Acta Scientiarum

http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v44i1.61181

Detection of retinal diseases from ophthalmological images based on convolutional neural network architecture

Safiye Pelin Taş[1], Sezin Barin[1] and Gür Emre Güraksin[2*]

[1]Biomedical Engineering Department, Engineering Faculty, Afyon Kocatepe University, Afyonkarahisar, Turkey. [2]Computer Engineering Department, Engineering Faculty, Afyon Kocatepe University, Ahmet Necdet Sezer Kampusü Gazligöl Yolu, 03200, Afyonkarahisar, Turkey. *Author for correspondence. E-mail: emreguraksin@aku.edu.tr

ABSTRACT. The retina is an eye layer that incorporates light- and color-sensitive cells as well as nerve fibers. It collects light and distributes it to the brain for image processing through the use of the optic nerve. Diseases that end up causing vision loss and blindness are generated by retinal ailments. As a result, it is imperative to diagnose and treat certain disorders as early as possible. Optical coherence tomography (OCT), an angiography imaging technique, is operated to help diagnose retinal disorders. Deep learning approaches, which are extensively utilized, have now become a convenient way for diagnosing retinal illnesses through OCT images as a result of their effective outcomes in interpreting medical images. To diagnose retinal disorders utilizing OCT scans, this investigation developed a hybrid methodology based on image pre-processing and convolutional neural networks (CNNs) (a deep learning method). Image pre-processing techniques including background filling, resizing, noise reduction, and highlighting are exercised at the pre-processing stage. The segmentation process provides a new CNN architecture with five convolution layers that does have a low computational cost. Compared to other publications using the same data set, the proposed method seems to have a success rate of 99.48 percent in the detection of retinal disorders, closing a significant gap in the literature. The proposed approach has the advantage of maintaining low computing costs in comparison to other studies in the literature. When the conclusions are regarded, it is noticed that the suggested method might be exerted as a decision support system to assist physicians in the clinical context during the diagnosis of retinal disorders.

Keywords: Convolutional neural network; image processing; opthalmological images; retinal diseases.

Received on October 12, 2021.
Accepted on January 2, 2022.

Introduction

The retina is an eye membrane that holds light- and color-sensitive cells as well as nerve fibers. It catches light and distributes it to the brain for image processing via the optic nerve (Kermany et al., 2018). Thus, diseases that cause vision loss and blindness may be driven by retinal abnormalities. Heart disease and hypertension can also triggered by retinal disorders (Miranda & Romero, 2019). Early identification and treatment are essential in preventing or eliminating such consequences. Age-related macular degeneration (AMD), drusen, diabetic retinopathy (DR), diabetic macular edema (DME), and myopic choroidal neovascularization are only a few of the serious disorders brought by retinal issues (CNV) (Taş, Barin, & Güraksin, 2021). DME is a retinal disorder that leads to damage in the blood vessels in the retina of the eye as well as consequent loss of vision (Jancy, Duela, Devi, & Lakshmi, 2021; Sharma, Khanna, & Bhargava, 2021). If DME is not treated, fluid leaks from the retina's blood vessels in the retina, causing swelling of the macula and eventually a sudden vision loss takes place (Jancy et al., 2021; Kaymak & Serener, 2018; Sharma et al., 2021). The higher amount of sugar in the blood, especially in Type II diabetic patients, increases the probability of DME (Jancy et al., 2021). Macular degeneration culminates in age-related macular degeneration (AMD). Dry and wet AMD are the two diverse types of AMD. Wet AMD is generated by the degeneration of retinal blood vessels, whereas Dry AMD is accompanied by profound retinal degeneration. While the dry form of AMD cannot be cured in general, the wet form of AMD can be managed in a variety of separate ways (Kaymak & Serener, 2018). Drusen is a term used to describe wet AMD. As the drusen swells, it bleeds and wounds the macula cells, triggering central vision loss (Sharma et al., 2021). In those over the age of 50, AMD is the leading cause of vision loss. Choroidal neovascularization (CNV) is a retinal disorder that produces vascular leakage or hemorrhage by permeating non-vascularized blood vessels' Retinal Pigment Epithelial

(RPE) cells (Rajagopalan, Venkateswaran, Josephraj, & Srithaladevi, 2021). CNV proceeds to the point where the RPE surface of the retina thickens (Gołębiewska et al., 2017). It is commonly noticed that age-related alterations or malignant myopic degeneration aggravate it (Sharma et al., 2021). Early detection of retinal abnormalities is critical for averting progression of disease and visual loss.

The imaging techniques of fundus fluorescein angiography (FFA) and optical coherence tomography (OCT) are frequently employed to investigate retinal disorders. A medical treatment in which a fluorescent dye is administered into the bloodstream to showcase the retinal blood vessels is known as fundus fluorescein angiography (Gupta, 2011). Optical coherence tomography (OCT) is a noninvasive imaging technique in which infrared rays are transmitted to the eye to capture high-resolution micron images of the retina (Rawat & Gaikwad, 2014.). These two methods are frequently performed in the diagnosis of retinal disorders. Each, though, has its own set of benefits and drawbacks. Fundus fluorescein angiography is a limited, slow, and invasive imaging procedure that is no longer exploited. Optical coherence tomography, on the other hand, is a novel, quick, and noninvasive technique that is better at evaluating a big capillary structure and delivering specific information on all of the eye's layers (Taş et al., 2021). When the history of the OCT technology is evaluated from its inception to the contemporary, it is noticeable that it has progressed at a rapid and successful rate (Leitgeb, 2019). OCT has become more and more widely practiced in the identification of retinal illnesses since it is a recent and still developing technology that delivers excellent results in imaging retinal layers. OCT comes in four different forms. They are known as Time Domain (TD-OCT), Spectral Domain (SD-OCT), Swept Source (SS-OCT) and Linear (L-OCT)(Cordes et al., 2021). TD-OCT and L-OCT fulfill A-scan in cartesian space. SD-OCT and SS-OCT carry out A-scan in Fourier space and require an FFT back to cartesian space to complete the A-scan. The article of Cordes et al (Cordes et al., 2021) can be referred to for more information about OCT scan systems. In retinal imaging, the varieties of TD-OCT and SD-OCT are more often exploited (Tayal et al., 2021). SD-OCT employs optical frequencies to measure the prevailed interferometric signal, facilitating imaging 50 times quicker than TD-OCT. It also presents a greater quantity of images per unit area. High-density raster scanning of retinal tissue is now possible thanks to the increased resolution and scanning speed (Forte, Cennamo, Finelli, & De Crecchio, 2009). These imaging techniques, as well as their own knowledge and experience, are benefitted by experts to screen for disorders and discover retinal diseases. Analyzing OCT and FFA ophthalmological images, on the other hand, is time-consuming, expensive, and subject to human error. Furthermore, late detection of retinal abnormalities prolongs, if not completely eliminates, the treatment procedure, resulting in a waste of time and effort. To address those problems in medical systems, sophisticated artificial intelligence (AI) approaches in medical diagnostics and image identification are extensively deployed (Avanzo et al., 2021; Born et al., 2021; Botwe et al., 2021; Law, Seah, & Shih, 2021).

AI is often applied in computational technologies that mimic mechanisms powered by human intelligence, such as thought, deep learning, adaptation, inheritance, and sensory understanding (Alzubi, Nayyar, & Kumar, 2018; Dong, Hou, Zhang, & Zhang, 2020; Son et al., 2019). It also comes in a range of applications in a variety of fields, particularly national security (Horowitz et al., 2018), health services (Calandra & Favareto, 2020; Li et al., 2021; Tan, Scheetz, & He, 2019), education (Chen, Chen, & Lin, 2020; Fadel, Holmes, & Bialik, 2019), and transportation (Kouziokas, 2017). Meanwhile, artificial intelligence's potential in healthcare has recently been proved by researchers. Modern computers' increased computational capacity and a massive reservoir of digital data have awarded AI the respect it deserves in the field of healthcare. Artificial intelligence models are benefitted in a variety of medical fields, involving risk modeling and stratification in biomedical research and clinical applications, personalized screening, diagnosis (including molecular disease subtype categorization), and therapeutic response prediction (Castiglioni et al., 2021). Artificial intelligence can receive, analyze, and report big data in divergent modalities to detect diseases and guide clinical decisions (Secinaro, Calandra, Secinaro, Muthurangu, & Biancone, 2021). Prominent in computer vision and a sort of artificial intelligence, deep learning algorithms can generate very good outcomes in fast and automatic classification of medical images, allowing for faster diagnosis with lower costs. As a result, deep learning approaches can help physicians overcome present drawbacks and detect and diagnose retinal illnesses more quickly. They can also make extremely accurate projections, saving both time and money.

Latest research mostly on identification of retinal disorders using OCT images have demonstrated that deep learning approaches are the most effective automatic retinal disease diagnosis techniques. Once the research are analyzed, preprocessing approaches are found to improve performance in studies using deep learning methods (Bhadra & Kar, 2020; Mezni, Slama, Mbarki, Seddik, & Trabelsi, 2021; Rajagopalan et al.,

2021; Tayal et al., 2021). Another conclusion made from the examination is that the architecture does not need to be large and complicated for a good classification success (Tayal et al., 2021). As a result of all of the examinations, the goal of this work was to classify as many OCT images as possible utilizing preprocessing techniques as well as a simple CNN architecture. Prioritization was assigned to either the preprocessing stage or the classification stage in the investigations when the preceding publications, whose details were supplied in the literature review chapter, were inspected. In addition, pre-trained deep neural networks with high computational cost architectures were generally preferred in the classification stage. This research was conducted to determine image quality through image processing as well as producing an ideal system by engaging in the classification architecture design. Another goal of the research was to keep the computing costs as low as possible. A hybrid system comprising of a CNN architecture with 5 convolution layers and preprocessing approaches has been described for this purpose. The findings were analyzed and compared to the most recent and noteworthy studies in the field. Image processing techniques such as resizing, space-filling, noise removal, and highlighting the region of interest were concentrated on throughout the image processing stage. A CNN architecture with low computational cost consisting of 5 convolution layers with 3x3 kernels was designed in the classification stage. 83.484 images achieved from the preprocessing step were used in the training of the designed CNN architecture. As a result of the training, 99.48% accuracy, 99.48% sensitivity, and 99.83% specificity were secured.

The remaining part of the paper is organized as follows: The introduction has commenced with a literature review. In the material and method section, the proposed hybrid system has been explained in detail after the technical details of the preprocessing, and CNN methods have been explained. The experimental results gained as a result of material and method have been explained and analyzed in the result and discussion section. In the result and discussion section, the suggested classification approach has also been compared to other state of art methods. Ultimately, the conclusions section has described all of the literature and discusses the study's future direction.

Literature review

The optical coherence tomography imaging technique is up-to-date and constantly evolving. Therefore, there has been a growing body of research on AI and deep learning in recent years. For example, Li et al. (Li et al., 2019) developed a CNN-based application to detect CNV, DME, and dry AMD diseases. To that end, they made use of 109,312 OCT images of 5319 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 the Beijing Tongren Eye Center between 2013 and 2017. They divided the images into four categories (CNV, DME, Drusen, and normal). They used the VGG-16 CNN architecture pre-trained in the ImageNet database. They reported that the proposed method had superior performance in retinal OCT images detection, with a prediction accuracy of up to 98.6%.

Mishra et al. (Mishra, Mandal, & Puhan, 2019) practiced the ResNet50 architecture to diagnose AMD and DME. They achieved an accuracy of 99.97% and a precision of 99.97% on the Duke database while they achieved an accuracy of 99.62% and a precision of 99.62% on the Neh database.

Alqudah (Alqudah, 2020) proposed a 19-layered CNN architecture called AOCT-NET to diagnose retinal diseases from OCT images. The AOCT-NET architecture achieved 97.78% accuracy, 97.78% sensitivity, and 97.778% specificity. Das et al. (Das, Dandapat, & Bora, 2019) proposed a multi-scale deep feature fusion-based classification approach to diagnose AMD, CNV, and DME from OCT images (n=83.484). The proposed method consisted of three steps, including a pre-processing. The method achieved 99.60% sensitivity, 99.87% specificity, and 99.60% accuracy.

Motozawa et al. (Motozawa et al., 2019) exercised two computational deep learning-based CNN models to diagnose AMD. The first model analyzed OCT images in two categories: AMD and healthy. The second model classified AMD images as exudative and non-exudative. The researchers also investigated the effect of transfer learning on the learning rate of the second model. The first model diagnosed AMD with 99.0% accuracy, 100% sensitivity, and 91.8% specificity. The second model detected exudative cases with 98.4% sensitivity, 88.3% specificity, and 93.8% accuracy.

Saha et al. (Saha et al., 2019) focused on the automatic classification of biomarkers causing AMD. They applied 19584 OCT B-scans from 153 patients to train and test three CNN models (Inception-V3, ResNet50, and Inception-ResNet50). The Inception-ResNet50 architecture outperformed the others. It was able to identify the subretinal drusenoid deposits with a sensitivity, specificity, and accuracy of 79, 92, and 86%,

respectively. It also managed to detect hyperreflective foci (hRF) within druseniod lesions with a sensitivity of 78% and a specificity of 100%, while it detected the hyporeflective foci (hRF) within druseniod lesions with 79% precision and 95% specificity. Researchers in general reported relatively successful results in different transfer learning architectures (Bhowmik, Kumar, & Bhat, 2019; Kamble et al., 2018; Wang et al., 2019).

Wang et al. (Wang et al., 2020) combined two CNN architectures and proposed a new method. The architectures distinguished images based on the presence of choroidal neovascularization (CNV) and identified CNV scans based on the presence of a CNV membrane. The first architecture was used for CNV membrane identification and segmentation, while the second was used for pixel-based vessel segmentation. The proposed model was able to diagnose all CNV with 100% sensitivity and 95% specificity.

Alqudah et al.(Alqudah, Alqudah, & AlTantawi, 2021) proffered a hybrid artificial intelligence system for multiclass classification of eye retina diseases using automated deep features extracted from OCT images. They used the OCT dataset published by Zhang Lab at the University of California at San Diego (UCSD) (Kermany et al., 2018) and the Farsiu 2013 Ophthalmology AMD dataset available at Duke University. The total dataset they used consisted of five classes: normal, AMD, CNV, DME, and Drusen. In the intended system, the features extracted from the CNN-based AOCT-NET architecture were classified using eight varying machine learning methods: Support Vector Machine with Linear kernel (LSVM), Support Vector Machine with Radial Basis Function kernel (RBF SVM), Artificial Neural Network (ANN), K Nearest Neighbor (KNN), Random Forest (RF), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Naïve Bayes (NB). CNN-based AOCTNET, previously proposed by Alqudah (Alqudah, 2020) for the diagnosis of retinal diseases, possessed a structure consisting of 19 layers. As a result of the evaluations, KNN machine learning displayed the best success with 99.44% accuracy, 99.44 sensitivity, 99,86 specificity and 99.45 precision.

Thomas et al. (Thomas et al., 2021) submitted a novel deep CNN architecture to automate AMD diagnosis from OCT images. AMD images taken from 4 different data sets were used in the study (Farsiu et al., 2014; Kermany et al., 2018; Rasti, Rabbani, Mehridehnavi, & Hajizadeh, 2018; Srinivasan et al., 2014). In this study, a CNN architecture with 6 convolution layers and two feed-forward routes was designed to extract features from OCT images and three machine learning methods (Multi-Layer Perceptron, SVM and Random Forest) were used to detect AMD disease using extracted features. The proposed technique with Random Forest has acquired the best performances with accuracy of 99.78% and sensitivity 99.8% via Kermany et al.'s dataset (Kermany et al., 2018) .

Najeeb et al. (2019) presented an image pre-processing algorithm to acquire the region of interest (ROIs) from retinal OCT images and a single layer convolutional neural network structure to detect retinal diseases from segmented OCT images. They used the OCT dataset containing four classes published by Zhang Lab at the University of California at San Diego (UCSD) (Kermany et. al. 2018). At the end of the study, they achieved 95.66% accuracy and 95% sensitivity.

Berrimi and Moussaoui (2020) introduced two new CNN architectures to detect retinal diseases from OCT scans. They used the OCT dataset containing four classes published by Zhang Lab at the University of California at San Diego (UCSD) (Kermany et. al. 2018), similar to the other studies. The two proposed CNN architectures also consisted of 3 convolution layers. The only difference between them was that the second CNN architecture included extra dropout and batch normalization layers. In this study, they also trained the VGG 16 and Inception V3 models to compare the proposed architectures. As a result of the trainings, Inception V3 architecture exhibited the best result with 99.27% accuracy. When the evaluation results were ranked from the best to the worst, the second CNN architecture designed by Berrimi and Moussaoui (2020) came after the Inception V3 architecture with an accuracy rate of 98.65%.

Bhadra and Kar (2020) proposed a deep multi-layered CNN for retinal disease detection and classification. They used the OCT dataset containing four classes published by Zhang Lab at the University of California at San Diego (UCSD) (Kermany et al., 2018). Their supplied methodology owned two sections: pre-processing and classification. The pre-processing section contained image processing methods to remove artifacts and segment ROIs. A CNN architecture consisting of six convolution layers were used to classify OCT scans in the classification section. Their proposed technique seemed to obtain accuracy of 96.5% for test dataset.

Mezni et al. (2021) tendered an advanced filtering and classification framework to detect AMD and DME disease using OCT images. The OCT image dataset consisting of 95 normal and 162 abnormal (AMD, DME) images was taken from the Hedi Raies Institute of Ophthalmology in Tunisia between February 2019 and August 2020. The proposed method included two parts: pre-processing and classification. OCT image pre-processing strategy was first applied by the use of the Block-matching and 3D filtering (BM3D) filter to reduce

the Gaussian noise in used OCT image data. Second, the appreciated classification classified the macular region using the Deep Belief Neural (DBN) methodology. The proposed method gained an accuracy of %92.54, sensitivity of 92.08% and specificity of 93.25%.

Tayal et al. (Tayal et al., 2021) presented a classification method based on 3 different CNN architectures consisting of 5,7,9 convolution layers to detect DME, Drusen and CNV using the OCT scans. The used OCT dataset prepared by (Kermany et al., 2018) was preprocessed for noise removal, contrast enhancements, contour-based edge, and detection of retinal layer extraction. The training was done with both raw and processed images and it was proven that the proposed preprocessing method increased the classification success. In the study, the architecture trained with pre-processed data gave the best accuracy (%97.14), while sensitivity (0.9447) and specificity (0.9816) values provided the worst outcome. Among the architectures trained with raw images, CNN architecture with five convolution layers achieved the best accuracy (% 96.54), while CNN architecture with seven convolution layers achieved the best results in sensitivity (0.9654) and specificity (0.9884).

