ABSTRACT

Retinal Disease Diagnosis using Deep Learning Is a system to process medical images using Deep Learning. It is a subset of Artificial intelligence (AI) which has the potential to revolutionize the disease diagnosis and management by performing classification difficult for human experts and by rapidly reviewing immense amounts of images. Despite its potential, the preparation and implementation of Artificial Intelligence algorithms is always challenging. In this Identifying medical diagnoses and treatable diseases by Image-based Deep Learning system, we sought to develop an effective transfer learning algorithm to process medical images to provide accurate predictions for diagnosis of the respective patient.

The retinal disease diagnosing process involves optical coherence tomography (OCT) images of the retina. The model (neural network) is trained on the OCT images to recognize the distinguishing features of specific classes of images of the human eye, faster and with fewer training examples and less computational power. Applying a conventional approach to a dataset of optical coherence tomography images, we can demonstrate performance comparable to that of human experts in classifying age related macular degeneration and diabetic macular edema. We can also provide a more transparent and interpretable diagnosis by highlighting the regions recognized by the neural network.

Deep learning is a branch of machine learning which is completely based on artificial neural networks, as neural networks are going to mimic the human brain so deep learning is also a kind of mimic of the human brain. It is a multidisciplinary field and aims to develop automated systems that are able to extract information from the unstructured data. Our aim is to classify a given Retinal OCT(Optical Coherence Tomography) image into one of the four different classes of Eye Diseases(CNV, DME, DRUSEN, NORMAL RETINA).

Keywords: Optical Coherence Tomography, Artificial Intelligence, Deep Learning, Neural Networks, Choroidal Neovascularization(CNV), Macular Degeneration, Diabetic macular edema

LIST OF FIGURES

CONTENTS

Drusen are like tiny pebbles of debris that build up over time. There are two different types of drusen: soft and hard.

"soft" drusen are large and cluster closer together

"hard" drusen are smaller and more spread out

Having a few hard drusen is normal as you age. Most adultsTrusted Source have at least one hard drusen. This type of drusen typically does not cause any problems and doesn't require treatment.

Soft drusen, on the other hand, is associated with another common eye condition called age-related macular degeneration (AMD). It's called "age-related" macular degeneration because it's more common in people older than 60.

As soft drusen get larger, they can cause bleeding and scarring in the cells of the macula. Over time, AMD can result in central vision loss. In other words, the condition can affect what you're able to see when you're looking straight ahead. Drusen can also occur in the optic nerve. Unlike drusen in the retina, optic nerve drusen can cause minor loss of peripheral (side) vision. Optic nerve drusen is not related to aging. They're more commonly seen in children.

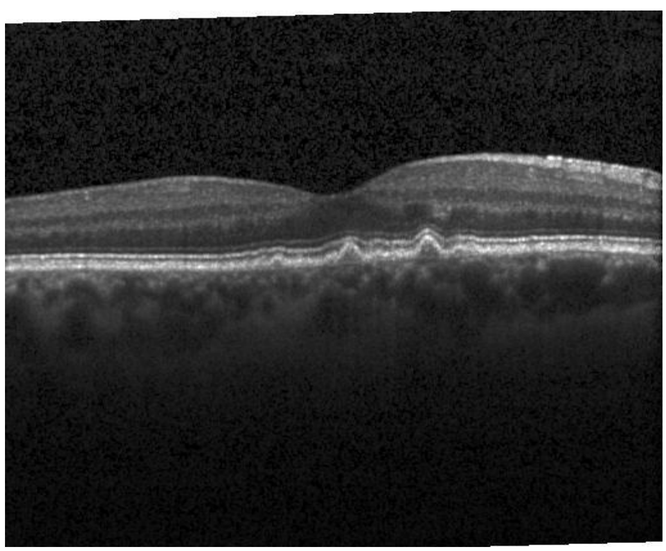

Figure 1.1.5 -DRUSEN (OCT Image)

1.2 Optical Coherence Tomography

Optical coherence tomography (OCT) is a noninvasive imaging technology used to obtain high-resolution cross-sectional images of the retina. OCT is similar to ultrasound testing, learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

Deep learning neural networks, or artificial neural networks, attempts to mimic the human brain through a combination of data inputs, weights, and bias. These elements work together to accurately recognize, classify, and describe objects within the data.

