university of groningen

faculty of science and engineering

Automatic classification of macular pathologies using deep learning techniques

Sina Rouzbahani (s4121589)

University of Groningen

Master's Thesis

To fulfill the requirements for the degree of Master of Science in
Computing Science: Data Science and Systems Complexity (DSSC)
at the University of Groningen under the supervision of

Dr. Steffen Frey (Scientific Visualization and Computer Graphics, University of Groningen)

Prof. dr. Jiri Kosinka (Scientific Visualization and Computer Graphics, University of Groningen)

Dr. Ulises Olivares Pinto (Distributed Computing, Optimization, 3D Computer Graphics, The National Autonomous University of Mexico (UNAM))

Sina Rouzbahani (s4121589)

October 31, 2022

Contents

Acronyms

AI Artificial Intelligence. 15

AMD Age-Related Macular Degeneration. 13, 55

ANNs Artificial neural networks. 15, 19, 20

BRB Blood-Retinal Barrier. 8

CE Cross entropy. 20, 27, 32

CNN Convolutional Neural Network. 7, 13, 15–17, 33

CNV Choroidal Neo-Vascularization. 13, 26, 55

DM Diabetic Maculopathy. 8, 57

DME Diabetic Macular Edema. 7–14, 20–28, 38, 39, 44–49, 51–55, 57, 58

DR Diabetic Retinopathy. 7, 8, 12, 20, 57

FA Fluorescein Angiography. 8

HoG Histogram of Oriented Gradients. 14

Kernel-SVM Kernel Support Vector Machine. 14

LBP Local Binary Pattern. 14

Linear-SVM Linear Support Vector Machine. 14

MSE mean squared error. 20

OCT Optical Coherence Tomography. 7–10, 12, 14, 21–24, 26, 28–30, 38, 40, 42, 44–49, 52–55, 57, 58

PCA Principal Component Analysis. 14

RF Random Forest. 14

SD-OCT spectral-domain optical coherence tomography. 8, 12, 14, 55, 57

SVM Support Vector Machine. 12

TNR True Negative Rate. 37, 54

TPR True Positive Rate. 37, 54

Acknowledgments

While working on this master's thesis, I received a lot of help, support, and assistance from many people.

I would like to express my deepest appreciation to my supervisors, Dr. Steffen Frey, Prof. dr. Jiří Kosinka, and Dr. Ulises Olivares Pinto, for their invaluable patience, feedback, and support that helped me to improve my level of work.

I am also thankful to the Scientific Visualization and Computer Graphics (SVCG) group and the University of Groningen (RUG) for providing me with the necessary tools and assistance. The National Autonomous University of Mexico (UNAM) for providing a dataset of real patients for this project.

Abstract

Diabetes is one of the most common diseases worldwide, affecting 9.3% of people during their lifetime [32]. Diabetic Retinopathy (DR) is a complication of diabetes caused by elevated blood sugar levels. It damages the retina, which can cause blindness if left undiagnosed and untreated. Partial or total vision loss is also observed in some patients [20]. Around 34.6% of patients suffering from diabetes experience DR [13]. This project focuses on learning-supported visual pathology analysis in Diabetic Macular Edema (DME), which is the most common reason for vision loss in DR patients [9]. Deep learning and Optical Coherence Tomography (OCT) technologies have brought trustworthiness and reliability in detecting retinal diseases. They have significantly enhanced the diagnostic performance of eye posterior segment disease [36].

Similar works have been done using a large dataset with labeled OCT images for retinal disease classification [26]. Models created by Kermany et al. [26] have depicted high accuracy, but their specificity and sensitivity display lower confidence than expected. Panozzo et al. [37] proposed a methodology using seven morphological features that increase the specificity and sensitivity of diabetic maculopathy diagnosis. In this project, deep learning methods are used to build Convolutional Neural Network (CNN) models that classify these seven morphological features plus an additional feature in a supervised learning setup. The trained models are later evaluated, and their metrics are studied in detail.

For this project, a dataset was prepared by the National Autonomous University of Mexico (UNAM), and clinical evaluation was manually done for each OCT image. As a result, ground truth labels for each OCT image are provided. Supervised learning is adopted using the ground truth labels to train the classification models to classify the features and predict the occurrence of DME. In this thesis, eight different morphological features are studied: foveal Thickness range, Macular Volume, presence of intraretinal Cysts, state of the Ellipsoid Zone and External Limiting Membrane, Disorganization of the Inner Retinal Layers, number of Hyperreflective foci, subfoveal Fluid, and Vitreoretinal relationship.

This project focuses on utilizing deep learning and machine learning algorithms to investigate the relationship between eight morphological features. The trained models are used to classify and detect the presence of individual and combined features against the novel data (test set). The results are further evaluated to establish the accuracy and reliability of the model to diagnose DME in OCT images with the usage of transfer learning. The results of the experiment illustrate the most substantial relationship between Thickening (T) and Macular Volume (MV). The results indicate that Cysts (C) is the prominent singleton feature that positively reinforces the accuracy of the trained DME detection model. Similarly, the combination of Thickening (T) and Macular Volume (MV) is the most prominent two-feature combination that positively reinforces the accuracy of the model. The average sensitivity and specificity for the prediction of DME is approximately 96% for both using single features and combinations of features.

***Keywords*—** Deep learning, Machine learning, Diabetic Retinopathy (DR), Diabetic Macular Edema (DME), Optical coherence tomography (OCT)

1 Introduction

Diabetes is one of the most prevalent chronic diseases, affecting around 415 million people globally. It is projected to be risen up to 750 million by 2030 [37]. DME is the most common cause of visual impairment in patients with diabetes mellitus. Roughly around 75,000 new patients in the United States show symptoms of diabetes mellitus annually [4]. The article published by the International Diabetes Federation shows that more than 21 million people are affected by DME globally and that approximately one in 14 people with diabetes has some degree of DME [51].

One of the consequences of diabetes on the human eyes is DR, which is caused by damage to the blood vessels of the light-sensitive tissue at the retina. DME is a progressive phase of DR and can result in permanent vision loss [38]. "The pathogenesis of DME is complex and multifactorial. It occurs mainly due to disruption of the Blood-Retinal Barrier (BRB), which leads to increased accumulation of fluid within the intraretinal layers of the macula [4]".

Currently, there are some techniques such as Fluorescein Angiography (FA) and spectral-domain optical coherence tomography (SD-OCT) for the assessment of Diabetic Maculopathy (DM). Specifically, SD-OCT presents both quantitive and qualitative information and provides high-resolution images in a non-invasive (not involving the introduction of instruments into the body) and repeatable way [37]. There are eight qualitative morphological parameters on SD-OCT images as follows [37]:

- Thickening (T)
- Macular Volume (MV)
- Size of intraretinal Cycts (C)
- State of Ellipsoid Zone (EZ) / External Limiting Membrane (ELM)
- The presence of disorganization of the inner retinal layers (DRIL)
- Hyperreflective foci (H)
- The presence or the absence of Subretinal Fluid (F)
- The vitreoretinal relationship (V)

The main focus of this project is instructed toward the learning-supported visual analysis of the DME pathology in the retina associated with diabetes. Different studies such as L. Giancardo et al. [10], I. Hecht et al. [31], and H. Y. Li et al. [54] have demonstrated the potential of using classical and deep learning algorithms to automatically extract the characteristics that determine the presence or absence of DME in the fundus and OCT images. However, this project focuses only on the OCT images as a specialized study.

This project has three main goals. The first goal is to classify and detect the eight mentioned morphological features separately and singularly on OCT images and study the relationship between them. The second goal is to simultaneously detect only two morphological features (combined features) on OCT images. The third goal is the detection of DME using the models trained on single features and various combinations of them on a new dataset of OCT images.

The data necessary for this project was not readily available. A team of medical experts was responsible for manually labeling the OCT images based on the morphological features observed. The labels are used as the ground truth while training the classification models to classify and identify the mentioned morphological features separately and their combinations. The models are then used to predict the occurrence of DME in novel data.

This thesis is structured as follows: Section 2 introduces related works and similar techniques in this field of study. Section 3 provides background information about the methodologies embraced in the project. The section further elaborates on the fundamentals of investigating the disease and medical images. Section 4 shows the approach used in this project and elaborates on consecutive steps to achieve the goals. Section 5 depicts the network architecture and its building blocks. The section further demonstrates the dataset used in the project in further detail and reveals some examples of the images used in this project. Section 6 addresses the hardware and software environment used in this project and represents the techniques and strategies used to implement the project. In addition, it represents the evaluation metrics used in this project to measure model performance. Section 7 addresses the results and analysis used from the experimentation and provides some insight into the obtained results. The section further discusses the results and makes a comparison with the related works. Section 8 concludes the thesis, and finally, Section 9 depicts the directions for future works in this field of study.

2 Related Work

2.1 OCT-based grading system

According to work done by Panozzo et al. [37], DME has been classified in several ways - according to its location, extent, or nature on OCT images. A grading protocol is proposed in their work to take seven morphological features (excluding Macular Volume) into account, mentioned in Section 1, and evaluate them. The grading protocol then classifies DME on OCT images in four distinct stages based on the seven mentioned morphological features. The proposed grading protocol system for classifying DME has been called TCED-HFV, in which each letter represents a morphological feature. The OCT images are graded based on the presence of each feature corresponding to its label. Then graded OCT images are individually categorized into one of the four different stages of DME as follows [37]:

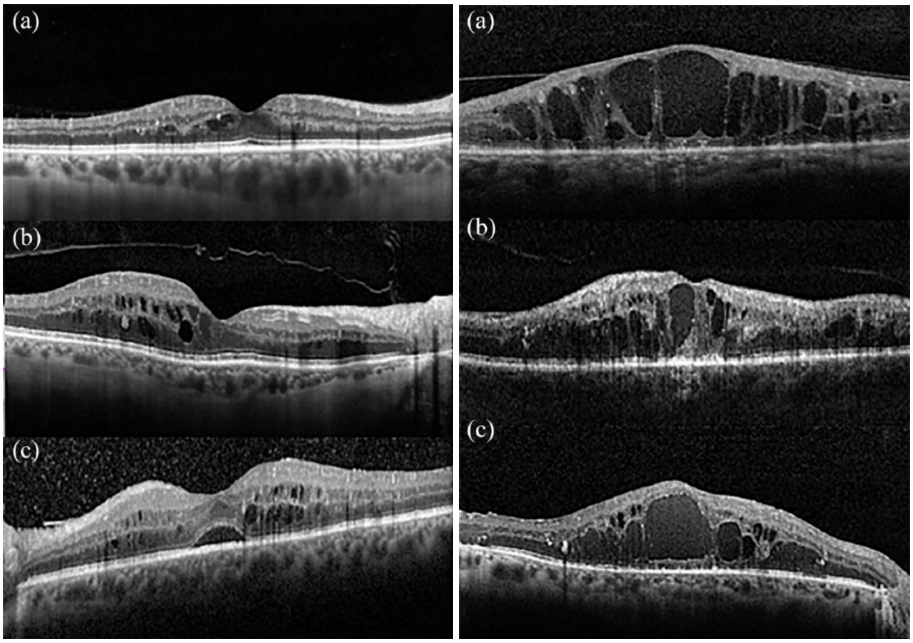
1. Early DME
2. Advanced DME
3. Severe DME
4. Atropathic maculopathy

Figures 1 and 2 show DME progression in the four stages separately. To better represent the TCED-HFV score in this thesis, a vector is defined consisting of seven values showing the label of each feature in order (T, C, E, D, H, F, V).

Figure 1a illustrates the early DME:
(a) Small cystoid spaces involving the temporal side of the fovea [37]. The TCED-HFV score in order is (1, 1, 0, 0, 0, 0, 0), in which each number shows the presence of each feature according to its label.
(b) Multiple perifoveal cystoid spaces in the outer nuclear layer, the outer plexiform layer, and the inner nuclear layer, with mild thickening of the temporal side of the macula [37]. The TCED-HFV score is (1, 2, 0, 0, 1, 0, 1).
(c) The retinal profile is preserved, and cystoid spaces in the outer plexiform and inner nuclear layers. The ellipsoid zone is not gradable due to subfoveal fluid, but the external limiting membrane is normal [37]. The TCED-HFV score is (1, 2, 0, 0, 1, 1, 4).

Figure 1b illustrates the severe DME:
(a) Multiple central coalescent macrocysts in the outer nuclear layer, the outer plexiform layer, and the inner nuclear layer with disorganization of the inner retinal layers (DRIL) [37]. The TCED-HFV score is (2, 3, 2, 1, 0, 0, 1), in which each number shows the presence of each feature according to its label.
(b) The central macrocyst is surrounded by large cystoid spaces involving the outer nuclear layer, the outer plexiform layer, and the inner nuclear layer [37]. The TCED-HFV score is (2, 3, 2, 1, 0, 0, 1).
(c) Central macrocyst and multiple large cysts surrounded by a few hyperreflective foci. The external liming membrane and the ellipsoid zone are not discernible subfoveal [37]. The TCED-HFV score is (2, 3, 2, 0, 1, 0, 0).

Figure 2a shows advanced DME:
(a) Cystoid spaces in the outer nuclear layer, the outer plexiform layer, and the inner nuclear layer, with thickening of the retina and central macrocyst [37]. The TCED-HFV score is (2, 3, 1, 0, 1, 0, 0),



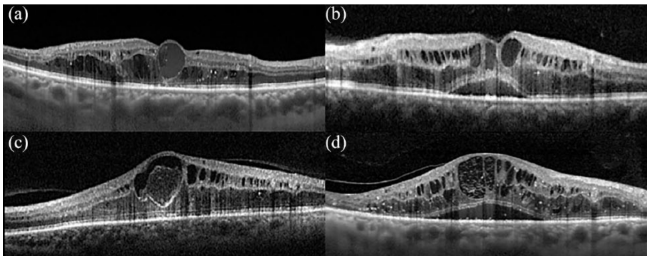
(a) Early DME (b) Severe DME

Figure 1: Two different stages of DME [37].

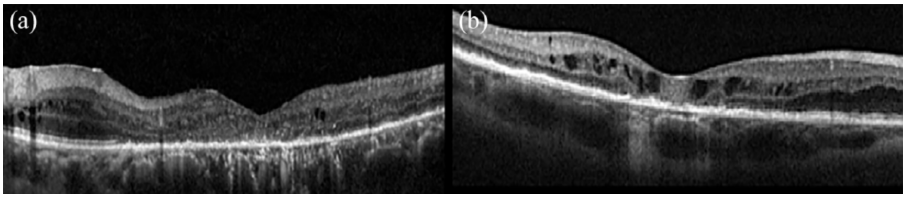
in which each number shows the presence of each feature according to its label.
(b) Intermediate cystoid spaces in the macula and the ellipsoid zone are not gradable, but the external limiting membrane is disrupted subfoveal [37]. The TCED-HFV score is (2, 2, 1, 0, 0, 1, 0).
(c) A large pseudocyst in the fovea with cystoid spaces in the parafoveal area. The ellipsoid zone and the external liming membrane are damaged subfoveal [37]. The TCED-HFV score is (2, 3, 1, 1, 0, 0, 1).
(d) Large cystoid spaces in the outer nuclear layer, the outer plexiform layer, and the inner nuclear layer with a shallow subfoveal detachment. Diffuse hyperreflective foci, non-gradable ellipsoid zone, but discontinuous external liming membrane [37]. The TCED-HFV score is (2, 3, 1, 0, 1, 1, 1).

Figure 2b shows Atrophic diabetic maculopathy:
(a) Central retinal thinning with disorganization of the inner retinal layers (DRIL). The external liming membrane and the ellipsoid zone are not discernible subfoveally, and the retinal pigment epithelium is atrophic [37]. The TCED-HFV score is (0, 1, 2, 1, 1, 0, 0).
(b) Central retinal thinning with DRIL. The external limiting membrane and the ellipsoid zone are not distinguishable subfoveally, and the retinal pigment epithelium is irregular and focally atrophic [37]. The TCED-HFV score is (0, 1, 2, 1, 0, 0, 0).

The work done by Pannozo et al. [37] proposes a method to classify DME solely on selected morphological features, which was a new way to classify DME in comparison to their prior work. Unfortunately, the study did not leverage machine learning methods' capabilities to classify and predict DME diseases.



(a) Advanced DME



(b) Atrophic DME

Figure 2: Two different stages of DME [37].

2.2 Deep learning in ophthalmology

According to A. D. Moraru et al. [36], the collaboration of deep learning and OCT technologies has brought trustworthiness in detecting retinal diseases. It has enhanced the diagnostic performance of eye posterior segment diseases. With the increase in population age and obesity, the occurrence of DR also increases, leading to many cases of vision loss. There are several means of screening and management in DR using machine learning classes and techniques to detect and classify different types of DR from medical images. Support Vector Machine (SVM), multiple layer perceptron classes, and radial basis function neural networks, can perform the same analysis as ophthalmologists for the retinal images [36].

The methodologies of medical image analysis are revolutionized by using deep learning algorithms using OCT and fundus images. The potential of AI in ophthalmology improves patient access to clinical diagnosis and reduces healthcare costs [36].

2.3 OpticNet architecture

S. Amit Kamran et al. [30] represents a highly accurate automated system to diagnose DR and other retinal diseases. Achieving an adequate diagnosis is feasible by using SD-OCT techniques, which show the morphology of retinal layers. DR resembles another wide variety of retinal diseases; it is not uncommon for misclassifications to occur during the diagnosis. A successful differentiation between various degeneration of retinal layers and their underlying causes has been achieved by S. Amit Kamran et al. [30], thanks to their novel convolution neural network architecture with a flawless accuracy of 99.8% and 100% for two separately available retinal SD-OCT datasets.

The retinal images are studied for five specific parameters - retinal thickness, augmentation of retinal thickening, macular volume, retinal morphology, and vitreoretinal relationship. These features act as the foundation for identifying the growth of macular density in the retinal layer to detect DME [30]. The CNN architecture proposed by S. Amit Kamran et al. [30] is shown below in figure 3. It depicts

three critical sections: a new residual unit subsuming Atrous Separable Convolution, a novel building block, and a mechanism to prevent gradient degradation. The mentioned architecture does not require any pre-trained weights, and it eases the training and deployment time of the model by many folds [30].

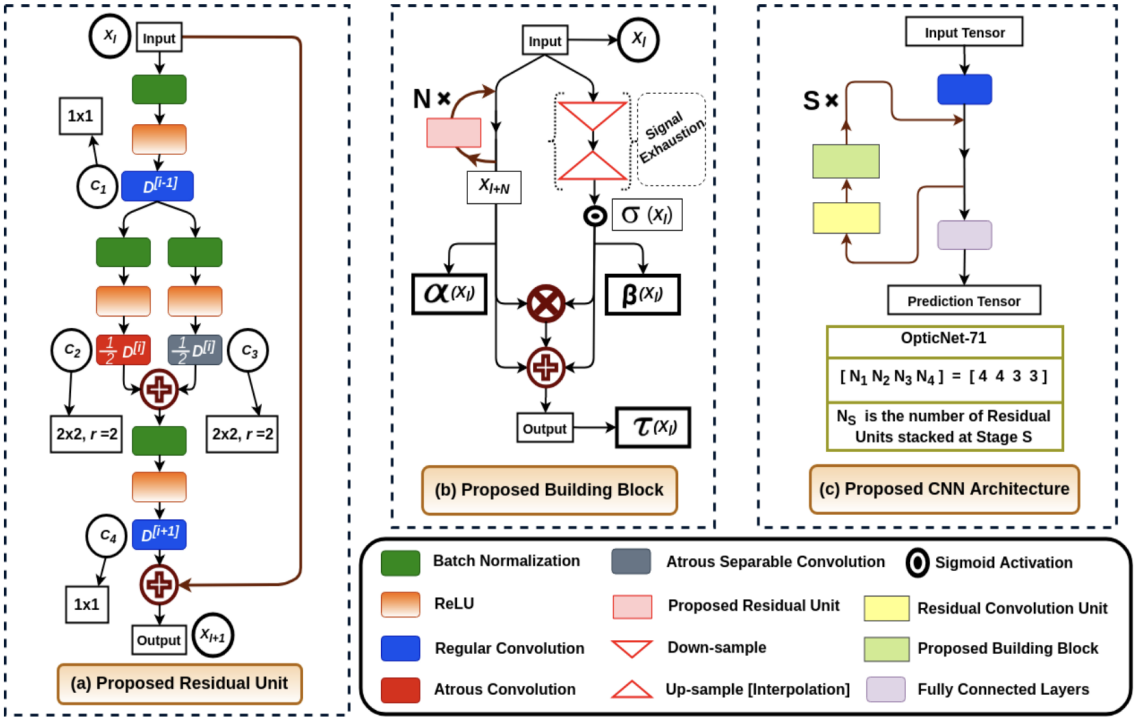


Figure 3: An illustration of the OpticNet architecture proposed by S. Amit Kamran et al. [30], which depicts three different sections as follows:
(a) illustrates how the proposed Residual Learning Unit enhances feature learning capabilities.
(b) shows how the mechanism handles gradient degradation.
(c) describes the whole CNN architecture.

The model has been benchmarked against two popular datasets of different sizes. The first dataset is OCT2017 [26] includes 84,484 images in which there are four distinct categories of four retinal conditions such as normal healthy retina, Drusen, Choroidal Neo-Vascularization (CNV), and DME. The other dataset, Srinivasan2014 [15], contains 3,231 images in which there are three classes, and the goal is to classify normal healthy specimens of the retina, Age-Related Macular Degeneration (AMD), and DME.