Sharma et al. (2021) designed a system consisting of preprocessing and CNN-based architecture to classify three retinal disorders (CNV, DME, Drusen). In the pre-processing stage, Random Under Sampling technique was applied for tackling the dataset's class imbalance. Then resizing was applied to bring the data to the same size, and augmentation was applied through Tensorflow ImageDataGenerator to increase the data again. A CNN architecture consisting of 10 convolution layers was used to classify retinal diseases. The proposed system achieved an accuracy of 99.38% with a precision amounting to 0.9938.

Rajagopalan et al. (Rajagopalan et al., 2021) planned a CNN architecture and a series of image pre-processing method to detect retinal disorders using OCT images. In pre-processing, four various filters were used for speckle noise reduction: mean filter, wiener filter, Kuan, and biorthogonal spline wavelet filter. Researchers used CNN architecture with 5 convolution layers and determined the training hyperparameters by random search method. Their system provided an accuracy of 97.01%, sensitivity of 93.43%, and 98.07% specificity.

Li, Cheng et al. (2021) proposed an end-to-end deep learning algorithmic framework based on domain adaptation Inception V3 to classify six categories of retinal OCT images using the prior knowledge of a similar domain. Using two different data sets, the researchers applied non-local mean filtering, laplacian filtering, and normalization processes to the images before the classification process, which was followed by applying data augmentation with the ImageDataGenerator, and then resized the images (229x229x3). Their method achieved 96.3% accuracy, 96.2% sensitivity, 99.2% specificity, and 96.2% F1-score.

Apon et al. (Apon, Hasan, Islam, & Alam, 2021) proposed a CNN model with four convolution layers to identify OCT images in four classes. The study aimed to design a significantly small and fast system to create a web application for a real-time OCT diagnostic system. In the study, the proposed CNN architecture with four convolution layers achieved %94.87 accuracy without preprocessing except resizing. The proposed CNN architecture held 423,460 parameters for training. In addition, the researchers examined the interpretability of the proposed model, which could provide major information to specialists for the diagnosis of retinal diseases with LIME and Grad-Cam methods.

Considering the literature review, it is clear that either the pre-processing stage or the categorization stage is given the utmost importance in these studies. Furthermore, at the classification process, pre-trained deep neural networks with high computational cost designs were frequently favored.

Material and methods

In this study, an automated computer-aided system was designed to diagnose retinal diseases from OCT images. The proposed system consisted of two main stages just like other OCT studies observed in the literature (Bhadra & Kar, 2020; Mezni et al., 2021; Rajagopalan et al., 2021; Tayal et al., 2021): They are the stages of pre-processing and deep learning architecture-based classification. The pre-processing part consists of 7 steps:

1 All images were converted to gray-scale.
2 White regions were filled in to remove background distortions.
3 All images were resized to 128x128.

Having completed all these stages, morphological image processing methods were used to reduce noise and highlight the essential points.

4 Contrast limited adaptive histogram equalization (CLAHE) was applied to the images.
5 Median filter was applied to the images.
6 Gaussian filter was applied to the images.
7 Non-local means denoising filter was applied to the images.

In the classification stage, a CNN architecture was designed based on AlexNet, which is known as a simple deep learning architecture. The CNN architecture consisted of five convolution layers and three fully connected layers. The flow chart of the proposed study is given in Figure 1.

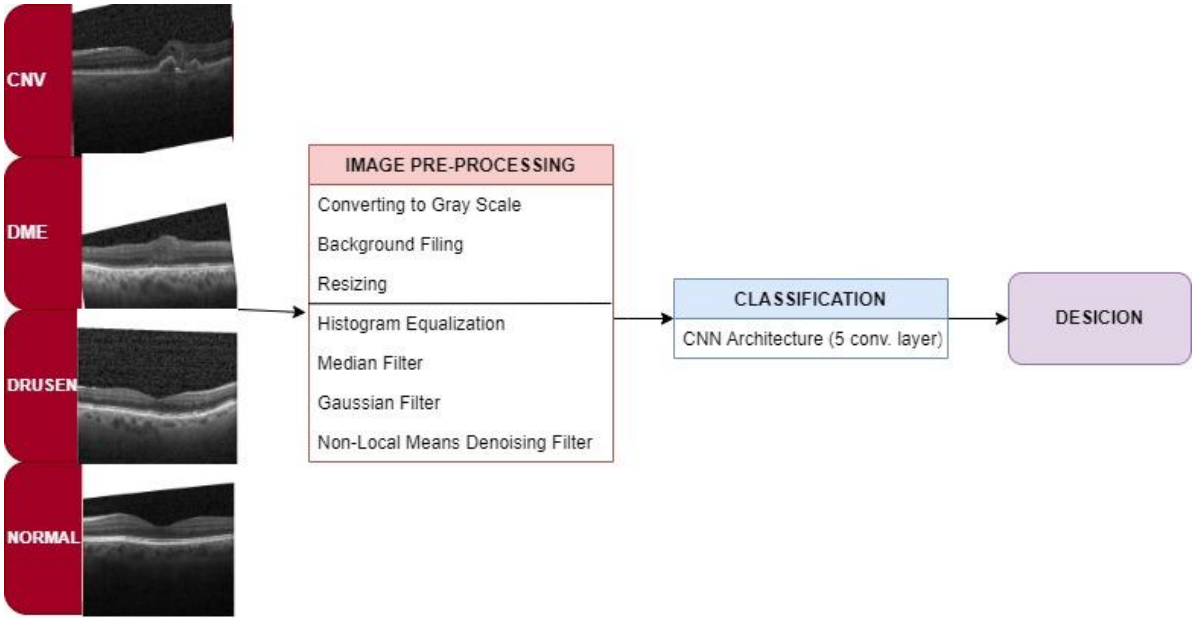


Figure 1. Flow chart of the study.

In addition, AlexNet architecture was trained with pre-processed images to compare the performance of the proposed CNN architecture with the AlexNet. Also, the proposed CNN architecture was trained using both raw images and pre-processed images to compare the effect of the pre-processing step on the performance.

Dataset

As for the dataset, a publicly available OCT dataset developed by Kermany et al. (2018) was used. 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 (Table 1). An example image of each class in the data set is given in Figure 2.

Table 1. Distribution table of data set by classes.

Data Set	Category	Number of Images
Training	CNV	37.205
	DME	11.348
	DRUSEN	8.616
	NORMAL	26.315
Validation	CNV	8
	DME	8
	DRUSEN	8
	NORMAL	8
Test	CNV	242
	DME	242
	DRUSEN	242
	NORMAL	242

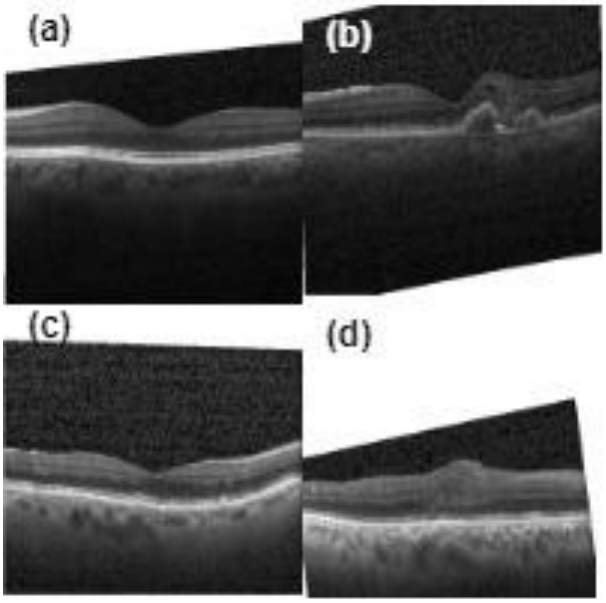


Figure 2. Retinal OCT images of (a) Normal (b) CNV (c) Drusen (d) DME.

Image preprocessing

The images used in this study are in various sizes and resolutions because they are from dissimilar periods of time and captured with different devices. Each image has a unique magnification rate. Most images have salt-and-pepper noise, while others have denser noise than normal OCT images. CNN architectures can be trained with original OCT images. However, they may require more convergence time and have a high error rate due to distortion (see Figure 2 for examples). In addition to these mentioned factors, studies marked that when image preprocessing methods were applied correctly, they increased the performance of deep learning methods (Li, Cheng et al., 2021; Mezni et al., 2021; Tayal et al., 2021; Uysal & Güraksin, 2021). In the study, some image processing methods were applied to image standardization and remove noise in images.

The data set used in the study included both gray-scale and RGB-scale images. For this reason, all images were converted to gray-scale, thus reducing the computational cost and standardizing the images. RGB image was represented by three color component intensities such as red, green, and blue. RGB image owned a 24 bits/pixel where 8 bits for each color (red, green and blue). Grayscale images were a monochrome image that contained brightness and intensities information only and no color information. A grayscale image contained a 8 bit/pixel allowing the image to represent (0-255) different brightness (gray) levels (Padmavathi & Thangadurai, 2016). RGB images involved more information than grayscale images. Although this increased the success of deep learning methods, it brought extra computational costs.