Deep neural networks consist of multiple layers of interconnected nodes, each building upon the previous layer to refine and optimize the prediction or categorization. This progression of computations through the network is called forward propagation. The input and output layers of a deep neural network are called visible layers. The input layer is where the deep learning model ingests the data for processing, and the output layer is where the final prediction or classification is made.

1.4 Introduction to Neural Networks

Artificial neural networks (ANNs) are software implementations of the neuronal structure of our brains. We don’t need to talk about the complex biology of our brain structures, but suffice to say, the brain contains neurons which are kind of like organic switches. These can change their output state depending on the strength of their electrical or chemical input. The neural network in a person’s brain is a hugely interconnected network of neurons, where the output of any given neuron may be the input to thousands of other neurons.

Learning occurs by repeatedly activating certain neural connections over others, and this reinforces those connections. This makes them more likely to produce a desired outcome given a specified input. This learning involves feedback – when the desired outcome occurs, the neural connections causing that outcome becomes strengthened. Artificial neural networks attempt to simplify and mimic this brain behavior. They can be trained in a supervised or unsupervised manner. In a supervised ANN, the network is trained by providing matched input and output data samples, with the intention of getting the ANN to provide a desired output for a given input.

An example is an e-mail spam filter – the input training data could be the count of various words in the body of the email, and the output training data would be a classification of whether the email was truly spam or not. If many examples of emails are passed through the neural network this allows the network to learn what input data makes it likely that an email is spam or not. This learning takes place by adjusting the weights of the ANN connections, but this will be discussed further in the next section. Unsupervised learning in an ANN is an attempt to get the ANN to “understand” the structure of the provided input data “on its own”.

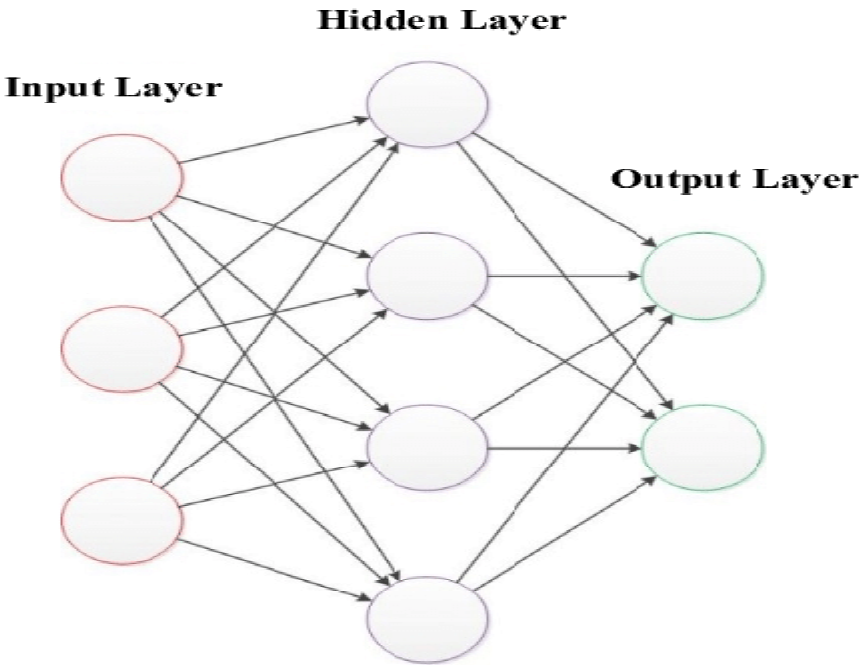

Figure 1.4.1 - Neural Network architecture

1.4.1 Structure of a node (Neuron)

The biological neurons are connected hierarchical networks, with the outputs of some neurons being the inputs to others. We can represent these networks as connected layers of nodes. Each node takes multiple weighted inputs, applies the activation function to the summation of these inputs, and in doing so generates an output.

