand, there are 3 classes, and a 3 x 3 matrix. The following are the values represented in confusion matrix

-True Positive (TP) - An outcome where the model correctly predicts the positive class.

-True Negative (TN) - An outcome where the model correctly predicts the negative class

-False Positive (FP) - An outcome where the model incorrectly predicts the positive class.

-False Negative (FN) - An outcome where the model incorrectly predicts the negative class.

4. Accuracy: Accuracy in a classification problem is the number of correct predictions made by the model over all kind's of predictions made.

   Accuracy = (TP + TN) / (TP + FP + FN + TN)

5. Precision: Precision calculates the rate of actual positives out of those predicted positive. It is given by the formula:

   Precision = TP / (TP + FP)

6. Recall/Sensitivity: Recall measures the rate of actual positives over all predicted values that are actually positive. It is given by the formula:

   Recall = TP / TP + FN)

7. Specificity: Specificity measures the True negatives over the sum of true negatives and false positives.

   Specificity = (TN) / (TN + FP)

8. f1-score: f1-score is the harmonic mean of the precision and recall and conveys the balance between the precision and the recall. It is calculated by the formula:

   f1-score = 2* [(Precision*Recall)/ (Precision+ Recall)]

9. Kappa Score: The Kappa statistic (or score) compares an observed accuracy with an expected accuracy (random chance). Observed Accuracy is the number of correctly classified instances throughout the entire confusion matrix. Expected accuracy is defined as the accuracy that any random classifier would be expected to achieve based on the confusion matrix and hence the Kappa score is given by

   Kappa = (observed accuracy - expected accuracy) / (1 - expected accuracy)

10. AUC ROC: which stands for Receiver Operating Characteristics curve is plotted with recall along the y-axis and the false positive rate (which is given by 1-Specificity) along the x-axis. Area under the ROC Curve also called AUC gives the degree of separability of classes which means it tells how much the model is capable of distinguishing between the classes. Higher value of AUC denotes a better model.
11. Macro-Precision, Macro-Recall and Macro-F1 measure: A macro-average will compute the metric independently for each class and then take the average (hence treating all classes equally). We find the precision and recall for each class separately and then find the average of all of the values, to get the Macro-Precision and Macro-Recall. Macro – F1 score is the harmonic mean of macro precision and macro recall.

Summary

In this chapter we try to explain the methodology adopted for conducting this research. We explain the dataset used, the preprocessing techniques and the model architecture of our network. We then highlight the hyperparameter tuning and the performance metrics used for our research.

CHAPTER 5: RESULTS AND DISCUSSIONS

In this chapter, we present the results of our experiments to get an understanding of the performance of our model on unseen data. We then compare them to previous related work and give our insights as discussions.

5.1 Experimental Setting

We first trained and validated MobileNetV2, DenseNet121 and VGG16 models using the APTOS dataset. We evaluated the models using accuracy, AUC and Kappa score. We found that MobleNetV2, even though is a lightweight network gives the best performance in terms of distinguishing between the classes (Figure 15, 16 and 17). Hence, we decided to conduct further experiments using MobileNetV2, since this is a multiclass classification problem.

We trained our MobileNetV2 network using 12 GB of RAM, 10 GB swap memory having i5 processor**.** We trained our MobileNetV2 model, applied the preprocessing and hyperparameter tuning settings mentioned earlier and achieved an accuracy of 91.68% on the test set mainly due to the good quality images that we fed into the network. The accuracy rapidly improved over the 20 epochs. (Figure 14)

We used a width multiplier of 1.4 and an input size of 224 x 224 x 3 for each input image. Table 4 shows the architecture of our network.