The models created by S. Amit Kamran et al. [30] yielded satisfactory results. In the OCT2017 dataset [26], the model achieved the most elevated test accuracy, 99.80%, among different existing models. It also showed impressive sensitivity and specificity metrics of 99.80% and 99.93%. In the Srinivasan2014 dataset [15], the model achieved a state-of-the-art result by scoring 100% accuracy, sensitivity, and specificity [30].

2.4 ML techniques for DME classification

The work done by K. Alsaih et al. [21] proposed an automatic supervised classification framework for SD-OCT images to identify DME versus normal images. DME can be classified based on the evaluation of features such as retinal thickening, hard exudates, intraretinal cystoid space formation, and subretinal fluid [21]. The features are evaluated individually and together as a set of different combinations of the same. A model is trained on a dataset consisting of 32 OCT volumes (16 DME and 16 normal cases). Each volume contains 128 B-scans OCT images, resulting in more than 3800 images.

The study by K. Alsaih et al. [21] investigated a generic pipeline including pre-processing, feature detection, feature representation techniques, and using classifiers to detect DME satisfactorily. OCT volumes are pre-processed through denoising, flattening, and cropping. Histogram of Oriented Gradients (HoG) and Local Binary Pattern (LBP) features are extracted from four levels using a multiresolution Gaussian image pyramid [1]. HoG is a feature descriptor widely used in computer vision tasks for object detection [2]. LBP is a straightforward and efficient texture operator that labels pixels of an image by thresholding the neighborhood of each pixel. Then it assesses the output as a binary number [1]. The LBP and the HoG features used a multiresolution image pyramid for feature representation, resulting in a high-dimensional feature space [21]. Principal Component Analysis (PCA) is further used to reduce the number of dimensions for visualization and feature selection. Three different classifiers are used in their work, such as Random Forest (RF), Linear Support Vector Machine (Linear-SVM), and Kernel Support Vector Machine (Kernel-SVM). The SD-OCT volume classification is performed based on the overall number of diseased scans detected in each volume, using the majority voting rule [21].

The classification done by K. Alsaih et al. [21] showed a promising result, but their study had some limitations. The assessment is done on a relatively small dataset. Their classification approaches investigated that the DME volume should consist of more than half of the slides having the presence of DME. Individual and combined features were assessed in their work, and the best classification performance with sensitivity and specificity of 87.5% was achieved. The results showed that their method is still not ready for clinical purposes because of an enormous false positive detection. Additional representation techniques and classifiers are evaluated and compared except comparing individual and combined features. The best results were obtained for LBP vectors while represented and classified using PCA and Linear-SVM [21].

[1] In an *image pyramid*, a signal or an image is subject to repeated smoothing and subsampling, which is a multiscale signal representation. Pyramid representation is an old way of scale-space representation and multiresolution analysis. Pyramid (image processing). (2022, June 27). In Wikipedia. https://en.wikipedia.org/wiki/Pyramid_(image_processing)

3 Fundamentals

This section demonstrates the essentials of machine learning models for the project and provides background information about the methodologies embraced. Also, it explains the fundamentals for the disease investigation and describes the information required for the medical images.

3.1 Neural networks

Artificial neural networks (ANNs) are computing systems inspired by biological neural networks of the brain. An ANNs is a group of linked nodes modeled by the neurons in a biological brain, and the connection between nodes plays the role of the synapse in a biological brain. It can transfer signals to the neurons in its vicinity. It is tagged as a supervised Deep-learning methodology in most worldly-wise AI systems [5]. Warren McCulloch and Walter Pitts first proposed the idea of merging multiple computing units into a network in 1943. In 1949, Hebb mentioned that a learning process could happen in the synaptic connection between neurons, and it is popularly known as Hebbian learning. After that, in 1957, Rosenblatt created the first neural network algorithm perceptron [28]. McClelland, Rumelhart, and the PDP research group introduced a back-propagation algorithm that allows multiple layers of perceptrons to be trained with feedback. This started the main idea of the hidden layers in the ANNs. The neural network is categorized into two principal types based on the dataset. It can be classified based on the learning method to train the network model, i.e., Supervised and Unsupervised learning [29][28].

3.1.1 Feed-forward neural networks

Neurons are the nonlinear components in the feed-forward neural network. They are placed in consecutive layers. The information flow in the feed-forward neural network proceeds through the input layer to the output layer through the hidden layers. Nodes in each layer are connected to the other layers, while there is no lateral connection between two nodes in one layer, and lateral feedback connections are impossible. The hidden layers are critical parameters in the network because they tweak their weights based on the feedback (learning) to drive the model toward a stable solution [14].

3.2 Convolutional neural networks

In 2012, the Convolutional Neural Network (CNN) was first proposed by Alex Krizhevsky [11]. CNN is defined as a solution to multiple problems in the field of computer vision, including pattern recognition and data inference problems. Also, medical imaging significantly benefits from it [42]. The central part of CNN is convolutional operations, and it can be demonstrated mathematically as:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]\ h[n-k] \tag{1}$$

where h is a set of filter coefficients of the system, and x is the input of the system [42].

CNN is popularly known for its image recognition and object detection potential. For example, it is able to recognize handwritten digits. It is a robust classification algorithm that can also handle high-dimensional datasets - CNN achieves this by utilizing convolution and pooling operations [27].

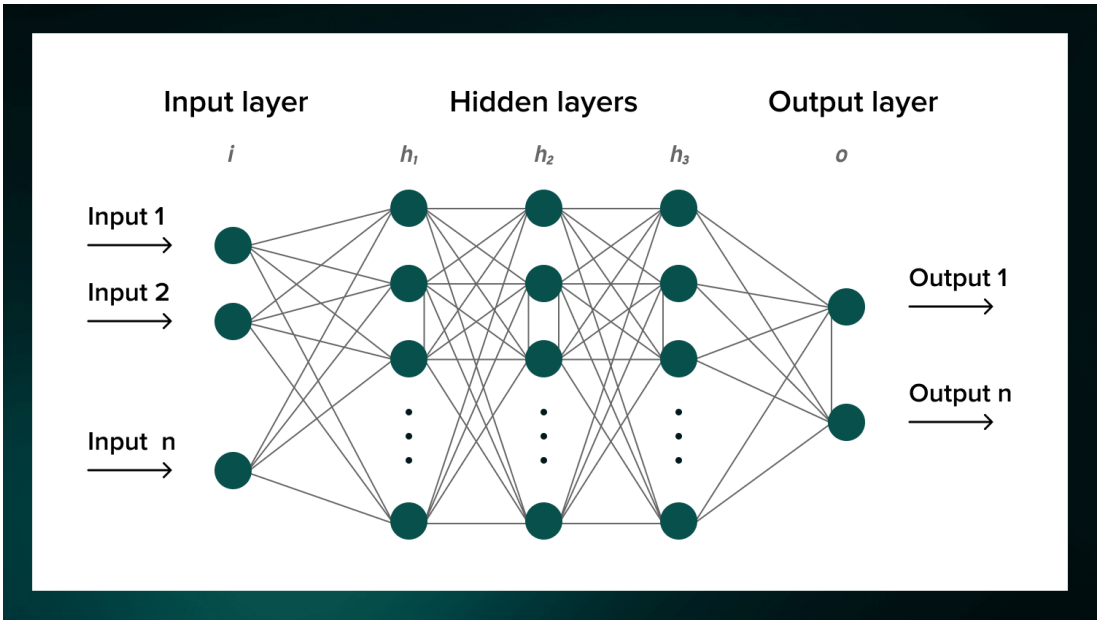


Figure 4: A fully connected convolutional neural network consisting of a sequence of connected layers that connect every neuron in one layer to every neuron in the other layer [59].

A CNN includes one input layer, hidden layers, and an output layer. The hidden layers consist of layers that perform convolutions which generally comprise a layer that performs a dot product of the convolution kernel with the layer's input matrix. In layman's terms, computers understand an image as arrays of pixels and depend on image resolutions. It can be shown as:

$$X = h \times w \times d \tag{2}$$

where X is the input image, h is the number of height pixels, w is the number of width pixels, and d is the dimension. For example, in an image of a 6 x 6 x 3 array of a matrix of RGB, 3 refers to RGB (Red, Green, Blue) values.

A 2-D convolutional layer is generally comprised of a combination of the following components: Kernel size, Stride, Padding, and number of in and out channels [42]. Figure 5 illustrates the different layers of the convolutional layers and shows that the layers in the CNN can be divided into two sections of the feature learning section in which the feature selection occurs, and the classification section in which dense layers are mainly located and provides the classified output.

3.2.1 Convolution layers

The convolution layer is the core and the most crucial part of the CNN model, which aims to decrease the image size for quicker computations of the weights and enhances its generalization. The first layer that gains features from the input image is the convolution layer, and it maintains the connection between pixels by learning the features of the image.

Figure 6 illustrates a convolution operation in which a kernel moves over the pixels of the image matrix and leads to an output value by performing a dot product. The table in figure 6 depicts the convolution of the 5x5 image matrix multiplied with a 3 x 3 filter matrix. The output will be the

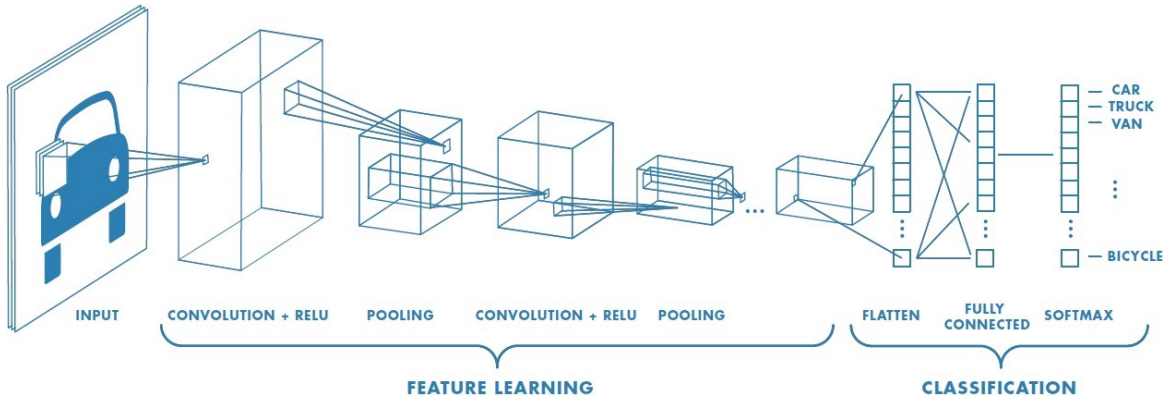
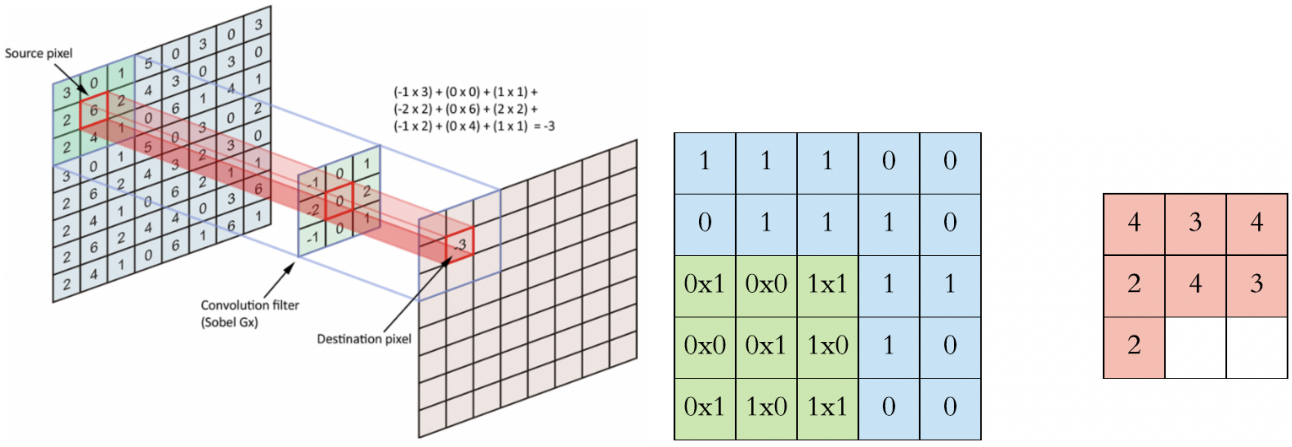


Figure 5: A complete flow of CNN to process an input image and classifies the objects based on values. The figure is adopted from [55].

convolution operation by sliding the 3x3 filter over the 5x5 input image. Invariably, the kernels (filters) must be smaller than the image size. An element-wise matrix multiplication for each pixel will be done, and the results will be sum. Convolution operations depicted in figure 6 is a 2D convolution using a 3x3 filter, while these convolutions are generally performed on 3D images. The images are a 3D matrix with height, width, and depth dimensions, where depth means three channels for RGB. There are two other attributes in CNN as Stride and Padding. The stride is the number of pixels that goes over the input matrix. For Stride = 1, the kernel shifts one pixel every step, and for Stride = 2, the filters shift two pixels per step time. Sometimes due to the filter and input image size, they do not fit perfectly, and then padding, which surrounds all the input edge pixels with zeros, can fix the issue [22]. The convolution operation can be shown mathematically as equation 3, where K is the kernel, X is the input image, and Y is the convoluted output. The original image X can be obtained by equation 4, considering the kernel is an invertible matrix.

$$Y = K \times X \tag{3}$$

$$X = K^{-1} \times Y \tag{4}$$



1	1	1	0	0
0	1	1	1	0
0x1	0x0	1x1	1	1
0x0	0x1	1x0	1	0
0x1	1x0	1x1	0	0

4	3	4
2	4	3
2		

Figure 6: Two different convolution operations with a kernel and the output [22].

3.2.2 Pooling

Pooling is used to perform downsampling that reduces the size of the spatial dimension of the convoluted feature or an image. Extracting only helpful information and discarding more inferior prominent details is the primary goal of the pooling layer [35].

The two most essential methods in pooling are average pooling and max pooling. In average pooling, the average value for rectangular pooling regions of a feature map can be calculated (figure 8b). Max-pooling is a technique of retaining only the max value within the pooled region. The pooling operation is typically performed after the convolution layer. The goal is to extract the most prominent feature within the pooled region, which is a characteristic of max-pooling [33][35].

Contrary to average pooling, Max pooling calculates the maximum value for rectangular pooling regions of a feature map and down-samples the input image (figure 8a). The convergence rate is faster using max-pooling, and it can yield a better generalization performance by electing superior invariant features [7].

There is another method in pooling called Min-pooling, which is not as crucial as the mentioned two and uses less than those. The min-pooling method calculates the minimum value for rectangular pooling regions of a feature map. Figure 7 shows the differences between the mentioned three methods for a real picture (top left) and illustrates how the various pooling methods can affect an image.

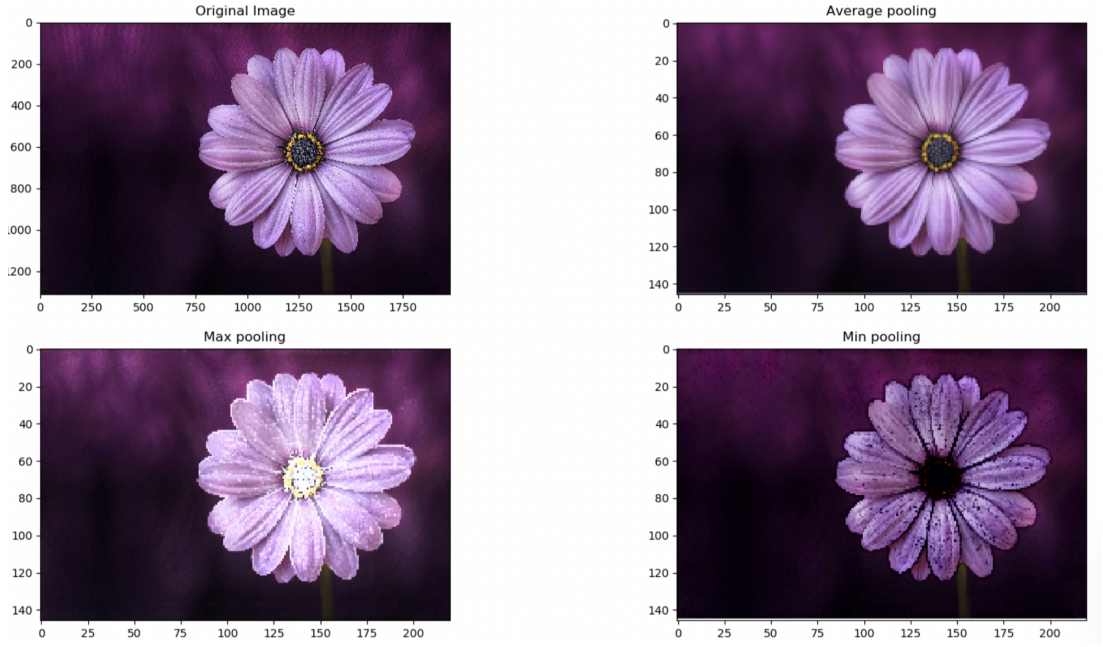
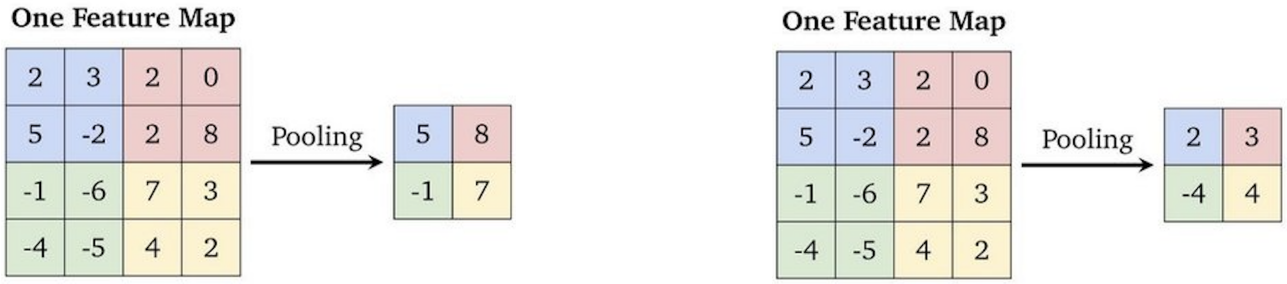


Figure 7: The results of three different pooling methods of size 9x9, applied to a real image (top left). The figure is adopted from [60].



(a) Max-pooling (b) Average-pooling

Figure 8: Example for the max-pooling and the average-pooling with a filter size of 2×2 and a stride of 2×2. Credit to the Stanford lecture CS231n [19].

3.2.3 Activation functions

Activation functions are critical in the architecture of neural networks by learning the abstract features via nonlinear transformation [45]. There is a considerable need for activation functions in neural networks, and the selection between the best activation functions for neural networks has a significant influence on the performance of ANNs. The output signal would only be a simple linear function if an activation function is not used and is just a polynomial of degree one. The activation function defines how the network accumulates the weighted sum of inputs to produce an output value. A polynomial of degree one is simple but cannot be learned due to its limitation and complexity. ANNs without an activation function work similarly to a Linear Regression model with limited performance and power [24]. The candidate activation function that can be used in the deep learning model to achieve the desired results are as follows:

1. Binary Step Function
2. Linear
3. Sigmoid
4. Tanh
5. ReLU
6. Leaky ReLU
7. Parametrized ReLU
8. Exponential Linear Unit
9. SoftMax [24]

3.2.4 Loss function

The learning process of an ANNs model is enhanced over a specific period. This is achieved by adapting the weights to push toward an optimal solution. It would be too complex to compute the ideal weights for a neural network model. That is why the learning process is cast as a search for the

optimal solution and leads to a satisfactory optimal stable state of the model where training and testing errors are reasonably minimized below the expected cut-off value. There are various methods to calculate loss functions, and the selection between different loss functions might be challenging and essential. By calculating the output error value of the models from the ground truth, the model can be trained to reach the ability to perform respectable predictions. The error values of the model are in their highest form when the variation of the model from the ground truth is the highest. Generally, error values should be minimized in neural networks to drive the model toward the optimal solution. The function used to calculate the error values of the model is called a loss function or cost function.

As mentioned, there are various loss functions, such as Cross entropy (CE) and mean squared error (MSE). For better training of a neural network model, the loss function will run in the background mathematically. The MSE is used as a target function in training and shows if training is progressing towards a stable solution. Many ANNs architectures use the MSE as a target function for training, and it can be displayed mathematically as [46]:

$$L_{MSE}(y,\hat{y}) = \frac{1}{N}\sum_{n=1}^{N}(y-\hat{y})^2 \tag{5}$$

The symbol $y, \hat{y}$, and N represent the target value, predicted value, and the number of the sample respectively [46].

The other important loss function is CE. In the field of information theory and upon entropy, CE has defined and typically calculates the distinction between two probability distributions. CE is also called logistic loss or log loss [12]. Suppose there is a target probability distribution as P and an approximation of the target distribution as Q. In that case, the cross-entropy of Q from P is the number of additional bits depicting an event using Q instead of P and can be expressed as H(P, Q), in which H is the CE function. If using the probabilities of the events from P and Q, then it can be computed and formulated as [12]:

$$H(P,Q) = -\sum_{x\in X} P(x)\ \log Q(x) \tag{6}$$

where P(x) is the probability of the event x in P, Q(x) is the probability of event x in Q, and the result is in bits [12].

3.3 History of DME

DME is an indication of DR, which predominantly leads to vision loss in DR patients. There are several systemic risk elements for the expansion of DME explained by Y. H. Yoon et al. [34], and the most important ones are such as a longer duration of diabetes, higher glycosylated hemoglobin (HbA1c) levels, and hypertension [34]. Various therapies have been identified to diminish the risk of DME among patients, aiming for glycemic control and management of hypertension and serum lipids [34].