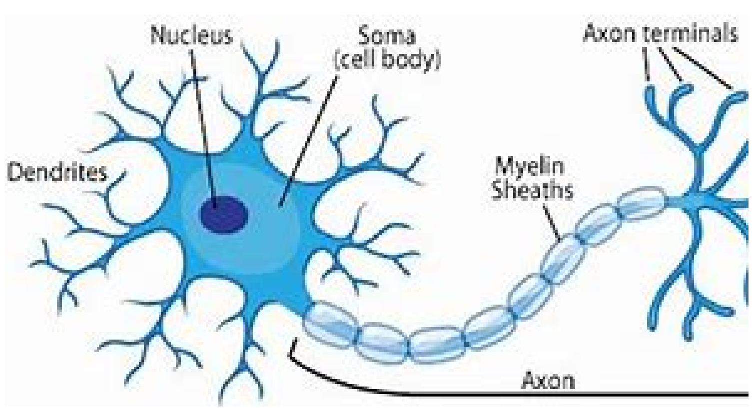

Figure 1.4.2 -Structure of a biological neuron

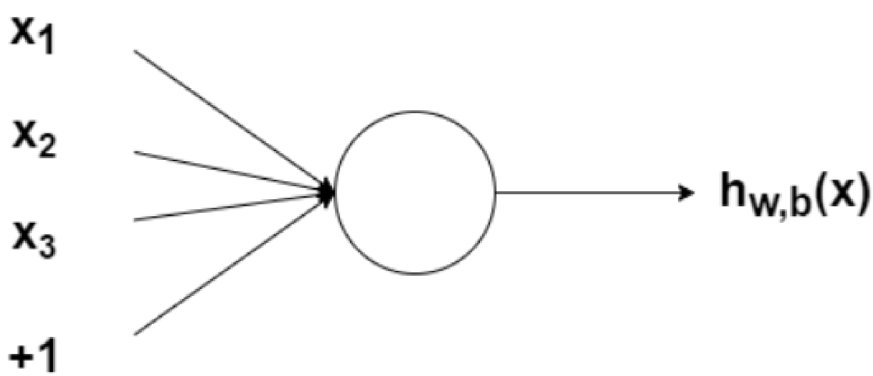

Figure 1.4.3 - Structure of an artificial neuron

Weighted input to the neuron = $x1w1 + x2w2 + x3w3 + b$

The circle in the image above represents the node. The node is the "seat" of the activation function, and takes the weighted inputs, sums them, then inputs them to the activation function.

1.4.2 Multilayer perceptron

Multi layer perceptron (MLP) is a supplement of feed forward neural network. It consists of three types of layers—the input layer, output layer and hidden layer. The input layer receives the input signal to be processed. The required task such as prediction and classification is performed by the output layer. An arbitrary number of hidden layers that are placed in between the input and output layer are the true computational engine of the MLP. Similar to a feed forward network in a MLP the data flows in the forward direction from input to output layer. The neurons in the MLP are trained with the back propagation learning algorithm. MLPs are designed to approximate any continuous function and can solve problems which are not linearly separable. The major use cases of MLP are pattern classification, recognition, prediction and approximation.

