Syrian Arab Republic

Ministry of Higher Education and Scientific Research

Syrian Virtual University

الجمهوريّة العربيّة السوريّة

وزارة التّعليم العالي والبحث العلميّ

الجامعة الافتراضيّة السوريّة

Deep learning Model to diagnose retinal diseases using Optical Coherence Tomography images (OCT)

(A thesis submitted as a fulfilment of requirements for a Master's degree in Bioinformatics)

By

Reema Othman Melhem

Supervised by

Dr. Yanal Ahmad Alkuddsi

2023 – 2024

Abstract:

Background: There are numerous eye conditions but the most two common retinal conditions are Age- Related Macular Degeneration (AMD), which the sharp, central vision and a leading cause of vision loss among people age 50 and aged, there are two types of AMD are wet AMD (Choroidal neovascularization (CNV)) and dry AMD (DRUSEN). Diabetic Macular Edema (DME), which is a complication of diabetes that can affect the fovea and it caused by fluid accumulation in the macula. thus, early discovery of conditions is a critical significance. The goal of this thesis is implementation of deep learning model used to detect four types of retinal cases (NORMAL, CNV, DME, and DRUSEN) by using Convolutional Neural Network (CNN) to Avoid manual diagnostic errors and help doctors make faster, and more accurate diagnoses.

Aim: The goal of this thesis is implementation of deep learning model used to detect four types of retinal cases (NORMAL, CNV, DME, and DRUSEN) by using Convolutional Neural Network (CNN) to Avoid manual diagnostic errors and help doctors make faster, and more accurate diagnoses.

Materials and Methods: To diagnose retinal disorders utilizing Optical Coherence Tomography (OCT) scans, this investigation developed a methodology based on image pre-processing and convolutional neural networks (CNNs) (a deep learning method). A publicly available OCT dataset containing four categories (normal, CNV, DME, drusen) was used. Data imbalance was addressed using oversampling techniques. Image pre-processing techniques including images resizing, noise reduction. The AlexNet architecture, a convolutional neural network (CNN), was utilized for image classification. The model comprised five convolutional layers, fully-connected layers, and dropout layers. Categorical cross-entropy loss function and Adam optimizer were employed. Overfitting was mitigated using techniques like learning rate reduction, early stopping, and model checkpointing. Training and validation metrics were logged for analysis.

Results: Our deep learning model achieved high accuracy in classifying retinal diseases from OCT images. The model achieved on the test set an accuracy of 99.38%, with minimal misclassifications between categories. These results are comparable to the best performing models in previous studies (96.54% - 99.48%). The proposed model has low computing costs in comparison to other studies in the literature

Conclusion: This study explored deep learning application to classify retinal diseases by using OCT images. We obtained a high test accuracy of 99.38% and this indicate the potential of deep learning models for aiding in the diagnosis of retinal diseases, potentially improving patient outcomes. Future work will focus on exploring hybrid algorithms for further accuracy improvement.

Keywords: Convolutional Neural Networks (CNN), image processing, retinal diseases, artificial intelligence, neural networks, Optical Coherence Tomography (OCT), Diabetic macular edema (DME), Choroidal Neovascularization (CNV), Drusen, Deep Learning, Alexnet

Acknowledgements:

I have received a lot of support and assistance Throughout the writing of this dissertation. This dissertation serves as the final step of my Master's degree at BioInformatics.

I would first like to express my gratitude to my thesis supervisor, Doctor Yanal Ahmad Alkuddsi, for his support, and the opportunity to work in the field of Deep Learning. It has been a challenging but rewarding experience to explore the realm of artificial intelligence in the field of medicine, particularly in ophthalmology.

I would like to convey my deepest appreciation to my family, as they are the fundamental catalysts behind my achievements. I am profoundly grateful to them for embodying unwavering work ethics and providing constant encouragement throughout my journey. A mere expression of gratitude cannot suffice to acknowledge the profound impact they have had on my life. It is through their unwavering support that I have been able to pursue and achieve my academic goals.

Table of Contents:

Table of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
DL	Deep Learning
AMD	Age-related Macular Degeneration
DME	Diabetic Macular Edema
OCT	Optical Coherence Tomography
CNV	Choroidal neovascularization
CNN	Convolutional Neural Network
NFL	Nerve Fiber Layer
FA	Fluorescein Angiography
ICG	Indocyanine Green Angiography
ERG	Electroretinography
ANI	Artificial Narrow Intelligence
AGI	Artificial General Intelligence
ASI	Artificial Super Intelligence
ML	Machine Learning
NLP	Natural Language Processing
MLP	Multi-layer Perceptron
DNNs	Deep Neural Networks
ReLu	Rectified Linear Unit
FC	Fully Connected
GPUs	Graphics processing units
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CCNN	Coherent Convolutional Neural Network
GANs	Generative Adversarial Networks

List of Figures

List of Tables

Introduction:

In the past decades, Artificial intelligence (AI) has revolutionized the disease diagnosis and the anatomization process. Many AI and subset Deep Learning (DL) networks are useful in medical image processing for prognosis and diagnosis of various ailments (e.g., breast cancer, lung cancer, brain tumor, and eye diseases), which are tedious and prone to human error if manually performed. In this study, the benefits of AI have been leveraged to classify and identify an ocular disease; the retina's structural complexity makes it inconvenient and time-consuming for accurate evaluation by the expert.

The retina is a thin layer of tissue at the back of the eye, near the optic nerve where photoreceptors are located, as shown in Figure 1, which are specialized cells that respond to light. The retina covers the entire back of the eye, so it is bowl-shaped[1]. The retina receives light, processes it through a layer of cells, detects characteristics such as color and light intensity, then converts this information into nerve signals and sends them to the optic nerve, which heads to the brain, which in turn decides What is the image?[2]

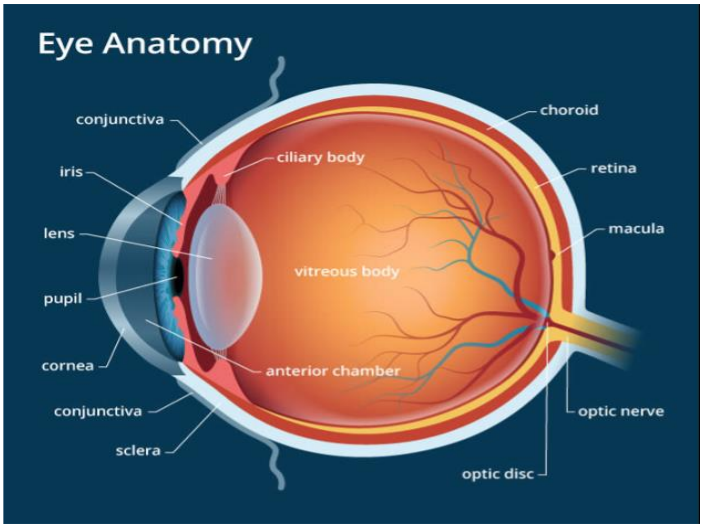

Figure 1: The Structure of the Eye[2]

The retina is nourished by two types of blood vessels: its own blood vessels that feed the inner layers of the retina, and choroidal blood vessels that feed the outer layers of the retina.[3]

Age-related Macular Degeneration (AMD) is a highly prevalent retinal disorder that accounts for 8.7% of blindness globally and affects the choroidal blood vessels, The inner blood vessels are the ones that are afflicted by Diabetic Macular Edema (DME).

Employing advanced AI technologies brought the much-needed headway in medical diagnosis. The automated detection of retinal diseases involves an image preprocessing, segmentation and sampling procedures, training of neural networks, and analysis of statistics. The researchers are currently focusing on improving the accuracy of classification and identification of the disease, reducing the computational time and memory utilization.

Problem Statement:

An examination known as Optical Coherence Tomography (OCT) has gained ground in the latest complementary clinical tests for the diagnosis of retinal, but the interpretation of these data, analysis, and interpretation takes up a significant amount of time. OCT images are also challenging in themselves due to the high noise and motion in the images. Furthermore, image resolution, filtering, and scanning patterns vary between optometrists.

Deep learning can help doctors make faster and more accurate diagnoses and predict disease in a timely manner.

In this research, we implemented a deep learning model to detect early abnormalities within the retina from a set of 84,495 OCT images (JPEG) and 4 categories (NORMAL, Choroidal neovascularization CNV, DME, DRUSEN) from retrospective cohorts of adult patients. If we can diagnose and treat eye conditions early, it gives us a great opportunity to maintain the vision of people before losing.

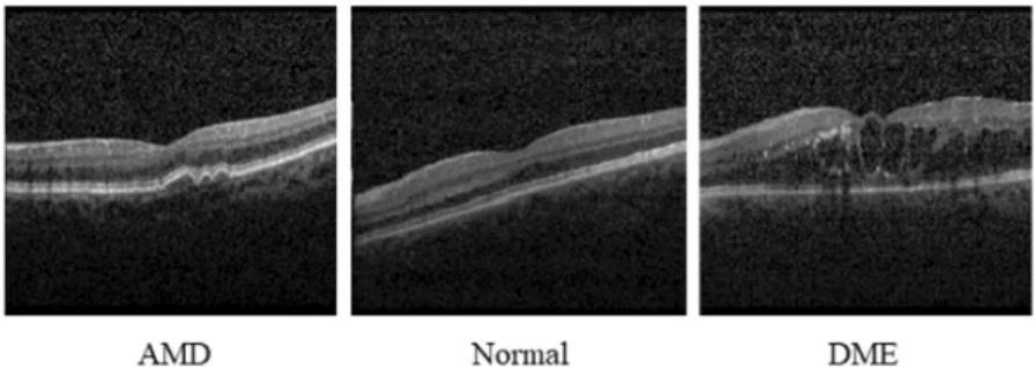

Figure 2: OCT scans of different macular eye diseases[4]

Aim of the study:

The goal of this thesis is implementation of deep learning model used to detect four types of retinal cases (NORMAL, CNV, DME, and DRUSEN) by using Convolutional Neural Network (CNN) to Avoid manual diagnostic errors and help doctors make faster, and more accurate diagnoses.

Theoretical Review

Chapter 1 – Retinal Diseases

1.1. Anatomy and function of the Human Eye:

It is estimated that sight accounts for 80% of all human perception. As the main organ for capturing, filtering and transmitting light information to the brain for processing.

The human eye is a spherical structure located on the front surface of the skull. The dimensions of the adult eye are relatively constant, with a sagittal diameter of about 24 mm and a transverse diameter of 24.5 to 25 mm, and it weighs about 7.5 grams. The main eye structures are formed during the fifth month of fetal development, and at birth, the eyes are about two-thirds the size of an adult's eyes. Eye growth gradually slows down from the second year until adulthood.

The human visual recognition process begins as soon as light penetrates the pupil and is directed through the cornea and lens: the pupil regulates the amount of light entering the eye as an aperture regulated by the surrounding iris, and the cornea and lens are responsible for this. To form the visual image on the retina. The retina then converts the image into electrical energy, which is then transmitted to the brain through complex neural pathways, connecting the eye to the visual cortex and other parts of the brain through the optic nerve. Figure 3 represents the anatomy of the eye.[5]

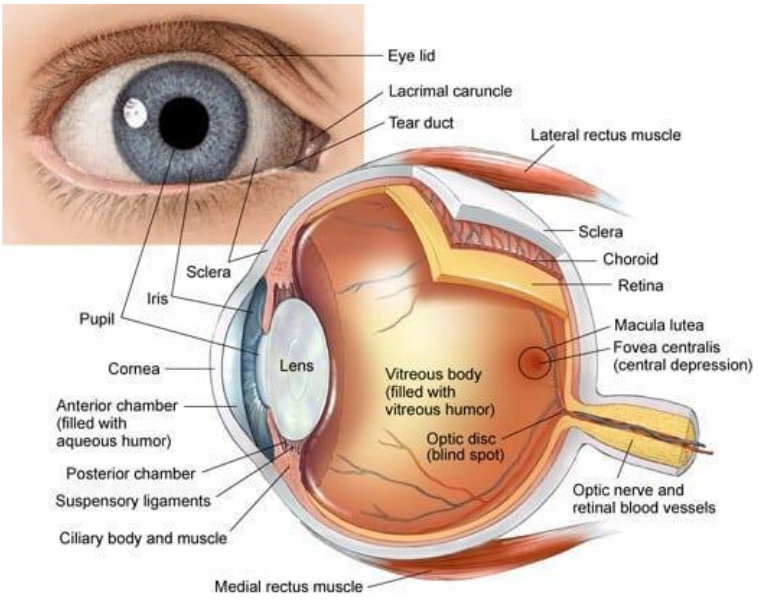

Figure 3: Sagittal and external human eye anatomy [6]

1.1.1. Retina Anatomy:

The retina plays a critical role in vision due to its responsibility for translating light into a biochemical message which is translated into electrical impulses and transmitted to the brain.

The retina exhibits a laminar organization of ten main layers and is approximately 0.5 mm thick. This segmentation is useful as it facilitates the identification of anomalous pathologies conditions that usually present a typical position within the retinal tissue[5]

Figure 4 and Figure 5 shows cross-sectional retinal scans obtained using OCT scans with the identification of all the layers. From the vitreous to choroid ,The retinal layers are:

Internal limiting membrane, Nerve Fiber Layer (NFL), Ganglion cell layer, Inner plexiform layer, Inner Nuclear Layer, Outer Plexiform Layer, Outer Nuclear Layer, External Limiting Membrane, Receptor layer and Retinal Pigment Epithelium.[7]

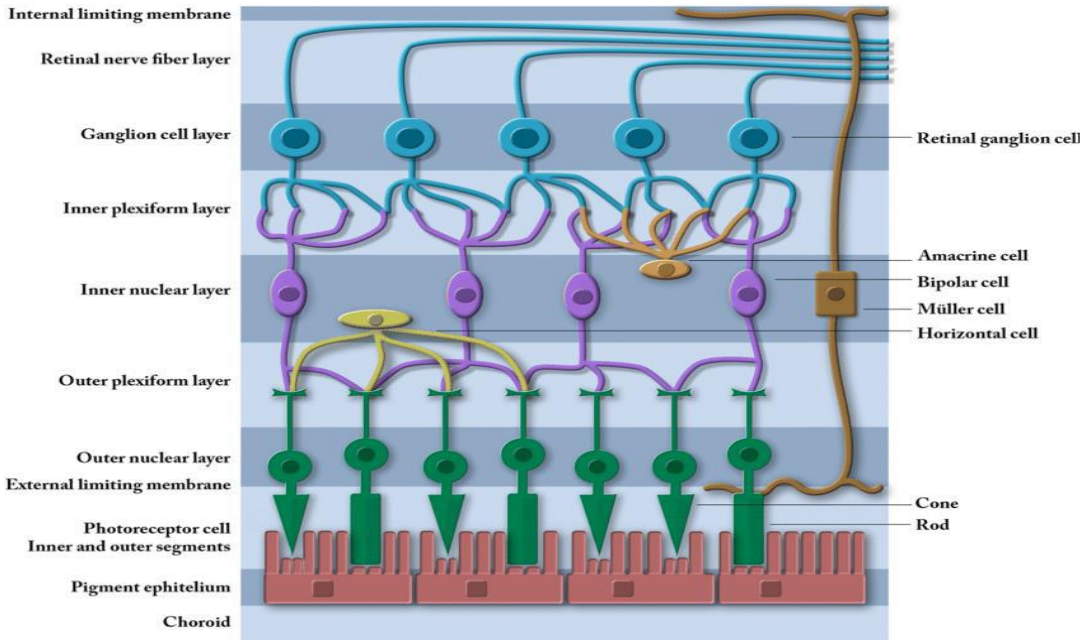

Figure 4: Retina layers [8]

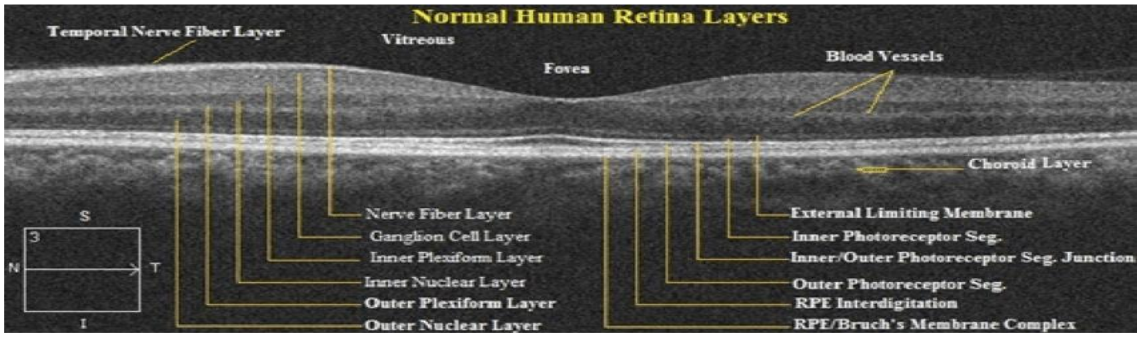

Figure 5: Labeled OCT Imaging [9]

1.2. Types of Retinopathy:

There are many types of retinopathy, and we studied three types of retinopathy CNV, DME, Drusen

1.2.1. Age-related macular degeneration AMD:

Age-related macular degeneration (AMD) is an eye disease that is the leading cause of vision loss in the elderly. It is prevalent in people aged 55 years and older. It occurs when aging causes damage to the macula the part of the retina that controls sharp, direct vision. [10]

It is important to have regular eye examinations because AMD occurs very slowly in some people and quickly in others, and the individual may not notice vision loss for a long time.[4]

1.2.1.1. Types of AMD:

The classification of AMD is divided into early, intermediate, and advanced nonneovascular AMD or advanced neovascular AMD (Table 1).[10]

Table 1: Classification of age-related macular degeneration based on the Age-Related Eye Disease[10]

Early AMD	**Intermediate AMD**	**Advanced Nonneovascular AMD (Advanced Dry AMD)**	**Advanced Neovascular AMD (Wet AMD)**
Presence of small drusen or few mediumsized drusen in 1 or both eyes Pigmentary changes	Presence of many medium-sized drusen, 1 large drusen, and/or Geographic atrophy not involving the central macula (fovea)	Geographic atrophy involving the central macula or fovea	Choroidal neovascularization in 1 eye

AMD is divided into two major categories: non-neovascular (dry), also called Drusen ,and neovascular (wet), also called CNV.[11]

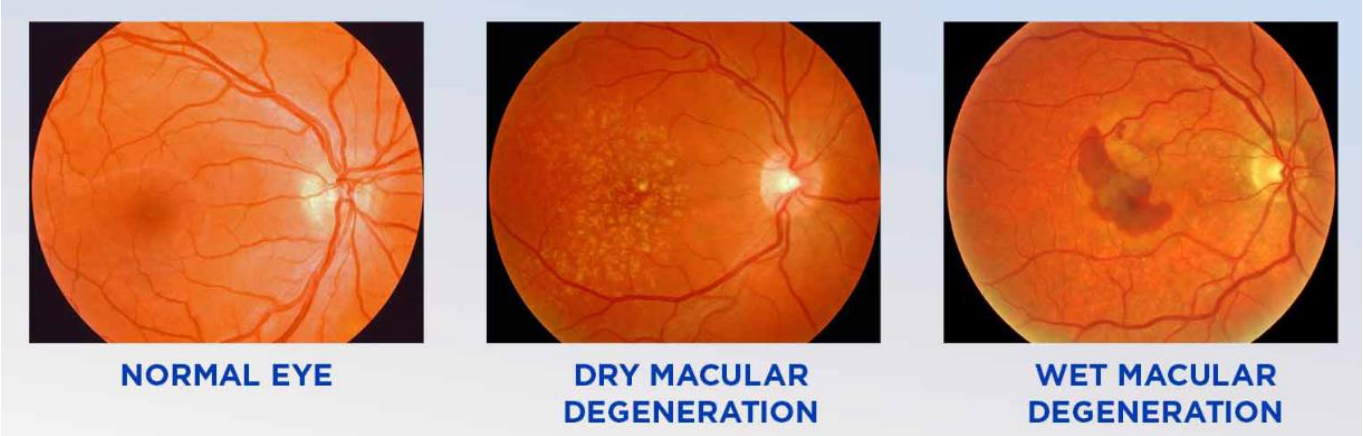

Figure 6: Age-Related Macular Degeneration (AMD) Types[12]

1.2.1.1.1. Drusen (non-neovascular dry AMD):

Prevalence: It is considered the most common, accounting for 85% of all AMD cases. **Pathophysiology:** The macula gradually weakens over time, leading to progressive blurring of central vision.

Clinical manifestations: small yellow or white deposits under the retina called druse, it progresses slowly. Vision loss can be mild to moderate, but rarely leads to complete blindness. [13]

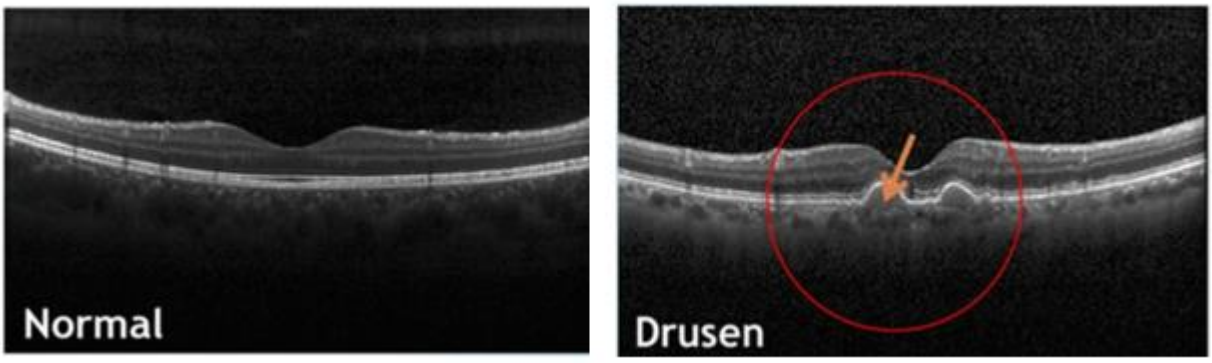

Figure 7: OCT scans of Drusen and normal retina[14]

1.2.1.1.2. CNV (neovascular wet AMD):

Prevalence: Wet AMD is less common (10-15% of cases) but more serious.

Pathophysiology: New, abnormal blood vessels grow under the retina, especially under the macula. These vessels leak blood and fluid, leading to damage to light-sensitive cells.

Clinical manifestations: It leads to rapid loss of vision and thus poses a greater threat to central vision.[13]

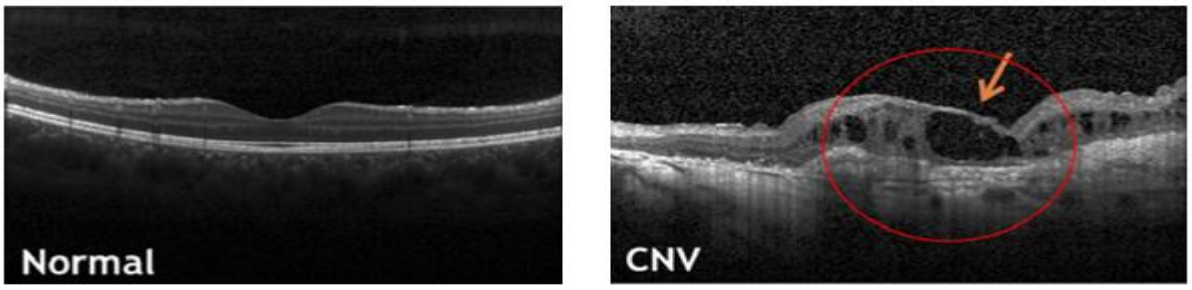

Figure 8: OCT scans of CNV and normal retina [14]

1.2.1.2. symptoms of AMD:

AMD often does not cause noticeable symptoms at first, but later causes the following symptoms:

- Mild blurriness or distortion in central vision.
- Difficulty perceiving colors as they appear less vibrant.
- Difficulty seeing clearly in dark environments.
- Straight lines may appear distorted (wavy or crooked).
- Difficulty distinguishing faces.
- A dark spot in the central vision caused by damage to the macula. [15]

1.2.2. Diabetic Macular Edema DME:

In DME, a buildup of fluid and proteins in the macula area is caused by prolonged high blood sugar and damage to blood vessels as the retina becomes swollen, and vision is greatly impaired. It's the most common cause of vision loss in patients with diabetes.[16]

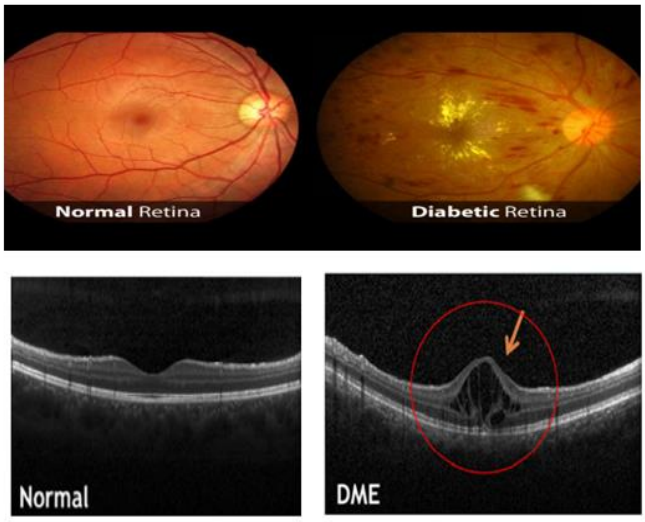

Figure 9: OCT scans of DME and normal retina[14]

1.2.2.1. symptoms of DME:

- Individuals may experience difficulty focusing clearly, blurry vision, or double vision, which leads to blurred or overlapping images.
- Floaters: These are small spots or specks that appear to be floating.
- Difficulty seeing colors, which makes colors appear faded or changed.
- Dark spots are areas where vision is lost.
- Seeing straight lines as curved or wavy.
- Difficulty seeing in bright light.
- The difference in the size of an object when viewed with one eye and then with the other.[17]

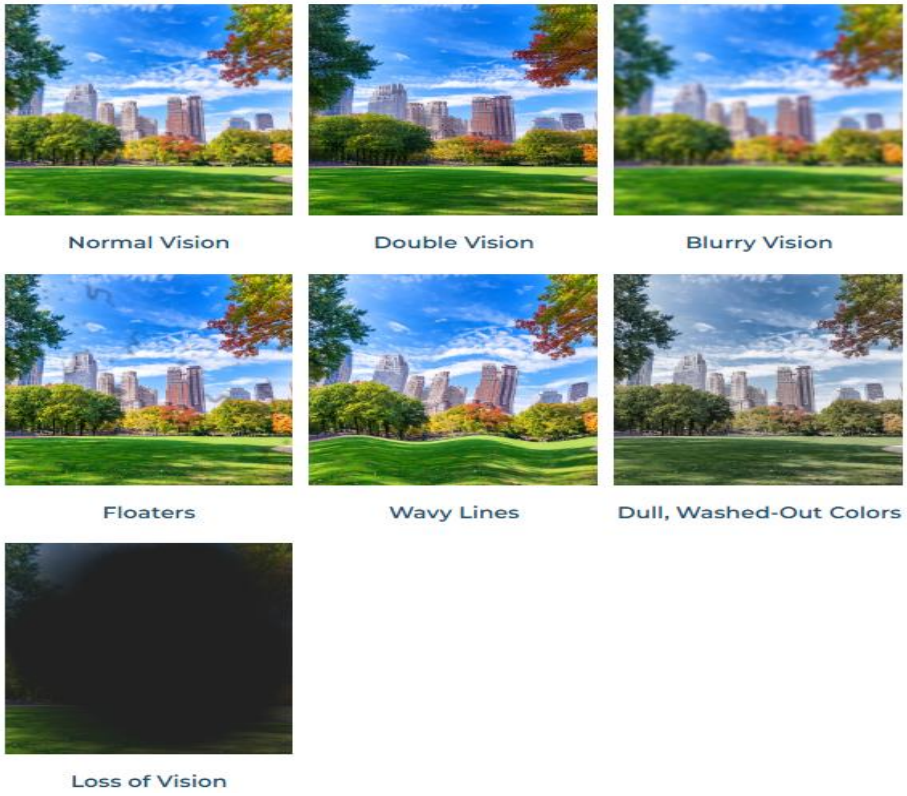

Figure 10: Symptoms of Diabetic Macular Edema[18]

1.3. Diagnosis of Retinopathy:

There are many traditional techniques for diagnosing retinopathy, including: Fundus Photography, Fluorescein Angiography (FA), Indocyanine Green Angiography (ICG), Electroretinography (ERG), OCT and future diagnostic methods in Development of artificial intelligence -based algorithms for automated analysis of retinal images, potentially improving diagnostic accuracy and efficiency.[19]

1.3.1. Fundus Photography:

Fundus imaging is defined as the process whereby reflected light is used to form a two dimensional and a high-resolution representation of the three-dimensional retina, the semi-transparent, layered tissue lining the interior of the eye that is projected onto the imaging

plane. Revealing blood vessels, the optic nerve, and the macula.[20] Its strengths are that it is a non-invasive technique, relatively inexpensive and widely available but provides only 2D images so gives limited detail and may miss subtle changes.[19]

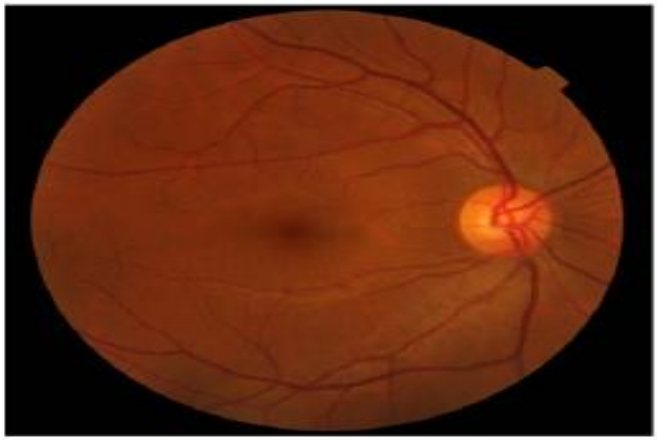

Figure 11: Fundus Photography of the eye [20]

1.3.2. Fluorescein Angiography (FA):

Fluorescein angiography (FA) is a landmark in the diagnosis of many systemic and ocular diseases involving retinal blood vessels.[21] Historically it is the gold standard for evaluating choroidal neovascularization in AMD.[10] It relies on injecting sodium fluorescein ($C_{20}H_{12}Na_2O_5$) into the patient's vein, followed by obtaining Serial images to visualize the patency and permeability of retinal vessels.[21] The main objective of this technique is to highlight the blood vessels to form a clear and visible image.[19] Although this method represents a turning point in retinal imaging, it suffers from several limitations. Several adverse effects may be observed including staining of the sclera, skin and urine. In addition, harmful drug interactions may occur, ranging from mild such as nausea and vomiting or moderate such as fainting to serious episodes such as respiratory difficulties, heart disorders, and tonic-clonic seizures. Therefore, the necessity of performing angiographic examination on several categories of patients should be carefully considered. [21], [22]

Figure 12: Fluorescein Angiography (FA) of the eye[23]

1.3.3. Indocyanine Green Angiography (ICG):

Indocyanine green angiography (ICG) is a similar method to FA in that it involves the injection of indocyanine green, a dye useful in assessing choroidal circulation. It is useful for diagnosing certain conditions such as choroidal neovascularization, but there is also the possibility of allergic reactions and it is not widely available such as FA.[10], [22]

Figure 13: Indocyanine green angiography (ICG) of the eye[24]

1.3.4. Electroretinography (ERG):

Electroretinogram (ERG) is a diagnostic test that measures the electrical response of the retina when it is exposed to a light stimulus. It arises from currents generated directly by neurons in the retina in combination with contributions from retinal glial cells. It is often done using a thin-fiber electrode placed in contact with the cornea. It helps in diagnosing retinal disorders and monitoring disease progression. [25], [26]