In image processing methods, thresholding is a widely used method for many purposes (Sezgin & Sankur, 2004). Thresholding is a simple but effective tool, especially for separating objects from the background. The thresholding method is so effective and widespread has also laid the groundwork for the continuous improvement of the applied methods, and many thresholding methods have emerged according to the application purpose. In images, the gray levels of the pixels belonging to the object are usually significantly dissimilar from the gray levels of the pixels belonging to the background. (Sezgin & Sankur, 2004). In the light of this information, an adaptive thresholding method, widely used in gray-scale images, whose formulation is given in Equation 1, is used.

$$dst(x,y) = \begin{cases} 0 & if\ src(x,y) > T(x,y) \\ maxValue & otherwise \end{cases} \quad (1)$$

where T(x,y) is a threshold calculated with adaptive methods individually for each pixel (*OpenCV: Image Thresholding*, December 18, 2021). The threshold value is a mean of the neighborhood of (x,y) .

In all deep learning architectures, the images applied to the input layer must be the same size. For this reason, all the images in the data set were resized to the same size by the interpolation method.

Histogram equalization is a simple and effective method for improving contrast in digital images. CLAHE is a type of adaptive contrast enhancement method and in previous studies it appeared to be effective in contrast enhancement methods of OCT images (Tayal et al., 2021). It aims to reduce the noise produced inhomogeneous areas and was initially developed for medical imaging (Rosenman, Roe, Cromartie, Muller, &

Pizer, 1993). In CLAHE, the image is divided into subsections. Each subsection has a histogram (Tayal et al., 2021). Each histogram is cropped using a crop threshold, and the contrast of each subsection is enhanced using histogram equalization. Those subsections are then combined to incur a contrast-enhanced image. This method is widely used in ophthalmology because it is pretty good at enhancing the visibility of hidden features of gray-scale images. (Uysal & Güraksin, 2021; Yadav, Maheshwari, & Agarwal, 2014). Two factors are determined in CLAHE algorithms: Clip limit and Tile size. ClipLimit can be expressed as the contrast enhancement limit, preventing image oversaturation, especially in homogeneous areas. Without ClipLimit, the adaptive histogram equalization technique may produce worse results than the original image. Tile size refers to the number of rows and columns that the image will be divided into.

Every image has a component of noise. Kumar and Sodhi (2020) categorized noises into four groups as Amplifiers or Gaussian noise, Salt and Pepper noise, Shot or Poisson noise, and Speckle noise. Outliers, known as "salt-and-pepper noise," result from bit errors during transmission or signal acquisition. Median smoothers are commonly used to remove noise-distorted images, especially outliers (Kumar & Sodhi, 2020). In addition, previous studies have also shown that the median filter is effective in removing noise in OCT images (Rajagopalan et al., 2021). The median filter is based on the logic of moving pixel-by-pixel on the image and replacing each pixel value with the median value of the pixels of neighbors in the specified window (Image Filtering).

The Gaussian smoothing operator is a 2-D convolution operator employed to blur images and remove detail and noise. In this sense, it is similar to the median filter, but it uses a discrete kernel representing the shape of the Gaussian hump (Wüthrich, Trimpe, Cifuentes, Kappler, & Schaal, 2017). The gaussian filter was generally used to suppress the speckle noise in OCT images (Devi, Ramkumar, Kumar, & Sasi, 2021). Probability Density Function (P(x)) of Gaussian distribution is represented by Equation (2)(Kumar & Sodhi, 2020).

$$P(x) = \frac{1}{\sqrt{2n\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (2)$$

Here, x is a gray level image. μ is mean value, and σ is the standard deviation. The standard deviation (σ) of the Gaussian determines the amount of smoothing (Ahmad, Khan, & Iqbal, 2019).

Buades et al. (Buades, Coll, & Morel, 2011) were the first to propose the non-local means denoising filter as a noise reduction algorithm. The method is based on the principle of generating a new value for a pixel by averaging similar pixel values (Buades et al., 2011). Similar pixel values may not always be neighbors or in close proximity. The method scans a large section of the matrix to find similar pixel values, which is why it is superior to other methods. In previous studies, the non-local mean filter was used to remove Gaussian noise in OCT images (Li, Cheng et al., 2021). Two factors were determined in non-local means denoising filter. Size in pixels of the template patch that was used to compute weights and Size in pixels of the window that was used to compute weighted average for given pixel(*OpenCV: Denoising*, December 12, 2021).

Classification

This stage consists of classification of retinal diseases through OCT images using CNN architectures. CNN architectures consist of convolution, fully connected layer, and pooling layers. CNN architectures are created by combining these layers by arranging them one after the other or by designing them as modules. The Convolution Layer is the most crucial layer in ESA. The primary purpose of convolutional layers is to extract features from inputs. Typically, edge, line, and corner detection features are reported in the first layers. Also, combining the dense layers later with the first layers allows for more specific features to be extracted, such as the appearance and occurrence of objects. In this layer, some filters are applied to the image to extract low- and high-level features from the image. The size of the applied filter is set by the architect of the model, but the pixel values of the filter used are optimized during the training of the architecture. Figure 3 displays the convolution operation of a 4x4 size 2D-image and a 2x2 size kernel.

Another momentous layer of the CNN architecture is pooling. In the pooling layer, spatial dimension reduction is performed gradually(Li, Karpathy, & Networks, 2016). The pooling layer not only preserves prominent properties, but also reduces space invariance (information loss). In this way, the number of learnable parameters for the model is reduced and problems such as learning slowness and excessive learning are prevented. In the pooling layer, the pooling process is applied to the pixels remaining in the window by sliding the window of a certain size on the property map. The most commonly used pooling process, which has three unalike application methods as Average, Global Average and Maximum, is maximum pooling(Ajit, Acharya & Samanta, 2020). Figure 4. manifests the maximum pooling applied to a 4x4 feature map. In the example given, the window size is set to 2x2.

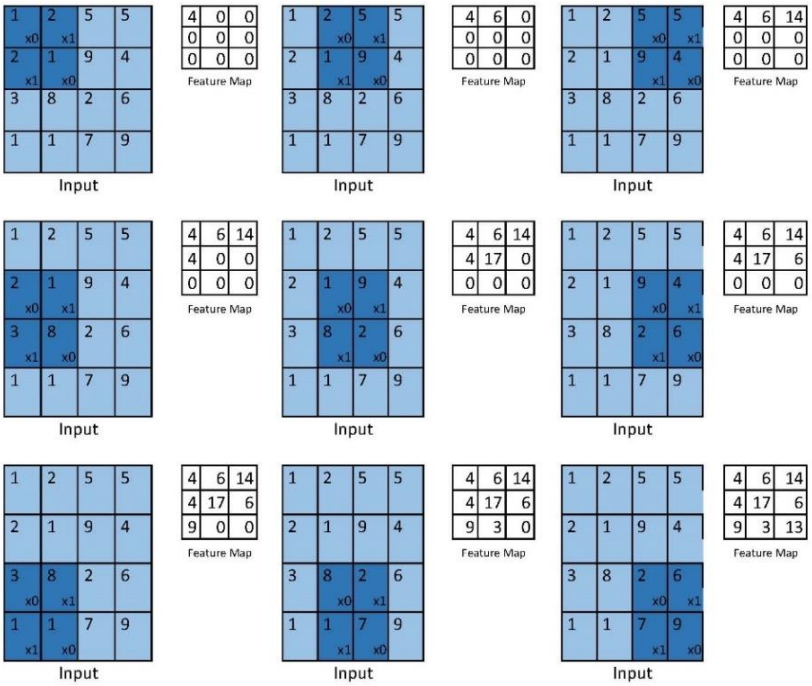


Figure 3. Example convolution operation.

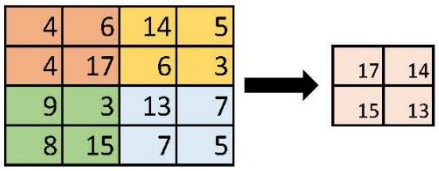


Figure 4. Example max-pooling operation.

Fully connected layers frame the last layers of a CNN architecture. Outputs from layers preceding these layers represent high-level features in the input data of the CNN. The fully connected layer flattens all neurons in the previous layer by connecting them to the neurons within it and then connects them to the output layer. This is a low-cost way to learn nonlinear combinations of features (Coşkun, Yildirim, Uçar, & Demir, 2017). A convolution layer cannot be used after the fully connected layer. The number of neurons of the last fully connected layer in a CNN architecture is equal to the number of classes. Figure 5 indicates an example of a fully connected layer.

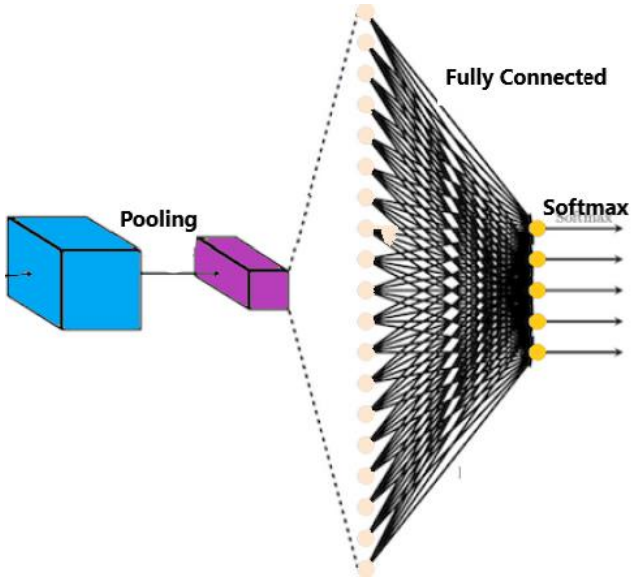


Figure 5. Example of fully connected layer.