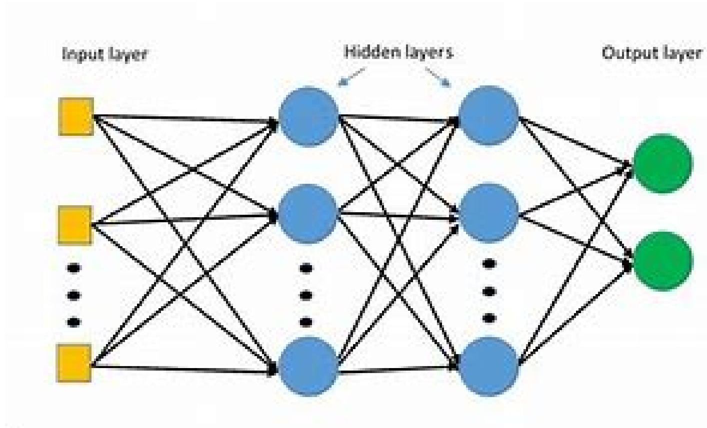

Figure 1.4.4 -Multilayer Perceptron (MLP)

1.4.3 Convolutional Neural Networks (CNN)

A convolutional neural network, or CNN, is a deep learning neural network designed for processing structured arrays of data such as images. Convolutional neural networks are widely used in computer vision and have become the state of the art for many visual applications such as image classification, and have also found success in natural language processing for text classification.

Convolutional neural networks are very good at picking up on patterns in the input image, such as lines, gradients, circles, or even eyes and faces. It is this property that makes convolutional neural networks so powerful for computer vision. Unlike earlier computer vision algorithms, convolutional neural networks can operate directly on a raw image and do not need any preprocessing.

A convolutional neural network is a feed-forward neural network, often with up to 20 or 30 layers. The power of a convolutional neural network comes from a special kind of layer called the convolutional layer.

The architecture of a convolutional neural network is a multi-layered feed-forward neural network, made by stacking many hidden layers on top of each other in sequence. It is this sequential design that allows convolutional neural networks to learn hierarchical features. The hidden layers are typically convolutional layers followed by activation layers, some of them followed by pooling layers. A simple convolutional neural network that aids understanding of the core design principles is the early convolutional neural network LeNet-5, published by Yann LeCun in 1998. LeNet is capable of recognizing handwritten characters.

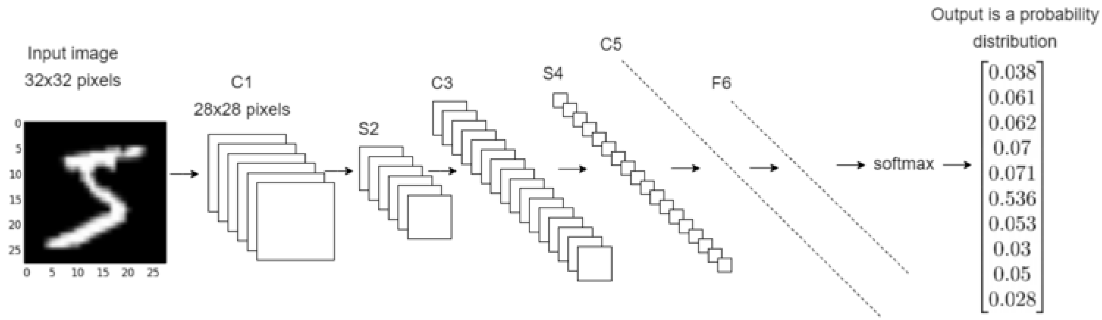

Figure 1.4.5 -Sample CNN architecture

1.4.4 Layers of CNN

Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

1. Convolutional layer
2. Pooling layer
3. Fully-connected (FC) layer

1.4.4.1 Convolutional Layer

The convolutional layer is the core building block of a CNN, and it is where the majority of computation occurs. It requires a few components, which are input data, a filter, and a feature map. Let's assume that the input will be a color image, which is made up of a matrix of pixels in 3D. This means that the input will have three dimensions—a height, width, and depth—which correspond to RGB in an image. We also have a feature detector, also known as a kernel or a filter, which will move across the receptive fields of the image, checking if the feature is present. This process is known as a convolution.

However, there are three hyperparameters which affect the volume size of the output that need to be set before the training of the neural network begins. These include: