CONTENTS

|  |  |
| --- | --- |
| Abstract | 1 |
| List Of Figures | 2 |
| 1. Introduction | 3 |
| 1.1 What is Diabetic Retinopathy ? | 3 |
| 1.1.1 Phases in Diabetic Retinopathy | 3 |
| 1.1.2 Motivation behind DL classification | 4 |
|  |  |
| 2. Literature Research and Review | 5 |
| 2.1 Related Datasets | 5 |
| 2.2 Related works in Diabetic Retinopathy | 5 |
| 2.3 Related works in Deep Learning | 6 |
|  |  |
| 3. Methodology | 9 |
| 3.1 Data Preprocessing | 9 |
| 3.2 Data Imbalance | 9 |
| 3.1.2 Data Augmentation | 10 |
| 3.1.2.1 Flip and Rotation | 10 |
| 3.1.2.2 Adding Noise | 10 |
| 3.1.2.3 Scaling and Cropping | 11 |
| 3.2 DenseNETS | 11 |
|  |  |
| 4. Implementation and Discussion | 13 |
| 4.1 Data Set | 13 |
| 4.2 Methodology for implementation | 14 |
| 4.3 Workflow of Model | 14 |
| 4.4 Evaluation of Model | 15 |
| 4.5 Model | 16 |
| 4.6 Effective learning rate identifier | 17 |
| 4.7 Test Results | 18 |
|  |  |
| 5. Summary and Prospectives | 21 |
| 6. Publication and References | 23 |
| 7. Appendix | 26 |
|  |  |
|  |  |

ABSTRACT

Diabetic Retinopathy (DR) is a situation coined for malfunctioning of human eye which make human vision blurry or even worst cases eye blindness. Usually properly trained ophthalmologists have to spend hours just to inspect each and every subtle features present and evaluate its significance and then arrive on conclusion. This whole process is very much time consuming and cumbersome from a human point of view. To reduce the human effort and window of error a lot of people tried to automate this whole process using state of the art techniques like Deep learning and computer vision. However there are many challenges like availability of balanced datasets, computationally efficient networks, proper fundus images to train a deep learning model. SO the primary objective in this research boils down to develop a DR method which is computation effective and reliable. On Kaggle EyePACS dataset our model achieves an accuracy of 92.1 % which is really good as compared to many state of the art available models.

LIST OF FIGURES

|  |  |
| --- | --- |
| *Figure 1. shows different stages of Diabetic Retinopathy* | 3 |
| Figure 2 represents various stages of DR eye along with the edges map representing vessels | 4 |
| Figure 3 Data Augmentation in play | 9 |
| Figure 4 Original Image, Horizontal flipped, Vertical flipped | 10 |
| Figure 5 Original Image, Added Gaussian Noise, Added alt and Pepper Noise | 10 |
| Figure 6 Original Image, Scaled 10% image, Cropped Image | 11 |
| Figure 7 Standard ConvNet Concept | 11 |
| Figure 8 ResNET Architecture Example | 12 |
| Figure 9 Softmax classifier doing final prediction | 12 |
| Figure 10 Displays a confusion matrix | 13 |
| Figure 11 Image describing performance for different learning rate | 17 |
| Figure 12 Demonstration of results for Random, Incorrect and Correctly Predicted | 19 |
| Figure 13: Distribution of Diabetic Retinopathy label values in Kaggle train | 18 |
| Figure 14 Random sample images from the Kaggle dataset | 19 |
| Figure 15 Test Results confusion Matrix | 23 |

# Chapter 1. Introduction – Diabetic retinopathy

## What is Diabetic Retinopathy?

Diabetes is a disease when blood sugar level in a human body imbalances. When the blood sugar level of human body increases a lot of side effects affects the proper functioning of human body, one of which is the most vital organ of human body that is *eye.*  The blood sugar level, after increasing by a certain amount start damaging the human eye blood vessels in retina. Due to the sugar getting deposited there, if affects the flow of blood and the blood vessels can overflow or leak. Sometimes they also can get closed in which case eye will not receive sufficient blood supply. Sometimes due to this abnormality, new vessels can come up in eye, and all the above situations can steal your vision. Diabetes retinopathy starts to affect eye slowly, if diagnosed early it can be completely cured. In the next section we discuss about the stages in diabetic retinopathy.

## Phases in Diabetic Retinopathy (DR)

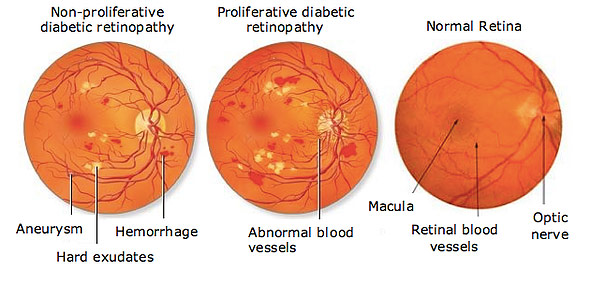
Diabetic retinopathy situation of human eye can be broken down into two main stages:

### NPDR (non – proliferative diabetic retinopathy)

This is the primary stage where few of the blood vessels swell. Sometime they may leak as well or get closed off, both the situation leads to the loss of vision of a patient in later stages.

### PDR (proliferative diabetic retinopathy)

This is the more advanced stage. In this stage abnormal growth is observed in retina. The newly gown blood vessels are quite fragile and often bleed, which affects the vision and sometimes blocks it as well. The newly grown fragile tissues looks like a scar in the eye.



**Figure 1 shows different stages of Diabetic Retinopathy.**

## Motivation behind Deep Learning based Classification of Disease

It is estimated by US Center for Disease Control and Prevention that 29.1 million people in the US are affected by diabetes and WHO estimated that 347 million have the same disease. Diabetic Retinopathy is a eye disease but is mainly caused by longstanding diabetes. Around 40-45 % of people that has diabetes are in some stage of DR which can e averted or slowed down if detected in time.

Detecting DR used to be a manual process some years back, which is a very complex and time taking process and involved human error as well. By the time Subject Matter Experts used to submit their review, often after one or two days, the delayed identification would lead to lost follow up, miscommunication and sometimes treatment as well.

DR can be identified using various clues like lesions, scars etc. The instruments needed to capture this information is actually a very high resolution camera and very costly thus making its multiple availability in poor area almost impossible. Thus a comprehensive and automated process to identify DR is very much needed, that too with good accuracy. This automated system can be developed with the help of image processing, computer vision and deep learning applications knowledge. A color fundus photograph of both eye will be needed for the purpose of algorithm development. Color fundus images has to be annotated for various stages of the disease asafollows:

1. Normal
2. Mild
3. Moderate
4. Severe
5. Very Severe

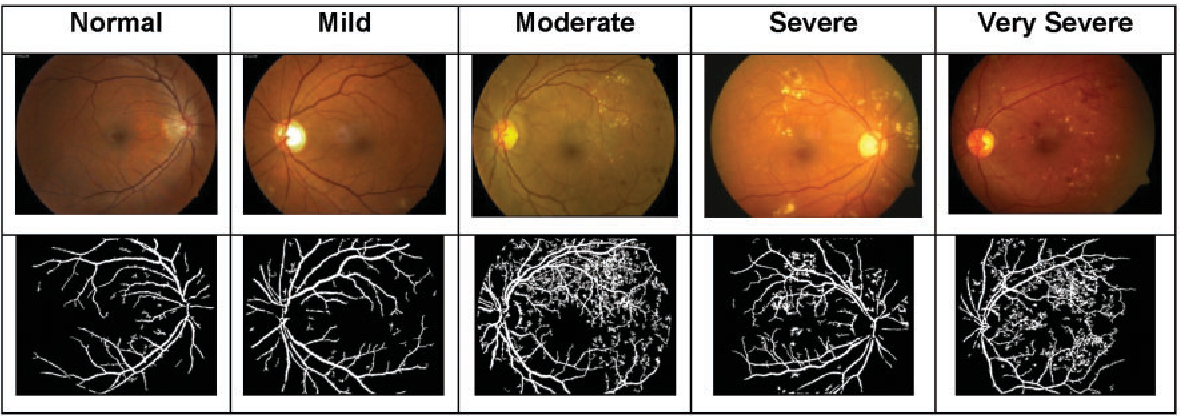


Figure 2 represents various stages of DR eye along with the edges map representing vessels

# Chapter 2.

# Literature Research and Review

## 2.1 Related Datasets

There exists several datsets of fundus images, developed by people and institutions all over the world. They are following:-

1. “Kaggle Diabetic Retinopathy Dataset” [14]
2. “E-Optha Dataset” [15]
3. “REVIEW - Retinal Vessel Image set for Estimation of Widths dataset” [12]
4. “STARE - Structured Analysis of the Retina dataset” [11]
5. “DRIVE - Digital Retinal Images for Vessel Extraction dataset” [14]
6. “MESSIDOR - Methods to Evaluate Segmentation and Indexing techniques in the field of Retinal Ophthalmology dataset” [14]

## 2.2 Related Works in Machine Learning

Before the development of deep learning algorithms, especially deep neuralaanetworks, a feature extraction step was needed for general computer vision tasks. Features, in this case, are distinguishing and significative small image patches. Highlights, for this situation, are recognizing and particular little picture pieces. As found in many literatures, a lot of work for specifically feature extraction domain has been tabled including the infamous Space Invariant Feature Transform and Speeded up Robust Features[1] and mark these names as they have been widely used and has been applied for retrieving medicinal images also called medical imaging retireival and object detection from scene. Following the element extraction step, conventional AI characterization alogorithms, for example, sopport vector machines, calculated logistic regression or decision trees, are prepared utilizing separated highlights to arrange the picture. This pipeline was applied to practically all conventional PC vision examination assignments including picture and recordings based order, picture and recordings based division, and so forth.

Programmed DR discovery dependent on conventional PC vision methods pursues a similar pipeline. To begin with, explicit highlights are separated from the fundus pictures utilizing single or blends of physically structured element extraction calculations. Classifiers are then prepared utilizing the removed element to characterize the DR stages. A few such works are quickly looked into beneath. A multilayer neural system model was prepared utilizing highlights removed by recursive area developing division calculations to perform paired DR arrangement task . [2] The affectability and particularity of the proposed framework is 80.2% and 70.7% separately. Recursive district developing division calculations was additionally embraced in to remove DR visual highlights, for example, exudates, hemorrhages, and microaneurysms. The highlights were then utilized for Healthy/Diabetic Retinopathy parallel classification.[3] The affectability and particularity of the proposed framework is 74.8% and 82.7% separately.

A very renowned scientist Lee[4] made a strong statement very early by bringing on table the methods mostly automated for assembling and measuring severity for different diabetic retinopathy lessions found in the human eye A calculation is prepared to arrange NPDR dependent on the nonappearance of these three sorts of injuries. The general precision of the proposed framework is 81.7%. In [27], a choice emotionally supportive network was proposed for early location of Diabetic Retinopathy (nearness of macroaneurysms) utilizing Bayes optimality criteria. The technique had the option to spot and pinpoint the beginning time of Diabetic Retinopathy (DR) with an affectability of 100% and particularity of 67%. The programmed DR grouping strategies examined in this segment utilize physically planned highlights. The performiance of such techniques is profoundly reliant on the element extraction process. AI calculations are utilized for numerical streamlining dependent on the highlights structured by people. Consequently, space information on the DR illness assumes a pivotal job in building successful DR characterization models. The impediment of applying ordinary PC vision calculations for Diabetic Retinopathy (DR) injury discovery or DR order is that the physically structured highlights are regularly overspecified, deficient or require quite a while and know-how and comprehension to plan and approve.

## 2.3 Related Works in Deep Learning

Rather than physically planning highlights and nourishing them to the classifier for DR screening, numerous analysts are currently fabricating start to finish profound learning models which gain proficiency with all the required highlights naturally. A few works which utilize profound neural system models to distinguish DR consequently are quickly investigated beneath. A 13-layer profound convolutional neural system (CNN) for screening DR was talked about in [29]. A few information enlargement systems (i.e., picture revolution, picture flipping and picture moving) were applied to beat the information imbalanced issue. 5,000 pictures from the EyePACS DR preparing set [5] were held as the test set and the remainder of the pictures from the EyePACS DR preparing set were utilized for preparing the proposed model. Deep learning base expertises are totally paying attention on properly learning and understanding many levels of abstraction and representation which that helps to make inference of the informational intelligence that is hidden in any kind of data such as colored images of dogs cats. So In this manner having a totally complete full set of properly and correctly classified images with annotations and without any having any presumptive perceptive of the available features required to do the classification activity, the system can be abled to look and learn the properties of the image and its underlying features which usually minimizes a defined cost or maybe a loss function that is indirectly or directly associated with the classification cost and score index that has to be optimized.

In [8], a Non-Local methods denoising technique was applied to the fundus picture to expel clamor from the EyePACS DR dataset. AlexNet [29] and GoogleNet[34] were used to arrange the pre-handled retinal pictures. An AUC score of 0.78 is accomplished by the GoogleNet model contrasted with an AUC score of 0.68 accomplished by AlexNet. In [35], a computationally effective CNN model, to be specific the MobileNet model [38], was utilized for double DR screening (DR V.S. Solid picture) utilizing the EyePACS DR dataset. An exactness of 0.73 was accomplished by the proposed strategy. Considering the incredible information unevenness nature of this dataset (73.4% of the picture are sound and 26.6% of the picture contains DIABETIC RETINOPATHY), just utilizing exactness score may not proper for showing the viability of the proposed model. For an outrageous model, a straightforward classifier which consistently predicts pictures as solid pictures will likewise accomplish an exactness of 0.73 utilizing the EyePACS DIABETIC RETINOPATHY dataset. In [36], GoogleNet and VGGNet were changed to two relating systems CKML (Combined Kernels with Multiple Losses Network) and VNXK (VGGNet with Extra Kernel). The retinal pictures were changed over to a half breed LGI shading space. Models prepared utilizing the LGI retinal pictures were then contrasted and models prepared utilizing RGB retinal pictures. The outcomes show that both CKML and VNXK accomplish higher exactness with LGI pictures. Be that as it may, the proposed model is likewise profoundly one-sided to the solid examples. The comparing precision from class 0(healthy) to class 4(proliferative DR) are 97.6%, 11.9%, 57.9%, 33.2% and 36.8% individually. In [9], a novel Zoom-in-Net model was proposed. It imitates the zoom-in procedure of an ophthalmologist who inspects the retinal pictures. Zoom-in-Net comprises of three sections, Inception-ResNet was received as the principle arrange (M-Net) for DIABETIC RETINOPATHY grouping. Two little CNNs, to be specific the Attention Network (A-Net) and the Crop Network (C-Net), were utilized for Attention Localization of the for DIABETIC RETINOPATHY pictures. The single Zoom-in-Net accomplishes a QWK score of 0.849 on the whole EyePACS DIABETIC RETINOPATHY test set. With a troupe of three Zoom-in-Net models, the QWK score was improved to 0.854.