3.4 Optical coherence tomography (OCT)

"OCT is a non-invasive imaging test using light waves to take cross-section pictures of the retina. With OCT, ophthalmologists can see each of the retina's distinctive layers, allowing them to map and measure their thickness" [44]. It is used in medical imaging that employs low-coherence light to capture micrometer-resolution, two and three-dimensional images from within optical scattering media (e.g., biological tissue) [47].

There are many types of OCT images, from the original time-domain OCT to spectral domain (SD-OCT) and swept-source OCT (SSOCT). In this project, images for the dataset are only spectral domain Optical coherence tomography (SD-OCT) images. Figure 9 explains the different layers of a normal retina in an OCT image.

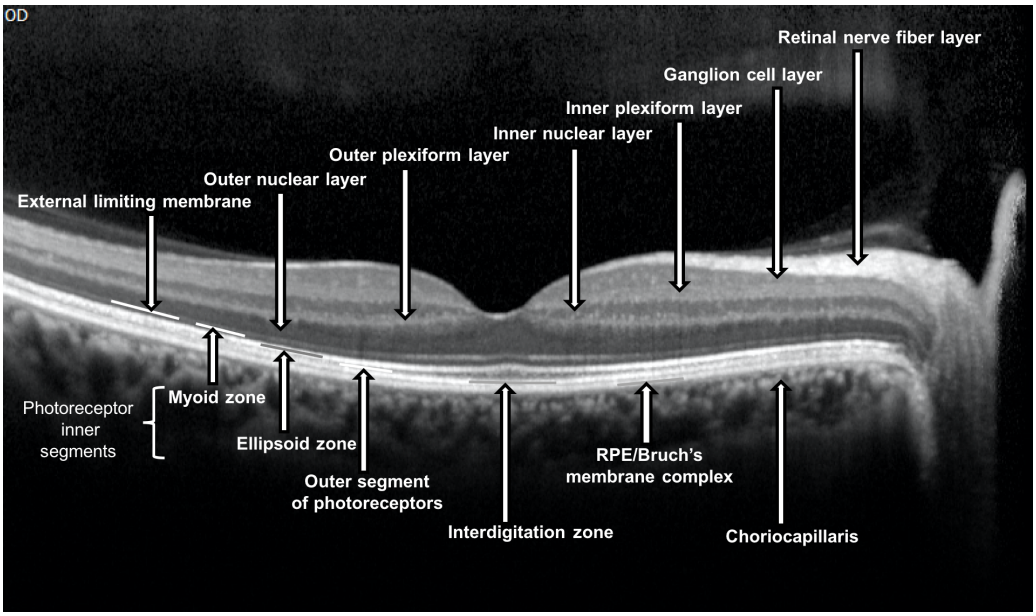


Figure 9: Explanation of different parts of retinal layers in OCT images [40].

3.5 DME in OCT images

OCT is an excellent tool for accurately assessing retinal layers because many layers underneath the surface of the retina are easily noticeable in OCT images. It helps to site changes to eye health earlier than just looking at the surface. It also quantifies retinal thickness and macular volume and qualitatively evaluates hyperreflective foci. OCT imaging plays an influential role and is widely used in diagnosing and detecting clinical outcomes of DME [43].

This project focuses on detecting morphological patterns on OCT and then detecting DME with high accuracy using OCT images. DME has crucial patterns when studying OCT images, such as diffused retinal thickening (DRT) and serous retinal detachment (SRD). Still, the most important pattern for the project is cystoid macular edema (CME) [43].

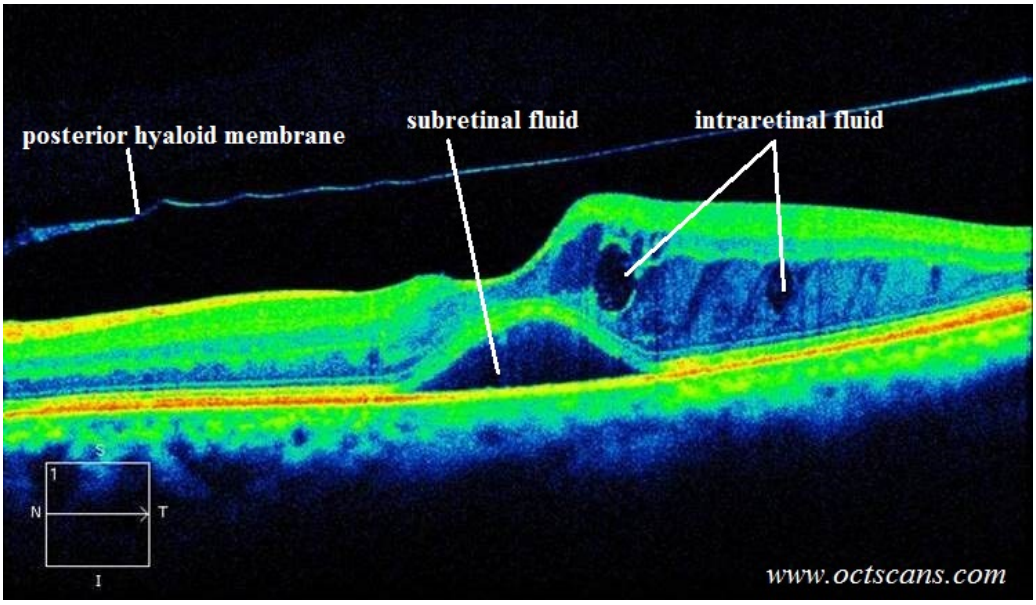


Figure 10: DME with the reason of intraretinal cysts [53].

Various deep learning models have been used to detect optical diseases and mainly to detect DME with a high accuracy using OCT and fundus images. However, there are not so many deep learning models to detect morphological patterns in OCT images in such a way that can help different therapeutic strategies for patients with DME based on OCT [43].

4 Proposed method

This section describes all three project goals mentioned in Section 1 in further detail. Furthermore, it proposes an approach in 5 steps to achieve the goals of this project.

4.1 Approach

As mentioned in Section 1, this project has three main goals. These three goals are accomplished in five steps. Figure 11 illustrates a high-level view of the goals and shows the steps done in this project and the remaining steps for future work. The five steps are addressed in this project as follows and are discussed in further detail in the following sections:

- **Step 1**: Find relationships between morphological features in linear and non-linear ways. In this regard, various correlation coefficient methods are tested. This step is discussed in details in Section 4.2, and its results are shown in Section 7.1.
- **Step 2**: Classification of individual morphological features to investigate which one has the potential for further investigation of DME detection. On this matter, individual features are classified separately based on their labels, with the highest possible number of images for each feature. This step is addressed in details in Section 4.3, and its results are indicated in Section 7.2.
- **Step 3**: Detection of the presence or absence of each morphological feature on the OCT images separately with an equal number of images for all the features. In this regard, The labels of features are modified to binary labels (presence or absence). This step is discussed further in Section 4.3, and its results are indicated in Section 7.3.
- **Step 4**: Detection of the presence or absence of two features simultaneously (combinations of features) on OCT images with an equal number of images for all the possible features. On this matter, all the features are selected two by two. Furthermore, one label as a binary label (presence or absence of both features simultaneously on OCT images) is assigned to them. This step is addressed further in Section 4.3, and its results are indicated in Section 7.4.
- **Step 5**: Detection of the presence or absence of DME using the trained models on individual features and combination of them with the usage of transfer learning. This step is discussed further in Section 4.4, and its results are indicated in Sections 7.2, 7.3, and 7.4.

4.2 Correlation between features

The correlation coefficient (ρ) indicates the strength of the linear relationship between two variables (X,Y) in a dataset which can be positive, zero, or negative. Zero correlation means no relationship between two different variables, and each variable has the highest correlation (1.0) with itself. The correlation coefficient is not a suitable measurement for the non-linear relationships between two different variables. The feasible range of values for the correlation coefficient is from -1.0 to 1.0. The correlation coefficient is calculated as follows:

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{7}$$

Pearson and Spearman's correlation are two popular and essential types of calculating a correlation coefficient. In this project, Spearman correlation is used for the linear relationship between the features in the dataset. Spearman correlation calculates the strength and direction of the monotonic relationship between two variables which is calculated as [3]:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{8}$$

ρ is the Spearman rank correlation coefficient, d_i is the difference between the two observation ranks, and n is the number of observations.

In addition, the Spearman correlation coefficient is used to understand which features are related and are close to each other. Also, a cluster map is made to check the relation of models based on their correlation results for the eight morphological features singularly.

Hierarchical clustering or a cluster map gives a better understanding of a dataset by improving the visual representation of correlation heatmaps and making finding groups of correlated features easier. It can group datasets into clusters based on the relationships among the different variables. Figure 18 is a cluster map using the Seaborn Python library, which illustrates how the different variables depend on each other and the relation between them.

4.3 Features and their combinations

As mentioned, the second step is to classify morphological features separately using the highest possible number of images for each feature. Figure 14 shows the difference in the number of available images for each label in each feature. Eight different models are created to classify each feature separately. The results of this step are shown in Section 7.2, and each feature is studied separately. The third step is to modify the labels of the eight morphological features to binary (presence or absence) labels. Then, eight models are created to detect each feature separately on OCT images. Each model uses an equal number of images for each feature in this step. The results of this step are shown in Section 7.3. The fourth step is to detect combinations of features on OCT images. All of features are selected two by two as combinations and one binary label is allocated to them (presence or absence of both features simultaneously on OCT images). Ten models are created for ten possible combinations to be detected. The results of this step are shown in Section 7.4.

4.4 Usage of transfer learning in DME detection

The fifth (last) step is to detect DME on OCT images with average sensitivity and specificity above 95 percent thanks to transfer learning using models trained on individuals and combinations of morphological features. All the weights and parameters of trained models are frozen and saved and then used for the training on a new dataset via transfer learning. One of the advantages of transfer learning is avoiding spending much time on training and verification processes. Only the last five layers of the models, three dense layers (also called classification layers) and two dropout layers, are unfrozen. DME diagnosis is implemented on a new dataset of OCT images with 2 labels (DME and normal). The results of the DME detection using transfer learning are shown in Sections 7.2, 7.3, and 7.4.

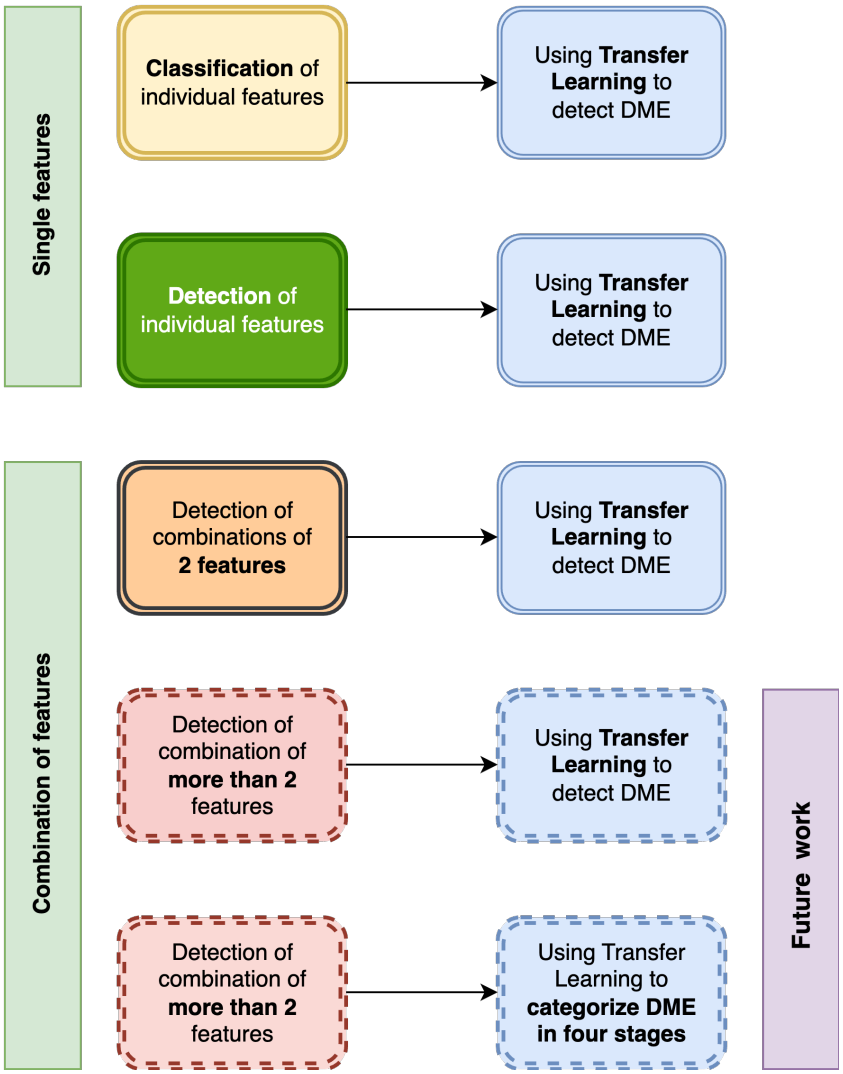


Figure 11: A high-level view of this project design shows the steps done in this project and the remaining steps for future work. Single features are classified and detected separately. Then, all the models are fed to transfer learning models to detect DME. Also, combinations of features are detected separately. Then, all the models are fed to transfer learning models to detect DME. Combinations of more than two features can also be implemented in later works.

5 Network architecture

This project is derived from the work done by Pannozo et al. [37] to investigate the presence or absence of DME in OCT images by utilizing deep learning and machine learning algorithms. Their study focuses on the grading protocol called TCED-HFV to categorize DME in four stages, mentioned in Section 2.1. Their grading protocol has been made from the seven morphological features (Section 1). In this thesis, one additional feature called Macular Volume (MV) is added to the seven mentioned morphological features. The central focus of this thesis is the diagnosis of DME on OCT images with average specificity and sensitivity above 95 percent with the usage of deep learning techniques, which has not been considered in work done by Pannozo et al. [37]. In addition, in this thesis, the mentioned morphological features are studied and diagnosed both individually and jointly to investigate how influential the features are for adequate detection of DME. This thesis emphasizes examining the features profoundly and does not categorize DME into the four stages, as done in the study by Panozzo et al. [37]. However, investigating the morphological features and combinations of them in the four stages mentioned by Panozzo et al. [37] is an interesting future work.

5.1 OCTNet

There is a deep neural network-based classifier called OCTNet that is used for the classification of DME, Drusen, and CNV on OCT images [42]. The OCTNet architecture introduced by Sunija A.P et al. [42] is a lightweight convolutional neural network with six convolutional blocks. The OCTNet is trained on a dataset with 83,484 OCT images for classification of DME, CNV, and Drusen. It depicted an impressive 99.69% accuracy, precision, and recall.

The OCTNet architecture used in this master project is inherited from the work done by Sunija AP et al. [42], and it is similar to the original architecture shown in Table 1. The network is also tuned for the datasets used in this project (Section 5.3).

In the architecture proposed by Sunija AP et al. [42], six convolutional blocks are used, followed by a ReLU activation function and a 2 x 2 max pooling operation for each block. After all six convolution blocks, an average pooling layer plays the feature extraction role in the architecture. The size of the output is a feature vector of 512 x 1. The feature vector is fed to three fully connected layers with two dropouts with a factor of (0.5). The original output is a probability vector of 4 x 1 [42]. Due to the several classification models with various labels in this project, different output layers corresponds to the class labels accordingly.

5.2 Model

The OCTNet architecture is used in this project as the base model for implementing the training process. The OCTNet architecture consists of six convolutional blocks and three dense layers with a ReLU activation function for each convolutional block. Also, CE, which is most typically used in medical image classification among loss functions, is used in this project. Due to the overfitting issue, a regularization technique is used in this project. Using dropout is influential in reducing the complexity of dense layers in neural network architectures by turning off some neurons in a neural network during training to avoid overfitting. The training and testing set ratio for all classification and detection of features models and DME detection models are 80% - 20% and 90% - 10% in this project.

5.2.1 Activation function - ReLU

Selecting an appropriate activation function is crucial in every neural network architecture. Tanh, ReLU, and sigmoid functions are the most commonly used activation functions in deep learning algorithms. The ReLU has some variations - LeakyReLU and PReLU; their usage is influenced by the study, data, and if the feedback from the ReLU variant provides better stability to the trained model [42]. In this architecture, the activation function used is a Rectified Linear Unit activation function (ReLU function) which is mathematically represented for an input x as:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & \text{otherwise} \end{cases} \tag{9}$$

5.2.2 Loss function

In deep learning, a model learns how to minimize the error by a mapping function via error back-propagation and update the model weights [41]. For the loss function in this architecture, a CE loss function is used to minimize the error, which is mathematically represented as:

$$J(\theta, X) = -weight[t]\ \log\left(\frac{\exp x[t]}{\Sigma_j \exp x[j]}\right) \tag{10}$$

where x is the output of the CE classification layer, t is the target class, weight is the 4 x 1 feature vector that stores the inverse class frequencies, X is the input batch of the images, and θ is the set of learnable model parameters [42].

In this project, three types of CE loss have been used depending on the number of classes. Categorical CE is adopted for the multi-class classification scenario, and binary CE or sparse categorical CE is used for binary classifications.

CE loss is the most commonly used in medical image classification among loss functions. Due to the class imbalance in medical images, mainly the normal class has more images, CE is the best option to deal with this issue in this project. During each learning process and classification, a modified CE loss function has been used with a different weight based on the number of images in each class to suppress the class imbalance problem [41].

5.3 Clinical data

For this project, a dataset was provided by the National Autonomous University of Mexico (UNAM), and a team of medical experts performed clinical evaluations for each OCT image and labeled them. As a result, the ground truth labels for each OCT image are provided in a CSV file. This dataset is used as the primary data for this project, aiming at classifying and detecting the eight mentioned morphological features.

Another dataset consisting of 3000 OCT images with binary labels (Normal and DME), including vertical and horizontal cuts, was provided aiming at the detection of DME.

The dataset contains a CSV file as ground truth for image labels consisting of eight columns (corresponds to the eight mentioned morphological features) for each OCT image, and a folder contains the corresponding OCT images. In the dataset, OCT images are provided in two different cuts, horizontal cuts, and vertical cuts. The dataset includes 897 OCT images from real patients.

The eight prominent morphological features on the OCT images are categorized into their labels based on their observed values and individual characteristics. The first five features are more influential than the others for DME detection, and the last three features are adjunctive features based on Pannozo et al. [37].

Eight mentioned morphological features are as follows:

- Thickening (T):
  - Label 0: increment below 10% of normal values (315-346.5 µm)
  - Label 1: increment above 10% and below 30% of normal values (347-409 µm)
  - Label 2: increment of more than 30% of normal values ($>$ 409 µm)
- Macular Volume (MV):
  - Label 0: $<=$ 0.26 µm
  - Label 1: $>$ 0.26 µm
- Size of intraretinal Cycts (C):
  - Label 0: Absent
  - Label 1: Mild (0-100 µm)
  - Label 2: Moderate (101-200 µm)
  - Lable 3: Severe ($>$200 µm)
- State of Ellipsoid Zone (EZ) / External Limiting Membrane (ELM):
  - Label 0: Intact
  - Label 1: Disruption
  - Label 2: Absent
- The presence of Disorganization of the Inner Retinal Layers (DRIL):

  - Label 0: Absent
  - Label 1: Present

- Hyperreflective foci (H):
  - Label 0: less than 30 in number
  - Label 1: higher than 30 in number

- The presence or the absence of subretinal Fluid (F):
  - Label 0: Absent
  - Label 1: Present

- The Vitreoretinal relationship (V):
  - Label 0: Absence of any visible adhesion or traction between vitreous cortex and retina
  - Label 1: IVD
  - Label 2: PVD
  - Label 3: VMT
  - Label 4: ERM

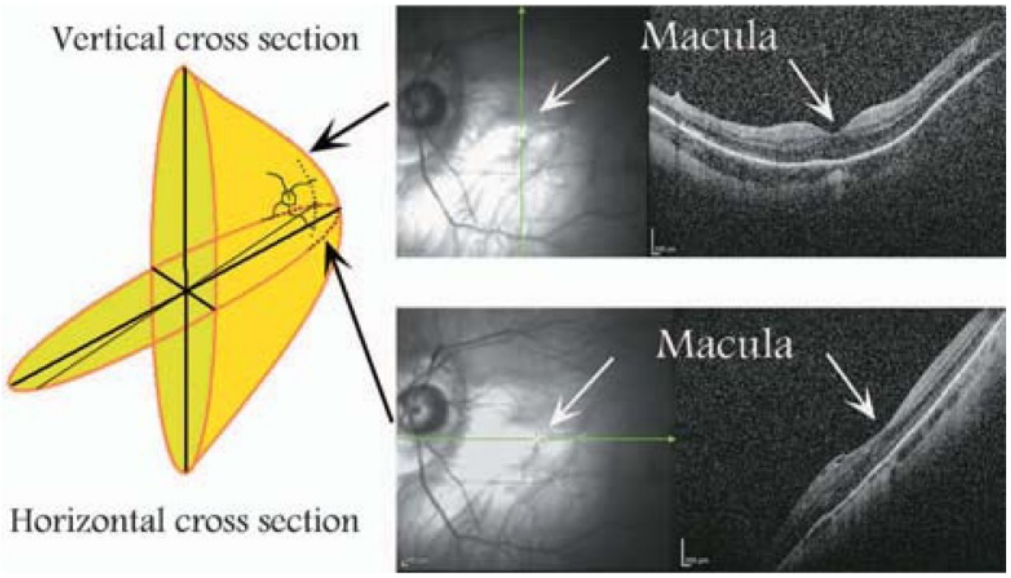
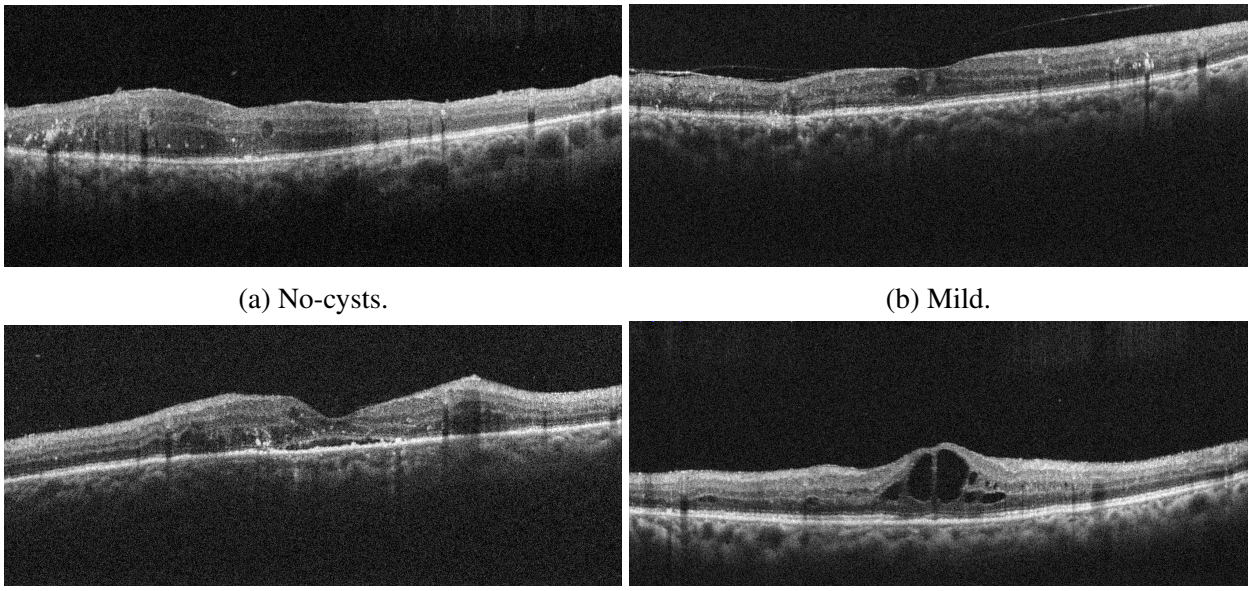


Figure 12: Difference in vertical and horizontal OCT cross-sections in a myopic eye with posterior staphyloma [6].