To build and train a CNN model from scratch is often very costly. Because one of the disadvantages of deep learning models is that they need a lot of data for training. For this reason, architectures such as ImageNet, which are trained with big data before and achieve very successful results, are mostly used in the studies (Berrimi & Moussaoui, 2020; Bhowmik et al., 2019; Kermany et al. 2018; Ye et al., 2019). These pre-trained models have learned how to analyze the image, and it is possible to retrain and exploit the model with less data for our problems by using the feature maps of these models. This method is called as transfer learning. AlexNet is one of the pre-trained transfer learning architectures. AlexNet is a CNN-based deep learning architecture designed by Krizhevsky, Sutskever, and Hinton (2012). AlexNet was first recognized by its GPU technology at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2012 and has paved the way for new CNN-based deep learning architectures since then. In general, AlexNet consists of an input layer (227x227x3), five convolutional layers, and three fully-connected layers. See (Krizhevsky et al., 2012) for more detailed information on AlexNet.

Performance metrics

In this study, the performance was assessed using accuracy, sensitivity, specificity, precision, and F1-Score was calculated using a confusion matrix, which summarizes predictions in a classification problem. A confusion matrix separates the number of true and false predictions by each class, thus showing the prediction profile of the architecture in a comparative fashion during classification. The following can be calculated using a confusion matrix:

. True Positive (TP) is a measure of the cases predicted as true to be actually true

. True Negative (TN) is a measure of the cases predicted as false to be actually false

. False Positive (FP) is a measure of the cases predicted as true to be actually false

. False Negative (FN) is a measure of the cases predicted as false to be actually true

. Accuracy is the ratio of the total number of correct predictions to the total number of predictions (Equation 3).

$$Accuracy = \frac{TP+TN}{FP+FN+TP+TN} \quad (3)$$

. Sensitivity: The proportion of true positives correctly predicted as positives (Equation 4).

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

. Precision: The ratio of true positive predictions to the overall number of positive predictions. (Equation 5).

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

. Specificity: The proportion of true negatives correctly predicted as negatives (Equation 6).

$$Specificity \ \frac{TN}{FP+TN} \quad (6)$$

. F1-Score: Harmonic mean of precision and sensitivity values (Equation 7).

$$F1-Score = \frac{2*TP}{FP+FN+(2*TP)} \quad (7)$$

. Gradient-weighted Class Activation Mapping (Grad-CAM): It uses gradients of any target concepts that progresses to the final convolution layer as to produce a rough localization map by highlighting the notable regions in the image for classification. To create a Grad-CAM heatmap, a model is created that is cut at the layer to be mapped, and fully connected layers are added for prediction. Next, the image to create a heatmap is given as input to the model. The layer where the heatmap will be created is taken and the gradient of this output is detected. Finally, the heatmap is overlaid with the original image.

Proposed method

As shown in the flowchart in Figure 1, this study consists of two parts as pre-processing and classification. Each section contains differing stages. This part of the article explains all the study steps in order.

The pre-processing stage consisted of seven steps (Figure 1).

1 All images were converted to gray-scale.

2 The white regions in the background were filled to eliminate distortions. For background filling, the images were converted to binary format using thresholding. This process can be used for varying purposes, such as reducing noise or identifying an object on an image. Thresholding updates pixels to black or white based on a given threshold value (Bradley & Roth, 2007). Afterward, the images were complemented, and black and white pixels were replaced to obtain a mask, which was then used to fill the background of the original images.

3 All images were resized to 128x128.

Four other pre-processing techniques were utilized to remove noise and highlight weighty points to diagnose diseases.

4 CLAHE with (5,5) tile size and 1.0 clip limit was applied to the images. Figure 6 compares the pre-CLAHE and post- CLAHE images. Because it is so effective at boosting the visibility of hidden features in gray level images, this approach is commonly exploited in medical imaging and ophthalmology (Uysal & Güraksin, 2021; Yadav et al., 2014).

5 The median filter with 33 kernel size was applied to the images to remove noise. Brightness and transparency were adjusted after applying the median filter.

6 There was an increase in salt-and-pepper noise after histogram equalization. However, masking failed to reduce that noise, either. Therefore, the last stage involved a Gaussian filter and non-local denoising means for noise reduction. The Gaussian smoothing operator with -1 kernel standard deviation and (3,3) kernel size was used to blur images and decrease noise. In this sense, It's close to the median filter, but it employs a distinct kernel to reflect the Gaussian hump's shape (Wüthrich et al., 2017).

7 Lastly, non-local means denoising filter with 31 search window size and 11 block size was applied to all images.

Figure 6 reveals the effects of all steps on images.

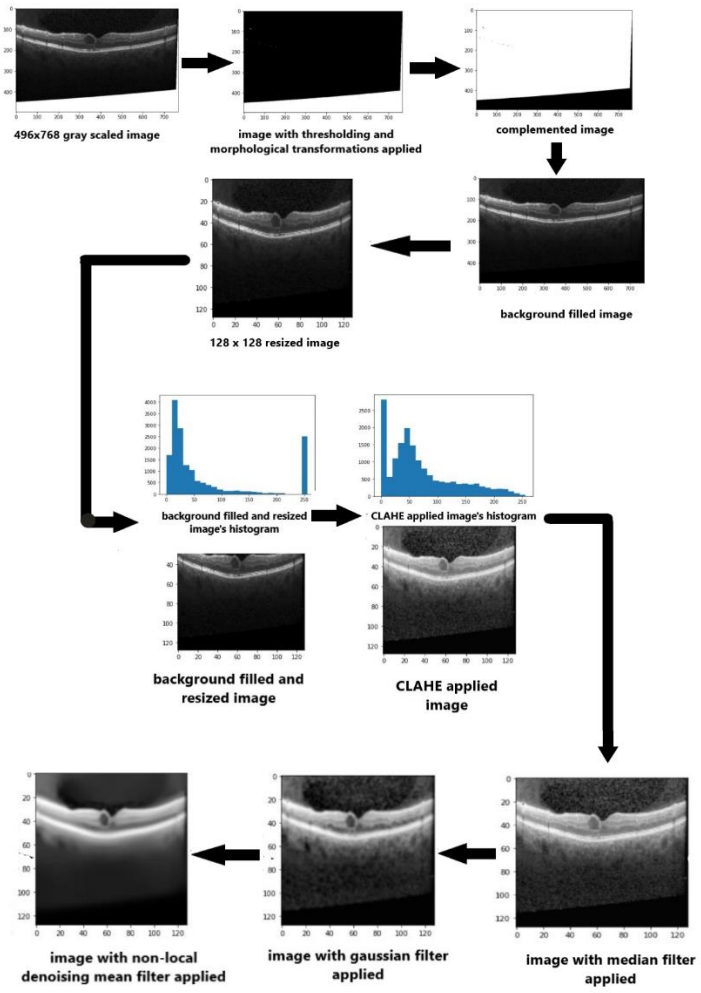


Figure 6. Images after the seven steps of pre-processing.

The classification step is based on the deep CNN architecture designed to classify the pre-processed images. The CNN architecture was designed according to the AlexNet architecture, which is a simple deep learning architecture. The reason why AlexNet was the architecture of choice is that it is not only cost-effective but also it performs well in classification. The AlexNet architecture has three pooling layers. However, the proposed CNN architecture had five pooling layers also. Also, the input layer size was set to 128x128, and the filter size of layers, especially fully connected layers, were reduced. These have led to a decrease in the number of parameters and the cost of computation. Parameter numbers of AlexNet and the proposed architecture are given in Table 2 for comparison. The AlexNet with pre-processed images and CNN architectures with raw images were also trained to evaluate the performance of the CNN architecture and to determine the effect of pre-processing on classification.

Table 2. Comparison of parameter numbers of AlexNet and proposed CNN architecture.

Architecture	Total Parameter Numbers	Trainable Parameter Numbers
AlexNet	343183956	343.163.788
Proposed CNN	3715332	3.715.332

The input layer of AlexNet was resized to 128x128 for training because the proposed CNN architecture had an input layer of 128x128.

In this study, the structure of the deep CNN architecture, which was designed based on AlexNet architecture, included 5 convolutions with 3x3 kernel in the feature extraction part and 4 fully connected layers in the classification part. ReLU activation functions were used in all convolution layers and 3 fully connected layers (except the last one). A softmax activation function was utilized in the last fully connected layer. Also, max-pooling layer with 2x2 window size was applied to each convolution layer output. Figure 7. expresses the block diagram of the proposed CNN architecture.

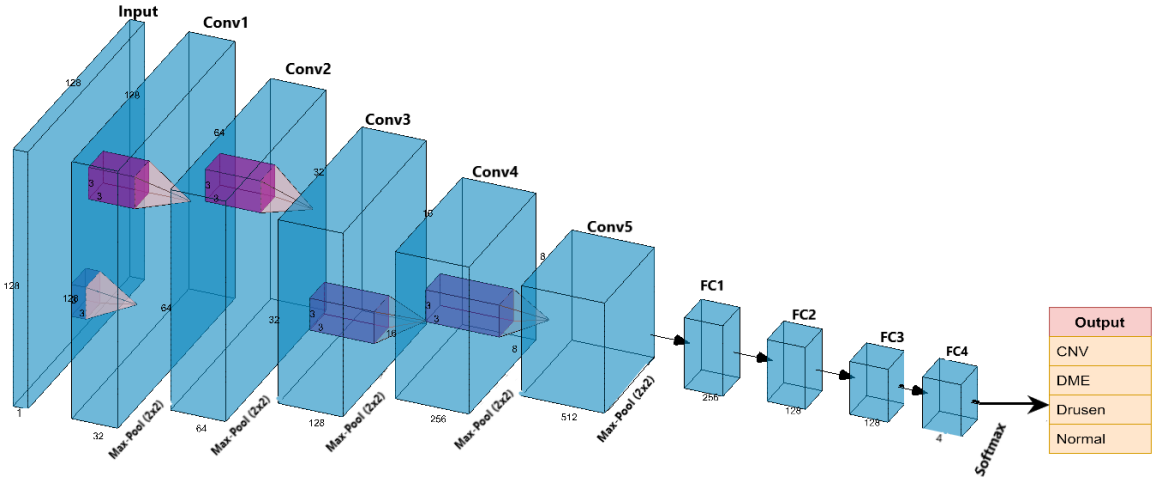


Figure 7. The block diagram of the CNN architecture.

A first-order gradient-based stochastic optimization algorithm, the ADAM optimization algorithm was exerted during the training of the architectures. The learning rate was determined as 0.0001. Batch-size 64, epoch number 48, and cross-entropy as loss function were preferred during the training process.

Results and discussion

This study involved two stages: image pre-processing and classification. The pre-processing stage consisted of seven steps: (1) converting images to gray-scale, (2) removing background distortions, (3) resizing, (4) histogram equalization, (5) median filter, (6) Gaussian filter, and (7) 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. In addition, The AlexNet with pre-processed images and CNN architectures with raw images were also trained to evaluate the performance of the CNN architecture and to determine the effect of pre-processing on classification.

The pre-processed images and the original images are displayed in Figure 8. The images indicate that the regions that are eminent for diagnosis became more prominent after pre-processing.

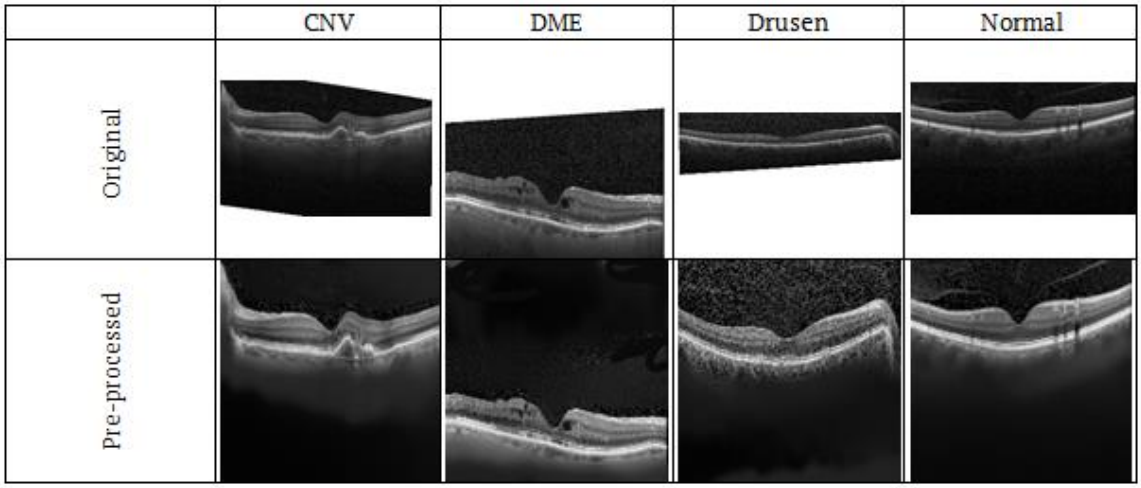


Figure 8. Original and pre-processed images.

The proposed CNN architecture (Figure 7) was trained with the pre-processed images. Figure 9 shows the accuracy and loss curves during training.

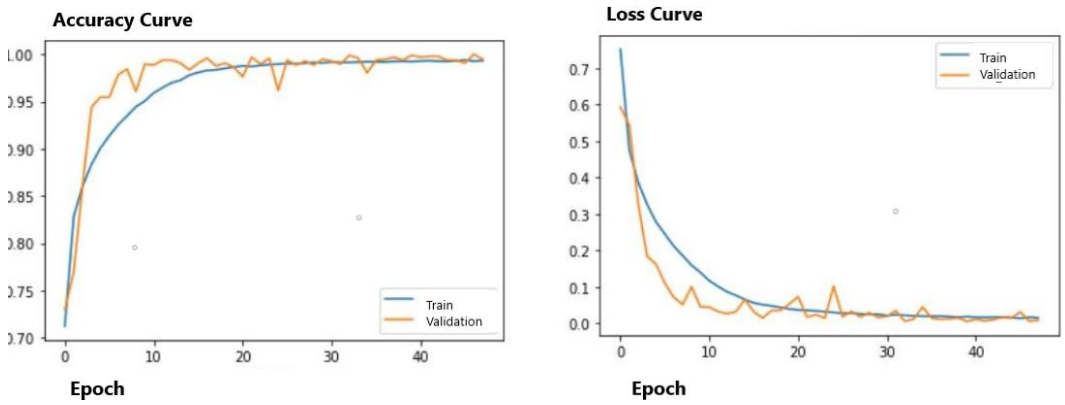


Figure 9. Accuracy and loss curves in CNN architecture trained with pre-processed images.

The CNN architecture had an accuracy of 99.40% and a loss of 0.0134, suggesting that the training was successful. Afterwards, the CNN architecture was tested to determine its performance. Figure 10 presents the confusion matrix of the classification of the test dataset consisting of 968 images (242 for each category).

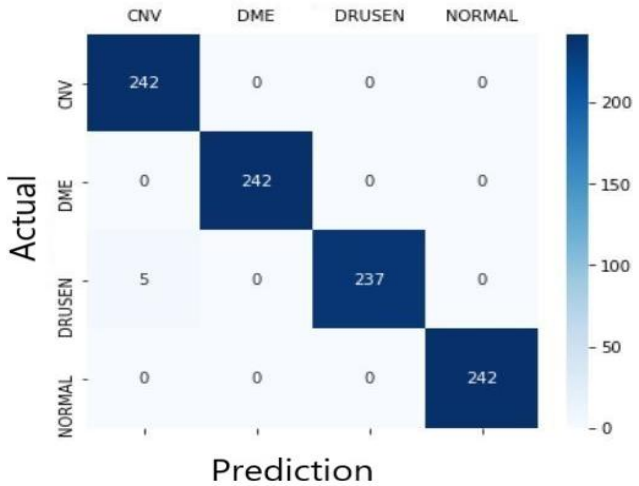


Figure 10. Confusion matrix of test data in CNN architecture trained with pre-processed images.

The confusion matrix shows that the CNN architecture detected all categories except for the Drusen with 100% accuracy. However, it evaluated five Drusen images incorrectly and labeled them as CNV. Table 3 discloses the class-based performance of the CNN architecture on the test data.

Table 3. Class-based performance of CNN architecture on test data.

	Accuracy	Precision	Sensitivity	Specificity	F1-Score	Data Number
CNV	0.9948	0.98	1.00	0.9931	0.99	242
DME	1	1.00	1.00	1	1.00	242
DRUSEN	0.9948	1.00	0.98	1	0.99	242
NORMAL	1	1.00	1.00	1	1.00	242

In short, 99.40% accuracy was incurred from the training data set with the proposed method, while 99.48% accuracy, 99.48% sensitivity and 99.83% specificity values were obtained from the test data set.

The proposed CNN architecture was trained and tested with raw images to determine the effect of pre-processing techniques on its performance. Figure 11 illustrates the accuracy and loss curves of the CNN architecture trained with raw images.

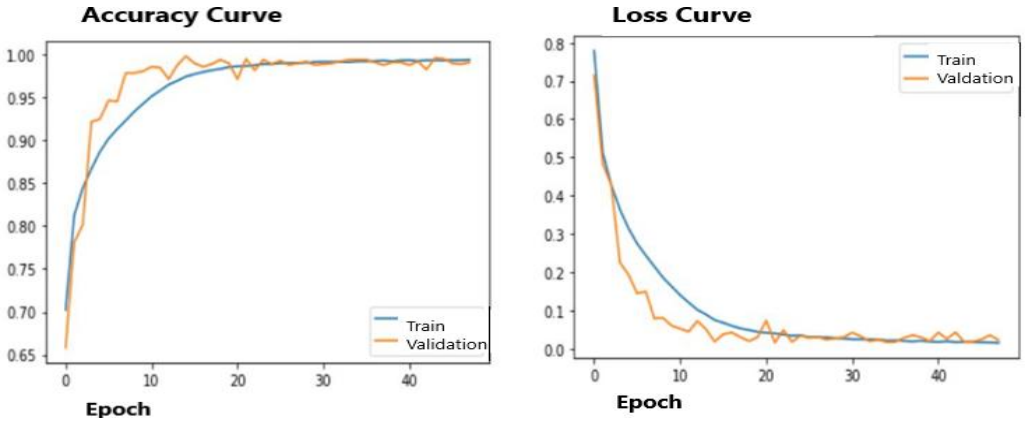


Figure 11. Accuracy and loss curves of CNN architecture trained with raw images.

The CNN architecture trained with raw images achieved 99.50% accuracy and 0.015 loss. The confusion matrix of the performance of the CNN architecture on the test dataset is shown in Figure 12.

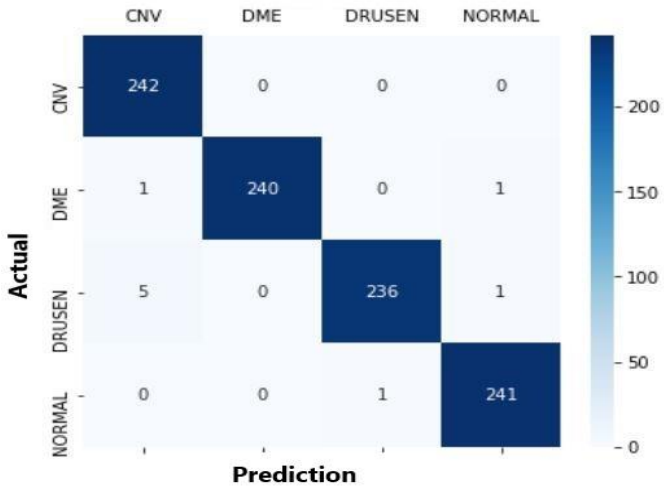


Figure 12. Confusion matrix of CNN architectures trained with raw images.

The CNN architecture trained with pre-processed images classified some Drusen images incorrectly. However, the CNN architecture trained with raw images made mistakes in other classifications as well. Table 4 denotes the performance values of the CNN architecture trained with raw images.

Table 4. Performance values of CNN architecture trained with raw images.

	Accuracy	Precision	Sensitivity	Specificity	F1-Score	Data Number
CNV	0.9758	0.98	1.00	0.9917	0.99	242
DME	1	1.00	0.99	1	1.00	242
DRUSEN	0.9958	1.00	0.98	0.9986	0.99	242
NORMAL	0.9918	1.00	1.00	0.9972	0.99	242

The CNN architecture trained with raw images achieved 99.50% accuracy during training but achieved 99.07% accuracy, 99.07% sensitivity, and 99.69% specificity during the test.

The results represented that the CNN architecture trained with pre-processed images outperformed the CNN architecture trained with raw images. Table 6 compares their performance. The CNN architecture trained with pre-processed images had more accuracy (by 0.41%), sensitivity (by 0.41%), and specificity (by 0.14%) compared to the CNN architecture trained with raw images.

Laying the basis for the creation of the CNN architecture, The AlexNet architecture was preprocessed and trained with images. In this way, the effect of the new model on the system success was observed The AlexNet architecture trained with pre-processed images achieved 99.07% accuracy, 99.07% sensitivity and 99.69% specificity during test. A confusion matrix was generated, and performance values were calculated for AlexNet. Figure 13 shows the confusion matrix and Table 5 shows the performance values.

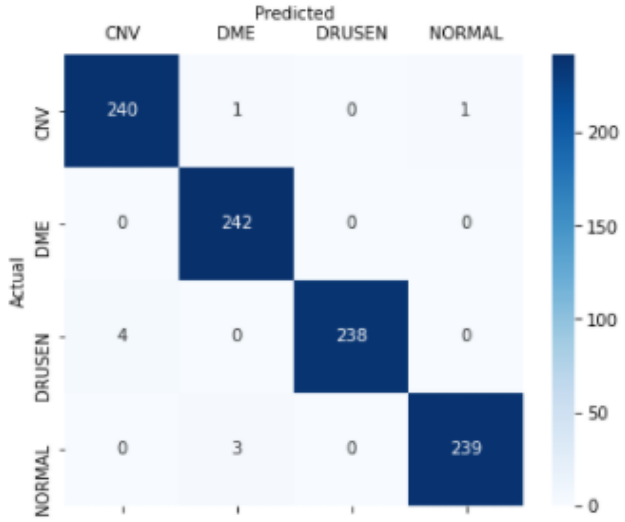


Figure 13. Confusion matrix of the AlexNet architecture.

Table 5. Performance values of the AlexNet architecture.

	Accuracy	Precision	Sensitivity	Specificity	F1-Score	Data Number
CNV	0.9917	0.9836	0.9917	0.9945	0.99	242
DME	1	0.9837	1	0.9945	0.99	242
DRUSEN	0.9834	1.00	0.9834	1	0.99	242
NORMAL	0.9876	0.9958	0.9876	0.9986	0.99	242

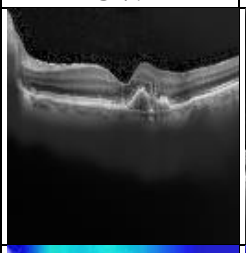
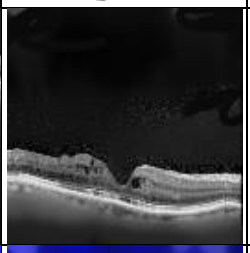
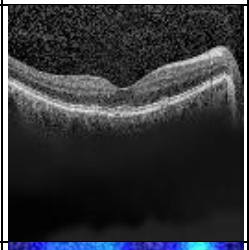
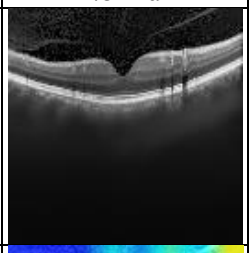
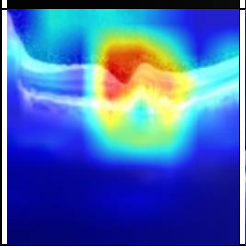
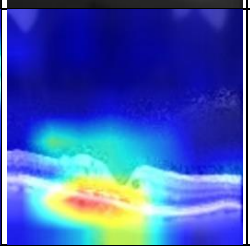
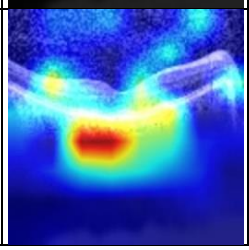
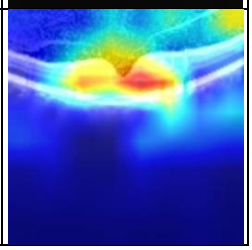
The results asserted that the CNN architecture trained with pre-processed images outperformed both the CNN architecture trained with raw images and the AlexNet architecture. On the other hand, CNN architecture trained with raw images and AlexNet architecture showed the same performance. This state supported the hypothesis proposed in this study. Thus, similar, or better results can be retrieved with pre-trained architectures with less computational cost. Performance comparisons of AlexNet and the proposed CNN architecture are given in Table 6.

When it comes to diagnosing retinal illnesses, even a single error can be highly costly. Heat maps were built to let specialists examine and see whether the CNN model focuses on the appropriate location, which helped to solve this problem. The heat maps picked up by the Grad-CAM method was attained by creating gradient maps of the last convolution layer of the architecture. Table 7 exhibits the heatmaps derived through using sample images for each class.

Table 6. Performance of CNN architecture trained with pre-processed and raw images.

	Accuracy	Sensitivity	Specificity
Preprocessed	99.48	99.48	99.83
Raw	99.07	99.07	99.69
AlexNet	99.07	99.07	99.69

Table 7. Heatmaps obtained from the proposed CNN for each class.

	CNV	DME	Drusen	Normal
Input Image				
Grad-CAM Heat Maps				

The proposed CNN architecture was compared to other systems proposed by earlier studies that employed the same dataset. Table 8 shows the results.

Table 8. Studies on the publicly available OCT dataset developed by Kermany et al. (2018).

Method	Classifier	Input Size	Dataset	Sensitivity
(Kermany et. al. 2018)	Inception V3 (94 conv. layers)	-	4 classes	Sensitivity: 97.8 Specificity: 97.4 Accuracy: 96.6
(Najeeb et al., 2019)	CNN (1 conv. layer)	128 x 128	4 classes	Sensitivity: 95.66 Accuracy: 95.66
(Li et al., 2019)	VGG-16 (13 conv. layer)	224 x 224	4 classes	Sensitivity: 97.8 Specificity: 99.4 Accuracy: 98.6
(Das et al., 2019)	Multi-scale spatial pyramid CNN (14 conv. layer)	496 × 512	4 classes	Sensitivity: 99.60 Specificity:99.87 Accuracy: 99.60
(Bhowmik et al., 2019)	VGG-16 (13 conv. layer)	224 x 224	4 classes	Sensitivity: 94 Accuracy: 94
(Berrimi & Moussaoui, 2020)	CNN (3 conv. layer)	224 x 224	4 classes	Sensitivity: 98.50 Accuracy: 98.65
	Inception V3 (94 conv. layer)	224 x 224	4 classes	Sensitivity: 99.25 Accuracy: 99.2
(Bhadra & Kar, 2020)	CNN (6 conv. layer)	128 x 128	4 classes	Sensitivity: 99.1 Specificity: 97.9 Accuracy: 96.5
(Alqudah, 2020)	AOCTNet (4 conv. layer)	256 x 256	5 classes	Sensitivity: 97.78 Specificity: 97.78 Accuracy: 97.78
(Alqudah et al., 2021)	AOCTNet (4 conv. layer) + KNN	256 x 256	5 classes	Sensitivity: 99.44 Specificity: 99.86 Accuracy: 99.44
(Thomas et al., 2021)	CNN (6 conv. layer) + Random Forest	96 x 96	2 classes	Sensitivity: 99.8 Accuracy: 99.78
(Tayal et al., 2021)	CNN (5 conv.layer)	128x128	4 classes	Sensitivity:94.47 Specificity:98.16 Accuracy: 97.14
(Li, Cheng et al., 2021)	Inception V3 based CNN (94 conv.	299x299	6 classes	Sensitivity:96.2%

	layer)			Specificity:99.2% F1-score:96.2% Accuracy:96.3%
(Sharma et al., 2021)	CNN (10 conv. layer)	128x128	4 classes	Accuracy: 99.38
(Apon et al., 2021)	CNN (4 conv. layer)	224x224	4 classes	Accuracy: 94.87
Proposed Method	CNN (5 conv. layer)	128x128	4 classes	Sensitivity: 99.48 Specificity: 99.83 Accuracy: 99.48