1.3.5. Optical Coherence Tomography (OCT):

Optical coherence tomography (OCT) is a non-invasive medical imaging technology that is playing role in the detection, diagnosis, and management of many eye conditions such as macular hole, macular edema, age-related macular degeneration, glaucoma, and diabetic retinopathy.it uses light waves to create a cross-sectional image of the retina, give detailed information about its structure and thickness, and detect early changes without touching the eye. Scanning takes about 5-10 minutes.[22], [27]

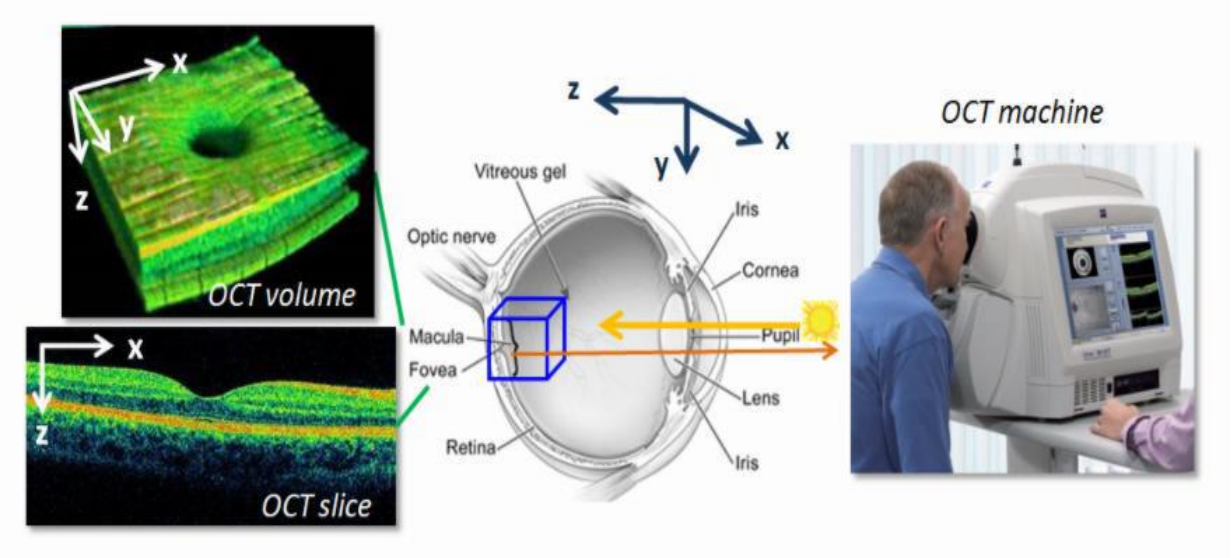

Figure 14: OCT 3D cross sectional images of eye,cross section of eye and OCT machine[28]

Chapter 2 – Deep Learning

2.1. Artificial intelligence AI:

Can machines think?" Alan Turing in his famous paper posed this question "Computing Machinery and Intelligence". Turing then introduced an indirect method to verify whether a machine can think, the Turing test, which examines a machine's ability to show intelligence indistinguishable from that of human beings. A machine that succeeds in the test is qualified to be labeled as AI.[29] AI uses computers to simulate human intelligent behaviors and it trains computers to learn human behaviors such as learning, judgment, and decision-making. AI is a compilation of computer science, logic, biology, psychology, philosophy, and many other disciplines, and it has achieved remarkable results in applications such as speech recognition, image processing, natural language processing, the proving of automatic theorems, and intelligent robots.[30]

There are three types of artificial intelligence: Artificial Narrow Intelligence (ANI) is where machines demonstrate intelligence in a specific domain such as playing chess, forecasting sales, making movie suggestions, and translating languages. Artificial General Intelligence (AGI) is concerned with making machines perform at the same level as humans. Artificial Super Intelligence (ASI) is concerned with By building machines that exceed human capabilities in several fields.[31]

Generally, the advances of AI application in industry have birthed various well-subfields of AI including: (1) computer vision (2) machine learning (3) natural language processing (4) knowledge-based systems (5) optimization (6) robotics (7) automated planning and scheduling as shown in Figure 15.[32]

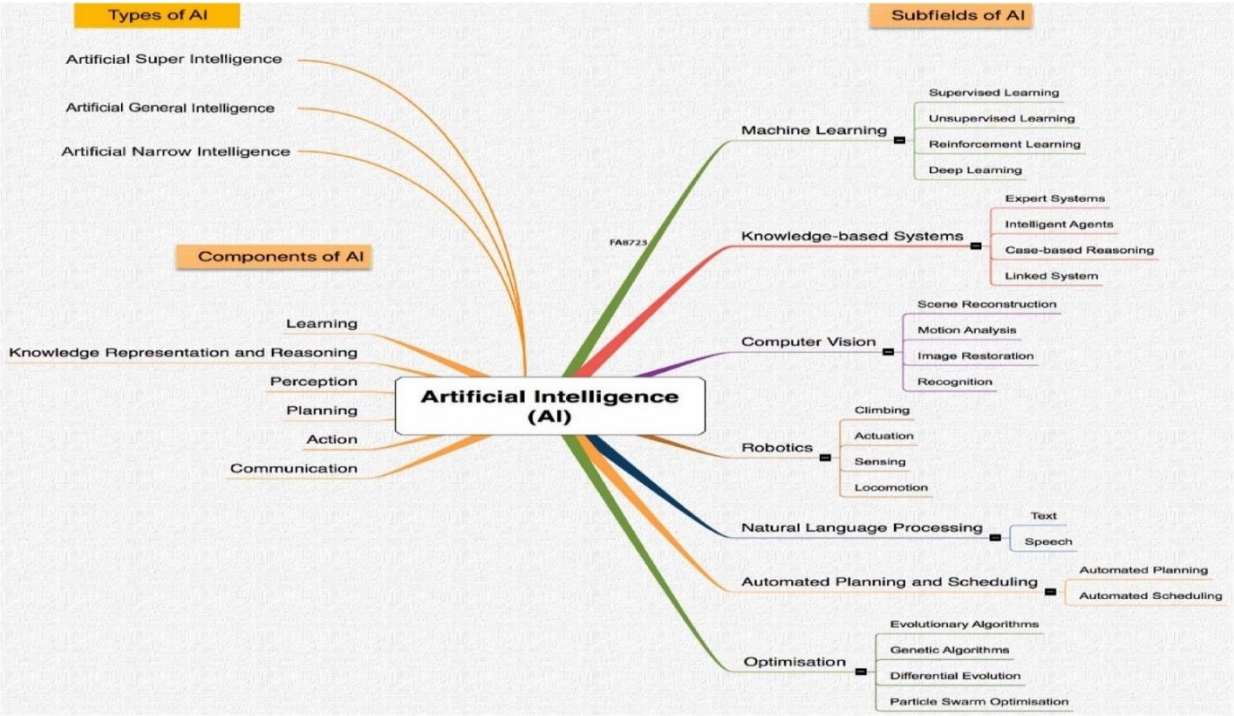

Figure 15: Components, types and subfields of AI [32]

2.2. Machine Learning (ML):

Machine learning is a branch of artificial intelligence that is based on the idea that a machine can learn data and make decisions without human intervention or assistance.

There are many machine learning applications, such as self-driving cars, how to distinguish between fruits, speech recognition, and web searching, by training on a large amount of data.

Machine learning is classified as supervised, unsupervised, semi-supervised, and reinforcement learning.

Supervised learning is a learning model designed to make predictions based on what it has learned in the past. The goal is to learn to map from input to output, such as forecasting sales. In contrast, in unsupervised machine learning algorithms, the goal is to find regularities in the input so that certain patterns occur more often than Others, such as document clustering.

Semi-supervised machine learning algorithms fall between supervised and unsupervised learning where the goal is to classify some unlabeled data using the labeled set of information.

Reinforcement machine learning algorithms learn by interacting with the environment and detecting errors and rewards.

Machine learning allows massive amounts of data to be analyzed but it also requires additional time and resources to train properly.[31]

2.3. Deep Learning:

Deep learning is a type of machine learning and AI that mimics the way humans acquire certain types of knowledge. [33] Human brains contain millions of interconnected neurons that work together to learn information, and deep learning is characterized by neural networks built from multiple layers of software nodes that work together so that the output of each layer is the input for the next layer as shown in Figure 16.[34]

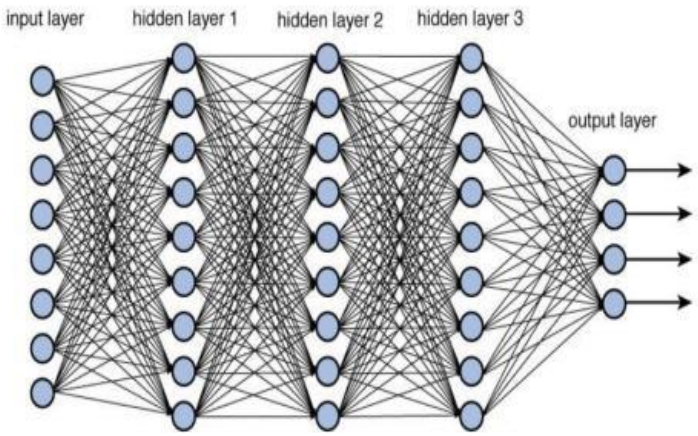

Figure 16: Deep Learning with Multiple Layers [35]

Deep learning models often require a large amount of training data to perform well because they have a large number of parameters that need to be tuned by the learning algorithm. Deep learning can be used in computer vision, speech recognition, natural language processing (NLP), and recommendation engines.[33]

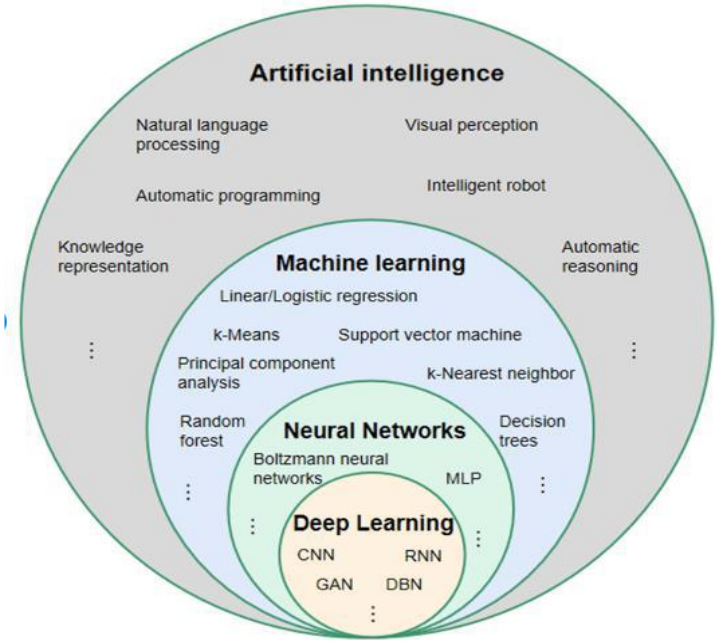

Figure 17: Relationship between artificial intelligence, machine learning, neural network, and deep learning[36]

We can find in literature two types of machine learning models: Supervised and Unsupervised.

The main difference between them in the way they are trained. While supervised models are trained with expected results (Input data are labelled), unsupervised models are only given input data without a set of labels they can learn from.

In deep learning, through the problem that the model helps to solve it ,we can distinguish its type:

- Supervised Deep Learning: Image classification, object detection, face recognition etc.
- Unsupervised Deep Learning: Word embedding, image encoding.

Deep Learning is a growing field with applications that span across several use cases. In each use case, a different model is predominant and gives the best efficiency.[37]

2.4. Neuron:

The biological nervous system is a network composed of many neurons. Similarly, neurons are also the basic processing unit of artificial neural networks. The principle of

operation is that multiple input values undergo mathematical transformation to obtain an output value (Figure 18). The mathematical transformation relationship between the input signal and the output value is

$$f\left(b+\sum_{i=1}^{n}(x_i \times w_i)\right)$$

f (.) is the activation function, there are many activation functions, such as ReLU, Sigmoid, Tanh, etc. ref [remotesensing-13-04712-v2]

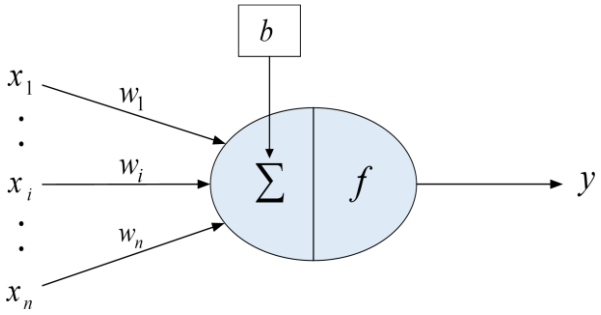

Figure 18: Neuron model, xi is the input signal, n is the number of signals, the weight value of the input signal is wi, bias is b and output of neurons is y [38]

2.5. Multi-layer Perceptron (MLP):

A Multilayer Perceptron Neural Network (or fully connected neural networks) is a Neural Network of the feed forward type. It uses the Backpropagation technique for learning. It has an input layer of neurons that act as receivers, one or more hidden layers of neurons that compute the data and undergo iterations and then the output layer which predicts the output [39]as shown in Figure 19.

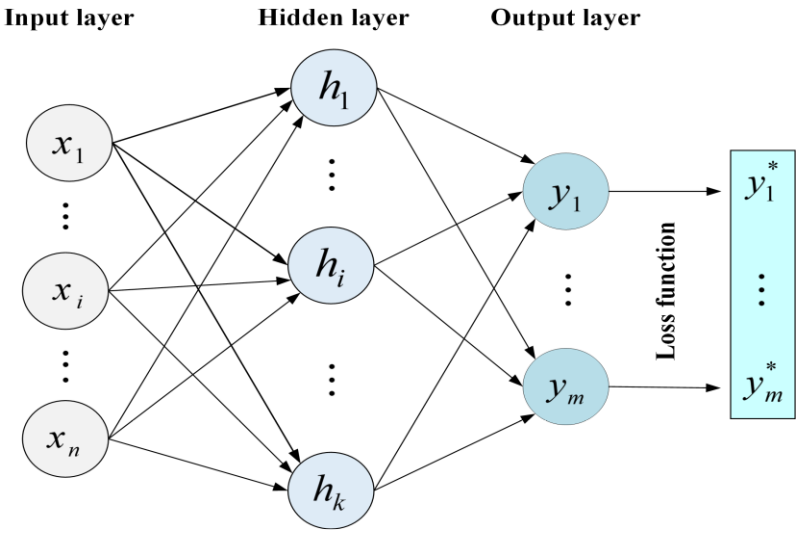

Figure 19: The structure of the MLP. x_n is the input value. The hidden unit is hk, it receives the input value of the previous layer. y_m is the output unit, and the real value is y*m [38]

A perceptron/neuron applies a linear model which takes multiple inputs and produce an output, usually multiplies that input with a vector of weights and add a bias.

$$Z = \vec{w}.X + b$$

It then applies a non-linear activation function. Which gives the real operation made by a neuron: [37]

$$Z = \sigma(\vec{w}.X + b)$$

2.6. Convolutional Neural Network:

CNNs are similar to traditional deep neural networks (DNNs): they are made up of layers of neurons that have learnable weights and biases. Each neuron receives some inputs, calculates a dot product, and optionally follows it with a nonlinear function. This process is repeated layer by layer until the output layer, where the network's prediction is generated. However, DNNs have an important limitation: they do not scale well in terms of computational resources. traditional neural networks use fully connected layers, where each neuron is connected to every neuron in the previous layer. This can easily lead to a network with thousands of parameters to be estimated. This type of configuration is for general purpose, i.e., it makes no assumptions about the properties of the input variables. For this reason, it tends to be expensive in terms of memory and computation.