(a) No-cysts. (b) Mild.

(c) Moderate. (d) Severe.

Figure 13: Four Cysts (C) classes concerning their experimental conditions. (a) is labeled as No-cysts, which means medical experts on that OCT image observed no cysts. (b) is labeled as Mild, which means the size of the cyst(s) observed by experts is(are) between 0 and 100 µm. (c) is labeled as Moderate, which means the size of the cyst(s) observed by experts is between 101 µm and 200 µm. (d) is labeled as Severe, meaning that the cyst(s) size is(are) bigger and is more than 200 µm.

5.3.1 Data loading and data pre-processing

For the separate classification or detection of each feature, the CSV file is cleaned based on the feature to an ID column and the feature columns. Firstly, some images in the dataset do not have any corresponding rows in the CSV because the CSV is still incomplete. Next, the invalid images are deleted; the IDs in the CSV file with the same image's name are matched together, and extra rows (empty rows) are deleted. For each model, based on the classifying feature, the OCT images are divided into classes with corresponding labels.

For example, for the feature Cysts (C), each row with the label 0 is separated into class No-cysts, and rows with the label 1 are split into class Mild. Rows with the label 2 are assigned to the Moderate class, and finally, the rows with the label 3 are divided into the class Severe. Four different folders with ground truth are set up to facilitate the training process.

The input data is first pre-processed by converting them into gray-scale equivalents. Raw images without pre-processing are fed to the training input based on the result needed. Then, from the Tensorflow_keras library, the function *ImageDataGenerator* is used to generate batches of tensor image data with real-time data augmentation. *ImageDataGenerator* function also helps to easily split the data to the train set, validation set, and test set. Also, some data augmentation is used, such as *shear_range* and *zoom_range*, to make the training set difficult for the model to avoid overfitting.

Then, the *flow* function of the *Tensorflow_Keras* is used to consume data and label arrays. It generates batches of augmented data and label arrays, generates batches of augmented data, and has two options for choosing labels from the directory or data frame. For images with the CSV label, the function *flow_from_dataframe* is used to take the data frame and the path to a directory and generate batches containing augmented and normalized data. For images with labels based on their directory, the function *flow_from_directory* is used to take the path to a directory and generate batches of augmented data.

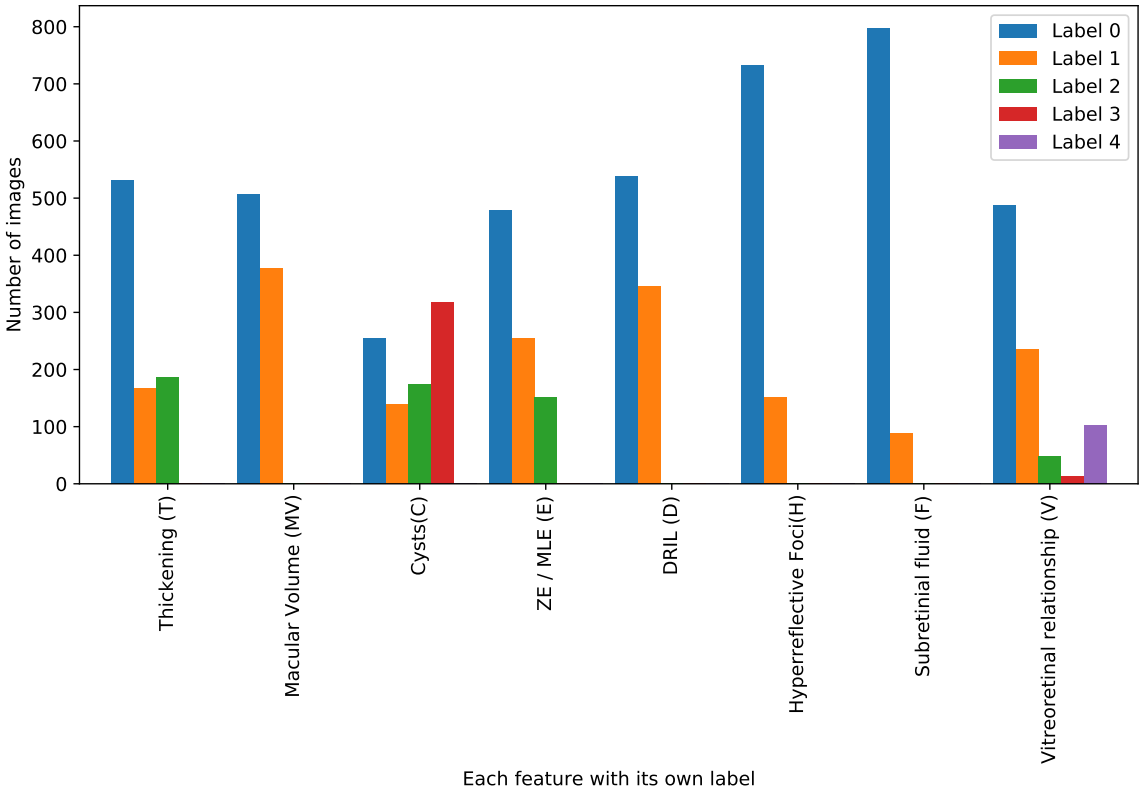


Figure 14: The number of available images for each label of corresponding feature separately in the data frame.

Table 1: The OCTNet architecture with the corresponding activation function for each convolutional block and each dense layer. Batch normalization is used in each convolutional layer to speed up training and make learning easier. The original output, calculated by the CE loss function, is a four-class vector.

Layer Name	Activations	Learnables
Input ($227\times227\times1$)	-	-
Convolution	$227\times227\times32$	Weights $7\times7\times1\times32$ *Bias* $1\times1\times32$
ReLU	$227\times227\times32$	-
Batch Norm	$227\times227\times32$	Offset $1\times1\times32$ *Scale* $1\times1\times32$
Max pool	$113\times113\times32$	-
Convolution	$113\times113\times32$	Weights $7\times7\times1\times32$ *Bias* $1\times1\times32$
ReLU	$113\times113\times32$	-
Batch Norm	$113\times113\times32$	Offset $1\times1\times32$ *Scale* $1\times1\times32$
Max pool	$56\times56\times32$	-
Convolution	$56\times56\times64$	Weights $5\times5\times64\times32$ *Bias* $1\times1\times64$
ReLU	$56\times56\times64$	-
Batch Norm	$56\times56\times64$	Offset $1\times1\times64$ *Scale* $1\times1\times64$
Max pool	$28\times28\times64$	-
Convolution	$28\times28\times128$	Weights $5\times5\times128\times64$ *Bias* $1\times1\times128$
ReLU	$28\times28\times128$	-
Batch Norm	$28\times28\times128$	Offset $1\times1\times128$ *Scale* $1\times1\times128$
Max pool	$14\times14\times128$	-
Convolution	$14\times14\times256$	Weights $3\times3\times256\times128$ *Bias* $3\times3\times256$
ReLU	$14\times14\times256$	-
Batch Norm	$14\times14\times256$	Offset $1\times1\times256$ *Scale* $1\times1\times256$
Max pool	$7\times7\times256$	-
Convolution	$7\times7\times512$	Weights $3\times3\times512\times256$ *Bias* $1\times1\times512$
ReLU	$7\times7\times512$	-
Batch Norm	$7\times7\times512$	Offset $1\times1\times512$ *Scale* $1\times1\times512$
Max pool	$3\times3\times512$	-
Average pool	$1\times1\times512$	-
Fully connected layer	$1\times1\times128$	128×512
Dropout (50%)	$1\times1\times128$	-
Fully connected layer	$1\times1\times32$	32×128
Dropout (50%)	$1\times1\times32$	-
Fully connected layer	$1\times1\times4$	4×32
Softmax	$1\times1\times4$	-
Cross-entropy	$1\times1\times4$	-

6 Implementation

This section explains the hardware and software environment used in this project and describes the techniques and strategies used to implement the project. It also demonstrates the evaluation metrics to measure how satisfactorily the models performed to enable the comparison between them.

6.1 Hardware environment

In this project, the experiments are performed and completed on the GoogleColab PRO, a Google Research product. It permits the implementation of python code through the browser and is particularly appropriate for machine learning and data analysis projects. This project implementation is executed on the TensorFlow 2.0 framework supported by a GPU with 25GB of RAM and 166GB of disc memory. GPUs in GoogleColab PRO are randomly selected, such as K80, P100, and T4.

6.2 Software environment

Python3 as the programming language and GoogleColab as an editor with a provided GPU and 25GB RAM are used for the project. TensorFlow and Scikit-learn are used as the libraries. CNN is used as the learning method. Visualization is performed by using the popular Matplotlib and Seaborn libraries.

6.3 Batching

Batching is one of the essential strategies for neural network learning to avoid running out of memory during training. The most common practice for the training in each iteration is to feed all the input data into the model. Input data can also be provided in batches of N components in each iteration, called mini-batch. The batch size can be explained as the number of training examples operated in one iteration. In the implementation, the mini-batch size is chosen differently for each experiment, but the most prominent mini-batch size is 16. The size of the mini-batch affects the training speed, and it can go beyond the CPU/GPU's memory capacity, but it has a more negligible effect on the final output.

6.4 Batch normalization

Batch normalization, also termed batch norm, is commonly used to train very deep neural networks that are computationally intensive. For each mini-batch, batch normalization standardizes the inputs to a layer, decreasing the number of training epochs and enhancing the speed of the learning process. The study by Kaiming He et al. [18] shows the effect of using batch normalization after the convolutional layers and before the activation functions on the ResNet architecture reached state-of-the-art results on the ImageNet dataset.

A big challenge in deep learning is that when the weights are refreshed in each mini-batch, the allocation of the inputs to the other layers might change. Using batch normalization can push the learning process toward the target when the target is moving. Batch normalization parametrizes a deep neural network, which can solve the problem of arranging updates across multiple layers.

6.5 Data augmentation

Augmented data differs from synthetic data because the whole synthetic data is generated entirely artificially using the original data. In contrast, augmented data is generated from original data by some geometric transformations such as rotation. Data augmentation is useful when there is not enough data and collecting a large amount of data is difficult. Data augmentation methods are helpful for multiple deep-learning tasks, such as image classification and object detection. It can enhance the performance of the deep learning model by making the training process more difficult for the model to avoid overfitting.

For example, for the classification of dogs and cats, data augmentation can generate new similar images of cats and dogs, which benefits in boosting the size of the dataset and sometimes balances the amount of data for each class in case of class imbalance. Also, it forces the model to see more images of different classes and learn for the goal of generalization and not overfit on a particular class.

6.6 Regularization

Regularization is any additional technique that aims to make the model generalize better and deliver more promising results on the test set [23]. In other words, regularization is used in neural networks to prevent overfitting and enhance the accuracy of a model when encountering new data.

One of the most prominent issues in training neural networks is overfitting which happens when the model in the training process performs too well in a particular class but cannot predict very well for the test. In other words, a model wrongly learns the unwanted detail and noise in the training data, negatively influencing the prediction of the model for the new data, meaning the model has not been generalized very well. Frequently, a complex network is more exposed to overfitting, and regularization is needed to avoid overfitting.

There are three regularization techniques: L1, L2, and dropout. The work done by N. Srivastava et al. [16] has shown that combinations of regularization methods can be used, and one of the minor test classification errors is when using L2 and dropout in a model. In this project, L2 and dropout are used the created models to decrease the error. Dropout is turning off some neurons in a neural network during training. It is mainly used for fully connected (dense) neural network layers because they are more complex than convolutional layers (figure 15). The loss function using L2 regularization is a loss function with squared L2 norm of the weights, and it is calculated as follows:

$$L_{MSE}(y,\hat{y}) = \frac{1}{N}\sum_{n=1}^{N}(y-\hat{y})^2 + \lambda\sum_{n=1}^{N}\omega_i^2 \tag{11}$$

where y is the target value, $\hat{y}$ is the predicted value, N represents the number of samples, ω shows the weight, and λ is the regularization parameter. In this project, in all models, $\lambda = 0.01$

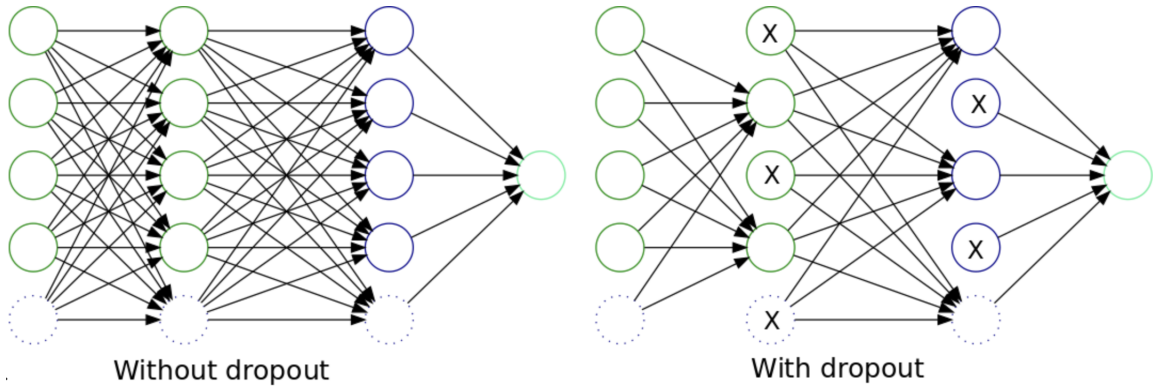


Figure 15: A neural network architecture with and without using dropout. The sign X shows the turned-off neurons during the training process [58].

6.7 Transfer learning

The concept of transfer learning comes from educational psychology proposed by Charles Judd to explain that the learning of transferring results from the generalization of experience. If there is a connection between two activities, transfer learning can happen. For example, someone who knows how to play the violin can learn to play the piano faster than others. In machine learning, transfer learning is a powerful problem-solving method that focuses on transferring knowledge across domains [39].

In deep learning, transfer learning happens where weights from a model trained on one task are taken and can be used in two ways. One way is to construct a fixed feature extractor, and the other is to initialize the weights and perform fine-tuning. In other words, transfer learning is training a neural network model on a new dataset with pre-trained weights acquired from training it on a different dataset, mainly used on a large dataset.

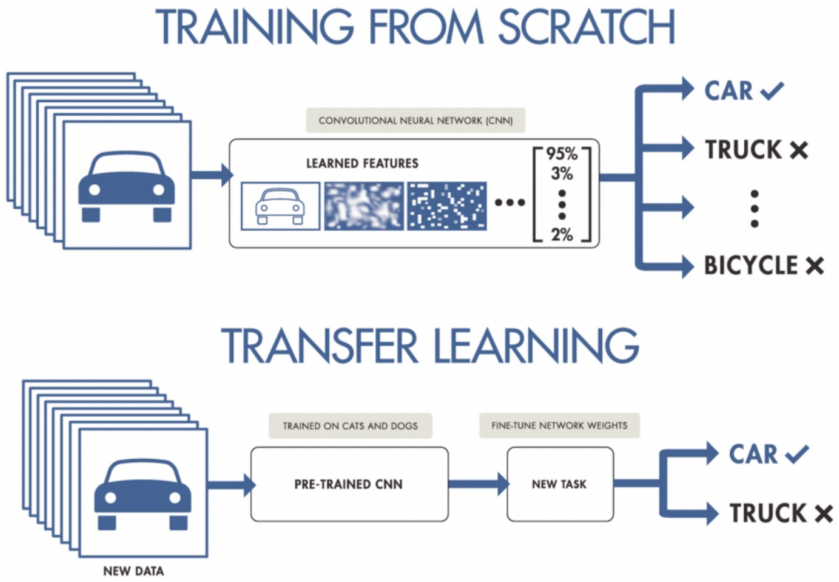


Figure 16: The difference between transfer learning and learning from scratch [57].

Two words are used interchangeably in deep learning - transfer learning and fine-tuning, but they are not the same. In transfer learning, the weights of all layers are frozen, and depending on the output, fully connected layers (dense layers) can be added to classify the output, and it is usually used for similar and small data sizes. In fine-tuning, a few layers can are frozen, and the remaining layers use another dataset usually used for large amounts of data [50].

6.8 Evaluation metrics

Evaluation metrics for the models allow the assessment of the accuracy of a model and measure the performance of the trained models, and describe the degree of effectiveness the models generalize on the unseen data. Selecting appropriate evaluation metrics for a model is also essential. The overall predictive capability of the models can be enhanced by using various evaluation metrics for the performance examination.

6.8.1 Accuracy and loss

One of the essential ways to evaluate machine learning models is to evaluate classification accuracy and logarithmic loss. The term classification accuracy or accuracy is the ratio of the number of correct predictions (True Positive plus True Negative) to the total number of input samples (summation of True Positive, True Negative, False Positive, and False Negative). Accuracy is different in the training set and test set, and high accuracy in the training set does not guarantee the model prediction very well [48].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

The other term, Logarithmic Loss or Log Loss, is evaluated by the false classification when a classifier assigns a probability to each class for all the samples. Suppose N samples belong to M classes, and there are:

$$LogLoss = \frac{-1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log p_{ij} \tag{13}$$

where y_{ij} shows whether the sample i belongs to class j and p_{ij} shows if the probability of sample i belongs to class j, the goal is to minimize the logarithmic loss in training [48].

6.8.2 Confusion matrix

A confusion matrix is an NxN matrix, where N is the number of prediction classes. For example, if there is a four-class classification task, a 4x4 confusion matrix can be used to evaluate how many samples are predicted correctly in each class. In confusion matrices, the matrix is divided into two dimensions, one is predicted values, and the other one is actual values along with the total number of predictions. There are four essential terms in the confusion matrix as follows:

- True Negative: When the model has the prediction No, the real or actual value is also No.
- True Positive: When the model predicted Yes, the actual value is also true (Yes).
- False Negative: When the model has predicted No, but the actual value is Yes (Type-II error).
- False Positive: When the model has predicted Yes, but the actual value is No (Type-I error).

6.8.3 Classification report

The classification report is another performance evaluation method in machine learning that shows a model's precision, recall, F1 Score, and support. It gives an insight into the overall performance of the model.

- Precision: The ratio of True Positives to the sum of True Positive and False Positives.

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

- Recall: The ratio of True Positives to the sum of True Positives and False Negatives.

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

- F1 Score: The weighted harmonic mean of precision and recall. For the F1 Score, the value of 1.0 is the best-expected performance of the model.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{16}$$

- Support: The number of actual class occurrences in the dataset. The more balance support, the structural strength of the classifier.

6.8.4 Sensitivity and Specificity

Sensitivity, recall, hit rate, or True Positive Rate (TPR) and Specificity, selectivity, or True Negative Rate (TNR) are two measures of the performance of a model in machine learning. Sensitivity is the proportion of true positives correctly predicted by the model, while Specificity is the proportion of True Negatives that the model correctly predicted.

- Sensitivity:

$$Sensitivity = \frac{TP}{TP + FN} \tag{17}$$

- Specificity:

$$Specificity = \frac{TN}{TN + FP} \tag{18}$$

7 Results and Discussion

This section explains the relationships between morphological features and shows how the features depend on each other by finding the correlation between them. Each feature is studied separately in this section. The features are trained and tested independently, and their results are shown. First, The morphological features are classified with the highest number of available images for each label of features. Second, the features are detected only in two classes with the same number of images for the training process. Then all the trained models (5 models for the classification, and 5 models for the detection of the features) are trained on a new dataset of OCT images to predict DME using on transfer learning. Their results are compared to see which feature is more effective for DME prediction.

There are 5 influential single features out of the 8 mentioned features in Section 5.3. The other 3 features, called adjunctive features, performed insufficiently, and only the 5 features are studied. In the next part, the combination of two features simultaneously is considered. There are 10 possible combinations for considering two features simultaneously, based on 5 influential features. Therefore, 10 models for combinations of features are created to detect every two features simultaneously with the same number of images for the training and testing. Then, their results are compared to find the most effective combination for a better prediction of DME.

The results are compared and discussed at the end of the section. This section further compares the results with the prior works.

7.1 Relevant morphological features

This project uses Spearman's correlation to find the relationship between variables (morphological features). There is another popular method for measuring the correlation coefficient called Pearson correlation. The difference between Pearson and Spearman's correlation is that Pearson only evaluates the linear relationship between two continuous variables, while Spearman's correlation can also assess the monotonic relationship and the linear relation [52].

Figure 17 depicts Spearman's correlation between eight features in a heatmap matrix in which the yellow color means the highest (strongest) correlation between every two features. The dark blue color indicates the lowest (weakest) correlation between every two features. The green squares depict a moderate correlation. The strongest correlation between two features, the greater relationship between them. Figure 18 illustrates hierarchical clusters to categorize data by similarity, which reorganizes the data and displays similar content next to one another for even more depth of understanding of the data besides the correlation heatmap.

By analyzing figures 17 and 18, the strongest correlation between all the features is the relationship between Thickening and Macular Volume. Figure 18, the clustering map also demonstrates the closest relationship between Thickening and Macular Volume, and then with Cysts. In addition, based on figure 17, DRIL also has a moderate correlation with Cysts. There are four correlations greater than 0.4 between the four features. Therefore, based on the correlation heatmap and clustering map, four features, Thickening, Macular Volume, Cysts, and DRIL, are more important and influential than other morphological features in this project.

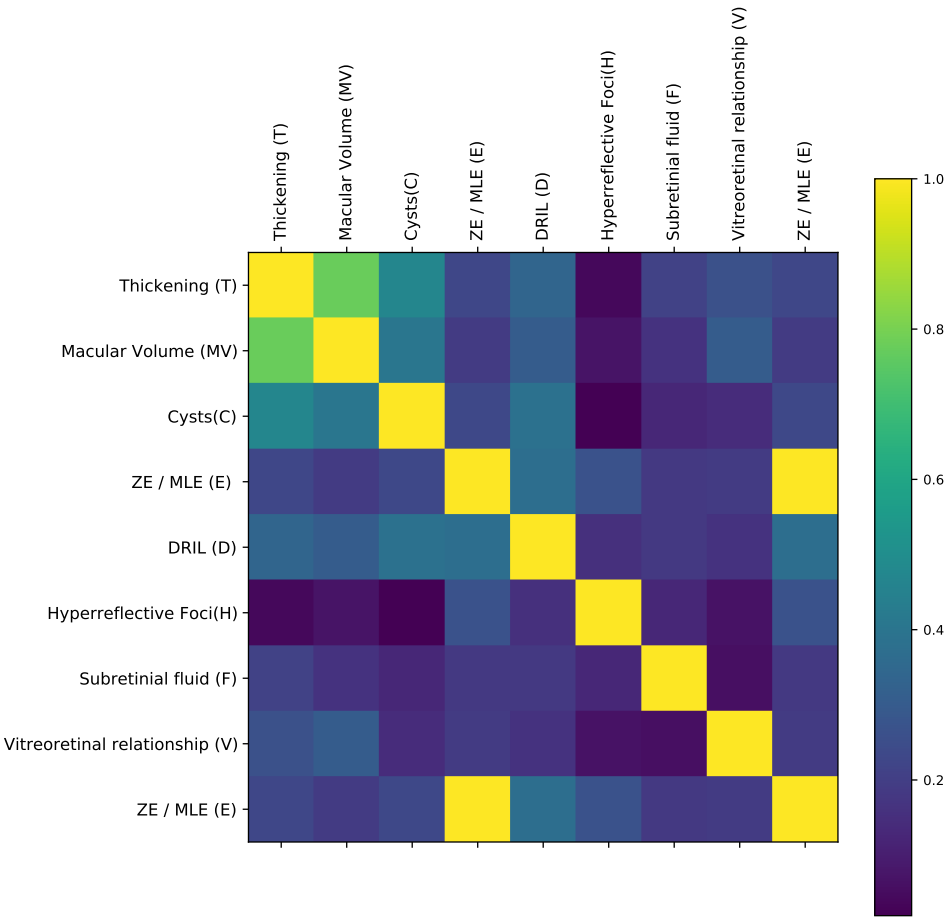


Figure 17: The correlation coefficient between the eight features in a correlation heatmap matrix. The color bar shows the range of the correlation values (0 to 1) with colors from dark blue (the weakest correlation) to yellow (the highest correlation). The correlation between each feature and itself is 1.

Some weak correlations can be seen between adjunctive features, and the lowest correlation between Hyperreflective foci (H) and the first three features (Thickening, Mocular volume, and cysts). Figure 18 demonstrates the lowest correlation coeffecienct between adjunctive features by cluster them taller than other features.

As the correlation coefficient are beneficial by showing the relationship between two variables, one downside of the correlation coefficient might be the deficiency in revealing the relationships of variables in other dimensions so that the features might be correlated non-linearly. Hence, other features might be correlated with each other in different ways, and there is a possibility to be correlated in higher dimensions.

In the following sections, more relationships between features can be investigated by examining the single features separately and their combination for DME detection.

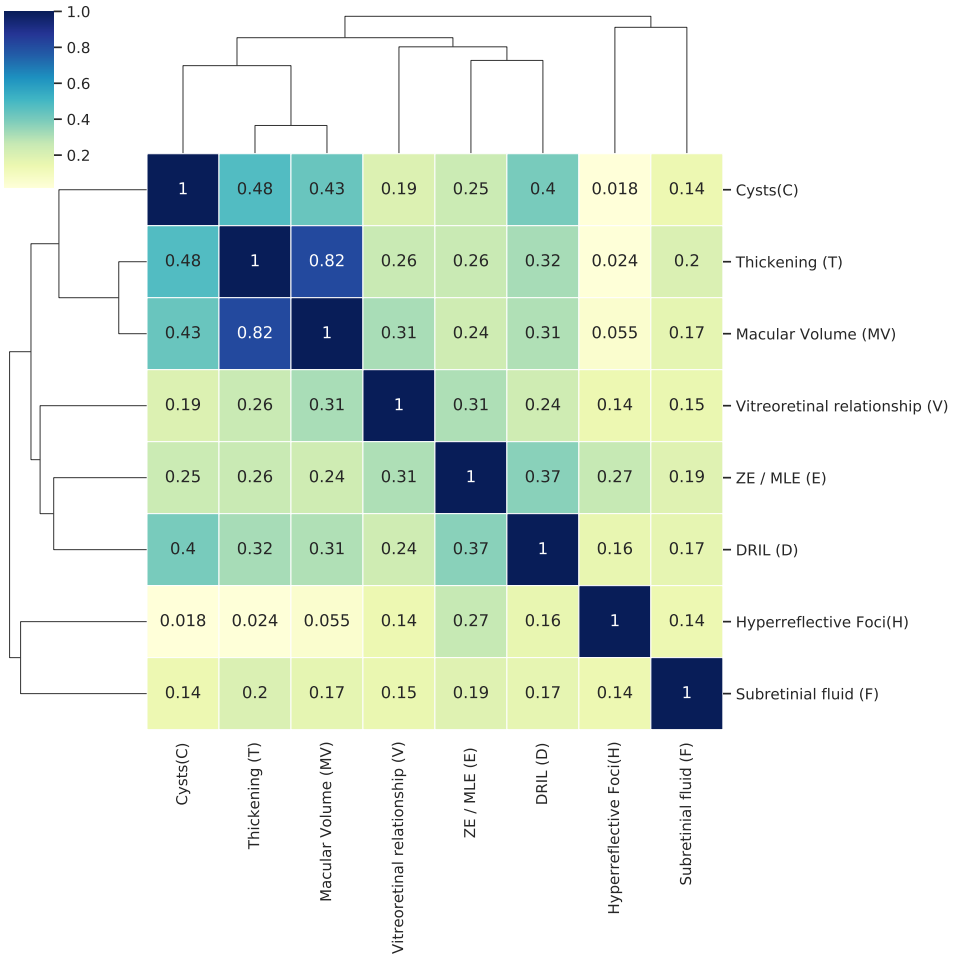


Figure 18: This figure illustrates the correlation between every two features and shows from the closest to the furthest relationship between features by clustering them.

7.2 Classification of individual features

In this section, the classification of each feature from the dataset is trained and tested separately with the highest possible number of images. Figure 14 shows that most of the features do not have the same number of images for their own classes, and some do not have a sufficient number of images in each class for the training process. Therefore, data augmentation techniques are done for their training and validation processes. The training and validation set ratio is 80% for the training set and 20% for the validation set. A pre-processing technique is done on all OCT images, which converts images to grayscale images before feeding them to the model.

7.2.1 Thickening (T)

Thickening (T) refers to foveal thickness in OCT images. The term fovea refers to ”a specialized retinal area that supports the highest visual acuity [8].” As mentioned in Section 5.3, the first morphological feature is Thickening (T) which has three labels. Figure 14 shows the number of images for each label, and the number of images for three classes of Thickening (T) differs significantly and is not balanced. For the Thickening (T) classification, 156 images are randomly picked for the training process of each class, and 15 images are split from each class for the test set. The implementation

has been performed for the Thickening (T) classification using normal images as label 3 and without normal images separately. In the end, the difference in results did not differ remarkably. Also, Based on the model's loss function, any overfitting has not been observed in the model, but the accuracy function of the validation fluctuates (figure 27a). After the classification performed for the feature thickening, the results are shown as follows:

Classification report	Precision	Recall	F1-Score	Support	Train set
Average accuracy	0.70	0.60	0.60	45	468

7.2.2 Macular Volume (MV)

The second feature is Macular Volume (MV), and it has two labels as mentioned in Section 5.3. Figure 14 illustrates that the number of images for each class of the Macular Volume does not differ much and is approximately balanced. For the classification of Macular Volume, 355 images are randomly picked of each class for the training set, and 25 images are split for the test set for each class. The classification of Macular Volume (MV) has been done, including normal images, and once without normal images. As the result, the difference in results does not vary remarkably. In the Macular Volume classification, no overfitting has been observed based on the loss function, but the accuracy function of the validation set fluctuated (figure 27b). After the classification performed for the feature Macular Volume, the results are shown as follows:

Classification report	Precision	Recall	F1-Score	Support	Train set
Average accuracy	0.70	0.68	0.65	50	710

7.2.3 Cysts (C)

The third feature is Cysts (C) which refers to the size of the intraretinal cysts. The round and minimally reflective spaces within the neurosensory retina can define intraretinal cysts, which can be located in different areas such as the outer nuclear layer, inner nuclear layer, or ganglion cell layer [37]. As mentioned in Section 5.3, the size of intraretinal cysts are categorized into four classes. Figure 14 shows the number of images of different labels of the Cysts (C) does not differ remarkably and is approximately balanced. The implementation is performed separately using normal images as label 4 and without normal images. In this feature, the results can differ significantly for classification, including and excluding normal images. The reason is the model can overfit on normal images and the images with label 0 (no-cysts). The reason is that the similarity between no-cysts and normal images can be so high. The label 0, no-cysts images, might have one of the other seven features, and distinguishing the little similarity between those are extremely difficult for the model with such a small dataset.

The model has been trained and tested many times, no overfitting has been observed. The accuracy function behaves naturally (figure 27c). One of the prominent issues in this model is overfitting on the result of the testing process on the label 0, which is no-cysts. The model cannot distinguish all four labels based on the size of the cysts. One possible reason is the fewer images of each class for such a big multi-label classification. For the classification of Cysts (C) with four labels, 15 images are split for the test set of each class, and 126 images are separated for the training and validation set. The results are shown as follows:

Classification report	Precision	Recall	F1-Score	Support	Train set
Average accuracy	0.36	0.30	0.25	60	504

7.2.4 State of Ellipsoid Zone (EZ) and External Limiting Membrane (ELM)

The fourth feature is the state of the Ellipsoid Zone (EZ) and External Limiting Membrane (ELM), which both together make one feature. As mentioned in Section 5.3, this feature has 3 labels. Considering the four outermost layers on OCT images, those layers are categorized as disrupted when they are not perfectly detectable and as absent when there is a complete loss of foveal reflectively in the first two bands on the four layers, and as intact when the layers are in the normal shape [37]. Figure 14 depicts that the number of images for each label of the EZ/ELM feature differs remarkably, which means the class imbalance happens when implementing classification. For the classification of EZ/ELM, 20 images are divided into the test set for each class, and 137 images of each class are split into the training set and validation set.

This model has been trained and tested several times. Both loss and accuracy functions behave naturally, and no overfitting is observed, but some slight fluctuations are seen in the accuracy function (figure 27e). In the confusion matrix of the testing process, fewer images are correctly predicted for label 1 rather than the other two labels. The reason would be the difficulty of predicting the disrupted retinal layers rather than just a prediction of their absence or being intact. The implementation is performed using normal images as label 3 and without normal images separately. In the end, the difference in results does not differ remarkably. For the classification of the feature EZ/ELM, the results are demonstrated as follows:

Classification report	Precision	Recall	F1-Score	Support	Train set
Average accuracy	0.66	0.55	0.52	60	411

7.2.5 Disorganization of the Inner Retinal Layers (DRIL)

The fifth feature is the Disorganization of the Inner Retinal Layers or, in short, DRIL, ”which is defined as the loss of clear demarcation between the ganglion cell-inner plexiform layer complex, the inner nuclear layer, and the outer plexiform layer in the central fovea [37]”. As mentioned in Section 5.3, the feature DRIL has 2 labels (absent and present). Figure 14 illustrates the number of images for each label of DRIL. For this binary classification, 50 images are split into the test set of each class, and 321 images are divided into the training and validation sets for each class.

The model has been trained and tested several times, and no overfitting is observed, but validation accuracy in the accuracy function fluctuated remarkably (figure 27d). In the testing process, the model does not overfit on any particular classes, but the prediction still is not frankly high for each class. The implementation is performed separately using normal images as label 2 and without normal images. In the end, the difference in results does not differ remarkably. For the classification of the feature DRIL, the results are shown as follows:

Classification report	Precision	Recall	F1-Score	Support	Train set
Average accuracy	0.65	0.62	0.60	50	642

7.2.6 Adjunctive features

There are three more features called adjunctive features based on the work by Pannozo et al. [37], which are not part of the main features. The sixth feature is Hyperreflective foci (H) which refer to the number of Hyperreflective foci measured by dividing all the scans into two groups of high HF and low HF. The arbitrary number 30 has been chosen for the feature as a cut-off value and is manually counted in each scan [37]. Figure 14 shows the number of images for each of the two labels for the feature Hyperreflective foci (H), and there is a huge difference in number between label 0 and label 1, which means there is a class imbalance for this feature. For the classification, once the implementation has been done, including normal images as label two, and once the implementation has been done without normal images. In both ways, it was difficult for the model to perform well. Furthermore, in the implementation, including normal images, the model was unable to distinguish between normal images and the label 0 (low HF) due to their feature similarity.

The model did not perform sufficiently for this feature due to the fewer images with label 1 (high HF). In addition, figure 18 illustrates this feature does not have a satisfactory relationship with other more important features.

The seventh feature is subretinal Fluid (F). As mentioned in Section 5.3, the feature has 2 labels that refer to its presence or absence. Figure 14 depicts the number of images for each label in this feature between all the features has the most different one, making the class imbalance even more complex and the prediction not accurate. The number of images for label 1 (presence of subretinal Fluid) is not really enough for accurate detection. The implementation has been done with only 73 images for label one, and the result is not acceptable. Additionally, same as the feature HF, it does not have a strong relationship with other features based on figure 18.

The eighth and last feature is the Vitreoretinal relationship (V) which is a simplified version of the international Vitreomacular Traction Study Group classification [37], and as mentioned in Section 5.3, the feature has five labels. ”PVD refers to posterior vitreous detachment and is defined as no residual vitreoretinal adhesion, demonstrated by a scan including the optic disk and IVD refers to incomplete posterior vitreous detachment, and VMT refers to residual macular vitreous attachment exerting anteroposterior traction, and also ERM was defined as evidence of epiretinal tissue adhering to the macular surface [37]”.

For the classification of this feature, figure 14 shows five various labels with different numbers of images for each. For some labels of this feature, there are not enough images for the training to make it possible. Instead of classification, detection is also done by changing the categorization to two labels (absence and presence). The implementation for the feature has been done, but the results still are not satisfactory enough, and the model did not perform sufficiently for this feature.

7.2.7 Comparison of the classified individual features

In the previous subsections, eight different models with respect to their examined features were separately implemented. Some features did not have a strong relationship with other features based on the correlation matrix. Furthermore, due to the class imbalance and insufficient training data, the model did not perform properly for adjunctive features. As the result, one issue in this project is the lack of enough images in the dataset for features such as Hyperreflective foci, Subretinal Fluid, and

Vitreoretinal relationship.

In this subsection, all the features with satisfactory results and those which are effective for accurately detecting the DME are compared. The best-performed feature is Macular Volume (MV) among the five influential features, and then Thickening (T) performed nicely. Due to the class imbalance, the last three features, Hyperreflective foci, Subretinal fluid, and Vitreoretinal relationship, are not considered in the table below, and also for DME detection neither.

On the one hand, one benefit of the classification of the features individually is showing their highest possible performance for training and testing with the number of available images. The DME detection using the classified features is done in the next section. On the other hand, one of the downsides of the results gained from the classification of single features separately is the deficiency of having the potential to compare with each other. It is expected that the models to contain some degree of bias due to the data sample imbalance. Hence, the comparison of their result are not accurate or reliable. One possible solution is to compare single features with the same number of images for the training and testing processes. Therefore, in Section 7.3, the features are detected singularly with the same number of images, then they are compared to each other.

Feature	Precision	Recall	F1-Score	Support	Trainset images	Labels
Thickening (T)	0.70	0.60	0.60	45	468	3
Macular Volume (MV)	**0.80**	**0.75**	**0.75**	**60**	**710**	**2**
Cysts (C)	0.36	0.30	0.25	60	504	4
EZ / ELM	0.66	0.55	0.52	60	411	3
DRIL	0.65	0.62	0.60	50	642	2
Hyperreflective foci (H)	-	-	-	-	-	2
Subretinal Fluid (F)	-	-	-	-	-	2
Vitreoretinal relationship (V)	-	-	-	-	-	5

7.2.8 DME detection with the usage of classified individual features

After training the model for each feature separately, the corresponding model is saved (wights and biases), and by using transfer learning, the pre-trained model is trained again for DME detection on a different dataset. The dataset for DME detection contains 2200 Horizontal and Vertical OCT images in total and in 2 classes of Normal and DME images. 750 images are divided into the training set for each class, 200 images are split into the validation set for each class, and 150 images are separated into the test set for each class. All the training processes for each feature are implemented in the same situation, with the same amount of images for each set, the same data pre-processing, and epochs to see the differences between the results and make the comparison more reasonable.

After loading each pre-trained model, the last five layers (3 dense and 2 dropout layers) are unfrozen, and the model has trained again on a different dataset to predict DME. For all 5 models, the loss and accuracy functions behaved naturally, and no overfitting are observed. Each model was trained and tested several times with different optimizers, and the best one for all models was the Adam optimizer. All of the models performed exceptionally well. The results show each feature's influence on DME detection separately, but they are not comparable because they are not trained with equal

number of images. The table below shows the performance of each influential feature with the usage of transfer learning in DME detection. Thickening performed the best for the single feature, which helped diagnose DME. However, Macular volume performed the second best, and its result is close to the thickening. The two mentioned features are more influential and prominent in diagnosing DME.

Feature	Precision	Recall	F1-Score	Support
Thickening (T)	**0.99251**	**0.99250**	**0.99250**	**300**
Macular Volume (MV)	0.98544	0.98500	0.98500	300
Cysts (C)	0.96012	0.95667	0.95659	300
EZ / ELM	0.95914	0.95667	0.95661	300
DRIL	0.96800	0.96667	0.96664	300

7.3 Detection of individual features

Due to a shortage in the number of images for different labels of each feature, the classification of single features is not accurate and reliable enough to compare with each other. Hence, in this subsection, only the detection of features singularly with an equal number of images for the training and testing process is considered.

For each feature separately, a range of 510 to 560 images is separated to be detected if the corresponding feature is available on each OCT image. Different labels of each feature are adjusted to 0 (absent) or 1 (present). For example, Cysts (C), with four labels that refer to the size of the intraretinal cysts, is modified with two labels if any cyst is detected on the OCT image or not. In the detection of single features, the ratio for the training and validation set is the same as in the previous section (80%-20%). In addition, data augmentation techniques are used for the training and validation processes, as mentioned in 5.3.1.

7.3.1 Comparison of detected individual features

Five influential features are detected separately, four of which used 560 images for training, and only Cysts (C) is trained on 510 images due to the dataset's shortage of label 0 (No-cysts). All five models used 40 images for the testing process. Thickening (T) performed the best among all five features, and Cysts (C) performed the second best. In the training process of all features, no overfitting is observed in their loss functions, and their accuracy functions behave naturally, but a slight fluctuation is seen in the validation accuracy. Due to the small number of validation samples, the model is unable to reach a true stable solution.

In the Thickening detection, it was more difficult for the model to predict label 1, which means the incrementation of retinal layers of more than 10% rather than label 0. The reason is that there is not a specific boundary on OCT images to distinguish the incrementation of the retinal layers specifically, and they are labeled in a range of less than 10% to above 30%.

The Macular Volume detection model did not overfit on a particular label and performed slightly above the average. That could be improved by adding more images to the training and validation sets for better results, as shown in the previous section (classification with the highest possible number of images).

In the Cysts detection, the model predicted label 0 (no-cysts) much more straightforwardly than label 1 (presence of cyst). The loss and accuracy function in the model performed naturally without overfitting. One possible issue is the fewer images for the no-cysts label, and the model limited to the label. For this issue, data augmentation improved performance moderately.

In the EZ/ELM, the model performed better than EZ/ELM model in the previous section, and modifying the labels for the feature enhanced the results. There is not any overfitting observed in the model.

In the DRIL detection, the model is overfitting on label 1 which means the disorganization of retinal layers. The loss and accuracy function does not show any overfitting based on their functions, but the test set results show that the model predicts wrong images on label 1. Applying data augmentation techniques could not make any remarkable improvement in the model. The model performed negligibly better with a higher number of images for this feature, but still, this feature could be the most challenging feature for the model to predict because there is not a specific boundary on OCT images when disorganization of retinal layers happens. The work by Pannozo et al. [37] explain that retinal layers are damaged but still visible in some cases, and it counts as no DRIL (label 0), which causes uncertainty for the model to predict.

The table below reveals the results of detecting the features with the same number of images. The results do not deliver the highest possible performance for each feature because the number of images for each feature is limited to the smallest one, but instead, the table shows suitable leverage for comparison between features in a comparable condition as follows:

Feature	Precision	Recall	F1-Score	Support	# of Train-set images
Thickening (T)	**0.76**	**0.75**	**0.75**	**40**	**560**
Macular Volume (MV)	0.63	0.62	0.62	40	560
Cysts (C)	0.73	0.65	0.62	40	510
EZ / ELM	0.68	0.65	0.64	40	560
DRIL	0.67	0.57	0.51	40	560

7.3.2 DME detection with the usage of detected individual features

After training a model for each feature individually, the corresponding model is saved, and by using transfer learning, the model is trained again for DME detection on a different dataset. The dataset for DME detection is the same one in Section 7.2.8 with the same ratio for training and validation sets and the same condition for the training and testing processes to make the comparison more suitable.

After loading all the models and unfroze their last 5 layers, all five models have trained again on a different dataset to predict DME. Each model's loss and accuracy functions behaved naturally, and no overfitting was observed. Each model is trained with a learning rate (0.001) using the Adam and SGD optimizers. It is observed that the Adam optimizer performed better in all models. All of the models performed exceptionally well, but the difference in their results is a term for comparing models trained on different features. The table below shows the performance of each influential feature with the usage of transfer learning in DME detection in which Cysts(C) performed surprisingly the best among features and Thickening (T) performed the second best. In addition, the results of Cysts(C) and Thickening(T) are so close together.

Feature	Precision	Recall	F1-Score	Support
Thickening (T)	0.96875	0.96667	0.96663	300
Macular Volume (MV)	0.94910	0.94333	0.94315	300
Cysts (C)	**0.97720**	**0.97667**	**0.97666**	**300**
EZ / ELM	0.92857	0.91667	0.91608	300
DRIL	0.95181	0.94667	0.94651	300

7.4 Detection of combined features

Aside from three adjunctive features that do not have sufficient images for the training process, five main features are classified and detected separately in the previous subsections. Hence, ten possible models trained on a combination of two features from the pool of five influential features are constructed. Another goal of this project is to find which combination of two features is more influential in detecting DME. Hence, In this subsection, the combination of two features simultaneously is considered and studied to investigate how the detection of two features at the same time on each OCT image can affect the detection of DME.

After the detection of each combination of features, ten combinations are compared and studied based on the correlation heatmap (figure 17), cluster map (figure 18), corresponding classification report, and the number of images for each training process. For each combination separately, a range of 380 to 400 images is separated for the training sets and 40 images for the test sets to be detected if the corresponding features are available on each OCT image simultaneously (combination of them). Furthermore, labels are adjusted to 1 (present) for OCT images having both features simultaneously or 0 (absent) for OCT images which have neither both corresponding features. For example, the combination of features Thickening (T) and Macular Volume (MV) is modified with two labels if both features are available on any OCT image or not. Label 1 for an image with both Thickening and Macular Volume, and label 0 for an image with neither Thickening nor Macular Volume. The training and validation set and data augmentation techniques are the same as in the previous section.

7.4.1 Thickening (T) and Macular Volume (MV)

The first two features among 10 possible combinations are Thickening (T) and Macular Volume (MV), which performed very well in the detection and classification of features singularly and significantly affect the detection of DME separately. The correlation between the two features, based on the cluster map in figure 18, is 0.82, which means these two features are highly correlated. In this subsection, a model trained on both features is a binary classification in which an image can either have Thickening and Macular Volume or none of these features. For this binary classification, 400 images are used for the training process, and 40 images are used for the test set. Each set is split equally for each class in the training and test sets. The loss and accuracy functions behaved naturally, and no overfitting was observed based on the loss function (figure 28a). The model predicted the images with label 0 slightly easier than images with label one. Then the model is trained and saved afterward to be used for DME detection via transfer learning. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.79762	0.75	0.73958	40

7.4.2 Thickening (T) and Cysts (C)

The second combination of features is Thickening (T) and Cysts (C), which performed singularly well in classification and detection individually, and are expected to perform satisfactorily as a combination on OCT images. The correlation between the two mentioned features, Thickening, and Cysts, based on the cluster map in figure 18 is 0.48, which means these two features are moderately correlated. As mentioned in the previous section, Thickening and Cysts separately affect the detection of DME. In this subsection, a model trained on both features is a binary classification in which an image can either have Thickening and cysts or none of these two features. For this binary classification, 390 images are separated for the training set, and 40 images are split for the test set (195 images for each class in training and 20 images for each class in the test set). Anything abnormal is not observed based on the accuracy and loss functions (figure 28b), and the model prediction performance is above average. Then the model is trained and saved afterward to be used for DME detection via transfer learning. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.72	0.70	0.69	40

7.4.3 Thickening (T) and DRIL

The third combination is Thickening (T) and DRIL, and the correlation between the two mentioned features, based on the cluster map in figure 18 is 0.31, which means these two features are moderately correlated. Thickening performed individually better than DRIL in the classification and detection of single features and significantly better for the detection of DME separately. In this subsection, a model trained on both features is a binary classification in which an image can either have Thickening and DRIL or none of these features. For this binary classification, 390 images are separated for the training set and 30 images for the test set (195 images for each class in training and 15 images for each class in the test set). The loss function illustrated no overfitting (figure 28c), and validation in the accuracy function fluctuated highly. Then the model is trained and saved afterward to be used for DME detection via transfer learning. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.71	0.63	0.60	30

7.4.4 Thickening (T) and EZ/ELM

The fourth combination is Thickening (T) and EZ/ELM, and the correlation between the two mentioned features, based on the cluster map in figure 18 is 0.26, which means these two features are weakly correlated. These two features were not expected to perform better than the other combinations as they are correlated weakly. Thickening performed significantly better than EZ/ELM in previous models for individual features. In this subsection, a model trained on both features is a binary classification in which an image can either have Thickening and EZ/ELM or none of these features. For this binary classification, 400 images are separated for the training set and 40 images for the test set. The loss and accuracy function behave naturally (figure 28d), but slight overfitting can be observed in the test set results because the model can predict well either on label 0 or label 1. The model is trained and saved afterward to be used for DME detection via transfer learning. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.68	0.65	0.64	40

7.4.5 Macular Volume (MV) and Cysts

The fifth combination is Macular Volume (MV) and Cysts (C), which are two crucial features as shown in the classification and detection of them on the OCT images in previous sections. Based on the cluster map in figure 18, the correlation between them is 0.43, which means these two features are moderately correlated. They also significantly affect the detection of DME individually, and it is expected to perform well as a combination. In this subsection, a model trained on both features is a binary classification in which an image can either have Macular Volume and Cysts or none of these features. For this binary classification, 400 images are separated for the training set and 30 images for the test set (200 images for each class in training and 15 images for each class in the test set). The loss and accuracy functions behaved naturally, and no overfitting was observed based on the functions and the test set results (figure 28e). Then the model is trained and saved afterward. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.78	0.77	0.76	30

7.4.6 Macular Volume (MV) and DRIL

The sixth combination is Macular Volume (MV) and DRIL. The correlation between the two mentioned features, based on the cluster map in figure 18, is 0.31, which means these two features are moderately correlated. As mentioned in the previous section, Macular Volume and DRIL significantly affect the detection of DME separately. In this subsection, a model trained on both features is a binary classification in which an image can either have Macular Volume and DRIL or none of these features. For this binary classification, 386 images are separated for the training set and 40 images for the test set (193 images for each class in training and 20 images for each class in the test set). The loss and accuracy functions illustrated some moderate fluctuations (figure 28f), and the issue could be solved by adding additional images. The model predicted the correct number of images above the average, but still not significantly well. Then the model is trained and saved afterward. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.63	0.62	0.62	40

7.4.7 Macular Volume (MV) and EZ/ELM

The seventh combination is Macular Volume (MV) and EZ/ELM, and the correlation between the two mentioned features, based on the cluster map in figure 18 is 0.26, which means these two features are weakly correlated. As mentioned in the previous section, Macular Volume and EZ/ELM significantly affect the separate detection of DME. In this subsection, a model trained on both features is a binary classification in which an image can either have Macular Volume and EZ/ELM or none of these features. For this binary classification, 330 images are separated for the training set and 30 images for the test set (165 images for each class in training and 15 images for each class in the test set). The correlation between the two features shows that they might not have a strong relationship and

good performance, but the model performed well. Figure 28g shows the loss function of the model. Then the model is trained and saved afterward. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.71	0.70	0.70	30

7.4.8 Cysts (C) and DRIL

The eighth combination is Cysts (C) and DRIL, and the correlation between the two mentioned features, based on the cluster map in figure 18, is 0.4, which means these two features are moderately correlated. In this subsection, a model trained on both features is a binary classification in which an image can either have Cysts and DRIL or none of these features. For this binary classification, 400 images are separated for the training set and 40 images for the test set (200 images for each class in training and 20 images for each class in the test set). The model was expected to perform well as these two features are moderately correlated and might have a strong relationship, but it performed surprisingly worse. The reason could be that the model is confused between images with cysts and disorganized retinal layers (because of their similarities), and the model predicted the wrong labels. Figure 28h shows that no overfitting is observed in this model. Then the model is trained and saved afterward. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.59	0.57	0.56	40

7.4.9 Cysts (C) and EZ/ELM

The ninth combination is Cysts (C) and EZ/ELM, and the correlation between the two mentioned features, based on the cluster map in figure 18 is 0.25, which means these two features are weekly correlated. In this binary classification, there are only 38 images that have both features simultaneously, and it is not a sufficient number of images for training a model. So, the result of this model is not available due to this issue, and it can be solved by adding a considerably more number of available images for the training and testing process. It is descoped from this research study.

7.4.10 DRIL and EZ/ELM

The last combination is DRIL and EZ/ELM, and the correlation between the two mentioned features, based on the cluster map in figure 18, is 0.37, which means these two features are moderately correlated. In this subsection, a model trained on both features is a binary classification in which an image can either have DRIL and EZ/ELM or none of these features. For this binary classification, 220 images are separated for the training set and 20 images for the test set (110 images for each class in training and 10 images for each class in the test set). Several fluctuations are observed in their loss and accuracy functions (figure 28i), and the reason is the few images in training and especially validation sets. The issue might be solved by adding more images to the datasets. Also, the result of the testing set might not be accurate and reliable due to the shortage in the number of images. Then the model is trained and saved afterward. The classification report of the model before transfer learning is as follows:

Classification report	Precision	Recall	F1-Score	Support
Average accuracy	0.74	0.70	0.69	20

7.4.11 Comparison of combined features

All 10 models with different feature combinations were studied in the previous subsection. The combination of Thickening (T) and Macular Volume (MV) performed the best among them. After that, Thickening (T) and Cysts (C) performed the second best because the results of the combination of DRILL and EZ/ELM are not undoubtedly reliable and accurate due to the shortage in the number of images for the test (20 images) and training process. The more images for the training and testing process, the more precise and reliable results will be. The result of the Cysts and EZ/ELM is also insufficient and not comparable with the other combinations, and was expected a better performance for the combination of Cysts and DRIL, and reason is the uncertainty of the model for the prediction between cysts and disorganized retinal layers. Then all 9 sufficient models are saved to be used via transfer learning to predict DME on a different and more extensive dataset.

Feature Combination	Precision	Recall	F1-Score	Support
Thickening & Macular Volume	**0.80**	**0.75**	**0.74**	**40**
Thickening & Cysts	0.72	0.70	0.69	40
Thickening & DRIL	0.71	0.63	0.60	30
Thickening & EZ / ELM	0.68	0.65	0.64	40
Macular Volume & Cysts	0.78	0.77	0.76	30
Macular Volume & DRIL	0.63	0.62	0.62	40
Macular Volume & EZ / ELM	0.71	0.70	0.70	30
Cysts & DRIL	0.59	0.57	0.56	40
Cysts & EZ / ELM	-	-	-	-
DRIL & EZ / ELM	0.74	0.70	0.69	20

7.4.12 DME detection with the usage of features combination

In the previous section, the combination of features was used, and 10 possible combinations were studied. Then, all 9 sufficient models were saved, and except their last 5 layers, all their layers and weights were frozen and trained on another dataset, as mentioned in Section 7.2.8. Models for each feature are implemented in the same situation, with the same amount of images for each set, the same data pre-processing, and the same epochs to see the differences between the results to compare them more accurately.

In this section, all 9 models are trained again with a low learning rate (0.001), and the loss and accuracy functions behaved naturally, and no overfitting can be observed. Each model is trained and tested many times with the SGD and the Adam optimizers, and the best one for all models was the Adam optimizer. All of the models performed remarkably well, but the difference in their results can be a term for comparing models trained on different features. The table below shows the performance of each influential feature with the usage of transfer learning in DME detection. Thickening (T) & Macular Volume (MV) combination performed the best among the 10 combinations. However, the combination of Thickening (T) and EZ/ELM performed the second best and close to the first. Furthermore, these two mentioned combinations are more influential and prominent in diagnosing DME. The number of images for the DME dataset can ensure reliability and accuracy for the results achieved. The table shows all possible combinations of features in a comparable condition as follows:

Feature Combination	Precision	Recall	F1-Score	Support
Thickening (T) & Macular Volume (MV)	**0.98382**	**0.98333**	**0.98333**	**300**
Thickening (T) & Cysts (C)	0.968	0.96667	0.96664	300
Thickening (T) & DRIL	0.95914	0.95667	0.95661	300
Thickening (T) & EZ / ELM	0.98077	0.98	0.97	300
Macular Volume (MV) & Cysts (C)	0.9717	0.97	0.96	300
Macular Volume (MV) & DRIL	0.95625	0.95333	0.95326	300
Macular Volume (MV) & EZ / ELM	0.97771	0.97667	0.97665	300
Cysts (C) & DRIL	0.96875	0.96667	0.96663	300
Cysts (C) & EZ / ELM	-	-	-	-
DRIL & EZ / ELM	0.95732	0.95333	0.95323	300

7.5 Discussion

The three goals of the thesis, mentioned in Section 1, are fulfilled separately to detect DME with average specificity and sensitivity of 96%. Relationships between morphological features are studied using Spearsman's correlation. The results show that the correlation between Thickening (T) and Macular Volume (MV) is the strongest, with the highest value (0.82). The classification of features individually with the highest number of available images for each feature in the dataset is done. The results indicate that Thickening (T) has the highest potential to enhance the accuracy of DME diagnosis. The identification of features by equally leveling the number of images of each feature is accomplished. The results demonstrate that Cysts (C) is the most noticeable singleton feature that boosts the DME detection if only the presence or absence of features is considered. The detection of 10 possible combinations of two features simultaneously is done with an equal number of images for each feature. The results show that the combination of Thickening (T) and Macular Volume (MV) is the most prominent combination of all possible combinations to increase the accuracy of DME detection.

The results for the three goals are achieved satisfactorily, but in some cases, uncertainty and unreliability are observed. Examining the 8 mentioned features allowed a more accurate prediction of DME with increased specificity and sensitivity for detecting DME. However, no specific and clear boundaries were observed for the diagnosis of DME because, in some cases such as DRIL (Section 7.3.1), the feature did not detect clearly due to some limitations, such as the shortage in the number of images for some features based on their labels, and the complexity of features in terms of their nature.

The classification of DME with the highest number of available images for each feature in the dataset indicates the best possible performance for each feature individually. However, it cannot guarantee a proper comparison among features because of the difference in the number of images for each feature. The results of the classification of features singularly is a method to show which feature plays the most significant role in detecting DME more accurately in the larger scale of images. In this thesis, Macular Volume (MV) and Thickening (T) performed closely and the best to be classified on each OCT image, and usage of those features in the detection of the DME. However, Macular volume got a better specificity of 77.4% and sensitivity of 75% than Thickening with a specificity of 65% and sensitivity of 60% in the classification of features individually. In addition, after transfer learning, the model used Thickening (specificity of 99% and sensitivity of 99%) performed slightly better than the

model that used Macular volume (specificity of 98.5% and sensitivity of 98.5%) in the DME diagnosis.

The detection of DME with the same number of images enables comparing the features to figure out the best individual feature among the available features with a limited and equal number of images. Having the same number of images for each feature does not show the best achievable result individually, but it provides a more reasonable comparison to study the features separately. Furthermore, this method enables valuable insights to study the DME accuracy in case only the presence or absence of features are considered. The results of the models showed that Thickening (T) was detected more accurately and straightforwardly than the other features individually (specificity of 75.5% and sensitivity of 75%) and had the highest potential for further diagnosing DME. However, The results show Cysts (C) (specificity of 97.6% and sensitivity of 97.6%) performed negligibly better than Thickening (T) (specificity of 96.7% and sensitivity of 96.6%), in DME detection.

Detecting combinations of features on OCT images with an equal amount of images can indicate the most reasonable comparison between all possible combinations considering the correlation between every two features. The results of the models in the combinations of features show that Thickening (T) & Macular Volume (MV) (specificity of 77.5% and sensitivity of 75%) are detected most accurately among all the combinations of features. In addition, their combination is the best among all combinations of features for further detection of DME (specificity of 98.3% and sensitivity of 98.3%). However, Thickening (T) & EZ/ELM performed surprisingly nearly the best (specificity of 98% and sensitivity of 98%). The essential role of Thickening (T) in the detection of DME is remarkable in considering singular features and combining them. In addition, the correlation between Thickening and Macular volume and the more accurate detection of them simultaneously on OCT images indicates their reliability for DME diagnosis in comparison with the combination of Thickening and EZ/ELM.

Table 2 compares the two best features in each of the three stages (classification and detection of features individually, detection of combined features). Furthermore, figure 19 illustrates a proper visualized relationship between the correlation, classification report, and the number of images used for training and testing processes for each combination of features and also in the detection of DME.

The work done by Panozzo et al. [37] studied seven features mentioned in Section 2.1 and proposed a grading protocol called TCED-HFV, in which each letter represents a feature. The grading protocol categorizes DME in four stages showing disease progression. A lack of the usage of machine learning techniques to identify and categorize DME is notable in their work. This thesis used machine learning techniques to classify the features automatically and compare them to investigate the most influential feature in detecting DME more accurately. In addition, all possible combinations of two features are used to diagnose DME. A lightweight deep learning architecture called OCTNet is used in all created models. The lightweight OCTNet architecture boosts the accuracy and reduces the computation time than heavy-weight architectures such as ResNet-50 or ImageNet [42].

Table 2: A comparison between sensitivity and specificity of DME detection among the usage of single features and combinations of them in this thesis.

Stage	Feature(s)	Sensitivity	Specificity
Single feature classification	Thickening (T)	99%	99%
	Macular Volume (MV)	98.5%	98.5%
Single feature detection	Cysts (C)	97.6%	97.6%
	Thickening (T)	96.6%	96.7%
Combination of features	Thickening(T) & Macular Volume (MV)	98.3%	98.3%
	Thickening (T) & EZ/ELM	98%	98%

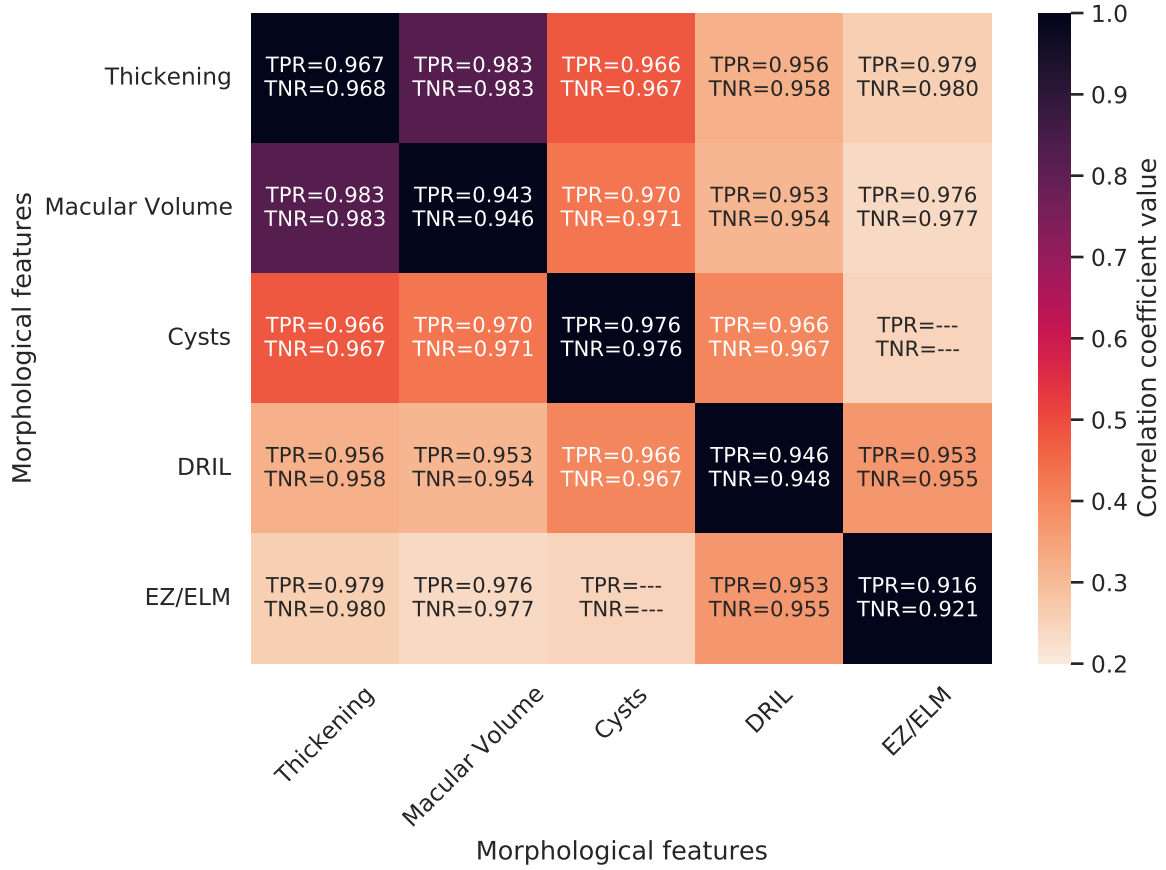


Figure 19: The performance of individual morphological features and their combinations for detecting DME on OCT images. The color bar on the right side indicates a value of the correlation coefficient between features. Each cell consists of two values: the top one is Sensitivity (TPR). The bottom one is Specificity (TNR). TPR and TNR are calculated for each feature and all combinations between them in the prediction of DME. The intersection cell of each feature with itself shows the performance of the model for the individual feature in the detection of DME.

Table 3: This table shows a comparison among related work done to detect or classify DME with this thesis.

Authors	Dataset	Sensitivity	Specificity	Classes
S. Amit Kamran et al. [30]	OCT2017 [26] 84,484 images	99.8%	99.9%	Normal - Drussen CNV - DME
S. Amit Kamran et al. [30]	Srinivasan2014 [15] 3,231 images	100%	100%	Normal - AMD DME
K. Alsaih et al. [21]	3800 images	87.5%	87.5%	DME - Normal
Sunija A.P et al. [42]	83,484 images [42]	99.6%	99.6%	Normal - Drussen CNV - DME
Current Study	3000 images	96%	96%	Normal - DME

The work by S. Amit Kamran et al. [30] demonstrated a highly accurate automated system to diagnose retinal disease using OCT images. As mentioned in Section 2.3, five morphological features are considered in their work to identify DME. They proposed an architecture that does not require any pre-trained weights, and it facilitates the training and deployment time of the model by many folds. Their work used two popular OCT datasets, and the model performed flawlessly for both datasets. The model achieved a sensitivity and specificity of 99.80% and 99.93% in the OCT2017 [26] dataset. Furthermore, the model achieved a state-of-the-art result by scoring 100% accuracy, sensitivity, and specificity in the Srinivasan2014 [15] dataset. For this thesis, in comparison to the work by S. Amit Kamran et al. [30], the dataset that used for DME diagnosis is 40 times smaller than the OCT2017 dataset, but still, the average specificity and sensitivity (96%) for both individual and combined features is this project, which is close to their work. On the other hand, their work achieved perfect specificity and sensitivity (100%) in the Srinivasan2014 dataset, which consists of 3,231 OCT images. However, the result of their work for the Srinivasan2014 dataset is more satisfactory than this thesis.

The work done by K. Alsaih et al. [21] proposed an automatic classification framework for SD-OCT images to identify DME versus normal images based on evaluating several morphological features, such as retinal thickening, hard exudates, intraretinal cystoid space formation, and subretinal fluid. Their study used more than 3800 OCT images for the DME detection. A pipeline has been proposed, including pre-processing, feature detection, feature representation techniques, and using classifiers to detect DME satisfactorily. Individual and combined features are assessed in their work, and the best classification performance with sensitivity and specificity of 87.5% was accomplished. Compared to this thesis, they used a bigger dataset to diagnose DME. However, the sensitivity and specificity of their work are lower. Furthermore, some morphological features, such as retinal thickening and subretinal fluid, are used in their work, similar to the features used in this thesis.

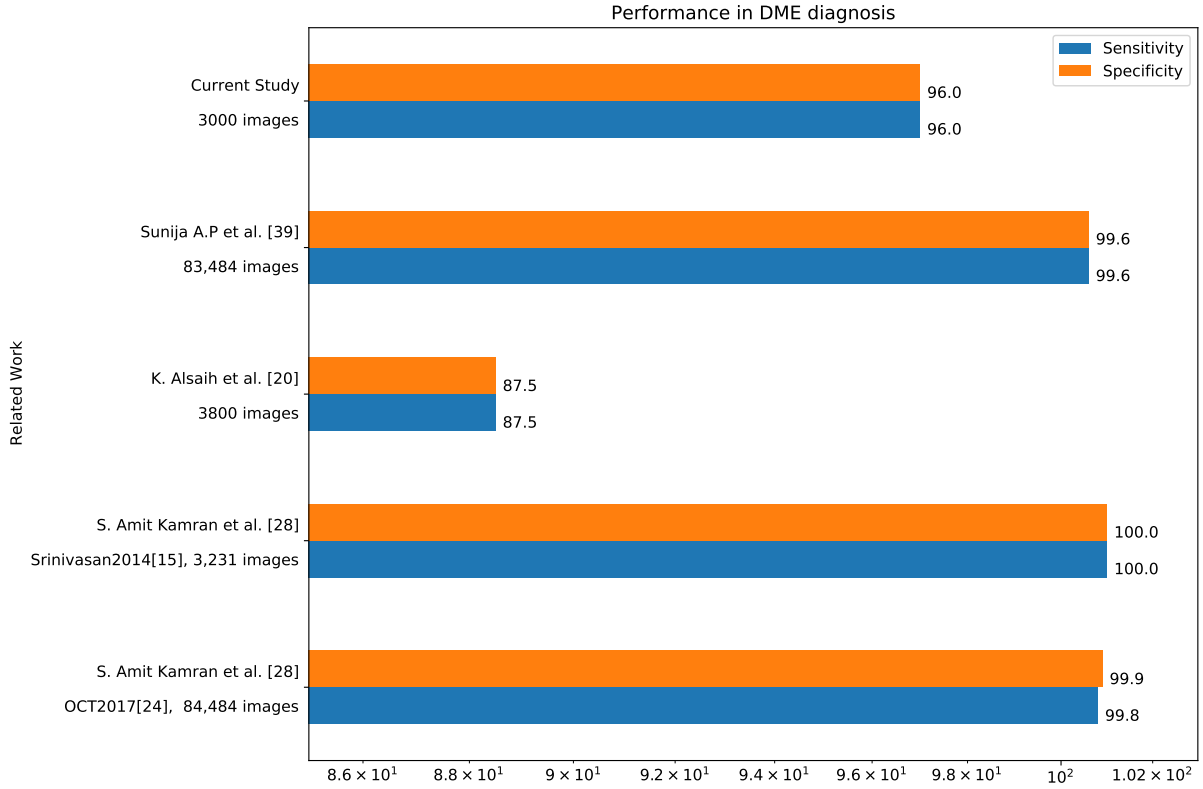


Figure 20: A comparison between the performance of different models in related work with this project.

8 Conclusion

Diabetes and its causes of visual impairment in people is one of the most common diseases globally. DME is one of the products of diabetes worldwide, influencing numerous people annually. There are various methods for examining DM, such as SD-OCT. The main focus of this thesis was instructed on a learning-supported visual investigation of DME pathology in the retina associated with diabetes using eight morphological features on OCT images.

Deep learning technology has great potential in classifying medical images, especially in ophthalmology, where different modalities such as OCT and fundus images can be used. Different disease methodologies have been detected and classified using deep learning techniques, including cataracts, glaucoma, age-related macular degeneration, and DR [25]. DR is a complication of diabetes that can lead to progressive stages such as DME [38]. OCTNet architecture introduced by Sunija A.P et al. [42], which is a lightweight and still one of the most potent architectures, is used in this thesis to classify and detect the morphological features and then diagnoses DME on OCT images.

The models mostly performed well in classifying and detecting single and combined features. Five single features with the highest possible available images are classified. The five single features are also detected on OCT images with an equal number of images in their training and test sets for an appropriate comparison among features. Also, ten possible combinations of two features are simultaneously detected on OCT images with an equal number of images for the training and test sets. Then the performance of all the fifteen single and combined features models are compared. All the models are used to diagnose DME using transfer learning in the same situation in their training processes.

The results indicate the significant role of Thickening (T) for both detected and classified OCT images on the dataset and also for the detection of DME. The Thickening (T) classification model performed the best among morphological features with an average sensitivity and specificity of 99%. The Thickening (T) detection model performed the second best (after the Cysts(C) detection model) with an average sensitivity and specificity of 96.7%. In DME detection using a combination of two features, Thickening (T) & Macular Volume (MV), which have the most robust correlation coefficient among the features, performed the best with the average sensitivity and specificity of 98.3%.

Although this thesis delivers some promising results, several limitations have to be raised, such as invalid labels for some images, a shortage of available images for adjunctive features, class imbalance for some features among their labels, and the limited number of images in the dataset. These difficulties caused some issues in the reliability and certainty of data and the results, but the data cleaning for invalid labels and data augmentation for class imbalance tried to improve the accuracy, reliability, and certainty of the results. Additional experiments need to be carried out on a larger dataset to increase the reliability of the results.

9 Future Work

This thesis concentrated on the detection of DME with the usage of eight morphological features on OCT images. There are other types of optical images, such as Fundus images. One of the other challenges can be detecting DME using Fundus and OCT images for each patient. DME can be detected by measuring the retinal thickness in both fundus and OCT images. As all the retinal layers are distinguishable in OCT images, detecting DME is more straightforward than in Fundus images. However, Fundus images can indicate the thickness and hard exudates in the macular region. Also, "the OCT image shows the same changes by giving a cross-sectional view of sub-retinal layers [17]". So, by having both OCT and fundus images, DME might be detected easier and more accurately. One of the disadvantages of fundus images is that observing cysts are so tricky, and they are not prominent in fundus images [17].

Another future challenge can be DME diagnosis on a huge scale of data. Classification or detection using machine learning techniques on large data is more accurate than a smaller amount of data. So, having a large amount of OCT images makes DME detection more accurate and reliable. The more images a model sees, the more accurate the prediction for unseen images will be. The number of images in a dataset is a reason for accurate results [56]. Some data augmentation techniques are used in this thesis. Still, various over-sampling techniques can increase the amount of data and solve class imbalance problems. In addition, OCTNet has been chosen as the deep learning architecture for this thesis due to its lightweight and still powerful architecture for the OCT images, but some other powerful architectures, such as OpticNet, can be used in future works.

One more crucial future challenge in continuing this project is the classification of DME on OCT images in four comparable stages mentioned by Pannozo et al. [37] with the usage of machine learning techniques. As mentioned in Section 2, DME can be categorized in four comparable stages, Early DME, Advanced DME, Severe DME, and Atropathic maculopathy. One of the advantages of the classification of DME into four stages is that its detection can have more insights into the progression of the disease rather than its detection only. In work done by Panozzo et al. [37], the classification of DME in the four stages is done without using machine learning techniques. Instead, a grading system called TCED-HFV was used. Hence, with the usage of deep learning and machine learning techniques, the proper stage of the DME can also be classified. DME is detected in this thesis using the features individually and in the combination of two features. Still, the detection can be done using three or more features simultaneously on OCT images. Besides that, some other features, such as central foveal thickness (CFT), have been used in their work in addition to the features mentioned in this thesis, and those features can be studied in future work.

References

[1] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *International conference on advances in pattern recognition*, Springer, 2001, pp. 399–408.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Ieee, vol. 1, 2005, pp. 886–893.

[3] Y. Dodge, *The concise encyclopedia of statistics*. Springer Science & Business Media, 2008.

[4] N. Bhagat, R. A. Grigorian, A. Tutela, and M. A. Zarbin, "Diabetic macular edema: Pathogenesis and treatment," *Survey of ophthalmology*, vol. 54, no. 1, pp. 1–32, 2009.

[5] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.

[6] H. Faghihi, F. Hajizadeh, and M. Riazi-Esfahani, "Optical coherence tomographic findings in highly myopic eyes," *Journal of ophthalmic vision research*, vol. 5, pp. 110–21, Apr. 2010.

[7] J. Nagi, F. Ducatelle, G. A. Di Caro, *et al.*, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *2011 IEEE international conference on signal and image processing applications (ICSIPA)*, IEEE, 2011, pp. 342–347.

[8] S. Tick, F. Rossant, I. Ghorbel, *et al.*, "Foveal shape and structure in a normal population," *Investigative ophthalmology & visual science*, vol. 52, no. 8, pp. 5105–5110, 2011.

[9] J. Ding and T. Y. Wong, "Current epidemiology of diabetic retinopathy and diabetic macular edema," *Current diabetes reports*, vol. 12, no. 4, pp. 346–354, 2012.

[10] L. Giancardo, F. Meriaudeau, T. P. Karnowski, *et al.*, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Medical image analysis*, vol. 16, no. 1, pp. 216–226, 2012.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[12] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[13] J. W. Yau, S. L. Rogers, R. Kawasaki, *et al.*, "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes care*, vol. 35, no. 3, pp. 556–564, 2012.

[14] S. Jorgensen and B. Fath, *Encyclopedia of Ecology*. Elsevier Science, 2014, ISBN: 9780080914565.

[15] P. P. Srinivasan, L. A. Kim, P. S. Mettu, *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical optics express*, vol. 5, no. 10, pp. 3568–3577, 2014.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[17] T. Hassan, M. U. Akram, B. Hassan, A. Nasim, and S. A. Bazaz, "Review of oct and fundus images for detection of macular edema," in *2015 IEEE International Conference on Imaging Systems and Techniques (IST)*, IEEE, 2015, pp. 1–4.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. arXiv: 1512.03385.

[19] F.-F. Li, A. Karpathy, and J. Johnson, "Cs231n: Convolutional neural networks for visual recognition," *University lecture*, 2015.

[20] M. M. Nentwich and M. W. Ulbig, "Diabetic retinopathy-ocular complications of diabetes mellitus," *World journal of diabetes*, vol. 6, no. 3, p. 489, 2015.

[21] K. Alsaih, G. Lemaitre, M. Rastgoo, J. Massich, D. Sidibé, and F. Meriaudeau, "Machine learning techniques for diabetic macular edema (dme) classification on sd-oct images," *Biomedical engineering online*, vol. 16, no. 1, pp. 1–12, 2017.

[22] Arden Dertat, *Part 4: Convolutional neural networks*, https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2, 2017.

[23] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: A taxonomy," *arXiv preprint arXiv:1710.10686*, 2017.

[24] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *towards data science*, vol. 6, no. 12, pp. 310–316, 2017.

[25] P. S. Grewal, F. Oloumi, U. Rubin, and M. T. Tennant, "Deep learning in ophthalmology: A review," *Canadian Journal of Ophthalmology*, vol. 53, no. 4, pp. 309–313, 2018.

[26] D. S. Kermany, M. Goldbaum, W. Cai, *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[27] B. B. Traore, B. Kamsu-Foguem, and F. Tangara, "Deep convolution neural network for image recognition," *Ecological Informatics*, vol. 48, pp. 257–268, 2018.

[28] S. Walczak, "Artificial neural networks," in *Encyclopedia of Information Science and Technology, Fourth Edition*, IGI global, 2018, pp. 120–131.

[29] Y.-c. Wu and J.-w. Feng, "Development and application of artificial neural network," *Wireless Personal Communications*, vol. 102, no. 2, pp. 1645–1656, 2018.

[30] S. Amit Kamran, S. Saha, A. Shihab Sabbir, and A. Tavakkoli, "Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images," *arXiv e-prints*, arXiv–1910, 2019.

[31] I. Hecht, A. Bar, L. Rokach, *et al.*, "Optical coherence tomography biomarkers to distinguish diabetic macular edema from pseudophakic cystoid macular edema using machine learning algorithms," *Retina*, vol. 39, no. 12, pp. 2283–2291, 2019.

[32] P. Saeedi, I. Petersohn, P. Salpea, *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas," *Diabetes research and clinical practice*, vol. 157, p. 107 843, 2019.

[33] M. Yani, S. Irawan, and C. Setianingsih, "Application of transfer learning using convolutional neural network method for early detection of terry's nail," *Journal of Physics: Conference Series*, vol. 1201, p. 012 052, May 2019.

[34] Y. H. Yoon, D. S. Boyer, R. K. Maturi, *et al.*, "Natural history of diabetic macular edema and factors predicting outcomes in sham-treated patients (mead study)," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 257, no. 12, pp. 2639–2653, 2019.

[35] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," *arXiv preprint arXiv:2009.07485*, 2020.

[36] A. D. Moraru, D. Costin, R. L. Moraru, and D. C. Branisteanu, "Artificial intelligence and deep learning in ophthalmology-present and future," *Experimental and Therapeutic Medicine*, vol. 20, no. 4, pp. 3469–3473, 2020.

[37] G. Panozzo, M. V. Cicinelli, A. J. Augustin, *et al.*, "An optical coherence tomography-based grading of diabetic maculopathy proposed by an international expert panel: The european school for advanced studies in ophthalmology classification," *European journal of ophthalmology*, vol. 30, no. 1, pp. 8–18, 2020.

[38] R. K. Singh and R. Gorantla, "Dmenet: Diabetic macular edema diagnosis using hierarchical ensemble of cnns," *Plos one*, vol. 15, no. 2, e0220677, 2020.

[39] F. Zhuang, Z. Qi, K. Duan, *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[40] Kaggle.com, *Retinal layers*, https://www.kaggle.com/datasets/basharalkuwaiti/oct-csv, 2021.

[41] S. Rajaraman, G. Zamzmi, and S. K. Antani, "Novel loss functions for ensemble-based medical image classification," *Plos one*, vol. 16, no. 12, e0261307, 2021.

[42] A. Sunija, S. Kar, S. Gayathri, V. P. Gopi, and P. Palanisamy, "Octnet: A lightweight cnn for retinal disease classification from optical coherence tomography images," *Computer methods and programs in biomedicine*, vol. 200, p. 105 877, 2021.

[43] Q. Wu, B. Zhang, Y. Hu, *et al.*, "Detection of morphologic patterns of diabetic macular edema using a deep learning approach based on optical coherence tomography images," *Retina (Philadelphia, Pa.)*, vol. 41, no. 5, p. 1110, 2021.

[44] David Turbert, *What is optical coherence tomography*, https://www.aao.org/eye-health/treatments/what-is-optical-coherence-tomography, 2022.

[45] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, 2022.

[46] Y. Park, "Concise logarithmic loss function for robust training of anomaly detection model," *arXiv preprint arXiv:2201.05748*, 2022.

[47] 2mel.nl, *Bio photonic sensing and imaging techniques*, https://www.2mel.nl/projects/photonic-sensor-laser-technology-to-test-treat-bladder-cancer/#:~:text=Optical%20coherence%20tomography%20(OCT)%20is,industrial%20nondestructive%20testing%20(NDT)..

[48] Aditya Mishra, towardsdatascience.com, https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234.

[49] Ganesh Ram, *Launching retinalyze oct – detect over 100 pathological signs in 45 seconds*, https://www.retinalyze.com/post/launching-retinalyze-oct-detect-over-100-pathological-signs-in-45-seconds.

[50] Greg Chu, www.medium.com, https://medium.com/deeplearningsandbox/how-to-use-transfer-learning-and-fine-tuning-in-keras-and-tensorflow-to-build-an-image-recognition-94b0b02444f2.

[51] International Diabetes Federation, *Diabetic macular edema*, https://www.idf.org/component/attachments/task=download&id=2153#:~:text=More%20than%2021%20million%20people,at%20particular%20risk%20for%20DME..

[52] Juhi Ramzai, *Pearson v/s spearman correlation coefficient*, https://towardsdatascience.com/clearly-explained-pearson-v-s-spearman-correlation-coefficient-ada2f473b8..

[53] Lee Vien, OD, FAAO, *Oct image for diabetic macular edema with intera retinal cysts*, https://www.octscans.com/diabetic-macular-edema.html.

[54] H. Y. Li, D. X. Wang, L. Dong, and W. B. Wei, "Deep learning algorithms for detection of diabetic macular edema in oct images: A systematic review and meta-analysis," *Available at SSRN 3824687*,

[55] Prabhu, *Medium.com*, https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.

[56] Sciforce, *Robust image classification with a small data set*, https://medium.com/sciforce/robust-image-classification-with-a-small-data-set-be4de9897495.