In the 80s the most used activation function was the sigmoid, a smooth non-linear function, that is continously differentiable. Despite its interesting properties, it has a very important concern, called gradient vanishing problem [11]. Sigmoid first derivative becomes flat not far from the origin, affecting network loss optimization due to near-to-zero gradients. In [12] ReLUs were used for improving Restricted Boltzmann Machines (RBMs), approximating stepped sigmoid units with ReLUs. In (Glorot, Bordes, and Bengio, 2011) the authors compared the performance of Sigmoid, Tanh and ReLU arriving to the conclusion that despite Sigmoid being more plausible biologically, Tanh and ReLU were more suitable to be used as activation function for training multi-layer perceptrons. ReLU networks have better performance in general, despite its non-differentiability at zero and its hard non-linearity. Furthermore, ReLU networks lead to sparse representations, being beneficial, both because information is represented in a more robust manner and because it leads to significant computational efficiency. Moreover, the simplicity of the function and its derivative reduces calculation time, being of significant importance when working with big networks. The constant value of the gradient, helps avoiding the gradient vanishing problem, allowing the design of deeper networks. For then on, ReLU has become the default activation function for deep learning. Many other activation functions have been published, like LeakyReLU, but not introducing significant performance gains.

Traditional models of pattern recognition in images since the late 50s have been based on extracting hand-crafted fixed engineered features or fixed kernels from the image and, then, using a trainable classifier on top of those features to get the final classification. Using this model, the problem of the DR detection has been tackled by feature extraction using on hand models targeted to the detection of microaneurisms, haemorrhages and exudates in retinal images (e.g. (Sudha and Thirupurasundari, 2014), (Torrents-Barrena et al., 2015), etc). This type of approach requires a good understanding of disease mechanisms to be able to find the important features present in the image. This knowledge is specific of the problem to be solved, requires a lot of labor. To automate the prediction of diabetic retinopathy a ResNet model was supposed. In that particular model 3 auxiliary layers of neural networks were added simultaneously by modifying the existing ResNet architecture to provide model auxiliary penalty called regularization during training the process. For obviously retaining as the testing set 20% images were sampled by making a random distribution. It was observed that if you donot consider the side output layers in ResNet results will demotivate you as ResNet have a multi focus phenomenon where it can do a lot of multi scale learning as well. The QWK score, specificity, sensitivity and accuracy of the proposed system on the test set is 0.73, 94%, 67% and 81% respectively.

Profound learning methods are not just successful for understanding general characterization assignments, yet additionally for forecast in restorative imaging. For our contextual analysis of diabetic retinopathy ailment evaluating, having enough information our technique can perform close to human level mastery. Lee et al. [26] presented strategies for naturally sectioning the reality of three early DIABETIC RETINOPATHY sores, to be specific hemorrhages, hard exudates, and cotton-fleece spots. A calculation is prepared to arrange NPDR dependent on the nonappearance of these three sorts of injuries. Therapeutic applications including picture arrangement, division and confinement are not rejected from the quick developing of profound learning . An outline of more than 300 ongoing productions, a bunches of picture examination put together synopsis particularly with respect to restorative picture investigation utilizing profound learning has been accounted for by [12]. A large portion of the ongoing profound learning calculations proposed for retinal picture investigation have utilized CNNs for shading fundus pictures [12]. A parcels and bunches of assortments of utilizations are tended to: picture division of anatomical structures , picture division investigation and picture discovery examination of human eye retinal variations from the norm , Correct and expedient making sense of human eye infections, and any sort of picture quality appraisal. In the last decade, several attempts to automatize the DR diagnosis through images of the eye fundus have been tested. They are basically focused on applying well-known pattern recognition models. In this work, we want to apply a Deep Convolutional Neural Network (DCNN) model, as it has been proven to be a very effective algorithm to solve general image classification problems. DCNN is a supervised learning model for automatic classification that does not require any pretreatment of the images, nor any feature analysis. Deep learning base expertises are totally paying attention on properly learning and understanding many levels of abstraction and representation which that helps to make inference of the informational intelligence that is hidden in any kind of data such as colored images of dogs cats. So In this manner having a totally complete full set of properly and correctly classified images with annotations and without any having any presumptive perceptive of the available features required to do the classification activity, the system can be abled to look and learn the properties of the image and its underlying features which usually minimizes a defined cost or maybe a loss function that is indirectly or directly associated with the classification cost and score index that has to be optimized. In this chapter we show that a DCNN is able to understand from data the most important features to make the classification of retinal images into the five DR categories, without the need of a hand-crafted feature extraction process. The thesis is organized as follows: we present the related work, next we explain the characteristics of the available data and why deep learning techniques can be applied over them, next we explain the methodology used for solving the problem, we show the obtained results and finally we expose the conclusions and further steps for improving the results.

In this chapter is shown that deep learning techniques are not only very effective for solving general classification tasks, but also for prediction in medical imaging. For our case study of diabetic retinopathy disease grading, having enough data our method is able to perform near human level expertise. Work of next chapters will be centered on testing the newer schemes, the use of alternative cost functions that encode the prior information of the ordering of the classes and other more elaborated methods for combining the information coming from both eyes.

# Chapter 3.

## Methodology

## 3.1 Data Preprocessing

## 3.1.1 Data Imbalance

Data Imbalance problem is one of the biggest challenge in image classification of medical images, or rare occurrence classes[16]. CNN are built to minimize errors through backpropagation. On sampling from training set in a imbalanced data set, probability of object belonging to particular class will be directly proportional to its imbalanced numbers, so in the quest of minimize errors model will be misled to believe that mostly observation belongs to set with high numbers than other. To negate the effects of data imbalance, different augmentation, and oversampling from minority class is tried out.

## 3.1.2 Data Augmentation

When number of data is limited but the CNN cannot learn properly with the amount of data available we cannot do anything but have to look for techniques which will help us increase our data and there we go for data augmentation as shown in *figure 3*.

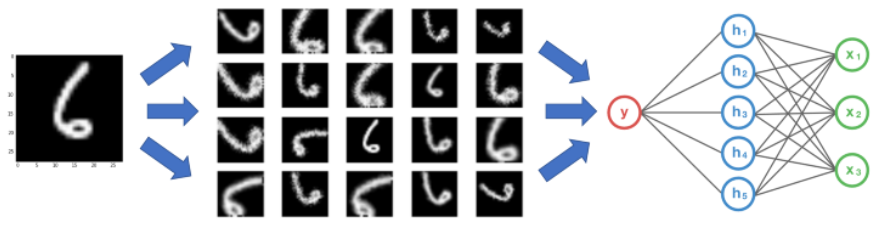


Figure 3 Data Augmentation in play

A CNN is said to be following the property called Invariance in *figure 3* as we can see that the above model will be invariant to rotation[17]. Similarly other properties like invariant to translation, cropping, illuminance also exists. So augmentation can be of different types, like offline augmentation and online augmentation. **Offline Augmentation**[18]is usually preferred for smaller datasets where we do the data augmentation a priori and store all the images for training. Other variety is **Online Augmentation**[18]**,** this variant is space efficient as it does augmentation at the time of training using various in house image processing functions and do not create new images to store them and does this process in batches otherwise it will lead to explosively increase in the size.

The following are the popular augmentation techniques with example:

### 3.1.2.1 Flip and Rotation

Images can be flipped horizontally as well as vertically as shown in *figure 4.* Horizontal flip can be seen as creating mirror image of an image. In some libraries or frameworks vertical flips option is not available, it can be taken as rotating image by 180 degree and performing horizontal flips[18].



Figure 4 Original Image, Horizontal flipped, Vertical flipped

### 3.1.2.2 Adding Noise

When neural network somehow learns very high frequency features or patterns that occurs a lot, it leads to overfitting of model. To cater this problem Gaussian noise[19] can be used as it has data points in all frequencies, and zero mean, which can effectively distort high frequency features. Another good suggestion is to use salt and pepper noise[20], it is random black and white pixels added to image as shown in *figure 5*



Figure 5 Original Image, Added Gaussian Noise, Added alt and Pepper Noise

### 3.1.2.3 Scaling and Cropping

Image can be scaled inward or outward, which also produces similar effect to that of cropping. All the process in scaling and cropping should be random with some preset values that should be taken. In the *figure 6* below following technique example is demonstrated



Figure 6 Original Image, Scaled 10% image, Cropped Image

## 3.2 DenseNet

CNN are playing a dominant roles in today’s world’s imaging technology [21]. From object detection, recognition to semantic and instance segmentation, CNN can be used to solve wide variety of problem statements [22]. Features from images are extracted in CNN by taking repeated convolutions using kernels, which are usually called standard ConvNets as shown [23] in *figure 7*.

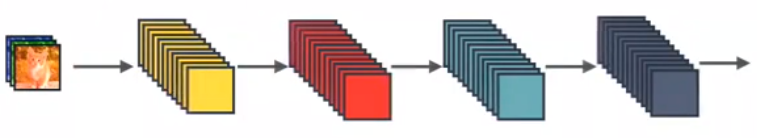


Figure 7 Standard ConvNet Concept

Gradient propagation was introduced in ResNet where identity is used as displayed in *figure 8*. Intitutively understood as a algorithm where a state is passed from one ResNet module to another one for better utilization of gradient values.

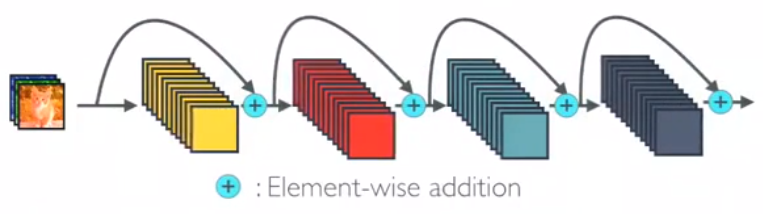


Figure 8 ResNET Architecture Example

In DenseNET all the layers pass its own and its preceding layer information to its succeeding layer. Thus each layers will have the information of the gradients which its previous layers have as shown in *figure9*.

Consider a Convolutional Neural Network of layers and a. Let the non linear transforms of an image is represented by where represents layer, which can be anything among Batch Normalization (BN), Rectified Linear Unit (ReLU), Pooling or Convolution. Output of layer is represented by .

Usually in CNN output of layer serves as a input for layer which gives rise to the transformations [27]:

In ResNETs a skip connection [28] is added which bypasses the non-linear transformation with a identity function as follows:

To further improvise the flow of information a different connectivity pattern in DenseNET[28] was introduced in which any subsequent layers are connected to all its previous layers. Lets assume that is output of DenseNET layer and it receives as inputs, so can be written as :

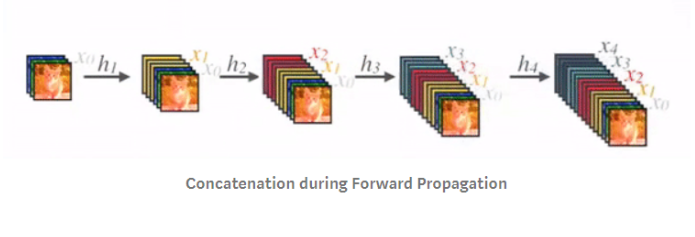


Figure 9 Represents Concatenation

# Chapter 4.

# Implementation and Discussion

This chapter deals with the models designed and implemented during working on this thesis. Aim of this exercise was to explore the problem statement in depth, compare performance between different models and to better understand the hyper parameter optimization process in various different state of the art models.

During past several years, most appropriately decade, a lot of people, companies, students and researchers are working towards the automatic diabetic retinopathy detection using image processing and computer vision. Most of the people have worked towards implementation of some state of the art classification models to get desired result.

## 4.1 Data Set

To obtain the fundus image datasets, we explored multiple websites and avenues but we found Kaggle[29] dataset as the best. The size of the total data set was 40 Gigabytes.

## 4.2 Methodology for Implementation

Convolution Neural Networks (CNN) gives us a lot of options for supervised deep learning for images example AlexNet[29] (Sutskever, Hinton and Krizhevsky, 2012), GoogleNet[30] (2015), VGGNet[31] (Zisserman & Simonyan 2015), ResNET[24] (Microsoft), all these architectures are based on special concept of neural network that is Convolution Neural Networks (CNNs).

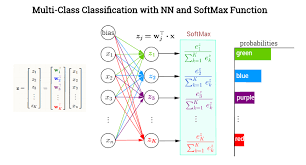
CNNs are made up to decompose or extract information from images and hold those information in different layers. To extract features which are nothing but statistical information from data, pooling layers are used which are also called dimensional down-size blocks. Information collected by one layer is passed to another succeeding layer which can also be called as features of features helps us in classification. To perform the final classification it is evident that a fully connected layer is required which collects information from all the layers on its behind and make mathematical prediction as shown in *figure 10*.

Figure 10 Softmax classifier doing final prediction

Usually different from our traditional algorithms, these CNN based algorithms provide us end-to-end learning scheme which are totally unaffected by human intervention or possibility of human error gets minimized here to the minimum if data sets is loaded properly.

Next we explain different aspects of the model used in this thesis that is DenseNET based Diabetic Retinopathy classification based on the available data from EyePACS, evaluation function, training procedure, model optimization and probabilistic combination of prediction from both the eyes.

## 4.3 Workflow of Deep Learning Model Implementation

In this area, we abridge the run of the mill work process of applying neural system models in a true application. The commonplace work process is condensed as pursues:

1. Data Acquisition: In the beginning a database for the application has to be made. Generally for training the whole model we keep aside 80 percentage of the overall data from the database that we have collected for the whole purpose. Now the remaining 20 percentage of the data is left for testing the model that we have trained on the 80% of the data, from which we usually some small portion of the data for backpropagation algorithm where we have to tune the hyper parameters.

2. Exploratory Data Analysis: Now since we have arranged a database for training the model we focus on exploratory analyzing the data. While analyzing the data we have to look for certain kinds of trands that can be present. Some of the trend re like there can be imbalance in data. to find this data imbalance we have to plot a histogram and check. There can be several kinds of features which might have very low latency and follow a certain trend. There can be some features which can help us detect the anomalies which we have to look upon and analyze closely

3. Data Preprocessing and Data Augmentation: Now since we have performed both the steps correctly we can perform certain things which can help us to correct the issues that we have addressed while looking at the exploratory data analysis report. If there are no issues then we can directly jump to step number 4. Well, to make neural network converge faster several data preprocessing followed by dataset augmentation methods are applied. Usually images data are not dealt within the range of 255 but are usually standardized in range of zero to one. Sometime some models deal with only single channel images so RGB channel images are converted into grayscale. To increase the number of sample in the under count classes, data augmentation techniques are applied to increase the headcounts of the number of samples from that particular class. By deploying these augmentation and preprocessing steps scientists has widely reduced problems like overfitting. It has been widely used for improving the model performance and reducing overfitting, especially for training with a small dataset or imbalanced dataset. Some of the very frequently deployed image augmentation techniques include cropping, zooming randomly, and sometimes splitting the images as well helps.

4. Architecture Design: Continuing in the sequence, we have to locate a lot of models which are reasonable for the applications. In the event that we have restricted computational assets or an ongoing forecast is required, we may utilize computationally effective CNN models for the undertaking. Then again, we may utilize profound CNN models to get great execution if satisfactory computational assets are given.

5. Hyperparameter Optimization and Training: Moving forward in the sequence comes the hyper parameter optimization. Hyperparameters can be termed as the brute force variables that are present in any deep learning model and has to be tuned wit trail and error methodologies. Hyperparameter plays a great role in the outcome of any deep learning model and are usually treated and stated and set by using a small set of data obtained from training set which is usually called as validation set. For hyperparameter optimization data is sampled from main train database and few values are checked against condition like testing accuracy should be more than training accuracy otherwise model is overfitting. For every step there are many possibilities of hyper parameter and there can be many possible combinations thus. So the best possible combination of hyperparameter has to be chosen.