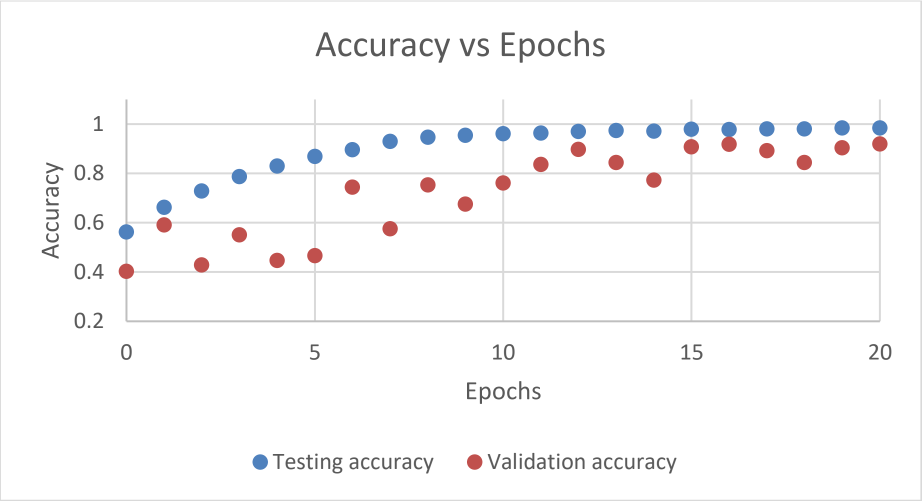

Figure 14. Validation and Testing accuracy improves over 20 epoch.

5.2 Results and Discussions

The results from analyzing the performance of the 3 classifiers i.e. *DenseNet121, VGG16, and a light weight MobileNetV2* are depicted in Table 5. Figure 15, 16 and 17 present the performance diagrammatically. We have analyzed the performance of these classifiers based on accuracies, kappa score and AUC score. These experiments were conducted on the APTOS dataset.

Table 5. Performance of DenseNet121, VGG16 and MobileNetV2 on APTOS Dataset

Models/Metrics	Accuracy	Kappa Score	AUC
DenseNet121	90.5	90.25	0.914
VGG16	79.3	79.01	0.7978
MobileNetV2	88.73	83.33	0.9289

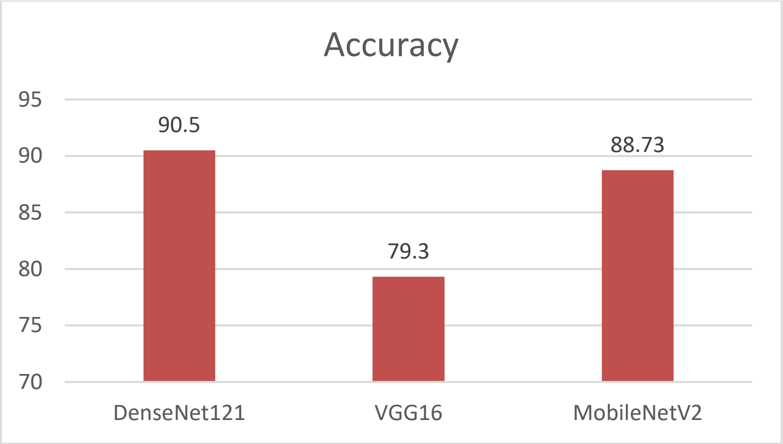

Figure 15. Graphical representation of accuracy on the APTOS Dataset.

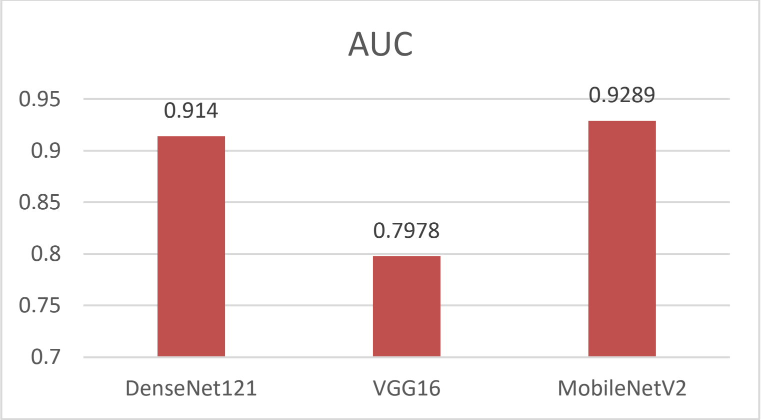

Figure 16. Graphical representation of AUC – Score on the APTOS Dataset.