In CNNs These networks are specifically designed to work with grid-structured inputs that have strong spatial dependencies in local regions of the grid. The most obvious example is the natural image. Natural images exhibit spatial dependencies where adjacent pixels often have similar color values. Therefore, the features within an image have dependencies among one another based on spatial distances. [40]

CNNs mainly has two parts: the first one is the feature extraction part and the second one is the classification part. In the first part, a series of convolution and pooling operations are performed for feature detection. For producing a feature map, using a filter, the convolution operation is applied. This feature map will contain negative pixel values and it should be replaced with zero. For that, a non-linear operation is performed after performing every convolution. Nonlinearity is introduced using ReLU. In the classification part, on top of these extracted features fully connected layers will act as classifiers. They assign a probability for the object on the image. When these images are too large, the pooling operation continuously reduces the dimensionality. This is done for reducing the number of computations and parameters in the network. This reduces training time and controls overfitting. [41]

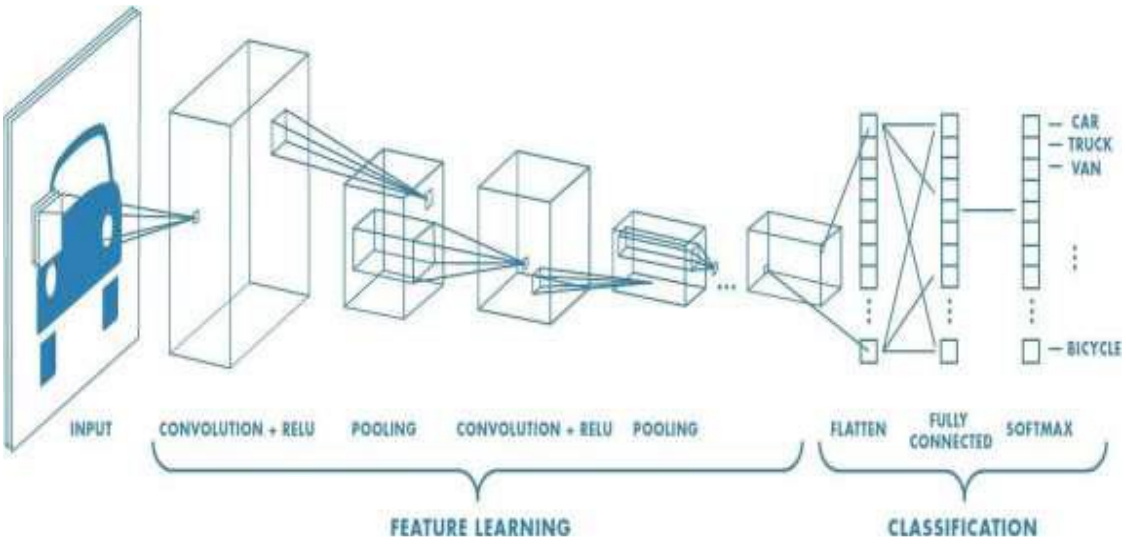

Figure 20: Architecture of CNN [41]

2.6.1. Convolutional Layer:

The main aim of a convolutional layer is to detect features or visual features in images such as edges, lines, color drops, etc. This is a very important property because, once it has learned a characteristic at a specific point in the image, it can recognize it later in any part of it.

Convolutional layer capture a feature map to predict the class probabilities for each feature through applying a filter that scans the whole image, few pixels at a time.

A filter is just a matrix of values, called weights, which are trained to detect specific features. The filter moves over each part of the image to check if the feature it is meant to detect is present. To provide a value representing how confident it is that a specific feature is present, the filter carries out a convolution operation, which is an element-wise product and sum between two matrices.[42], [43], [44]

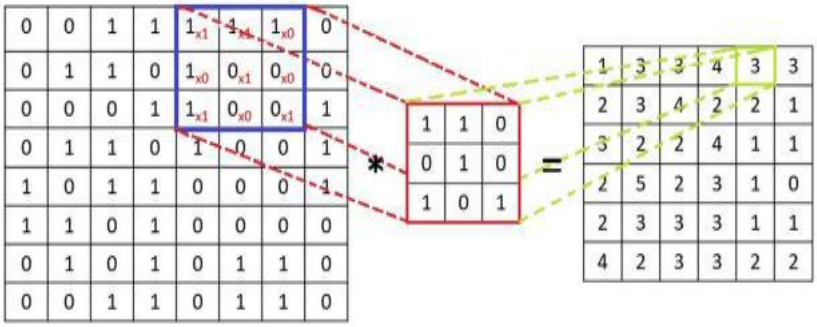

Figure 21: Convolutional Layer [45]

2.6.2. Pooling Layer:

the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality

reduction. Besides, it helps extract the dominant features and the most useful ones. There are many types of pooling operations :

1. Max pooling: returns the maximum value from the portion of the image covered by the Kernel.

2. Average pooling: returns the average value from the portion of the image covered by the Kernel.

3. Sum pooling: returns the sum of the values covered.[37]

Average pooling and maximum pooling are the two most widely used pooling methods. [38]

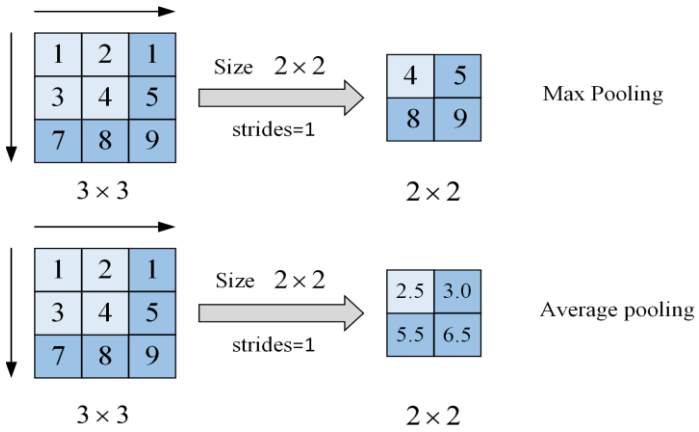

Figure 22: Max Pooling and Average pooling [38]

2.6.3. Nonlinearity Layer (Activation Function):

An activation function plays an essential role in CNN layers. The filtered output provides another mathematical function called activation. ReLU, abbreviated from the Rectified Linear Unit, is the most common activation function in feature extraction using CNN. The main objective of the activation function is to decide the final output of a neural network, such as 'yes' or 'no'. The activation function maps the output values between -1 and 1, 1 and 0, and so on.

The activation function can be differentiated into two categories, which are:

1. Linear Activation Functions. A simplified mathematical expression of linear activation functions can be written as $F(x) = CY$. The input values are multiplied with the constant parameter, C, which is the weight of each neuron. The process results in an output that is proportional to the input. Linear functions can perform more than the step function since they only give a single final answer of yes or no and not multiple choices.

2. Non-linear Activation Functions. Non-linear activation functions are used in modern neural networks. They allow the model to design a complicated mapping between the network's input and output, which is essential for complex learning and modeling systems.[46]

2.6.4. Activation functions:

The activation function must also have the ability to differentiate, which is an extremely significant feature, as it allows error back-propagation to be used to train the network. The following types of activation functions are most commonly used in CNN and other deep neural networks[47]

2.6.4.1. Sigmoid function:

this activation function uses real numbers as inputs and limits the output between 0 and 1. The curve of the sigmoid function is S-shaped and can be mathematically represented as

$$f(x)sigm = \frac{1}{1 + e^{-x}}$$

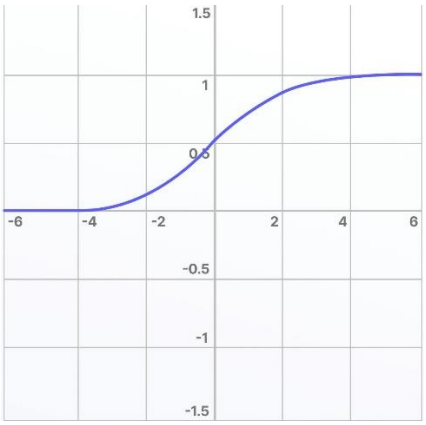

Figure 23: Sigmoid Graph

2.6.4.2. Tanh Function:

Apparently, the tanh function is similar to sigmoid since both use real numbers as their inputs. However, the tanh function limits its output in -1 and 1. The tanh function can be mathematically represented as

$$f(x)tanh = \frac{e^{x} + e^{-x}}{e^{x} + e^{-x}}$$

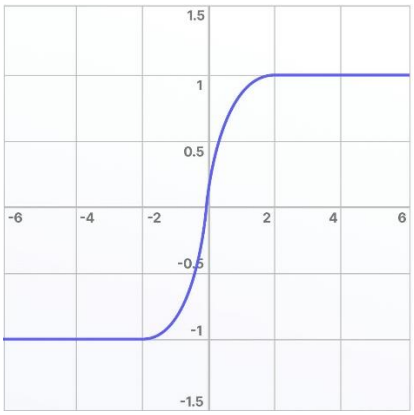

Figure 24: Tanh Graph

2.6.4.3. Rectified Linear Unit (ReLu) Function:

ReLU is the most common function used in CNN. All inputs are converted into positive numbers. The computational load of ReLU is relatively lower than other functions. Mathematically, the representation of the ReLU function is presented as

$$f(x)ReLU = \max(0, x)$$

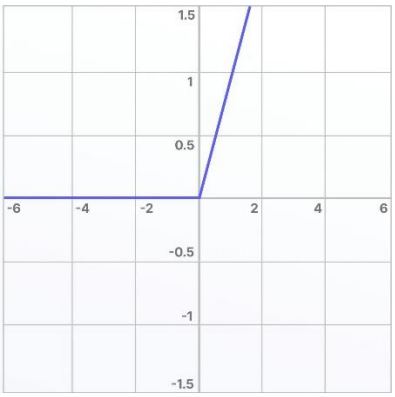

Figure 25: ReLu Graph

2.6.4.4. Leaky ReLU:

If the ReLU function is responsible for down-scaling the negative inputs, the Leaky ReLU function ensures that inputs are never ignored. This function is used to solve a dying issue in ReLU. A mathematical representation of Leaky ReLU is presented as

$$f(x)LeaklyReLU = \begin{cases} x, & \text{if } x > 0 \\ mx, & \text{if } x \leq 0 \end{cases}$$

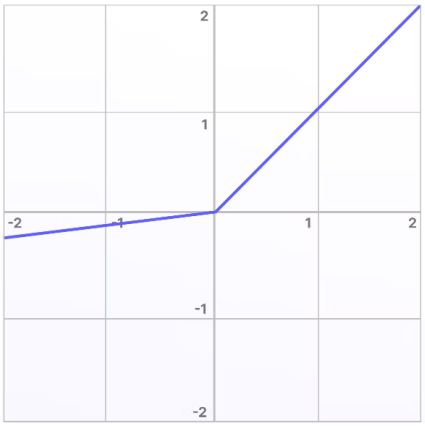

Figure 26: Leaky ReLU Graph

2.6.5. Fully Connected Layer:

Commonly, this layer is located at the end of each CNN architecture. Inside this layer, each neuron is connected to all neurons of the previous layer, the so-called Fully Connected (FC)

approach. It is utilized as the CNN classifier. It follows the basic method of the conventional multiple-layer perceptron neural network. The input of the FC layer comes from the last pooling or convolutional layer. This input is in the form of a vector, which is created from the feature maps after flattening. The output of the FC layer represents the final CNN output, as illustrated in Figure 27 [47]. The output layer is where we get the predicted classes. The information is passed through the network and the error of prediction is calculated. The error is then backpropagated through the system to improve the prediction [48], [49]

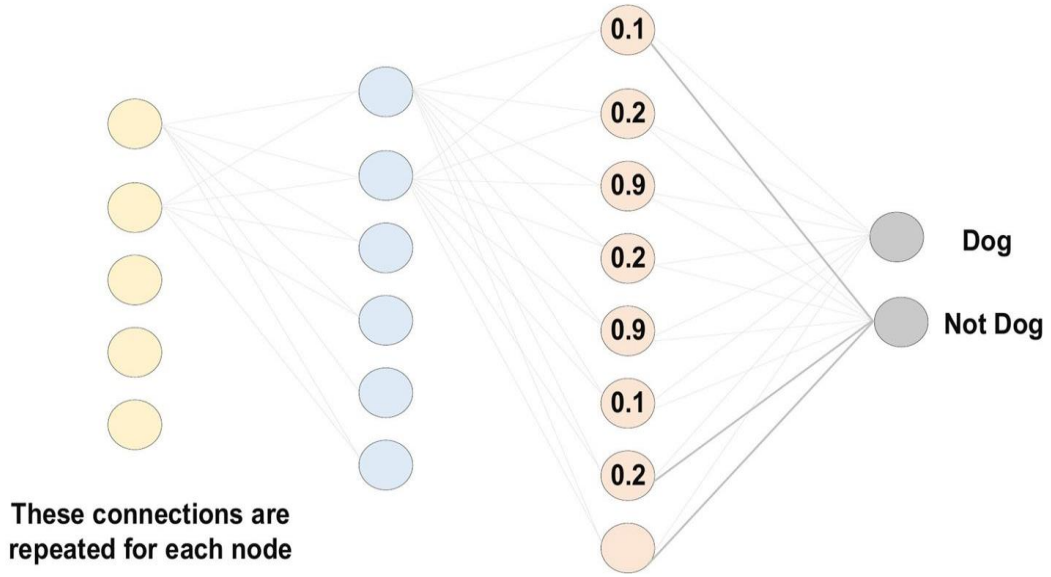

Figure 27: Fully connected layer [47]

2.6.6. Softmax:

The Softmax regression is a form of logistic regression that normalizes an input value into a vector of values that follows a probability distribution whose total sums up to 1. The output values are between the range [0, 1] which is nice because we are able to avoid binary classification and accommodate as many classes or dimensions in our neural network model. This is why softmax is sometimes referred to as a multinomial logistic regression.[50], [51]

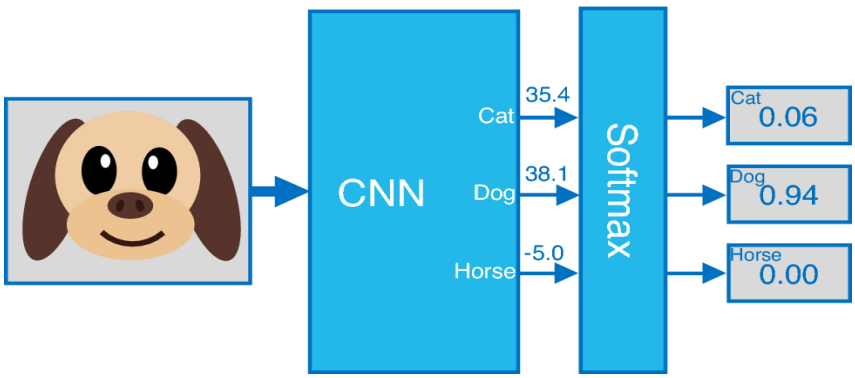

Figure 28: Example of a three classes CNN: Cats, Dogs, and Horses[52]

2.6.7. Dropout Layer:

To mitigate the risk of overfitting in neural networks, reducing the number of trainable model parameters can be an effective approach. Overfitting often occurs when neural networks have larger sizes, with more layers and nodes per layer.

A regularization technique called dropout. It involves randomly disabling neurons with a predefined dropout-rate probability P during network training. Each unit in the neural network has a chance of being "dropped out" in each training iteration. By doing so, the network is encouraged to learn more robust and generalizable representations, enhancing its ability to generalize and avoiding overfitting scenarios. Figure 30 represents a neural network using dropout layers [53]

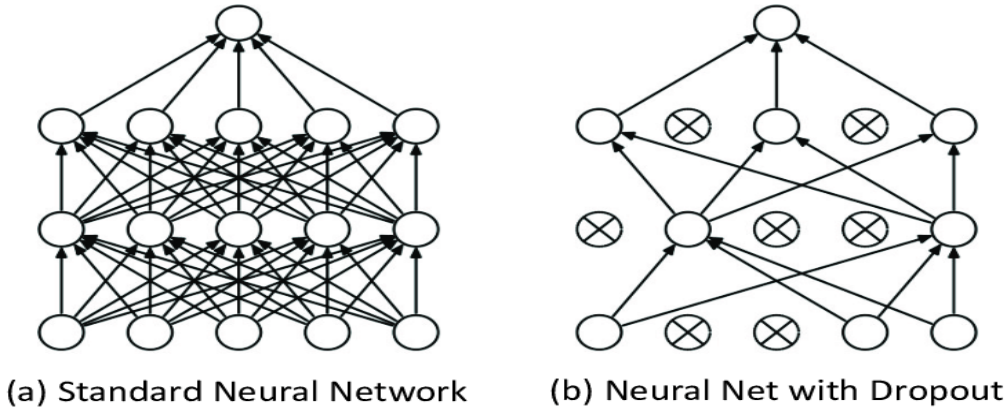

Figure 29: Neural Network Dropout [53]

2.6.8. Backpropagation:

The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. The weights that minimize the error function is then considered to be a solution to the learning problem. Backpropagation, short for "backward propagation of errors," is an algorithm for supervised learning of artificial neural networks using gradient descent. Given an artificial neural network and an error function, the method calculates the gradient of the error function with respect to the neural network's weights. It is a generalization of the delta rule for perceptrons to multilayer feedforward neural networks, as shown in Figure 31

The "backwards" part of the name stems from the fact that calculation of the gradient proceeds backwards through the network, with the gradient of the final layer of weights being calculated first and the gradient of the first layer of weights being calculated last. Partial computations of the gradient from one layer are reused in the computation of the gradient for the previous layer. This backwards flow of the error information allows for efficient computation of the gradient at each layer versus the naive approach of calculating the gradient of each layer separately.

Backpropagation's popularity has experienced a recent resurgence given the widespread adoption of deep neural networks for image recognition and speech recognition. It is considered an efficient algorithm, and modern implementations take advantage of specialized Graphics processing units (GPUs) to further improve performance.[54]

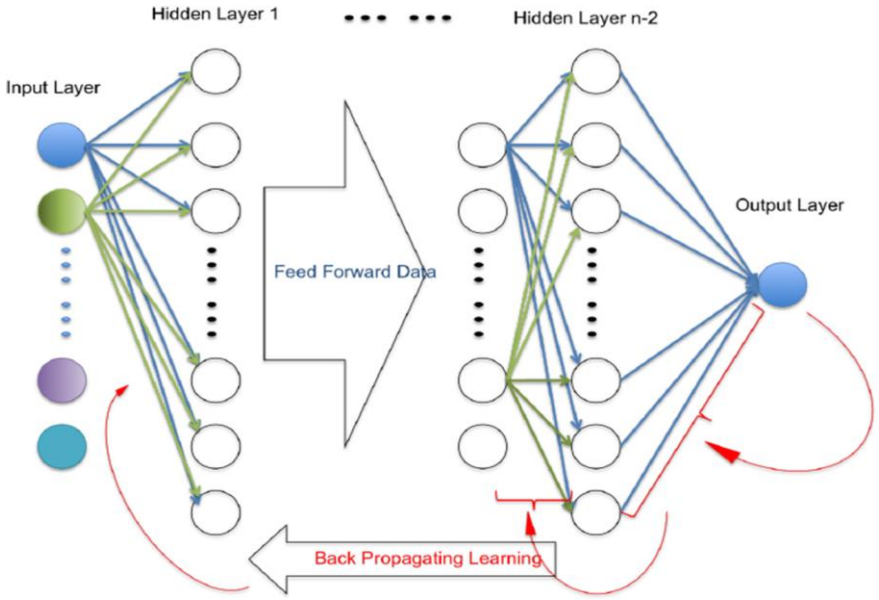

Figure 30: Backpropagation of Weights[55]

2.7. Popular CNN Architecture:

Architecture in CNN is influenced by the organization and function of the visual cortex. The design is made to resemble neuron connections in human brains. After knowing several layers in CNN, we will discuss some popular CNN architectures in this section.[45]

2.7.1. LeNet-5:

This version is a gradient-based CNN learning structure and was first introduced for digital handwriting character recognition. The structure diagram of LeNet-5 is presented in Figure 32. The input of LeNet-5 is grayscale images with a dimension of 32×32×1, which then pass six feature maps of a convolutional layer with a 5×5 filter and a stride. Those six feature maps are pre-processed image channels from the 28×28×6-sized convolutional operation. Stride is used as sliding control of a filter when passing through the dataset. The sliding control uses the tanh activation function. The second pooling layer has a 2×2 filter, six feature maps, and two strides. The tanh function on the second layer results in a 14×14×6 image.

The third step is a second convolutional layer with 16 feature maps, a 5×5 filter, and a stride, resulting in an image with a dimension size of 10×10×16. The fourth layer is a pooling layer with a 2×2 filter, two strides, and 16 feature maps. Four hundred nodes exist in the fourth layer, resulting in an output image with a dimension of 5×5×16. Then, there is a fully connected layer with 120 feature maps using the tanh activation function in the next layer; each has a dimension of 1×1. On this fifth layer, there are 120 nodes connected to 200 nodes on the fourth layer. The sixth layer is fully connected with 84 nodes, resulting in 10164 nodes of trained output parameters. The last layer in LeNet-5 is a fully connected layer with a 5-sized softmax activation function, resulting in a classified output image.[45]

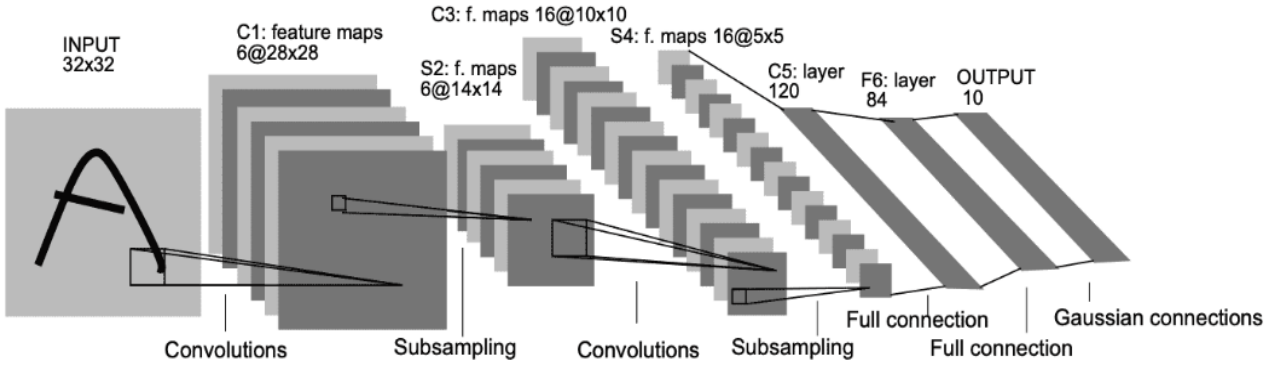

Figure 31: LeNet-5 Architecture [45]

2.7.2. AlexNet:

Alex Krizhevsky introduced AlexNet in 2012 on a research project called ImageNet LargeScale Visual Recognition Challenge. This architecture is one of CNN architectures with a basic, simple, yet effective layer design. AlexNet has five convolutional layers, followed by a pooling layer on its fourth layer and three layers of a fully connected layer on its fifth. In AlexNet architecture, the convolutional kernels are extracted during the back-propagation optimization procedure by optimizing with the stochastic gradient function. The convolutional layer acts with the sliding convolutional kernel, creating convolved feature maps to gain information within a given neighborhood window. The following equation is the function used in AlexNet as a half-wave rectifier, which significantly fastens the training phase and avoids overfitting.

$f(x)$=max(x,0)

The dropout technique in Alexnet is used as a stochastic regulator in determining the number of input neurons with 0 values to reduce co-adaptation neurons, which is commonly used in the fully connected layer. The architecture of Alexnet can be seen in Figure 33 [56].

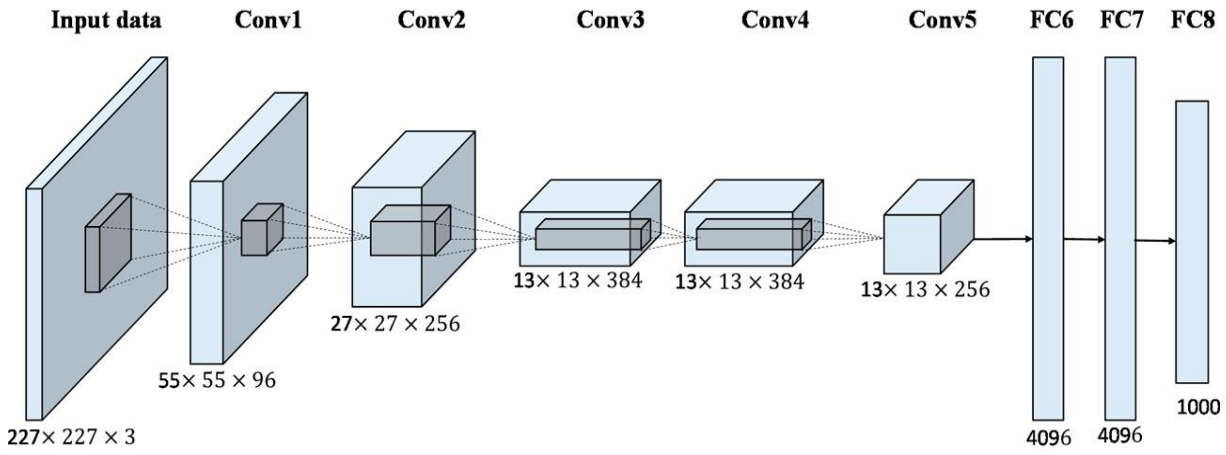

Figure 32: AlexNet Architecture [56]

2.7.3. VGG-16:

This architecture employs 13 convolutional layers and 3 fully connected layers.[57]The convolutional layer in VGG-16 has a size of 3×3 with a 1-sized stride and padding. Meanwhile, the pooling layer has a size of 2C2 with a 2-sized stride. The resolution of the input image in VGG-16 is 224×224. After each pooling layer is run, the size of the feature map will be reduced by 50%. The last feature map made before the fully connected layer is 7×7 with 512 channels and continues to be expanded to a vector with a size of 7×7×512 channels[58].The architecture of AVGGNet-16 is represented in Figure 34.

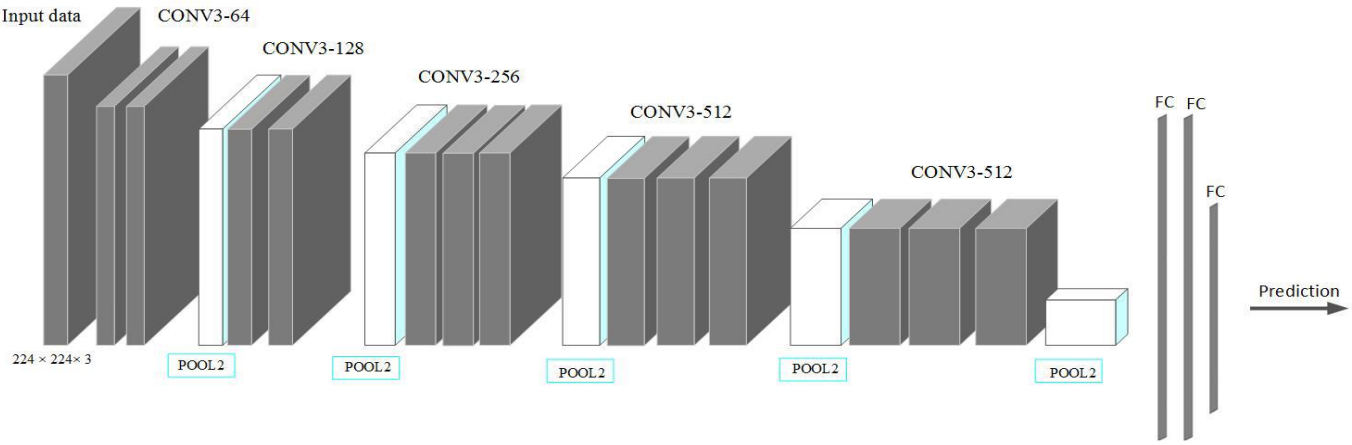

Figure 33: VGGNet Architecture [57]

2.8. Model Evaluation:

There are several metrics to evaluate the performance of a classification model on automatic image classification using deep Learning. The traditional accuracy metric may not be sufficient if the distribution of class labels is imbalanced. Precision-Recall metrics are especially useful when dealing with highly imbalanced classes, where precision measures the ability of the model to identify relevant data points while recall measures the ability of the model to find all relevant cases. Furthermore, the confusion matrix is a suitable method for summarizing the performance of a classification algorithm in the context of imbalanced classes [59]

The value of true positive (TP) and true negative (TN) and false positive (FP) and false-negative (FN) can be derived from the confusion matrix and are explained below:

- TP: correctly predicted positive class;
- FP: incorrectly predicted positive class;
- FN: incorrectly predicted negative class;
- TN: correctly predicted negative class.

2.8.1. Accuracy:

It is the measure of how accurately the classifier can classify the data. The following equation provides the accuracy:[60]

$$\text{Accuracy(range0} - 1) = \frac{TP + TN}{FP + FN + TP + TN}$$

2.8.2. Precision:

It defines the relation of total positive results that are correct to the classifier's total positive results, as provided by[60]

$$\text{Precision (range } 0 - 1) = \frac{TP}{TP + FP}$$

2.8.3. Sensitivity (or recall):

It corresponds to the TP rate of the considered class and is computed using [60]

$$\text{Senstivity (range } 0 - 1) = \frac{TP}{TP + FN}$$

2.8.4. F1 score:

The F1 score is defined as the harmonic mean of the model's precision and recall. The F1 score can be formulated in Equation[60]

$$\textbf{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

2.8.5. Specificity:

It corresponds to TN rate of the considered class (i.e., the proportion of negatives that have been correctly identified). The following equation provides the specificity:[61]

$$\text{Specificity (range } 0 - 1) = \frac{TN}{TN + FP}$$

2.8.6. Confusion Matrix:

A clean and unambiguous way to present the prediction results of a classifier is to use a confusion matrix [62]

	Predictive Values		
Actual Values		**Positive**	**Negative**
	Positive	TP	FN
	Negative	FP	TN

Figure 34: Confusion Matrix[62]

Chapter 3 – Review of Literature

Over the past decades, there has been tremendous research in the field of medical imaging, and many machine learning and deep learning models have been developed by researchers to get a more accurate result. The ease of use of AI has proven to be beneficial in the medical industry. For the early detection of retinal diseases, optical coherence tomography (OCT) is preferred.

3.1. First study:

"Coherent convolution neural network based retinal disease detection using optical coherence tomographic images" A study conducted by Somil et al[63]. The main objective of this study was to propose a deep learning model to improve the accuracy of retinal disease detection and reduce the efforts of medical experts. To achieve this, the researchers proposed coherent convolutional neural network (CCNN), an organized and efficient network with five layers, which used 84,495 optical coherence tomographic (OCT) images, while 83,484 images were used in the training of the model. The proposed network (CCNN) has been developed by pruning the pretrained model of VGG-16 and also included the batch normalization layers and the dropout layers for a better accuracy and prevention overfitting. The performance of the proposed model achieved an accuracy of 97.19% for the detection of the retinal disease.

3.2. Second study:

"A Deep Learning-Based Framework for Retinal Disease Classification" A study conducted by Savita et al [61].This study addressed the problem of automatic detection of retinal pathological conditions.

In order to solve the above problem, this study developed an artificial intelligence model. The model was based on a custom 19-layer deep convolutional neural network called the VGG-19 architecture. The model (VGG-19 architecture) was enabled by transfer learning and use the weights of the pre-trained ImageNet dataset. The researchers trained the model over 25 epochs. The model was designed to learn from a large set of images captured using optical coherence tomography (OCT) and classify them into four retinal conditions: (1) CNV, (2) clear drusen, (3) DME, and (4) natural form. The training datasets (had taken from publicly available sources) consisted of 84,568 copies of retinal OCT images. The proposed model achieved classification accuracy (99.17%) with sensitivity (99.00%), specificity (99.50%), and area under the curve (0.99917).

3.3. Third study:

"Detection of retinal diseases from ophthalmological images based on convolutional neural network architecture" A study conducted by Sezin et al [64] To diagnose retinal disorders utilizing OCT scans, this investigation developed a hybrid methodology based on image pre-processing and CNNs (a deep learning method). A publicly available OCT dataset developed by Kermany et al. (2018) was used[65]. The images were selected from retrospective cohorts of adult patients from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People’s Hospital, and Beijing Tongren Eye

Center between 2013 and 2017. The dataset consisted of four categories and 84.484 images (83.484 training and 1.000 test). Thirty-two images were exercised for validation, 83.484 images were exerted for the training, and 968 images were applied for the test. The pre-processing part consisted of 7 steps, including converting images to grayscale, resizing, histogram equalization, and noise reduction using filters such as mean, Gaussian, and nonlocal means. The deep learning architecture, based on the AlexNet architecture, consists of five coevolutionary layers with three fully connected layers. The architecture was trained and tested with pre-processed images and garnered 99.48% accuracy, 99.48% sensitivity, and 99.83% specificity. The researchers emphasized that data augmentation procedures were not applied in this study since they would add to the cost of system calculation. However, for classes with less data than other groups, a balance can be achieved to accumulate information by using other deep learning methods, such as data augmentation or Generative Adversarial Networks (GANs).

3.4. Fourth study:

"DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases" A study conducted by Jivansha et al [14]. This study presented a diagnostic tool based on a deep-learning framework for four-class classification of ocular diseases by automatically detecting diabetic macular edema, drusen, choroidal neovascularization, and normal images in optical coherence tomography (OCT) scans of the human eye. Images of retinal OCT scans for DME, drusen, CNV, and normal retina were taken from the public dataset (Mendeley database) published in Kermany et al[65]. Images were captured from the dataset and divided into training, test, and validation folders, each containing subfolders for four typical categories (CNV, DME, drusen, and normal), with a total of 84,495 b-scan views of OCT images in .jpeg format. The proposed study used a deep learning framework, the three different CNN models using five, seven, and nine layers were used for analysis of the OCT images and showed significant results. The five-layer CNN model showed an accuracy of 96.54% with precision of 97.19% in normal classes. The seven-layer CNN model exhibited slightly a higher precision for the DME and drusen classes. The nine-layer CNN model had a higher sensitivity and specificity value for accuracy in precision of 98.53% for the CNV class. The paper underlined the importance of image enhancement techniques, such as contrast enhancement and edge detection, which improved the accuracy and performance of the CNN models.

Practical Review

Chapter 4 –Materials and Methods

4.1. Dataset:

a publicly available OCT dataset developed by Kermany et al. (2018) was used. Optical coherence tomography (OCT) images (Spectralis OCT, Heidelberg Engineering, Germany) were selected from retrospective cohorts of adult patients from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center between July 1, 2013 and March 1, 2017.The dataset consisted of four categories (NORMAL,CNV,DME,DRUSEN) and 84.484 images. The dataset was organized into 3 folders (train, test, val) and contains subfolders for each image category as shown in Table 2.

Table 2: OCT images Dataset

Retinal Disorder Classes	No.-Training	No.-Validate	No.-Test
CNV	37205	8	242
Normal	26315	8	242
DME	11348	8	242
DRUSEN	8616	8	242

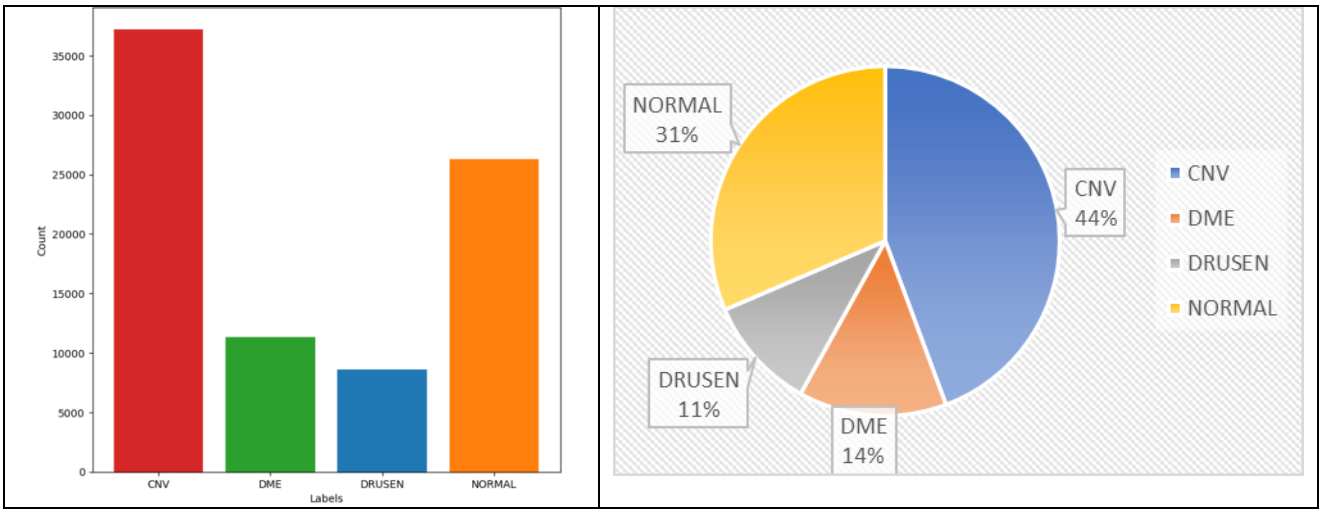

Figure 35: Distribution of Image Classes - Kermany Database

Link to download dataset : https://www.kaggle.com/datasets/paultimothymooney/kermany2018

Figure 36 shows an image for each disease:

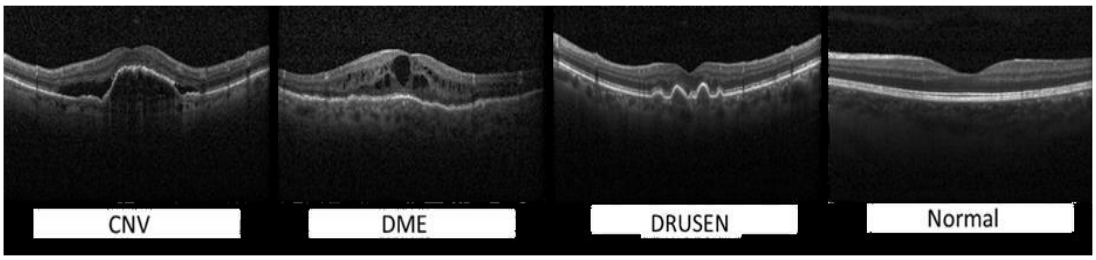

Figure 36: Comparison of OCT Scans for Different Diseases – Kermany Database

Due to data imbalance, we took a balanced set of data shown in table 3, and we used oversampling techniques by augmentation data to increase the number of images which are just copies of existing images.

Table 3: balanced OCT images Dataset

Retinal Disorder Classes	No.-Training	No.-Validate	No.-Test
CNV	11000	8	242
Normal	11000	8	242
DME	11000	8	242
DRUSEN	11000	8	242

4.2. Language and tool used:

This model was designed using Python language with Keras & Tensorflow libraries, an opensource deep learning framework, and tested on core system architecture with a Windows 10 Pro having 64-bit operating system, AMD Ryzen7 5800H with Radeon Graphics 3.20 GHz Processor and 16 GB of RAM.
Additionally, essential libraries such as Numpy and Matplotlib are implemented to perform crucial tasks, such as numerical computations and data visualization.
The most significant libraries are reviewed in Table 4:

Table 4: The most important libraries used in the project

Library	Description
pandas	Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data. Pandas support operations like Sorting, Re-indexing, Iteration, Concatenation, Conversion of data, Visualizations, Aggregations, etc.

numpy	The name “Numpy” stands for “Numerical Python”. It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data. It consists of in-built mathematical functions for easy computations. Even libraries like TensorFlow use Numpy internally to perform several operations on tensors. Array Interface is one of the key features of this library.
seaborn	Provides a high-level interface for drawing attractive and informative statistical graphics.
sklearn	It is a famous Python library to work with complex data. Scikit-learn is an open-source library that supports machine learning. It supports variously supervised and unsupervised algorithms like linear regression, classification, clustering, etc. This library works in association with Numpy and SciPy.
skimage	scikit-image is a collection of algorithms for image processing. It is available free of charge and free of restriction.
tensorflow	This library was developed by Google in collaboration with the Brain Team. It is an open-source library used for high-level computations. It is also used in machine learning and deep learning algorithms. It contains a large number of tensor operations. Researchers also use this Python library to solve complex computations in Mathematics and Physics.
keras	It is an open-source neural network library written in Python designed to enable fast experimentation with deep neural networks. With deep learning becoming ubiquitous, Keras becomes the ideal choice as it is API designed for humans and not machines, according to the creators.. Before installing Keras, it is advised to install the TensorFlow backend engine.
Matplotlib	This library is responsible for plotting numerical data. And that’s why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, histograms, scatterplots, graphs, etc.

4.3. Data preprocessing & Data Augmentation:

The images used in this study are of different sizes and resolutions because they were taken with different devices and are from different time periods, but the images must match the network input size to train the network. Also, most of the images contained high noise, and this leads to a high error rate. Previous studies had indicated that using proper pre-processing of images increases the performance of the deep learning model. Therefore, in this study, some image processing methods were applied:

- All images were resized to 224x224.
- Removing noise in images by using a non-local means filter: Compared to simpler denoising techniques like averaging, non-local means filters aim to strike a balance between noise reduction and detail preservation. By searching for similar patches within a larger window, the filter can replace noisy pixels while maintaining sharp edges and textures in the image.

Data Augmentation is the most widely used technique in a deep learning project while working with image data. This is done by applying domain-specific techniques to images from the training data that create new and different training images, and then adding those altered images into the dataset to increase its size.

Apart from increasing the size of the data, data augmentation also helps prevent overfitting, improves the model's performance on unseen data.

Transforms include a range of operations from the field of image manipulation, such as shifts, flips, rotation, zoom range, and brightness adjustments and much more.[66]

Table 5, shows the techniques used to train the proposed model.

Table 5: Augmentation techniques and values

Augmentation Techniques	Value
shear_range	0.2
zoom_range	(0.73, 0.9)
horizontal_flip	True
rotation_range	10
width_shift_range	0.10
height_shift_range	0.10
brightness_range	(0.55, 0.9)
fill_mode	constant

Figure 37 reflects the effects of these steps on the images

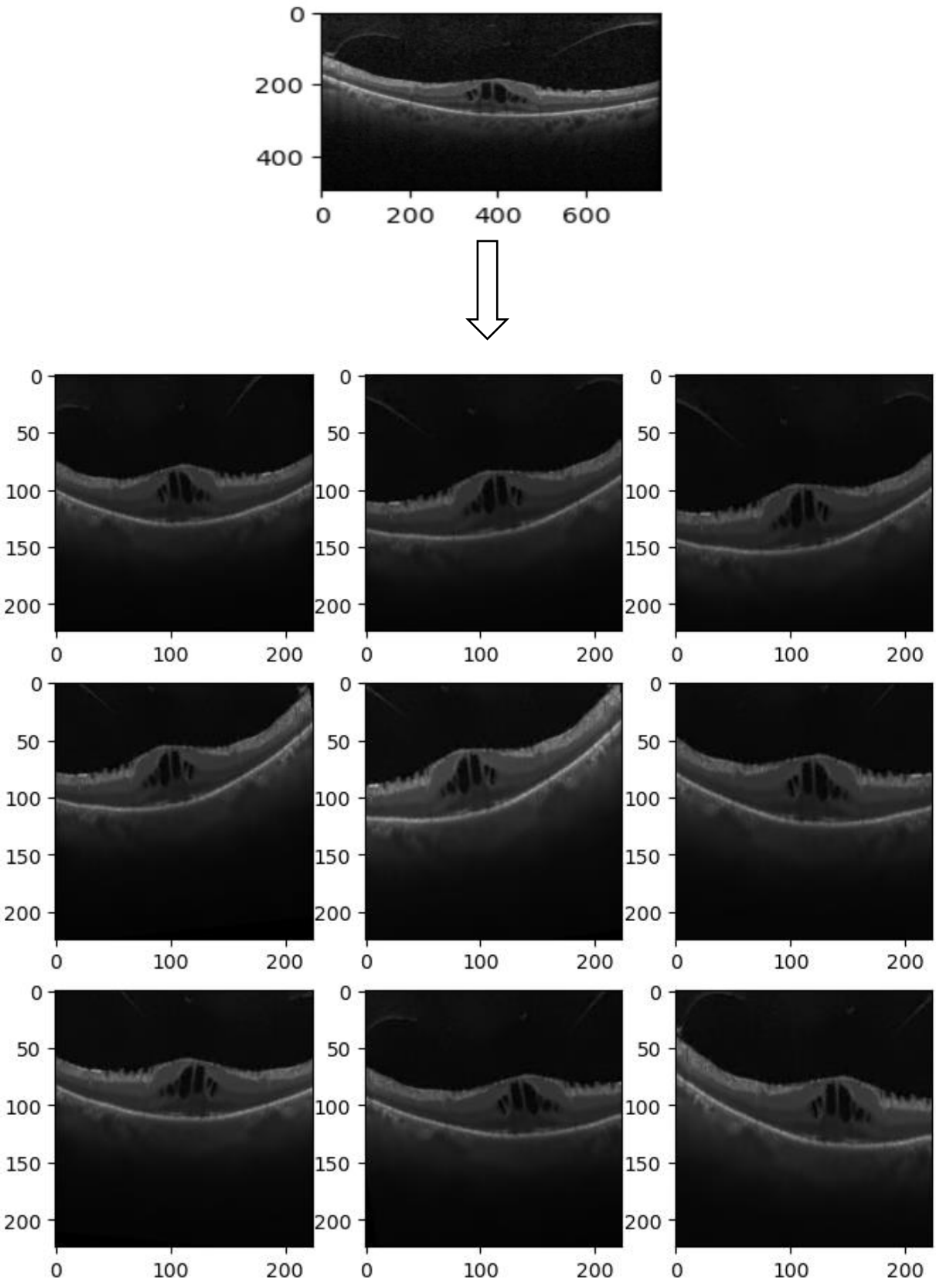

Figure 37: The effects of data preprocessing & data augmentation on the images

4.4. Model Architecture:

We used AlexNet architecture, a convolutional neural network (CNN) for image classification.

We used AlexNet architecture, a convolutional neural network (CNN) for image classification.We used tensorflow, keras, layers libraries to build and create layers for the model. The model consisted of:

- ✓ **Input Layer:**
  - layers.Conv2D: This is the first layer, performing convolution with 96 filters of size (11, 11) and a stride of (4, 4). The activation function is relu.
  - input_shape: This specifies the expected input shape of the image data (img_width, img_height, 3 color channels)
- ✓ **Convolutional Blocks:** The model had five convolutional layers, batchnormalization layers, and three maxpooling layers
  - layers.Conv2D: layer that extracts features from the input using filters. Parameters like filter size, stride, and padding define how the features are captured.
  - layers.BatchNormalization: layer improves training stability by normalizing the activations of the previous layer.
  - layers.MaxPool2D: layer downsamples the feature maps, reducing computational cost and potentially improving model generalization.
- ✓ **Fully-Connected Layers:**
  - layers.Flatten: This layer transforms the high-dimensional output of the convolutional layers into a 1D vector.
  - layers.Dense: This layer creates fully-connected layers with neurons for processing the flattened data. The first layer has 4096 neurons, and the final layer has num_classes neurons (4 in this case) for multi-class classification.
    - The first fully-connected layer uses the "relu" activation.
    - The final output layer uses the "softmax" activation to produce probabilities for each class.
  - layers.Dropout(0.5): This layer randomly drops 50% of the neurons during training to prevent overfitting.

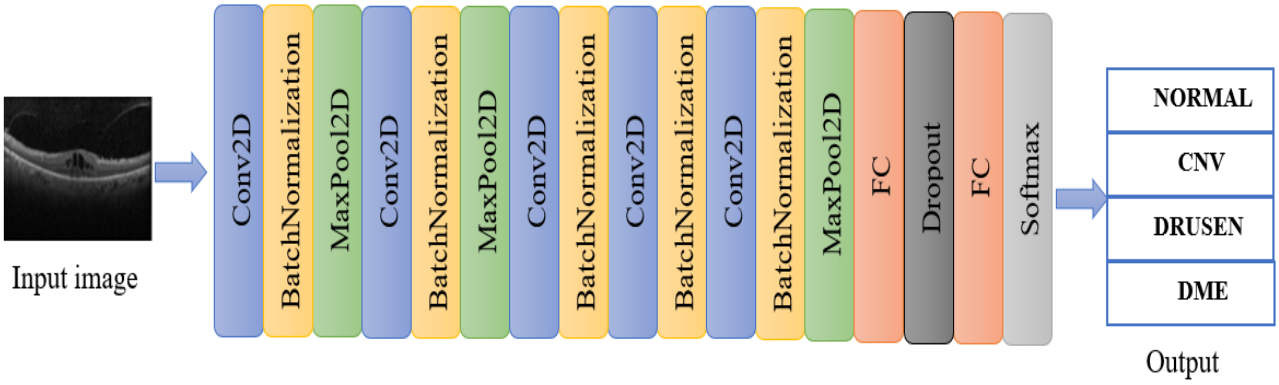

Figure 38: The block diagram of the CNN architecture.