[57] www.medium.com, https://adiksoni095.medium.com/transfer-learning-is-the-reuse-of-a-pre-trained-model-on-a-new-problem-a71559072a10.

[58] www.oreilly.com, https://www.oreilly.com/library/view/machine-learning-for/9781786469878/252b7560-e262-49c4-9c8f-5b78d2eec420.xhtml.

[59] Yulia Gavrilova, *Convolutional neural networks for beginners*, https://serokell.io/blog/introduction-to-convolutional-neural-networks..

[60] Madhushree Basavarajaiah, *Max-pooling vs min-pooling vs average-pooling by madhushree basavarajaiah*, https://medium.com/@bdhuma/which-pooling-method-is-better-maxpooling-vs-minpooling-vs-average-pooling-95fb03f45a9, Feb 8, 2019.

Appendices

A Figures

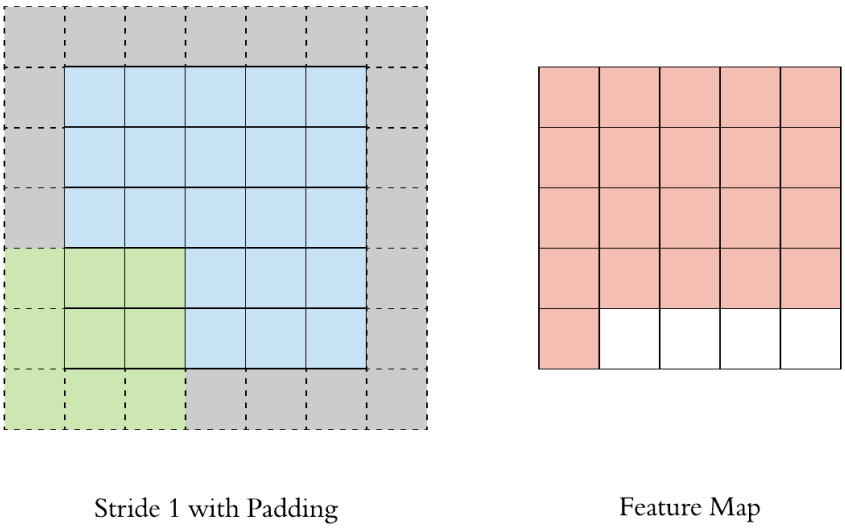


Figure 21: The gray area around the input is the padding. Either pad with zeros or the values on the edge, Then the dimensionality of the feature map matches the input [22].

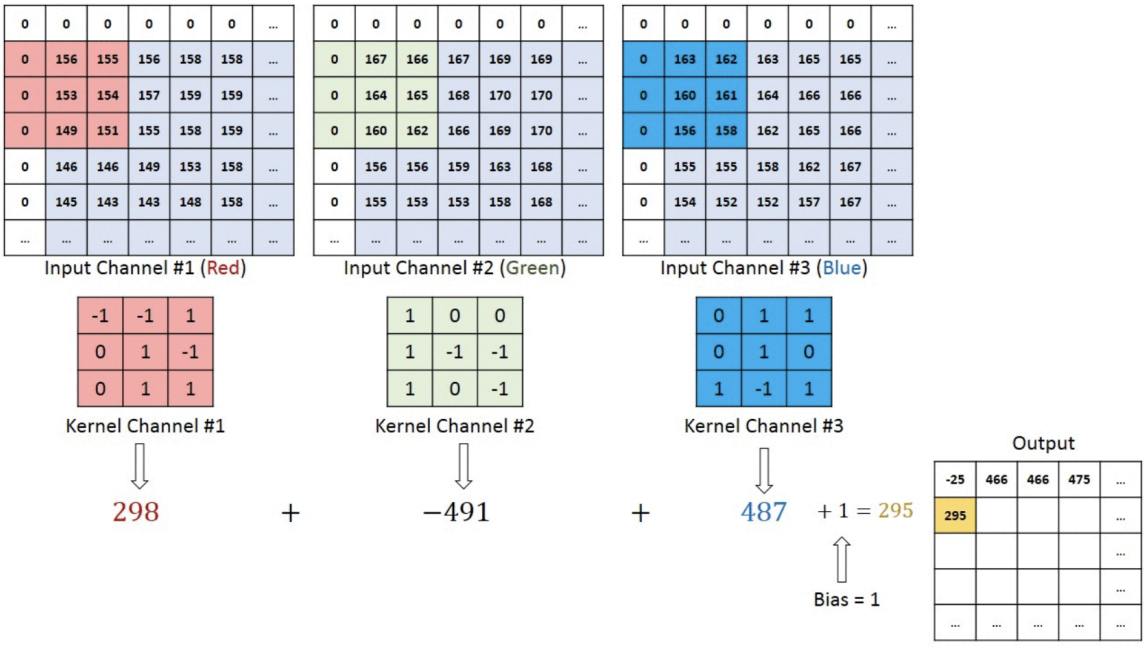


Figure 22: Illustrates the convolution operation for an multi-color image with 3 channels (RGB); The output is squashed or aggregated as the output. image credit: (www.guru99.com, 2021) .

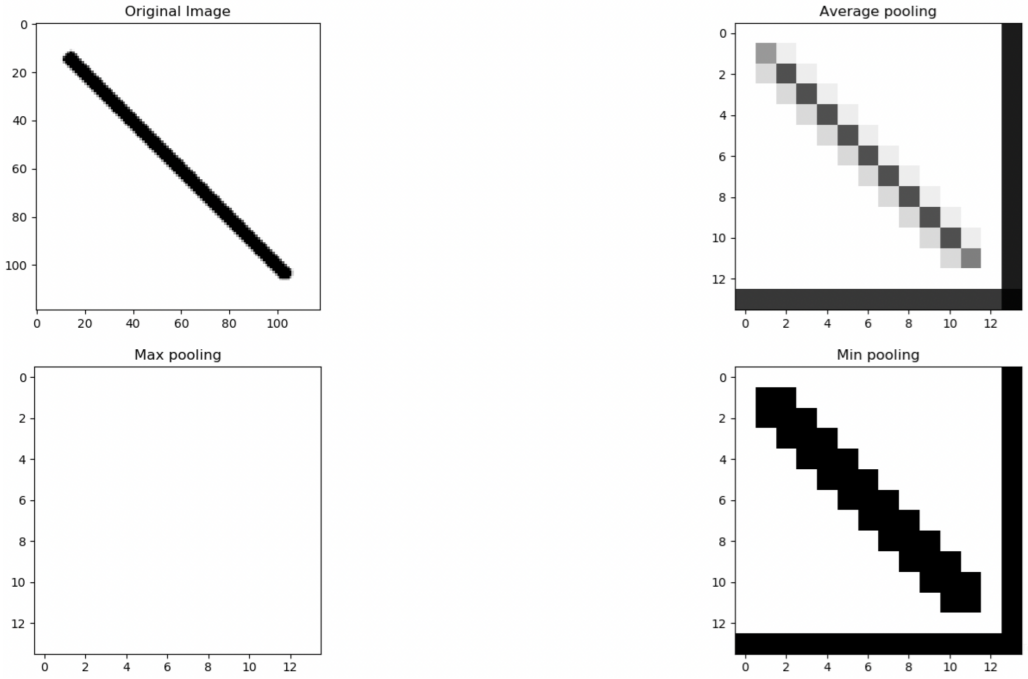


Figure 23: This figure illustrates how four different pooling method perform for a black and white image and Min-pooling yields a better result for images with white background and black object in it. [60].

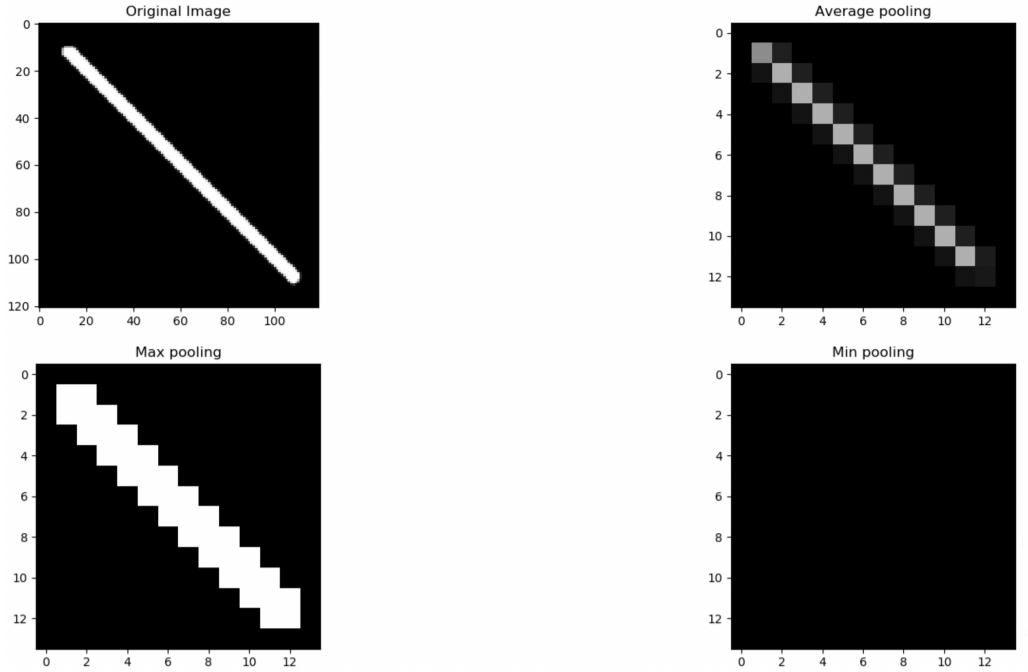


Figure 24: This figure illustrates how four different pooling method perform for a black and white image and Max pooling gives better result for the images with black background and white object (Ex: MNIST dataset). [60].

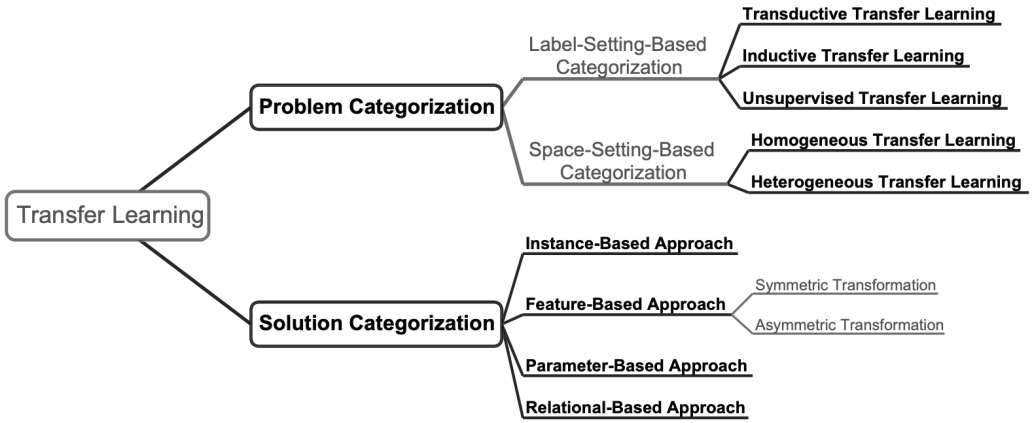


Figure 25: This figure shows the categorizations of transfer learning [39].

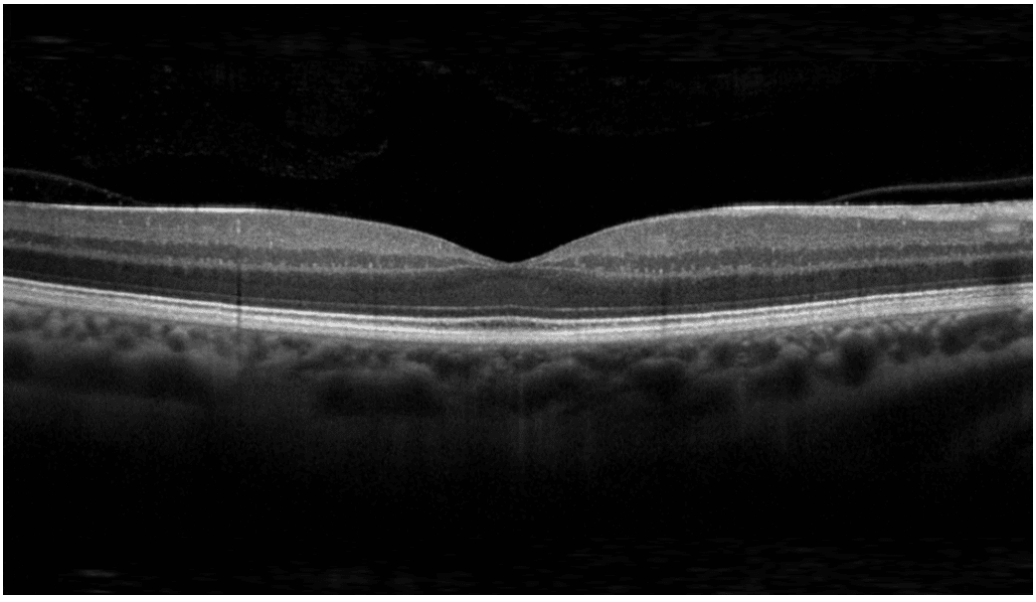
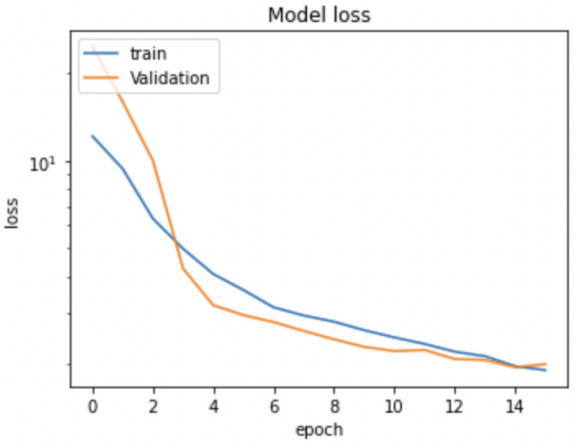
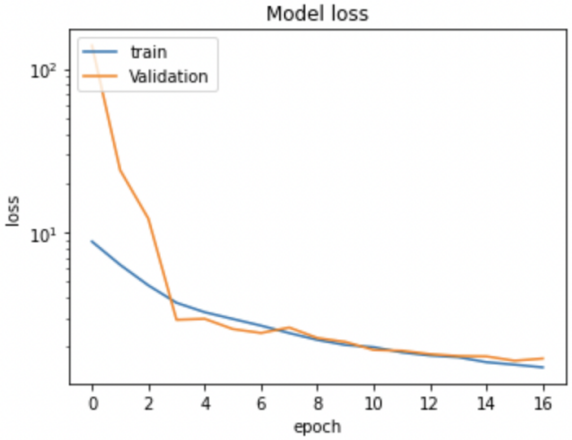


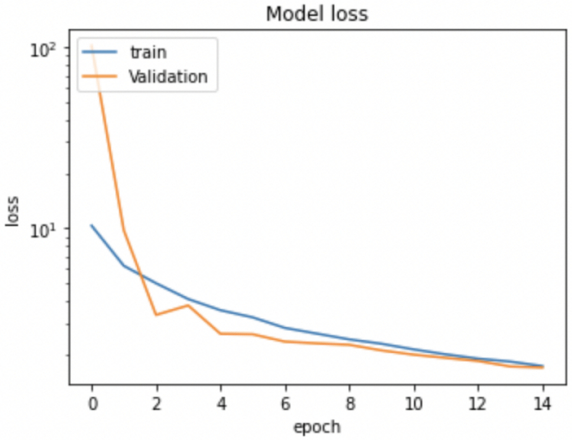
Figure 26: This figure shows a healthy retinal layers in an OCT images in which ophthalmologist can see all the deeper retinal layers to see if a patient has any disease or not [49].



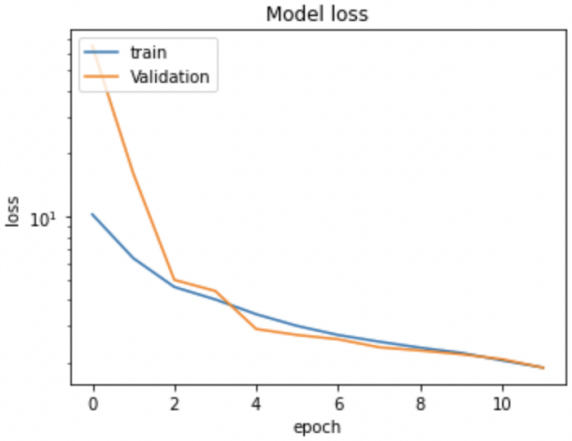
(a) Thickening (T) loss function.



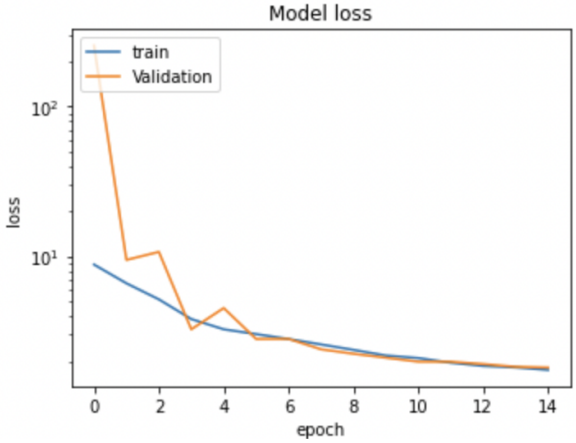
(b) Macular Volume (MV) loss function.



(c) Cysts (C) loss function.

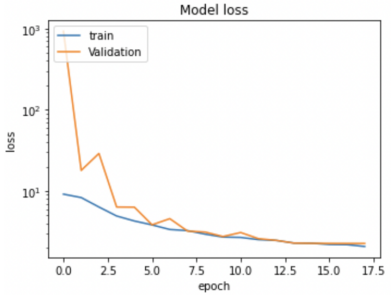


(d) DRIL loss function.

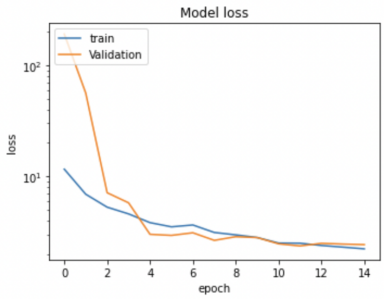


(e) EZ/ELM loss function.

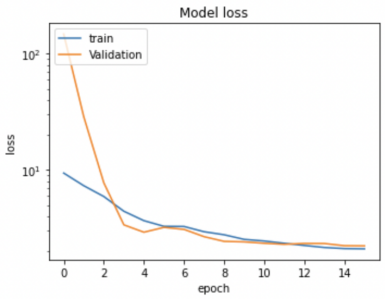
Figure 27: The training and validation loss functions of the classification of each influential feature.



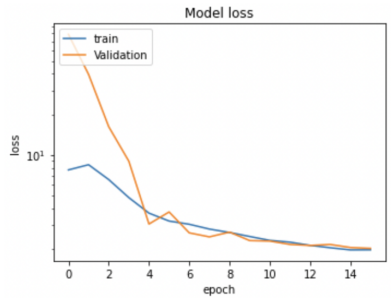
(a) Thickening and Macular Volume loss.



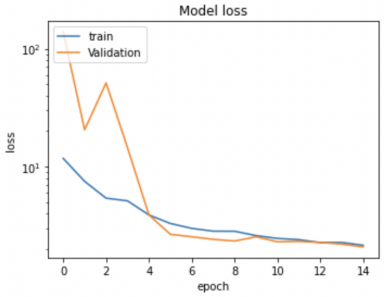
(b) Thickening and Cysts loss.



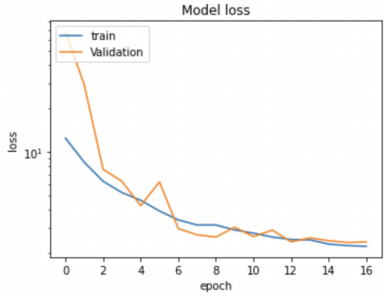
(c) Thickening and DRIL loss.



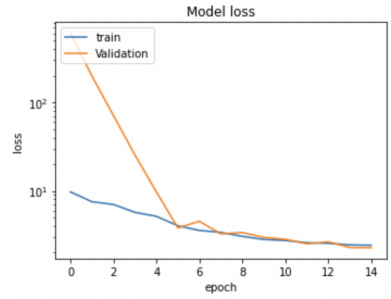
(d) Thickening and EZ/ELM loss.



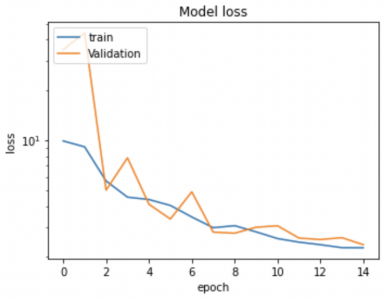
(e) Macular volume and Cysts loss.



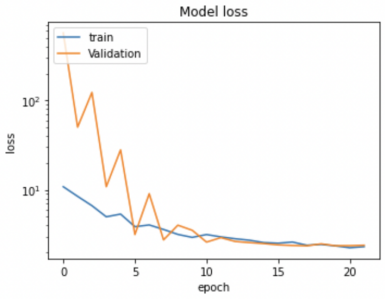
(f) Macular volume and DRIL loss.



(g) Macular volume and EZ/ELM loss.



(h) Cysts and DRIL loss.



(i) EZ/ELM and DRIL loss.

Figure 28: The training and validation loss functions of the detection of combined features.