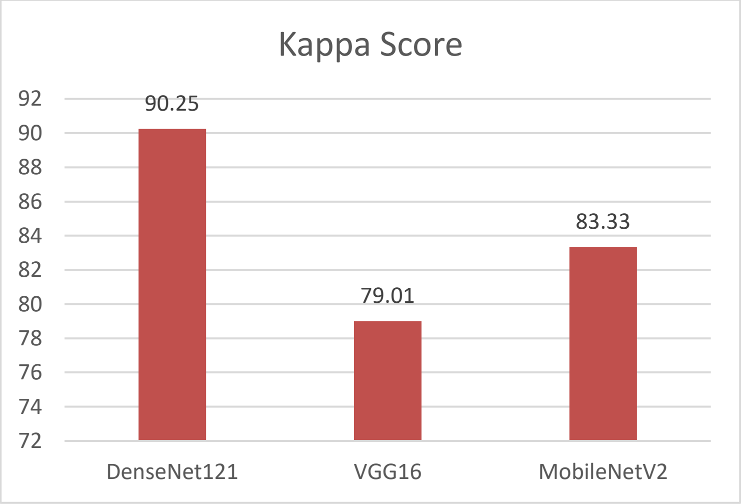

Figure 17. Graphical representation of Kappa – Score on the APTOS Dataset

Following these experiments, as depicted by table 5 we clearly see that although MobileNetV2 gives less accuracy that DenseNet121 and much more than VGG16, both of which are heavy and dense architecture; and MobileNetV2 gives the best results in terms of AUC score which is a metric to compare the degree of separability between the classes of the model, we therefore conclude that light weight architecture MobileNetV2 gives comparable results to dense architecture of DenseNet121. Hence, we conclude that even though MobileNetV2 is a lightweight architecture, it still gives better results that the other heavy architecture. Since we aim to build mobile efficient models we choose to build our refined model by training MobielNetv2 on the custom dataset.

Hence, we tested the performance of lightweight MobileNetV2 architecture on our custom dataset which was created using handpicked images and roughly same amount of good quality images where kept for each class (i.e. approximately 3400 + images in each class) to remove any kind of biases. Using the various preprocessing techniques mentioned in section 4.2. We tested the performance of our classifier on unseen data which is 50% of Messidor2 dataset kept aside solely for testing purpose.

Table 6 presents the confusion matrix obtained from testing the model on an unseen dataset which is the test set for the MobleNetV2 architecture. Table 7 presents the precision, recall and f1-score per class and Table 8 presents the key performance measures of our MobileNetV2 classifier. The results demonstrate the model to have good generalization abilities with 91.68% testing accuracy and is compared with other related work in the literature.

Table 6. Confusion matrix - Test set Messidor2 dataset – MobileNetV2.

		Predicted labels				
		0	1	2	3	4
Actual labels	0	576	11	14	3	2
	1	9	130	9	1	1
	2	3	5	143	7	3
	3	0	1	3	19	3
	4	1	1	1	2	13

Table 7. Performance metrics per class - MobileNetV2.

	0	1	2	3	4
Precision	0.9779	0.8784	0.8412	0.5938	0.5909
Recall	0.9505	0.8667	0.8882	0.7308	0.7222
f1-score	0.9640	0.8725	0.8640	0.6552	0.6500

Table 8. Performance measures for MobileNetV2.

Performance measure	Value
Macro Precision	77.64%
Macro Recall	83.17%
Macro f1-score	80.11%
Training accuracy	98.45%
Validation accuracy	91.90%
Testing accuracy	**91.68%**

On comparing with other related previous work, we see that our model has been able to achieve promising results. The previous related work [63] achieved 90.8% accuracy using the same architecture of MobileNetV2 but on a different dataset. We see that Gao et.al. in [63] divided the problem into a 2-class problem of DR and RDR (referable DR). According to them, grade 0 and 1 of the Messidor dataset form the DR category while grade 2 and 3 are considered the referable DR category which needs urgent attention. Their MobileNetV2 model has achieved and accuracy of 90.8% for class DR and 92.3% for class RDR. Our work on the contrary has studied the problem as a 5-class problem with classes 0, 1, 2, 3 and 4 as explained in section 3.1 and we have achieved an average accuracy of 91.68% and an AUC of 0.9 using our custom dataset for training and testing it on 50% of the Messidor2 dataset which is our unseen test set.