The model was compiled using categorical cross-entropy as the loss function and Adam optimization algorithm was exerted during the training of the architectures. The learning rate was determined as 0.0001. Batch-size 32, epoch number 30.

```
def create_alexnet_model(img_width, img_height, num_classes):
    model = Sequential()
    model.add(layers.Conv2D(
            filters=96,
            kernel_size=(11, 11),
            strides=(4, 4),
            activation="relu",
            input_shape=(img_width, img_height, 3),
        )
    )

    model.add(layers.BatchNormalization())
    model.add(layers.MaxPool2D(pool_size=(3, 3), strides=(2, 2)))

    model.add(layers.Conv2D(filters=256, kernel_size=(5, 5), strides=(1, 1), activation="relu", padding="same"))
    model.add(layers.BatchNormalization())
    model.add(layers.MaxPool2D(pool_size=(3, 3), strides=(2, 2)))

    model.add(layers.Conv2D(filters=384, kernel_size=(3, 3), strides=(1, 1), activation="relu", padding="same"))
    model.add(layers.BatchNormalization())

    model.add(layers.Conv2D(filters=384, kernel_size=(3, 3), strides=(1, 1), activation="relu", padding="same"))
    model.add(layers.BatchNormalization())

    model.add(layers.Conv2D(filters=256, kernel_size=(3, 3), strides=(1, 1), activation="relu", padding="same"))
    model.add(layers.BatchNormalization())
    model.add(layers.MaxPool2D(pool_size=(3, 3), strides=(2, 2)))

    model.add(layers.Flatten())
    model.add(layers.Dense(4096, activation="relu"))
    model.add(layers.Dropout(0.5))
    model.add(layers.Dense(num_classes, activation="softmax"))  # Output for 4 classes

    model.compile(
        optimizer=tf.keras.optimizers.Adam(learning_rate=0.0001),
        loss="categorical_crossentropy",
        metrics=["accuracy"],
    )
```

To prevent overfitting, three callbacks were implemented. The first callback, ReduceLROnPlateau, was used to dynamically reduce the learning rate when the model's performance plateaued. This callback had a patience value of 3. The second callback employed was the early stop method: this callback was designed to halt the training process if the model's performance did not improve for a specified number of epochs. In this case was used an early stopping of 7 epochs. The third callback employed was the model checkpoint
callback function to save only best performance weights of the proposed model.
We used CSVLogger also. This callback logs the training and validation metrics (usually accuracy and loss) to a CSV file during each epoch. This allows you to track the training progress and analyze it later.

```
checkpoint = ModelCheckpoint( filepath="weightfinal.best.keras",monitor='val_acc', verbose=1, save_best_only=True, mode='max')
early_stopping = EarlyStopping(patience=7, mode='min', monitor='val_loss')
learning_rate_scheduler = ReduceLROnPlateau(monitor='val_loss', factor=0.1, patience=3, verbose=1,mode='min')
call=CSVLogger('my_trainfinall.csv',separator=',',append=False)
callbacks_list = [checkpoint, early_stopping, learning_rate_scheduler, call]
```

Chapter 5 – Results and Discussion

This study involved two stages: image pre-processing and classification. The pre-processing stage consisted of resizing and non-local denoising mean filter. In the classification stage, a five-layer CNN architecture was designed based on the AlexNet architecture. The pre-processed images were resorted to train and test the designed CNN architecture.
The proposed CNN architecture (Figure 38) was trained with the pre-processed images. Figure 39 and Figure 40 show the accuracy and loss curves during training.

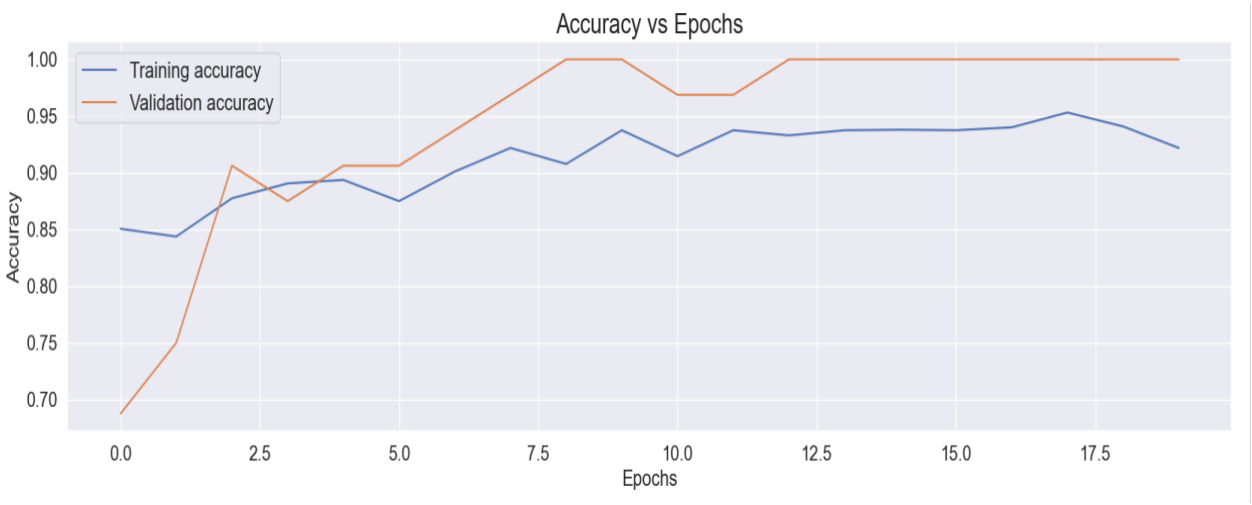

Figure 39: Accuracy curves in CNN architecture trained with pre-processed images.

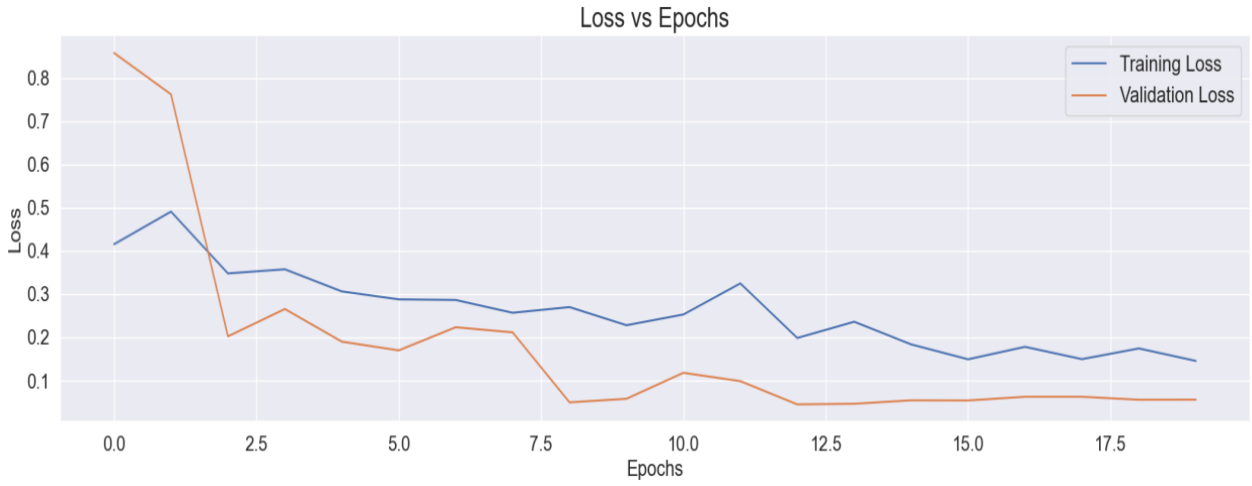

Figure 40: Loss curves in CNN architecture trained with pre-processed images.

The CNN architecture had an accuracy of 95.31% and a loss of 0.19 for training , and it had an accuracy of 100% and a loss of 0.03 for validation, suggesting that the training and validation were successful. Afterwards, the CNN architecture was tested to determine its performance as shown in Figure 41 and it had an accuracy of 99.38% and a loss of 0.039. Figure 42 presents the confusion matrix of the classification of the test dataset consisting of 968 images (242 for each category).

```
(eval_loss, eval_accuracy) = new.evaluate(test_generator, batch_size= batch_size, verbose= 1)
print('Test Loss: ', eval_loss)
print('Test Accuracy: ', eval_accuracy)
```

```
16/16 ━━━━━━━━━━━━━━━━━━━━ 7s 412ms/step - accuracy: 0.9921 - loss: 0.0359
Test Loss:  0.039779167622327805
Test Accuracy:  0.9938016533851624
```

Figure 41: Testing data in CNN architecture trained with pre-processed images.

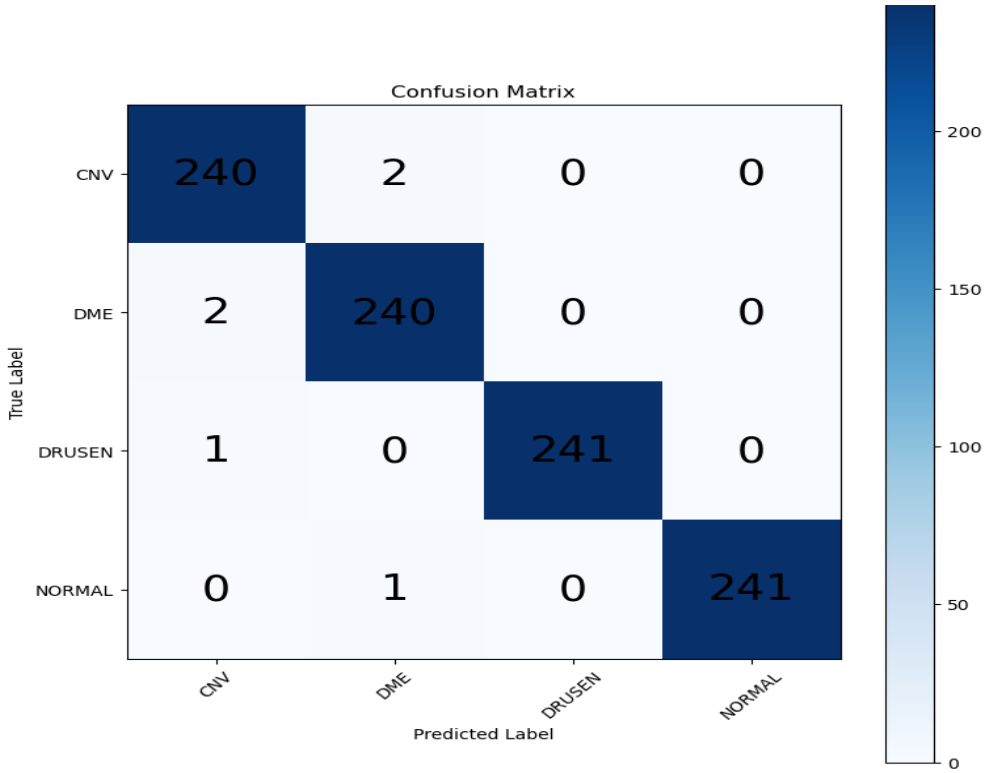

Figure 42: Confusion matrix of test data in CNN architecture trained with pre-processed images.

The confusion matrix shows that the CNN architecture evaluated 2 CNV images only incorrectly and labeled them as DME, it evaluated 2 DME images only incorrectly and labeled them as CNV, it evaluated 1 DRUSEN images only incorrectly and labeled them as CNV, it evaluated 1 NORMAL image only incorrectly and labeled them as DME.

Table 6 discloses the class-based performance of the CNN architecture on the test data.

Table 6: Class-based performance of CNN architecture on test data.

	Accuracy	Precision	Sensitivity	Specificity	F1-score	Data Number
CNV	0.9948	0.99	0.9917	0.9959	0.99	242
DME	0.9948	0.99	0.9917	0.9959	0.99	242
DRUSEN	0.9990	1.00	0.9959	1.0000	1.00	242
NORMAL	0.9990	1.00	0.9959	1.0000	1.00	242

the results obtained by using pre_processing and Keras techniques such as Adam optimization, data augmentation, learning rate, and others, helped to obtain these results.
Also, when we used the Oversampling technique to modify the unequal data classes to create balanced datasets, Dropout Layer in Alexnet architecture, and Earlystopping techniques, These techniques had a very significant impact on the results by preventing overfitting.
By comparing the results we obtained with the previous studies mentioned in this study[14], [61], [63], [64] , in which the testing accuracy results ranged between 96.54% and 99.48%, we noticed that the method proposed in this study ranks the second in terms of performance with 99.38% accuracy. The study that obtained the highest accuracy, which conducted by Sezin et al, relied on the Alexnet architecture, but with more layers than what was used in this study, and this will add to the cost of system calculation.

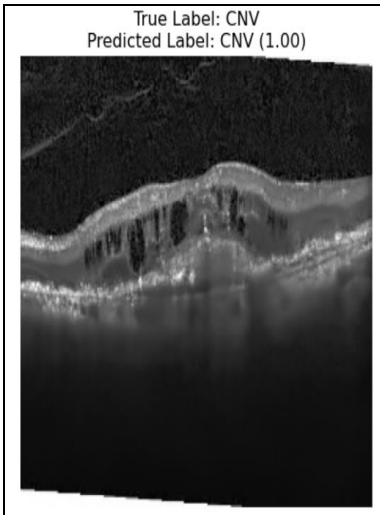

Figure 43: CNV model predict

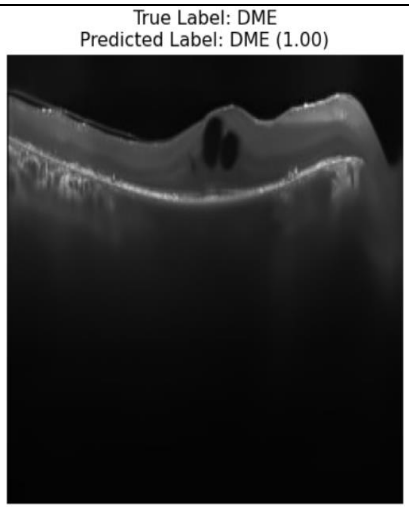

Figure 44: DME model predict

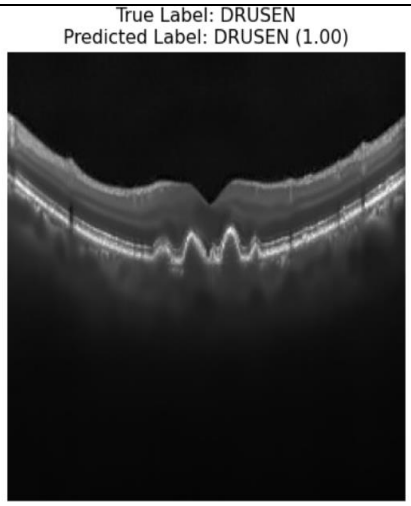

Figure 45: DRUSEN model predict

Chapter 6 – Conclusions and Future Work

Nowadays, eye diseases are among the most common reason for vision loss, these diseases can cause blindness and affect the lives of patients.

In this thesis, some of the techniques were used deep learning-based, included Alexnet architecture to classify four types of retinal (CNV, DME, DRUSEN, in addition to normal cases) from OCT images. Furthermore, the proposed models were also trained and tested using a huge dataset of images and we obtained testing accuracy rate of 99.38%.

Despite several challenges need to be resolved to increase DL adoption in medicine, deep learning solutions are among the most promising solutions to improve the general health of a people, and to lower the costs of healthcare.

In the future, the aim is to reduce the size of deep models for portable machines like mobiles to implement an online application to diagnose OCT images for helping doctors, patients, and clinics, also extension of this work will focus on developing hybrid algorithms that mean a combination of two models to increase the recognition rate of the final classification process underscoring the advantages of hybrid algorithms, and enhancing the performance of the model.

References

[1] M. Ptito, M. Bleau, and J. Bouskila, “The Retina: A Window into the Brain,” *Cells*, vol. 10, no. 12, p. 3269, Nov. 2021, doi: 10.3390/cells10123269.

[2] S. Muchuchuti and S. Viriri, “Retinal Disease Detection Using Deep Learning Techniques: A Comprehensive Review,” *J. Imaging*, vol. 9, no. 4, p. 84, Apr. 2023, doi: 10.3390/jimaging9040084.

[3] W. S. Wright, R. S. Eshaq, M. Lee, G. Kaur, and N. R. Harris, “Retinal Physiology and Circulation: Effect of Diabetes,” in *Comprehensive Physiology*, 1st ed., R. Terjung, Ed., Wiley, 2020, pp. 933–974. doi: 10.1002/cphy.c190021.

[4] S. S. Mishra, B. Mandal, and N. B. Puhan, “MacularNet: Towards Fully Automated Attention-Based Deep CNN for Macular Disease Classification,” *SN COMPUT. SCI.*, vol. 3, no. 2, p. 142, Mar. 2022, doi: 10.1007/s42979-022-01024-0.

[5] C. E. Willoughby, D. Ponzin, S. Ferrari, A. Lobo, K. Landau, and Y. Omidi, “Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review,” *Clinical Exper Ophthalmology*, vol. 38, no. s1, pp. 2–11, Aug. 2010, doi: 10.1111/j.1442-9071.2010.02363.x.

[6] “Anatomy of the Eye - Columbia Eye.” Accessed: Apr. 29, 2024. [Online]. Available: https://columbiaeye.com/anatomy-eye/

[7] N. Mahabadi and Y. Al Khalili, “Neuroanatomy, Retina,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2024. Accessed: Apr. 29, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK545310/

[8] A. Cerveró, A. Casado, and J. Riancho, “Retinal changes in amyotrophic lateral sclerosis: looking at the disease through a new window,” *J Neurol*, vol. 268, no. 6, pp. 2083–2089, Jun. 2021, doi: 10.1007/s00415-019-09654-w.

[9] G. Mohandass, R. A. Natarajan, and S. Sendilvelan, “Retinal Layer Segmentation in Pathological SD-OCT Images Using Boisterous Obscure Ratio Approach and its Limitation,” *Biomedical and Pharmacology Journal*, vol. 10, no. 3, pp. 1585–1591, Sep. 2017.

[10] C. J. Thomas, R. G. Mirza, and M. K. Gill, “Age-Related Macular Degeneration,” *Medical Clinics of North America*, vol. 105, no. 3, pp. 473–491, May 2021, doi: 10.1016/j.mcna.2021.01.003.

[11] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, and H. Soltanian-Zadeh, “Multi-scale convolutional neural network for automated AMD classification using retinal OCT images,” *Comput Biol Med*, vol. 144, p. 105368, May 2022, doi: 10.1016/j.compbiomed.2022.105368.

[12] “Age-Related Macular Degeneration (AMD): Types, Cause, Symptoms, and Treatment.” Accessed: Apr. 29, 2024. [Online]. Available: https://www.oscarwylee.com.au/glasses/eye/age-related-macular-degeneration

[13] A. Stahl, “The Diagnosis and Treatment of Age-Related Macular Degeneration,” *Deutsches Ärzteblatt international*, Jul. 2020, doi: 10.3238/arztebl.2020.0513.

[14] A. Tayal, J. Gupta, A. Solanki, K. Bisht, A. Nayyar, and M. Masud, “DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases,” *Multimedia Systems*, vol. 28, no. 4, pp. 1417–1438, Aug. 2022, doi: 10.1007/s00530-021-00769-7.

[15] “Age-Related Macular Degeneration (AMD) | National Eye Institute.” Accessed: Apr. 29, 2024. [Online]. Available: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/age-related-macular-degeneration

[16] A. Saraiva *et al.*, “Classification of Optical Coherence Tomography using Convolutional Neural Networks:,” in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, Valletta, Malta: SCITEPRESS - Science and Technology Publications, 2020, pp. 168–175. doi: 10.5220/0009091001680175.

[17] C. D. Haydinger, L. B. Ferreira, K. A. Williams, and J. R. Smith, “Mechanisms of macular edema,” *Front. Med.*, vol. 10, p. 1128811, Mar. 2023, doi: 10.3389/fmed.2023.1128811.
[18] “Symptoms of Diabetic Macular Edema (DME) - DME and ME.” Accessed: Apr. 29, 2024. [Online]. Available: https://dmeandme.com/symptoms-of-diabetic-macular-edema-dme/
[19] S. A. Hassan, S. Akbar, A. Rehman, T. Saba, H. Kolivand, and S. A. Bahaj, “Recent Developments in Detection of Central Serous Retinopathy Through Imaging and Artificial Intelligence Techniques–A Review,” *IEEE Access*, vol. 9, pp. 168731–168748, 2021, doi: 10.1109/ACCESS.2021.3108395.
[20] G. Lim, V. Bellemo, Y. Xie, X. Q. Lee, M. Y. T. Yip, and D. S. W. Ting, “Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review,” *Eye and Vis*, vol. 7, no. 1, p. 21, Dec. 2020, doi: 10.1186/s40662-020-00182-7.
[21] M. C. Savastano *et al.*, “Fluorescein angiography versus optical coherence tomography angiography: FA vs OCTA Italian Study,” *European Journal of Ophthalmology*, vol. 31, no. 2, pp. 514–520, Mar. 2021, doi: 10.1177/1120672120909769.
[22] Y. Deng *et al.*, “Age-related macular degeneration: Epidemiology, genetics, pathophysiology, diagnosis, and targeted therapy,” *Genes & Diseases*, vol. 9, no. 1, pp. 62–79, Jan. 2022, doi: 10.1016/j.gendis.2021.02.009.
[23] “Fluorescein Angiography (FA).” Accessed: Apr. 29, 2024. [Online]. Available: https://www.retinaconsultants.md/contents/diagnostic-testing/fluorescein-angiography-fa
[24] “Indocyanine Green (ICG) Angiography - Des Plaines, IL & Libertyville, IL: Retina Consultants, Ltd.” Accessed: Apr. 29, 2024. [Online]. Available: https://www.retinaconsultants.md/contents/diagnostic-testing/indocyanine-green-icg-angiography
[25] E. E. Cornish, A. Vaze, R. V. Jamieson, and J. R. Grigg, “The electroretinogram in the genomics era: outer retinal disorders,” *Eye*, vol. 35, no. 9, pp. 2406–2418, Sep. 2021, doi: 10.1038/s41433-021-01659-y.
[26] J. J. McAnany, O. S. Persidina, and J. C. Park, “Clinical electroretinography in diabetic retinopathy: a review,” *Survey of Ophthalmology*, vol. 67, no. 3, pp. 712–722, May 2022, doi: 10.1016/j.survophthal.2021.08.011.
[27] K. Yojana and L. T. Rani, “Optical coherence tomography based diabetic – ophthalmic disease classification and measurement using bilateral filter and transfer learning approach,” *Acta IMEKO*, vol. 12, no. 3, pp. 1–7, Sep. 2023, doi: 10.21014/actaimeko.v12i3.1345.
[28] G. R. Dias, “The Site for Healthcare Professionals: Optical Coherence Tomography (OCT),” The Site for Healthcare Professionals. Accessed: Apr. 30, 2024. [Online]. Available: https://learnbiomedengine.blogspot.com/2018/05/optical-coherence-tomography-oct.html
[29] Y. Xu *et al.*, “Artificial intelligence: A powerful paradigm for scientific research,” *The Innovation*, vol. 2, no. 4, p. 100179, Nov. 2021, doi: 10.1016/j.xinn.2021.100179.
[30] C. Zhang and Y. Lu, “Study on artificial intelligence: The state of the art and future prospects,” *Journal of Industrial Information Integration*, vol. 23, p. 100224, Sep. 2021, doi: 10.1016/j.jii.2021.100224.
[31] S. D. Baum, A. M. Barrett, and R. V. Yampolskiy, “Modeling and Interpreting Expert Disagreement ...”.
[32] S. O. Abioye *et al.*, “Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges,” *Journal of Building Engineering*, vol. 44, p. 103299, Dec. 2021, doi: 10.1016/j.jobe.2021.103299.
[33] L. Alzubaidi *et al.*, “A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications,” *J Big Data*, vol. 10, no. 1, p. 46, Apr. 2023, doi: 10.1186/s40537-023-00727-2.
[34] A. M. Barhoom, S. Abu-Naser, E. Alajrami, B. Abu-Nasser, M. Musleh, and A. Khalil, “Handwritten Signature Verification using Deep Learning,” Dec. 2019.

[35] “Figure 2 The structure of a Deep Neural Network with three hidden layers,” ResearchGate. Accessed: Apr. 30, 2024. [Online]. Available: https://www.researchgate.net/figure/The-structure-of-a-Deep-Neural-Network-with-three-hidden-layers_fig2_352996743
[36] “Figure 3. Relationship between artificial intelligence, machine...,” ResearchGate. Accessed: Apr. 30, 2024. [Online]. Available: https://www.researchgate.net/figure/Relationship-between-artificial-intelligence-machine-learning-neural-network-and-deep_fig3_354124420
[37] Hadjer Benmeziane, “Master Thesis - Comparison of Deep Learning Frameworks and Compilers,” 2020, doi: 10.13140/RG.2.2.15094.22085.
[38] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, “Review of Image Classification Algorithms Based on Convolutional Neural Networks,” *Remote Sensing*, vol. 13, no. 22, p. 4712, Nov. 2021, doi: 10.3390/rs13224712.
[39] M. Desai and M. Shah, “An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN),” *Clinical eHealth*, vol. 4, pp. 1–11, 2021, doi: 10.1016/j.ceh.2020.11.002.
[40] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, “Convolutional neural networks,” in *Machine Learning*, Elsevier, 2020, pp. 173–191. doi: 10.1016/B978-0-12-815739-8.00010-9.
[41] P. Scholar, “Diabetic Retinopathy Stage Classification using CNN,” vol. 06, no. 05, 2019.
[42] H. Z. Belbeisi, Y. S. Al-Awadi, M. M. Abbas, and S. S. Abu-Naser, “Effect of Oxygen Consumption of Thylakoid Membranes (Chloroplasts) From Spinach after Inhibition Using JNN,” 2020, Accessed: Apr. 30, 2024. [Online]. Available: https://philpapers.org/rec/BELEOO
[43] M. A. H. A. Bakr, H. M. Al-Attar, N. K. Mahra, and S. S. Abu-Naser, “Breast cancer prediction using JNN,” *International Journal of Academic Information Systems Research (IJAISR)*, vol. 4, no. 10, 2020.
[44] I. A. Alshawwa, H. Q. El-Mashharawi, M. Elkahlout, M. O. Al-Shawwa, and S. S. Abu-Naser, “Analyzing Types of Cherry Using Deep Learning,” *International Journal of Academic Engineering Research (IJAER)*, vol. 4, no. 1, 2020.
[45] P. Purwono, A. Ma’arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, “Understanding of Convolutional Neural Network (CNN): A Review,” *IJRCS*, vol. 2, no. 4, pp. 739–748, Jan. 2023, doi: 10.31763/ijrcs.v2i4.888.
[46] G. Wei, G. Li, J. Zhao, and A. He, “Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses,” *Sensors*, vol. 19, no. 1, p. 217, Jan. 2019, doi: 10.3390/s19010217.
[47] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[48] A. Saleh, R. Sukaik, and S. S. Abu-Naser, “Brain tumor classification using deep learning,” in *2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech)*, IEEE, 2020, pp. 131–136. Accessed: Apr. 30, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9328072/
[49] A. H. Alfarra, L. F. Samhan, Y. E. Aslem, M. M. Almasawabe, and S. S. Abu-Naser, “Classifications of Pineapple using Deep Learning,” 2021, Accessed: Apr. 30, 2024. [Online]. Available: https://philpapers.org/rec/ALFCOP-2
[50] H. R. Almadhoun and S. S. Abu-Naser, “Classification of alzheimer’s disease using traditional classifiers with pre-trained cnn,” 2021, Accessed: Apr. 30, 2024. [Online]. Available: https://philpapers.org/rec/ALMCOA
[51] B. El-Habil and S. S. Abu-Naser, “Cantaloupe Classifications using Deep Learning,” *International Journal of Academic Engineering Research (IJAER)*, vol. 5, no. 12, 2021, Accessed: Apr. 30, 2024. [Online]. Available: https://philarchive.org/archive/ELHCCU-2

[52] G. C. Cardarilli *et al.*, “A pseudo-softmax function for hardware-based high speed image classification,” *Sci Rep*, vol. 11, no. 1, p. 15307, Jul. 2021, doi: 10.1038/s41598-021-94691-7.
[53] M. A. Zaidan *et al.*, “Mutual Information Input Selector and Probabilistic Machine Learning Utilisation for Air Pollution Proxies,” *Applied Sciences*, vol. 9, no. 20, p. 4475, Oct. 2019, doi: 10.3390/app9204475.
[54] M. A. Aish, S. S. Abu-Naser, and T. N. Abu-Jamie, “Classification of pepper using deep learning,” 2022.
[55] M. Avand, A. N. Khiavi, M. Khazaei, and J. P. Tiefenbacher, “Determination of flood probability and prioritization of sub-watersheds: A comparison of game theory to machine learning,” *Journal of Environmental Management*, vol. 295, p. 113040, Oct. 2021, doi: 10.1016/j.jenvman.2021.113040.
[56] X. Han, Y. Zhong, L. Cao, and L. Zhang, “Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification,” *Remote Sensing*, vol. 9, no. 8, p. 848, Aug. 2017, doi: 10.3390/rs9080848.
[57] U. Muhammad, W. Wang, S. P. Chattha, and S. Ali, “Pre-trained VGGNet Architecture for Remote-Sensing Image Scene Classification,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing: IEEE, Aug. 2018, pp. 1622–1627. doi: 10.1109/ICPR.2018.8545591.
[58] Q. Guan *et al.*, “Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study,” *J. Cancer*, vol. 10, no. 20, pp. 4876–4882, 2019, doi: 10.7150/jca.28769.
[59] Md Jahid Hasan, Md. Shahin Alom, and U. Fatema, “Classification Performance Analysis of Retinal OCT Image using Handcrafted and Deep Learning Feature with Support Vector Machine,” 2021, doi: 10.13140/RG.2.2.36346.82888/2.
[60] L. Ali, F. Alnajjar, H. A. Jassmi, M. Gocho, W. Khan, and M. A. Serhani, “Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures,” *Sensors*, vol. 21, no. 5, p. 1688, Mar. 2021, doi: 10.3390/s21051688.
[61] A. Choudhary, S. Ahlawat, S. Urooj, N. Pathak, A. Lay-Ekuakille, and N. Sharma, “A Deep Learning-Based Framework for Retinal Disease Classification,” *Healthcare*, vol. 11, no. 2, p. 212, Jan. 2023, doi: 10.3390/healthcare11020212.
[62] S. R. Qwaider, S. S. Abu-Naser, and I. S. Zaqout, “Artificial Neural Network Prediction of the Academic Warning of Students in the Faculty of Engineering and Information Technology in Al-Azhar University-Gaza,” 2020.
[63] P. K. Upadhyay, S. Rastogi, and K. V. Kumar, “Coherent convolution neural network based retinal disease detection using optical coherence tomographic images,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9688–9695, Nov. 2022, doi: 10.1016/j.jksuci.2021.12.002.
[64] S. P. Taş, S. Barın, and G. E. Güraksın, “Detection of retinal diseases from ophthalmological images based on convolutional neural network architecture,” *Acta Sci Technol*, vol. 44, p. e61181, Jul. 2022, doi: 10.4025/actascitechnol.v44i1.61181.
[65] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
[66] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, “A review of medical image data augmentation techniques for deep learning applications,” *J Med Imag Rad Onc*, vol. 65, no. 5, pp. 545–563, Aug. 2021, doi: 10.1111/1754-9485.13261.