6. Performance Evaluation: So when the proper hyperparameter combination has been identified model has to be tested and it performance has to be evaluated which is being done in this step.

7. Ensemble learning (optional): In the quest of improvising model more and more there are various methodologies of processes can be evaluated, out of the one is ensemble learning. It’s a great deal of deep learning model where various models are being stacked and they are asked to make the prediction, based on the popular vote a result is usually selected.

## 4.4 Evaluation Of Model

The problem to be modeled is essentially a classification problem and we define certain performance assessment criteria to evaluate the performance of each applied algorithm and derive inferences from it. The evaluation metrics for our models are listed as follows:

We tackle the classification problem using supervised learning methods due to the availability of true class labels of the available data. Hence it is possible to compute the correct and incorrect class predictions for the samples of data by comparing prediction results with the true labels and hence determine the misclassification error as the proportion of incorrectly classified samples. Conversely, the prediction accuracy is a measure of the proportion of samples that have been correctly classified by the models.

We determine the model performance using two types of validation methods. One being the hold-out accuracy and the other being cross-validation accuracy. The hold-out procedure consists of splitting the data into two sets namely the training and the testing set. The two sets of data are disjoint or mutually exclusive of each other and commonly the training set is the larger of the two sets. Hence, in this method the classifier is learned or trained on the training set without any exposure to the test set. The trained model is then applied to the test set and outputs predictions for each of the test sample classes which is compared with the true labels. Cross validation on the other hand involves splitting data into folds and training the classifier on data except a particular fold which is left for validation. This procedure is repeated across all folds and the final accuracy is representative of the average performance accuracy over all folds. We are interested to see how hold-out and cross validation accuracies of our chosen models compare against each other

In cases when data is imbalance, using accuracy as sole metric to evaluate the model is misleading. For example lets consider a example:

Let the number of images for 5 classes as following:

A – 5000

B – 500

C – 150

D – 53

E – 5

So here lets consider that our model is giving accuracy of 98 %.

2 % error – 2/100 \* 5708 ~ 114 erros

So even if out model completely misclassifies class D and E still accuracy will look quite good which is not good and misleading.

So like in our cases where data is imbalanced we use metrics like confusion matrix as shown in *figure 11*

Figure 13 Displays a confusion matrix

In confusion matrix for every class we have information of True Positives, True Negatives, False Positives and False Negatives. As you can see in the *figure 11* that the rows and columns are named as actual and predicted. And the values in each cell is the intersection of its row name and column name. Several metrics has been made out of this confusion matrix which are described as follows[33]:

Table 1 Different metrics from Confusion Matrix

|  |  |
| --- | --- |
| Metric Name | Formula |
| aAccuracy | a (TP+TN)/total |
| aMisclassification Rate | a (FP+FN)/total |
| aTrue Positive Rate, aSensitivity, aRecall | aTP/actual yes |
| aFalse Positive Rate: | aFP/actual no |
| aPrecision | aTP/predicted yes |
| aPrevalence | aactual yes/total |
| aTrue Negative Rate, aSpecificity | aTN/actual no |

Precision and recall are metrics which give us a picture about model performance and can be evaluated from the confusion matrix. Precision is a measure of how many of the predicted values in a class are correctly classified or actually part of the true labels of the class. Hence it is a measure of positive prediction by the model. Recall on the other hand is a measure of the amount of information correctly retrieved or in other words the number of samples correctly predicted. Models can be optimized based on a measure which combines or balances both precision and recall. This is called F-measure which is a weighted average of the precision and recall of the model. Precision and recall can be computed from the confusion matrix to determine model performance. In a binary sense, the confusion matrix portrays the number of true positives, true negatives, false positives and false negatives.

## 4.5 Data Pre-Processing and Augmentation

Exploratory Data Analysis (EDA) [64] might prove to be the best man to bet on as any process which will help you a lot in general to make a genral deep learning model. It refers to the primary inspection process of the data in order to discover underlying patterns such as imbalanced data distribution, identify anomalies or outliers using statistical methods and illustrates data properties using graphical representations. As the Kaggle Diabetic Retinopathy dataset [5] contains fundus retinal images captured from a fundus camera and the corresponding DR severity labels, we perform basic Exploratory Data Analysis on both the retinal images and the labels

The Kaggle dataset [5] is a large Diabetic Retinopathy image based dataset with very high-resolution images taken under different lighting and imaging conditions. There are in total 17653/5453 overall both eye pairs of colorful images for the totally training/test set and the corresponding Diabetic Retinopathy severity levels are provided. The severity level of the eyes in diabetic retinopathy are rated in the range of zero to four by an expert oncologist. While doing exploratory analysis and on plotting a histogram you can clearly see that histogram is highly skewed which clearly represent data imbalance. On investigating further we found out that healthy eyes images constituted of around 73 percentage, while a little worse than that constituted of around six percentage. Going down to next class we get 15 percentage labelled as moderate. Rest 4 percentage was divided equally in between the much less talked about classes of severe and proliferative dr. High amount of correlation was found out in between the data of pair of eyes. In 87 percentage of the cases it was observed that both the eyes have similar DR level, while in rest of the eyes it was observed that DR level differentiated by atmax one which is superb.  
  
Please scroll down to next page to look at the figure 14 where different images from kaggle datasets are being displayed. As you can see, it is clearly visible that there is a variation in the illumination condition of various images not just that, images were taken from different cameras as well. Since different cameras were used in some cases you can see that the size of the captured images varies. The size of the captured images can vary from 400 x 315 to a bigger size like 5184 x 3456. For normalizing the difference caused due to illumination effects images are resized to 256 x 256 using interpolation bilinear. Some of the images can be too dark and some of the images can be much highlighted as well. The ones that are dark are usually called under exposed images and the one that are lighted are called over exposed images. But the catch is poorly exposed images do not let model learn properly and thus has to to be deleted from the database. To banish the poorly exposed data we calculated and plotted histogram values of the images and screened them. To classify an image to be underexposed we used to look whether 90% of its pixel values are under 30 value, and similarly to look at the over exposed images we used to look that if 90% of the pixel values are near or greater than 230. This values of threshold has to be choose carefully as some of the images also had black borders. There are other methodologies as well to screen the poorly lit images like Laplacian of Gaussian method to determine whether an images is blurred or sharp enough to be considered for training of the model for diabetic retinopathy.

DenseNETs usually because of its skip connection requires a larget datasets of images to avoid overfitting. But a large dataset can be imbalanced as well so we require to have a balanced large datasets in general to avoid overfitting. But usually a datasets mainly of diseases can be hugely imabalanced and needs to be corrected usually through a technique discusses above called augmentation. Here since the dataset is not considerably large we have used offline segmentation. In order to perform image augmentation using image processing defined function following are the methods used :

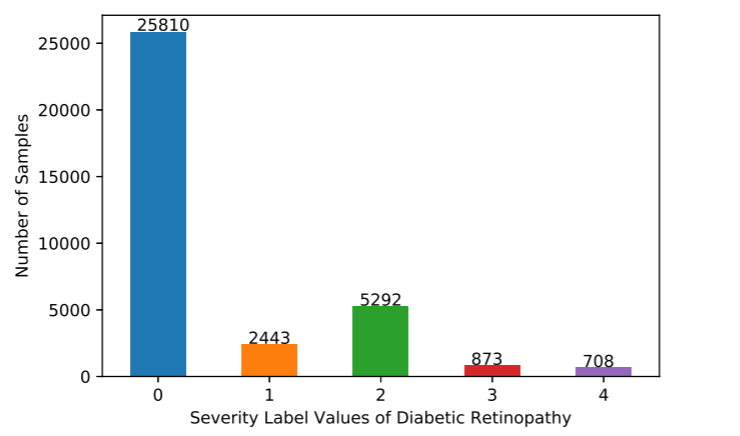


Figure 13: Distribution of Diabetic Retinopathy label values in Kaggle train

In the 80s the most used activation function was the sigmoid, a smooth non-linear function, that is continuously differentiable. Despite its interesting properties, it has a very important concern, called gradient vanishing problem [11]. Sigmoid first derivative becomes flat not far from the origin, affecting network loss optimization due to near-to-zero gradients. In [12] Reclus were used for improving Restricted Boltzmann Machines (RBMs), approximating stepped sigmoid units with Reclus. In (Godot, Bores, and Benjie, 2011) the authors compared the performance of Sigmoid, Tan and Relax arriving to the conclusion that despite Sigmoid being more plausible biologically, Tanah and relax were more suitable to be used as activation function for training multi-layer perceptron. Relook networks have better performance in general, despite its non-differentiability at zero and its hard non-linearity. Furthermore, relook networks lead to sparse representations, being beneficial, both because information is represented in a more robust manner and because it leads to significant computational efficiency. Moreover, the simplicity of the function and its derivative reduces calculation time, being of significant importance when working with big networks. The constant value of the gradient, helps avoiding the gradient vanishing problem, allowing the design of deeper networks. For then on, relax has become the default activation function for deep learning. Many other activation functions have been published, like Leaky relay,

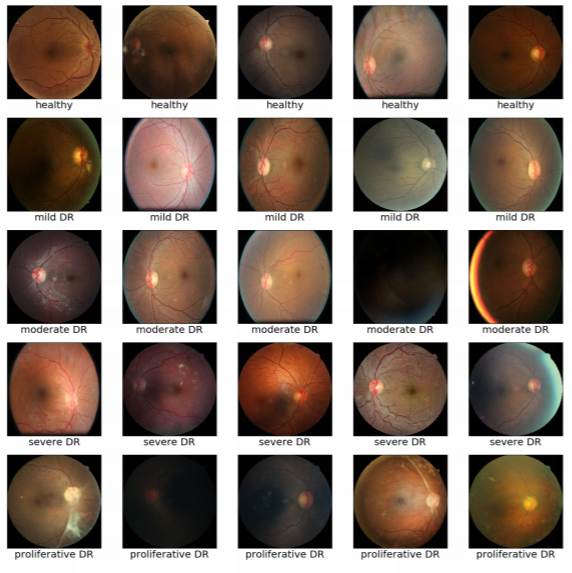
1. Cropping, random uniform

Figure 14 Random sample images from the Kaggle dataset

2. Rotation (0\_ to 360\_), random uniform

3. Mirroring ( X and Y axes), random uniform

4. Brightness correction, random gaussian (*s* = 0.1)

5. Contrast correction, random Gaussian[34] (*s* = 0.1)

Offline segementation was performed in two stages, in first stage images were oversampled to create a kind of a balance between the numbers of images per class. Then a series of transforms described above were applied on randomly selected images from the minority classes based on the defined random numbers. The transforms make sure that after oversampling the images are not repetitive but are different from each other even maybe their sources were same making the final prediction invariant to rotation, translation, intensity and contrast over the training set. Also to make the convergence faster for DenseNET a mean normalization for pixel values was performed per image.

## 4.6 Model

A very popular upgradation of Resnet is DenseNET 201 [35], which is considered for classification of diabetic retinopathy. It consists of series of convolutional layers, activation blocks and max-pooling dimensional reduction blocks. Its end is attached to a fully connected layer on which a softmax function is applied which gives probability estimation of every class as a output. Architecture has been fully described in Appendix. Some parameters of DenseNET are activation function, optimization and regularization methods to be used, number of layers, number and size of classification layers, number of filters per convolution layer, size of the convolution, input size, number and size of classification layers. There donot exists any unique solution to this problem as there are very high number of parameters are involved. There can be different condition co-existing to give the same outputs.

## 4.7 Training and Testing Procedure

Since it’s a multi-class classification problem, for optimization in the learning stage a log-loss function is used. Original training set was split into 2 randomly selected subsets with 80% data and 20% data respectively. The later smaller set was used to cross validate sets of obtained hyper parameter. Leaky ReLU[37] was selected as the activation function. To compensate the vanishing gradient problem, which occurs very frequently in deep neural networks, batch normalization was applied in all the layers, which reduced the internal covariance shift and regularization. As an additional regularizer dropout[36] with probability = 0.5 was introduced. There was no effect on using L2-regularization. To randomly initialize the weights Kaiming and Hee approach was used. Biases were initialized to zero.

### 4.5.1 Effective Learning Rate Identification Results

To identify the correct learning rate different values were tried out, results of which are as follows in form of *figure 12*:

In the 80s the most used activation function was the sigmoid, a smooth non-linear function, that is continuously differentiable. Despite its interesting properties, it has a very important concern, called gradient vanishing problem [11]. Sigmoid first derivative becomes flat not far from the origin, affecting network loss optimization due to near-to-zero gradients. In [12] Reclus were used for improving Restricted Boltzmann Machines (RBMs), approximating stepped sigmoid units with Reclus. In (Godot, Bores, and Benjie, 2011) the authors compared the performance of Sigmoid, Tan and Relax arriving to the conclusion that despite Sigmoid being more plausible biologically, Tanah and relax were more suitable to be used as activation function for training multi-layer perceptron. Relook networks have better performance in general, despite its non-differentiability at zero and its hard non-linearity. Furthermore, relook networks lead to sparse representations, being beneficial, both because information is represented in a more robust manner and because it leads to significant computational efficiency. Moreover, the simplicity of the function and its derivative reduces calculation time, being of significant importance when working with big networks. The constant value of the gradient, helps avoiding the gradient vanishing problem, allowing the design of deeper networks. For then on, relax has become the default activation function for deep learning. Many other activation functions have been published, like Leaky relay,



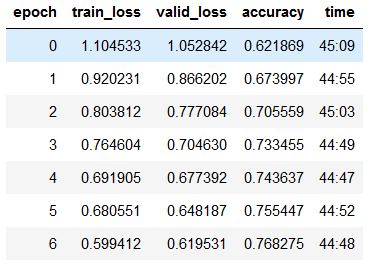
Figure 11 Image describing performance for different learning rate

As you can clearly see that corresponding to learning rate = 1e-02 the Loss is minimized so it is a good starting point while training the model.

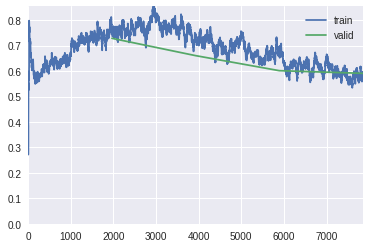
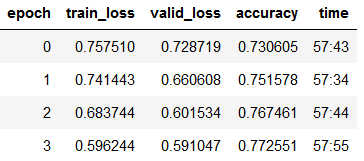
In the 80s the most used activation function was the sigmoid, a smooth non-linear function, that is continuously differentiable. Despite its interesting properties, it has a very important concern, called gradient vanishing problem [11]. Sigmoid first derivative becomes flat not far from the origin, affecting network loss optimization due to near-to-zero gradients. In [12] Reclus were used for improving Restricted Boltzmann Machines (RBMs), approximating stepped sigmoid units with Reclus. In (Godot, Bores, and Benjie, 2011) the authors compared the performance of Sigmoid, Tan and Relax arriving to the conclusion that despite Sigmoid being more plausible biologically, Tanah and relax were more suitable to be used as activation function for training multi-layer perceptron. Relook networks have better performance in general, despite its non-differentiability at zero and its hard non-linearity. Furthermore, relook networks lead to sparse representations, being beneficial, both because information is represented in a more robust manner and because it leads to significant computational efficiency. Moreover, the simplicity of the function and its derivative reduces calculation time, being of significant importance when working with big networks. The constant value of the gradient, helps avoiding the gradient vanishing problem, allowing the design of deeper networks. For then on, relax has become the default activation function for deep learning. Many other activation functions have been published, like Leaky relay,

In the 80s the most used activation function was the sigmoid, a smooth non-linear function, that is continuously differentiable. Despite its interesting properties, it has a very important concern, called gradient vanishing problem [11]. Sigmoid first derivative becomes flat not far from the origin, affecting network loss optimization due to near-to-zero gradients. In [12] Reclus were used for improving Restricted Boltzmann Machines (RBMs), approximating stepped sigmoid units with Reclus. In (Godot, Bores, and Benjie, 2011) the authors compared the performance of Sigmoid, Tan and Relax arriving to the conclusion that despite Sigmoid being more plausible biologically, Tanah and relax were more suitable to be used as activation function for training multi-layer perceptron. Relook networks have better performance in general, despite its non-differentiability at zero and its hard non-linearity. Furthermore, relook networks lead to sparse representations, being beneficial, both because information is represented in a more robust manner and because it leads to significant computational efficiency. Moreover, the simplicity of the function and its derivative reduces calculation time, being of significant importance when working with big networks. The constant value of the gradient, helps avoiding the gradient vanishing problem, allowing the design of deeper networks. For then on, relax has become the default activation function for deep learning. Many other activation functions have been published, like Leaky relay,

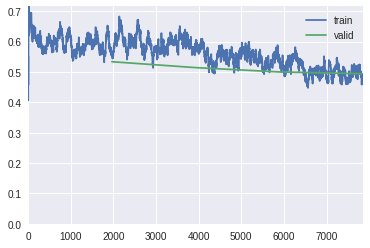
#### Training with Learning Rate = 1e-03 and weight decay 1e-06



#### Training with Learning Rate = max learning rate, /3, /9 and weight decay 1e-06



#### Training with Learning Rate = max learning rate/3, /27, /9 and weight decay 1e-06



## 4.6 Test Results - Confusion Matrix

There are several machine learning algorithms but over the years of development general characteristics of learning and performance evaluation criteria have been established. The general idea is to divide the currently present data into training data, validation data and test data parts. The training and validation parts are used in the learning of the algorithm and construction of the model while the test dataset is utilized to evaluate the model performance on novel data that it has never observed beforehand in its learning. There are various performance measures of the model depending on whether it is a classification or regression problem. These measures include mean square error, absolute error and cross validation techniques for regression problems and prediction error, precision, recall, F-measure (weighted average of precision and recall of the model), confusion matrix, ROC (sensitivity vs specificity curves) and AUC curves for classification problems. The initial work by C Shang in 1996 involved developing an absolute error based algorithm for evaluating communication channel equalisers adapting single-layer perceptron models. Research by Paolo Sonego in 2008 shows the application of ROC curve estimates to evaluate performance in biomedical application models and classification of biological sequences and 3D structures. Recently, precision, recall and F-measures have found their application in evaluating models in text classification for sentiment analysis and social media analytics.

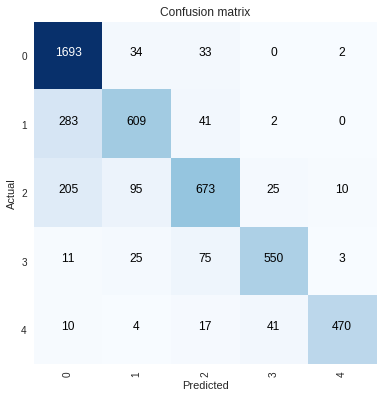


Figure 15 Test Results confusion Matrix

Table 2 Demonstrating Final Result of DenseNET Classification

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class/Metric (%)** | **Precision** | **Sensitivity** | **F1-Score** | **Accuracy** |
| **Normal** | 96.08 | 76.89 | 85.00 | 88.23 |
| **Mild** | 65.13 | 79.40 | 72.00 | 90.24 |
| **Moderate** | 66.77 | 80.22 | 73.00 | 90.00 |
| **Severe** | 82.83 | 89.00 | 86.00 | 96.29 |
| **Very Severe** | 86.72 | 96.90 | 92.00 | 98.23 |
| ***Total*** | 80.00 | 85.00 | 81.6 | **92.6** |

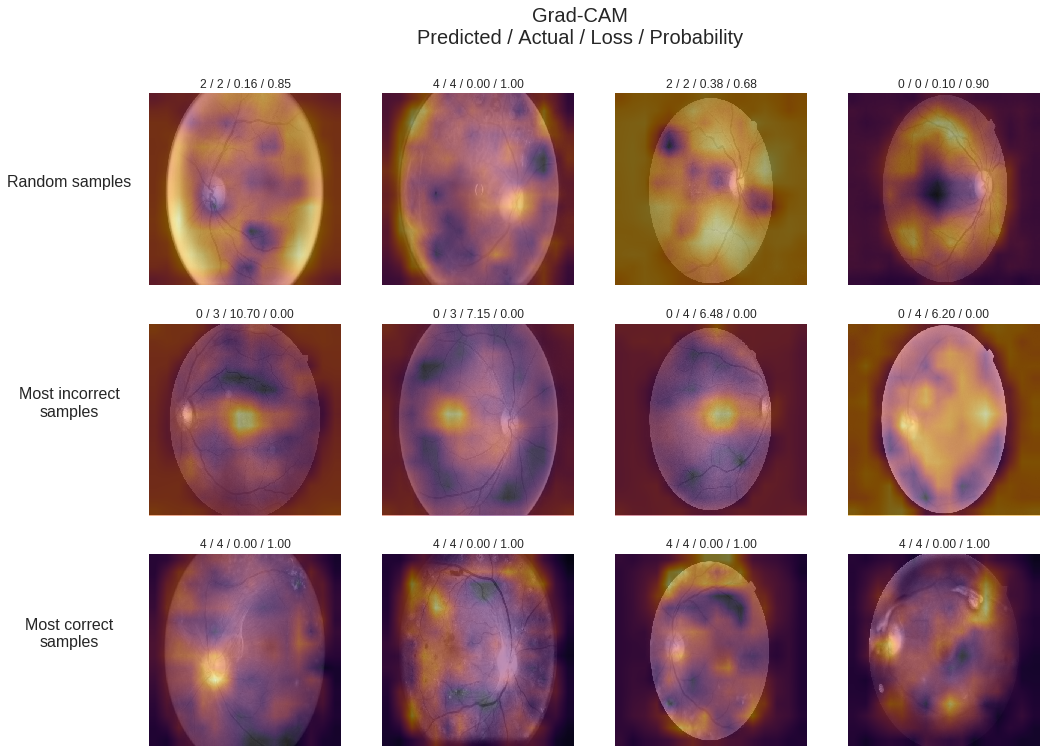


Figure 12 Demonstration of results for Random, Incorrect and Correctly Predicted

# Chapter 5.

# Summary and Prospectives

Medical diagnosis is heavily supported by Medical Imaging. The analysis of medical images requires highly specialized expertise that is provided by specialized doctors. The advent of advanced machine learning techniques, like deep learning, is facilitating the design of high performance automatic classifiers in a broad range of applications, also for medical diagnosis. The purpose of this thesis is the exploration of new automatic diagnostic methods for medical diagnosis, concretely for diabetic retinopathy disease grading. As stated before in this work, Diabetic Retinopathy is one of the main causes of blindness in the world. Early detection can reduce disease progression and consequently also the incidence of blindness. The diagnostic of DR is done primarily by retina fundus image analysis, that is done by ophthalmologists specifically trained for this purpose.   
  
Automatic diagnostic systems for DIABETIC RETINOPATHY can reduce dramatically not only the costs associated to diagnostic, but also the probability of developing blindness in the general population. For this purpose we use supervised deep learning techniques. Given a class differentiation based on objective properties present in data, these parameterized models are able to learn the statistical regularities that have to be taken into account to separate the images. In this thesis, convolutional neural networks are used, which are efficient neural networks designed for exploiting local high correlations present in images.   
Machine learning methods in general and deep learning in particular are models that learn from data. Thus, a key factor for the success is having a statistically representative dataset of the population from which we want to predict a concrete property, in this case, a disease class. For this purpose it is required to have a labeled dataset with enough samples (that are of the order of magnitude of thousands elements per class) in order to create models with good generalization. For this purpose, we use a public dataset of EyePACS.

After the work done in this thesis, we can conclude that designed models can be used successfully as a high confidence diagnostic tool that can help to reduce costs in diagnostic and also to reduce the incidence of the DIABETIC RETINOPATHY disease in the general population.

Future research lines can be focused on the next different directions, that we explain below:

1. *Reinforcement Learning*: Adding to the models the possibility of enhancing its performance, designing online learning methods that allow continuous learning of networks from the corrections done by ophthalmologists on inference time. Hospital Universitari Sant Joan de Reus wants to use the classification model developed in this thesis in its daily work. This needs a previous process of integration of computer systems as well as the permission of Catalan Health Care authorities. If it is finally done, learning from online work could be very interesting.
2. *Increase the number of classes to predict*: One of the lines to explore is increasing the number of properties to predict of the original model. Instead of only predicting diabetic retinopathy class, other classes can also be included.
3. Use of Interpretation Model in other applications: The designed interpretation model is domain independent, therefore, it would be interesting the study of its application in other domains, ie. for interpretation of other medical imaging classification tasks like for example, radiology, ultrasound, or other related imaging applications.

# References

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 110, 3 (June 2008), 346-359. DOI=http://dx.doi.org/10.1016/j.cviu.2007.09.014

[2] Shah AR, Gardner TW. Diabetic retinopathy: research to clinical practice. Clin Diabetes Endocrinol. 2017;3:9. Published 2017 Oct 19. doi:10.1186/s40842-017-0047-y

[3] Gardner TW, Chew EY. Future opportunities in diabetic retinopathy research. Curr Opin Endocrinol Diabetes Obes. 2016;23(2):91–96. doi:10.1097/MED.0000000000000238

[4] Lee R, Wong TY, Sabanayagam C. A two-stage filter for removing salt-and-pepper noise using noise detector based on characteristic difference parameter and adaptive directional mean filter. Eye Vis (Lond). 2015;2:17. Published 2015 Sep 30. doi:10.1186/s40662-015-0026-2

[5] Ankita Gupta, Rita Chhikara, Diabetic Retinopathy: Present and Past, Procedia Computer Science, Volume 132, 2018, Pages 1432-1440, ISSN 1877-0509

[6] S. izza Rufaida and M. I. Fanany, “Residual convolutional neural network for diabetic retinopathy,” in Proceedings of IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2017, pp. 367–374.

[7] D. Zhang et al., “Review of Video and Image Defogging Algorithms and Related Studies on Image Restoration and Enhancement,” 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 (SmartWorlda/SCALCOM/UICa/ATC/ CBDCom/IOP/SCI). IEEE, 2017.

[8] M. Alban and T. Gilligan, “Empirical Evaluation of Rectified Activations in Convolutional Network.” Report of Standford Education, 2016.

[9] Z. Wang et al., “A Comprehensive Study of Alzheimer's Disease Classification Using Convolutional Neural Networks,” in Proceeding of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2017, pp. 267–275

[10] Xavier Glorot, Yoshua Bengio ; Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9:249-256, 2010.

[11] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10), Johannes Fürnkranz and Thorsten Joachims (Eds.). Omnipress, USA, 807-814.

[12] Xu, Bing, Naiyan Wang, Tianqi Chen and Mu Li. “Empirical Evaluation of Rectified Activations in Convolutional Network.” ArXiv abs/1505.00853 (2015): n. pag.

[13] Websites for data : [http://www.ces.clemson.edu/~ahoover/stare/](https://www.researchgate.net/deref/http%3A%2F%2Fwww.ces.clemson.edu%2F~ahoover%2Fstare%2F) , https://www.researchgate.net/deref/http%3A%2F%2Fwww5.cs.fau.de%2Fresearch%2Fdata%2Ffundus-images%2F

[14] Websites for data : <https://www.kaggle.com/c/diabetic-retinopathy-detection>

[15] Websites for data : <http://www.adcis.net/en/third-party/e-ophtha/>

[16] Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J Big Data* **6,**27 (2019) doi:10.1186/s40537-019-0192-5

[17] Alberto Bietti and Julien Mairal. 2017. Invariance and stability of deep convolutional representations. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Ulrike von Luxburg, Isabelle Guyon, Samy Bengio, Hanna Wallach, and Rob Fergus (Eds.). Curran Associates Inc., USA, 6211-6221.

[18] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6,**60 (2019) doi:10.1186/s40537-019-0197-0

[19] Hussain Z, Gimenez F, Yi D, Rubin D. Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. *AMIA Annu Symp Proc*. 2018;2017:979–984. Published 2018 Apr 16.

[20] Ma H, Nie Y (2018) A two-stage filter for removing salt-and-pepper noise using noise detector based on characteristic difference parameter and adaptive directional mean filter. PLoS ONE 13(10): e0205736

[21] Yamashita, R., Nishio, M., Do, R.K.G. et al. Insights Imaging (2018) 9: 611. <https://doi.org/10.1007/s13244-018-0639-9>

[22] Chen H, Zhang Y, Zhang W et al (2017) Low-dose CT via convolutional neural network. Biomed Opt Express 8:679–694

[23] Martin Gjoreski, Stefan Kalabakov, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. 2019. Cross-dataset deep transfer learning for activity recognition. In Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers

[24] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[25] Y. Zhu and S. Newsam, "DenseNet for dense flow," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 790-794.

[26] Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. arXiv, 2015; arXiv:1511.07122

[27] Li Y, Xie X, Shen L, Liu S. Reverse active learning based atrous DenseNet for pathological image classification. BMC Bioinformatics. 2019;20(1):445. Published 2019 Aug 28.

[28] EyePacs Dataset <https://www.medicmind.tech/resources-2/>

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, 1097-1105.

[30] Pedro Ballester and Ricardo Matsumura Araujo. 2016. On the performance of GoogLeNet and AlexNet applied to sketches. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16). AAAI Press 1124-1128.

[31] [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) [cs.CV]

[32] Cyril Goutte and Eric Gaussier. 2005. A probabilistic interpretation of precision, recall and *F*-score, with implication for evaluation. In Proceedings of the 27th European conference on Advances in Information Retrieval Research (ECIR'05), David E. Losada and Juan M. Fernández-Luna (Eds.). Springer-Verlag, Berlin, Heidelberg, 345-359. DOI=http://dx.doi.org/10.1007/978-3-540-31865-1\_25

[33] Yong Xu, Jie Wen, Lunke Fei, Zheng Zhang, "Review of Video and Image Defogging Algorithms and Related Studies on Image Restoration and Enhancement", Access IEEE, vol. 4, pp. 165-188, 2016.

[34] S. Pei and T. Lee, "Nighttime haze removal using color transfer pre-processing and Dark Channel Prior," 2012 19th IEEE International Conference on Image Processing, Orlando, FL, 2012, pp. 957-960.

[35] Guan, Ziqiang, Ritesh Kumar, Yi Ren Fung, Yeahuay Wu and Madalina Fiterau. “A Comprehensive Study of Alzheimer's Disease Classification Using Convolutional Neural Networks.” ArXiv abs/1904.07950 (2019): n. pag.

[36] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1 (January 2014), 1929-1958.

[37] Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-rectified-linear-unit-ReLU-the-leaky-ReLU-LReLU-a-01-the-shifted-ReLUs\_fig1\_284579051 [accessed 21 Dec, 2019]

[38] A. G. Howard et al., “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.