Another paper [25] proposed a network called Zoom-in-Net which does two tasks simultaneously i.e. it mimics the zooming in of the clinician to examine retinal images by developing attention maps and highlights suspicious regions and make predictions based on these suspicious regions and also the whole image. Due to a difference in the annotation scales used in the datasets used for the study they have also used similar technique of Gao et.al. [63] and transformed the problem into a binary classification task of referable vs non-referable and achieved an accuracy of 91.1% on the Messidor dataset and an accuracy of 90.5% on the EyePacs dataset. Thus, we see that the performance of our model is comparable to the state-of-the-art and can be applied in clinical settings using mobile applications for testing purposes.

Since the papers in the literature point out that MobileNetV2 architecture (as explained in chapter 2 and chapter 3) is much more lightweight that the other heavy and dense architectures, like DenseNet121, even though, we have not conducted any experiments to validate these findings of the efficiency we conclude that our model is much more lightweight and efficient than other models in terms of speed and memory consumption and is ideal for resource constrained devices like mobile environment.

Summary

In this chapter we present the results and discussions for our research work and give deeper insights into the results obtained. We demonstrate that the results are comparable to previous works in the literature and could be further improved and clinically tested.

CHAPTER 6: CONCLUSION

In this research, we have been able to achieve promising results on our DR severity classification system using a custom-made dataset with several pre-processing and image augmentation techniques. Our model used the computationally efficient architecture of MobileNetV2 which is known to be fast and computationally efficient algorithm. The model was tested on unseen data to test the generalizability of the model. Our results show that the model has been able to achieve good performance due to the various techniques that we used at every stage of the ML pipeline. In future, we aim to deploy and test the model in a smartphone and thereby test its effectiveness as a point-of-care technology for grading the severity of DR. We may also experiment with other variations of the MobileNet architecture and compare their effectiveness amongst each other and hence improve the state-of-the-art.

We have studied the various trends in the literature for diabetic retinopathy detection using artificial intelligence and have studied the behavior of lightweight mobileNetV2 architecture with other transfer learning architectures. We then tested the performance of the classifiers and demonstrate that light weight architectures give comparable results with the added benefits of faster computation and efficiency in computational cost.

In our experiments we see that we have obtained the testing accuracy of 91.68% on unseen data and the model demonstrates good generalizability ability using light weight mobile friendly architecture of MobileNetV2.

6.1 Future Research Directions

A number of future directions can be listed out to benefit the research community and they are explained below:

1 Improve the existing model: The existing model can be fine-tuned using various hyperparameter tuning and could also be improved using different preprocessing techniques. This could help in building models which can have better performance than the existing system. Improving the model to have good generalizability is the key for any good deep learning model.

2 Build an android application and test in clinical setting: We could test the performance of this model by incorporating it into an android application and hence test it in a real clinical setting with real patients. This would give a better insight into the effectiveness of such systems in clinical settings.

3 Further research: Since the MobileNetV2 is used for its lightweight architecture, speed and computational efficiency, it would be worthwhile to do a comparative study of the models based on these criteria as future work.

6.2 Related Publications

[1] S. Sheikh and U. Qidwai, “Using MobileNetV2 to Classify the Severity of Diabetic Retinopathy”, published in UKSim-AMSS 22nd International Conference on Modelling & Simulation, Cambridge, UK, 25-27 March 2020.

[2] S. Sheikh and U. Qidwai, "Smartphone-based Diabetic Retinopathy Severity Classification using Convolution Neural Networks", [Accepted] in Intelligent Systems Conference (IntelliSys) 2020, Amsterdam, September 2020.

REFERENCES

1. IDF Diabetes Atlas 9th edition 2019, https://www.diabetesatlas.org/, last accessed 2020/1/14.
2. Nijalingappa, P., B, S.: Machine learning approach for the identification of Diabetes Retinopathy and its stages. International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2015, IEEE. (2015).
3. Bhaskaranand, M. et al.: Automated Diabetic Retinopathy Screening and Monitoring Using Retinal Fundus Image Analysis. Journal of Diabetes Science and Technology. 10, 2, 254-261 (2016).
4. PMC, E.: Europe PMC, http://europepmc.org/abstract/MED/26720694, last accessed 2020/1/14.
5. Rajalakshmi, R. et al.: Validation of Smartphone Based Retinal Photography for Diabetic Retinopathy Screening. PLOS ONE. 10, 9, e0138285 (2015).
6. Russo, A. et al.: Comparison of Smartphone Ophthalmoscopy With Slit-Lamp Biomicroscopy for Grading Diabetic Retinopathy. American Journal of Ophthalmology. 159, 2, 360-364.e1 (2015).
7. Ryan, M. et al.: Comparison Among Methods of Retinopathy Assessment (CAMRA) Study. Ophthalmology. 122, 10, 2038-2043 (2015).
8. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861. 2017.
9. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018. p. 4510–4520.

10. F. Chentli, S. Azzoug, and S. Mahgoun, “Diabetes mellitus in elderly,” Indian journal of endocrinology and metabolism, vol. 19, no. 6, p. 744, 2015.

11. Y. Zheng et al., “The worldwide epidemic of diabetic retinopathy,” Indian Journal of Ophthalmology, vol. 60, no. 5, pp. 428–431, 2012.

12. “Facts about diabetic eye disease,” https://nei.nih.gov/health/diabetic/ retinopathy, accessed: 2019-03-20.

13. Early Treatment Diabetic Retinopathy Study Research Group et al., “Classification of diabetic retinopathy from fluorescein angiograms: ETDRS report number 11,” Ophthalmology, vol. 98, no. 5, pp. 807–822, 1991.

14. T. Kauppi et al., “The diaretdb1 diabetic retinopathy database and evaluation protocol.” in Proceedings of Medical Image Understanding and Analysis (MIUA), vol. 1, 2007, pp. 1–10.

15. C. I. Sánchez et al., “Retinal image analysis to detect and quantify lesions associated with diabetic retinopathy,” in Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 1. IEEE, 2004, pp. 1624–1627.

16. V. Esmann et al., “Types of exudates in diabetic retinopathy,” Acta Medica Scandinavica, vol. 174, no. 3, pp. 375–384, 1963.

17. R. Klein et al., “The Wisconsin epidemiologic study of diabetic retinopathy: VII. diabetic nonproliferative retinal lesions,” Ophthalmology, vol. 94, no. 11, pp. 1389–1400, 1987.

18. J. Kaur and D. H. Sinha, “Automated detection of diabetic retinopathy using fundus image analysis,” International Journal of computer Science and Information Technologies, vol. 3, p. 4794, 2012.

19. Taylor, D. (2012). Diabetic Eye Screening Revised Grading Definitions.

20. American Academy of Opthalmology Staging Guideline, American Academy of Opthalmology, October 2002

21. Williams, R., Airey, M., Baxter, H., Forrester, J., Kennedy-Martin, T. and Girach, A. (2004). Epidemiology of diabetic retinopathy and macular oedema: a systematic review. *Eye*, 18(10), pp.963-983.

22. Dutta, S., Manideep, B., Basha, S., Caytiles, R. and Iyengar, N. (2018). Classification of Diabetic Retinopathy Images by Using Deep Learning Models. *International Journal of Grid and Distributed Computing*, 11(1), pp.99-106.

23. D. Zhang et al., “Diabetic retinopathy classification using deeply supervised ResNet,” in Proceedings of IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2017.

24. M. Alban and T. Gilligan, “Automated detection of diabetic retinopathy using fluorescein angiography photographs,” Report of Standford Education, 2016.

25. Z. Wang et al., “Zoom-in-net: Deep mining lesions for diabetic retinopathy detection,” in Proceeding of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2017, pp. 267–275.

26. S. Agatonovic-Kustrin and R. Beresford, “Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research,” Journal of Pharmaceutical and Biomedical Analysis, vol. 22, no. 5, pp. 717–727, 2000.

27. New York University Computer Vision (CSCI-GA.2271-001) course notes. Slides: Neural nets 1 and 2, Convolutional Networks 1 and 2, Object Classification 1 and 2. https://cs.nyu.edu/~fergus/teaching/vision/index.html Accessed March 5, 2020.

28. Goodfellow I, Bengio Y, Courville A. Deep Learning (online book). http: //www.deeplearningbook.org/. Accessed March 5, 2020.

29. Y. LeCun, L. Bottou, Y. Bengio, et al. Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998

30. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25.2012.

31. Online notes for CS231n: Convolutional Neural Networks for Visual Recognition. http://cs231n.github.io/transfer-learning/. Accessed March 5, 2020.

32. "Color Fundus Photography | Department of Ophthalmology", *Ophthalmology.med.ubc.ca*, 2020. [Online]. Available: https://ophthalmology.med.ubc.ca/patient-care/ophthalmic-photography/color-fundus-photography/. [Accessed: 20- Mar- 2020].

33. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV].

34. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs.CV].

35. Huang G, Liu Z, van der Maaten L, et al. Densely Connected Convolutional Networks. arXiv:1608.06993 [cs.CV]

36. Williams GA, Scott IU, Haller JA, et al. Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. Ophthalmology 2004;111(5):1055-62.

37. Lin DY, Blumenkranz MS, Brothers RJ, et al. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. Am J Ophthalmol 2002;134(2):204-13.

38. Larsen N, Godt J, Grunkin M, et al. Automated detection of diabetic retinopathy in a fundus photographic screening population. Invest Ophthalmol Vis Sci 2003;44(2):767-71.

39. Leese GP, Ellis JD, Morris AD, et al. Does direct ophthalmoscopy improve retinal screening for diabetic eye disease by retinal photography? Diabet Med 2002;19(10):867-9. 157

40. Ahmed J, Ward TP, Bursell SE, et al. The sensitivity and specificity of nonmydriatic digital stereoscopic retinal imaging in detecting diabetic retinopathy. Diabetes Care 2006;29(10):2205-9.

41. Velez R, Haffner S, Stern MP, et al. Ophthalmologist vs retinal photographs in screening for diabetic retinopathy. Clinical Research 1987;35(3):A363.

42. Pugh JA, Jacobson JM, Van Heuven WA, et al. Screening for diabetic retinopathy. The wide-angle retinal camera. Diabetes Care 1993;16(6):889-95.

43. Lawrence MG. The accuracy of digital-video retinal imaging to screen for diabetic retinopathy: an analysis of two digital-video retinal imaging systems using standard stereoscopic seven-field photography and dilated clinical examination as reference standards. Trans Am Ophthalmol Soc 2004;102:321-40.

44. Abramoff MD, Folk JC, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol 2013;131(3):351-7.

45. Brady CJ, Villanti AC, Pearson JL, et al. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res. 2014;16(10):e233.

46. Mitry D, Peto T, Hayat S, et al. Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC Norfolk Cohort on behalf of the UKBiobank eye and vision consortium. PLoS One. 2013;8(8):e71154.

47. Mudie LI, Wang X, Friedman DS, et al. Crowdsourcing and Automated Retinal Image Analysis for Diabetic Retinopathy. Curr Diab Rep. 2017 Sep 23;17(11):106. doi: 10.1007/s11892-017-0940-x. Review.

48. L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet and J. Langlois, "Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening", *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1116-1126, 2016. Available: 10.1109/tmi.2015.2509785.

49. M. Haloi, et al., "A Gaussian scale space approach for exudates detection, classification and severity prediction", 2015. [Accessed 4 December 2019].

50. P. Prentašić and S. Lončarić, "Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion", *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 281-292, 2016. Available: 10.1016/j.cmpb.2016.09.018.

51. S. Joshi and P. Karule, "A review on exudates detection methods for diabetic retinopathy", *Biomedicine & Pharmacotherapy*, vol. 97, pp. 1454-1460, 2018. Available: 10.1016/j.biopha.2017.11.009.

52. S. Kumar, and B. Kumar, "Diabetic Retinopathy Detection by Extracting Area and Number of Microaneurysm from Color Fundus Image.", *5th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE,* pp. (pp. 359-364)., 2018. [Accessed 4 December 2019].

53. Avijit Dasgupta and Sonam Singh. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, pages 248–251. IEEE, 2017.

54. T. R. Rakshitha, D. Devaraj and S.C. Prasanna Kumar, “Comparative study of imaging transforms on diabetic retinopathy images”, Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on. IEEE, (2016).

55. H. H. Vo and A. Verma, “Discriminant color texture descriptors for diabetic retinopathy recognition”,Intelligent Computer Communication and Processing (ICCP), 2016 IEEE 12th International Conference on. IEEE, (2016).

56. K. Maninis, J. Pont-Tuset, P. Arbelaez and L. Gool, "Deep Retinal Image Understanding", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016.

57. S. Mohammadian, A. Karsaz and Y. Roshan, "Comparative Study of Fine-Tuning of Pre-Trained Convolutional Neural Networks for Diabetic Retinopathy Screening", in *2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, Iran, 2017.

58. J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation", *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 12, pp. 2181-2193, 2017. Available: 10.1007/s11548-017-1619-0.

59. R. Mansour, "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy", *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 41-57, 2017. Available: 10.1007/s13534-017-0047-y.

60. S. Dutta, B. Manideep, S. Basha, R. Caytiles and N. Iyengar, "Classification of Diabetic Retinopathy Images by Using Deep Learning Models", *International Journal of Grid and Distributed Computing*, vol. 11, no. 1, pp. 99-106, 2018. Available: 10.14257/ijgdc.2018.11.1.09

61. X. Xu et al., "Smartphone-Based Accurate Analysis of Retinal Vasculature towards Point-of-Care Diagnostics", *Scientific Reports*, vol. 6, no. 1, 2016. Available: 10.1038/srep34603.

62. R. Sarki, S. Michalska, K. Ahmed, H. Wang and Y. Zhang, "Convolutional neural networks for mild diabetic retinopathy detection: an experimental study", 2019. Available: 10.1101/763136 [Accessed: 24-Feb-2020].

63. J. Gao, C. Leung and C. Miao, "Diabetic Retinopathy Classification Using an Efficient Convolutional Neural Network", in *2019 IEEE International Conference on Agents (ICA)*, 2019.

64. "dimitreOliveira/APTOS2019BlindnessDetection", *GitHub*, 2019. [Online]. Available: https://github.com/dimitreOliveira/APTOS2019BlindnessDetection. [Accessed: 04- Dec- 2019].

65. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/data. [Accessed: 24-Feb-2020].

66. [Online]. Available: http://www.adcis.net/en/third-party/messidor2/. [Accessed: 24- Feb- 2020].

67. "Operations on Arrays — OpenCV 2.4.13.7 documentation", *Docs.opencv.org*, 2020. [Online].

Available:https://docs.opencv.org/2.4/modules/core/doc/operations_on_arrays.html#void%20convertScaleAbs(InputArray%20src,%20OutputArray%20dst,%20double%20alpha,%20double%20beta). [Accessed: 24- Feb- 2020].

68. "OpenCV: Bioinspired Module Retina Introduction", *Docs.opencv.org*, 2020. [Online]. Available: https://docs.opencv.org/3.4/d2/d94/bioinspired_retina.html. [Accessed: 24- Feb- 2020].

69. "ImageNet", *Image-net.org*, 2020. [Online]. Available: http://www.image-net.org/. [Accessed: 20- Mar- 2020].

70. T. Mitchell, *Machine learning*. London: McGraw-Hill Pub. Co. (ISE Editions), 1997.