As the comparison table is reviewed, it may be noticed that the method proposed in this study ranks the second in terms of performance. The most successful study seems to be the multi-scale deep feature fusion (MDFF) based on the classification system designed by Das et al. (Das et al., 2019). The MDFF-based classification architecture consists of 14 convolution layers, and the computational cost is high compared to the proposed method, and the difference between the evaluation metrics is only 0.05-0.12%. Taking the other studies into consideration, it is seen that pre-trained CNN architectures are also frequently included in studies. However, it is remarked that pre-trained CNN architectures cannot provide the expected high performance in the classification of retinal diseases, and they show lower performances or close performances to each other than the CNN architectures designed in terms of diagnosing the retinal diseases. Among the studies compared, there are methods whose calculation cost is close to or lower than the proposed method (Alqudah, 2020; Apon et al., 2021; Berrimi & Moussaoui, 2020; Bhowmik et al., 2019; Najeeb et al., 2019; Tayal et al., 2021). However, surveying the results attained, it may be spotted that the performance loss is high. The approach suggested by Alqudah et. Al (Alqudah et al., 2021) is the closest to the performance values procured in this study, and the computational cost is relatively low. However, the input image sizes are high in this method, and a more complex method was suggested in the classification stage. Therefore, it is seen that the method proposed employed in this study has an essential place in the literature in terms of computational cost and performance while classifying the OCT images.

Conclusion

Retinal diseases can lead to blindness all around the world. Heart disease and hypertension can also be triggered by retinal disorders. As a result, early detection and treatment are critical for preventing or reducing the prevalence of those diseases and disabilities. To identify retinal illnesses, optical coherence tomography (OCT) and fundus fluorescein angiography (FFA) are frequently employed. Analyzing OCT and FFA ophthalmological images, on the other hand, is time-consuming, costly, and prone to human mistake. Experts use imaging techniques and their own knowledge and experience to screen for and detect retinal diseases. However, experts sometimes overlook symptoms or misdiagnose someone with another condition. As a result, computer-assisted systems are required to assist specialists in evaluating OCT and FFA images. In recent years, significant progress has been made in image analysis thanks to deep learning models (a subbranch of AI) as they are becoming more and more complex with advances in technology. Deep learning models have outperformed most computer-aided diagnostic tools. Thus, the current studies were analyzed and the gaps in the literature were identified. Upon analysis, a deep learning-based method (a five-layer CNN architecture) for detecting eye disorders from OCT images was developed. The images in the publicly available OCT dataset developed by Kermany et al. (2018) were pre-processed [1] and then the CNN architecture was trained with those images. The five-layer structure allowed the researchers to prevent losses and reduce the convergence time. The CNN architecture had 99.48% accuracy, 99.48% sensitivity, and 99.83% specificity. The CNN architecture trained with pre-processed images achieved a 0.41% increase in accuracy compared to the CNN architecture trained with raw images.

This study has come up with five conclusions. To begin, it is vital to diagnose retinal problems as soon as possible. Second, a computer-aided approach could benefit physicians in detecting and diagnosing retinal illnesses. Third, deep learning algorithms are effective decision support systems that can guide specialists in diagnosing retinal illnesses due to their high accuracy rates. Finally, when properly constructed, simplified CNN architectures can give superior performance at lower computational costs than more sophisticated or pre-trained systems. Fifth, image pre-processing procedures can improve the performance of simple deep learning architectures while also lowering their computational costs.

There are some drawbacks to this study. SD-OCT scans are limited by the resolution of CCD cameras. Furthermore, the study exclusively put OCT images to use to diagnose retinal disorders. Although the study's results are highly accurate, using images acquired using several imaging modalities will boost confidence in the system and the classification's success. The data set operated for the study has an unbalanced data distribution. 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 GAN.

In the future, the proposed method can be trained using 3D images or it may earn more precise results by using OCT data together with other imaging techniques. The proposed model can be submitted to experts by integrating it with the desktop or mobile application. In addition, results can be improved regularly by retraining the system with up-to-date data.

References

Ahmad, K., Khan, J., & Iqbal, M. S. U. D. (2019). A comparative study of different denoising techniques in digital image processing. In *8th International Conference on Modeling Simulation and Applied Optimization* [ICMSAO] (p. 1-6). DOI: https://doi.org/10.1109/ICMSAO.2019.8880389

Ajit, A., Acharya, K., & Samanta, A. (2020, February). A review of convolutional neural networks. In *International Conference on Emerging Trends in Information Technology and Engineering* [Ic-ETITE] (p. 1-5). DOI: https://doi.org/10.1109/IC-ETITE47903.2020.049

Alqudah, A., Alqudah, A. M., & AlTantawi, M. (2021). Artificial intelligence hybrid system for enhancing retinal diseases classification using automated deep features extracted from OCT images. *International Journal of Intelligent Systems and Applications in Engineering*, *9*(3), 91–100. DOI: https://doi.org/10.18201/IJISAE.2021.236

Alqudah, A. M. (2020). AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Medical and Biological Engineering and Computing*, *58*(1), 41–53. DOI: https://doi.org/10.1007/s11517-019-02066-y

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: an overview. *Journal of Physics: Conference Series*, *1142*(1), 012012. DOI: https://doi.org/10.1088/1742-6596/1142/1/012012

Apon, T. S., Hasan, M. M., Islam, A., & Alam, M. G. R. (2021). Demystifying Deep Learning Models for Retinal OCT Disease Classification using Explainable AI. Retrieved from: http://arxiv.org/abs/2111.03890

Avanzo, M., Porzio, M., Lorenzon, L., Milan, L., Sghedoni, R., Russo, G., ... Mettivier, G. (2021). Artificial intelligence applications in medical imaging: a review of the medical physics research in Italy. *Physica Medica*, *83*, 221–241. DOI: https://doi.org/10.1016/J.EJMP.2021.04.010

Berrimi, M., & Moussaoui, A. (2020). Deep learning for identifying and classifying retinal diseases. In *International Conference on Computer and Information Sciences* [ICCIS 2020] (p. 1-6). DOI: https://doi.org/10.1109/ICCIS49240.2020.9257674

Bhadra, R., & Kar, S. (2020). Retinal disease classification from optical coherence tomographical scans using multilayered convolution neural network. In *The IEEE Applied Signal Processing Conference* [ASPCON] (p. 212–216). DOI: https://doi.org/10.1109/ASPCON49795.2020.9276708

Bhowmik, A., Kumar, S., & Bhat, N. (2019). Eye disease prediction from optical coherence tomography images with transfer learning. *In International Conference on Engineering Applications of Neural Networks* (p. 104–114). Springer.

Born, J., Beymer, D., Rajan, D., Coy, A., Mukherjee, V. V, Manica, M., ... Rosen-Zvi, M. (2021). On the role of artificial intelligence in medical imaging of COVID-19. *Patterns*, *2*(6), 100269. DOI: https://doi.org/10.1016/j.patter.2021.100269

Botwe, B. O., Akudjedu, T. N., Antwi, W. K., Rockson, P., Mkoloma, S. S., Balogun, E. O., ... Arkoh, S. (2021). The integration of artificial intelligence in medical imaging practice: perspectives of African radiographers. *Radiography*, *27*(3), 861-866. DOI: https://doi.org/10.1016/J.RADI.2021.01.008

Bradley, D., & Roth, G. (2007). Adaptive thresholding using the integral image. *Journal of Graphics GPU and Game Tools*, *12*(2), 13–21. DOI: https://doi.org/10.1080/2151237X.2007.10129236

Buades, A., Coll, B., & Morel, J.-M. (2011). Non-local means denoising. IPOL - *Image Processing On Line*, *1*, 208–212. DOI: https://doi.org/10.5201/IPOL.2011.BCM_NLM

Calandra, D., & Favareto, M. (2020). Artificial Intelligence to fight COVID-19 outbreak impact: an overview. *European Journal of Social Impact and Circular Economy*, *1*(3), 84–104. DOI: https://doi.org/10.13135/2704-9906/5067

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., ... Sardanelli, F. (2021). AI applications to medical images: from machine learning to deep learning. *Physica Medica*, *83*, 9–24. DOI: https://doi.org/10.1016/J.EJMP.2021.02.006

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: a review. *IEEE Access*, *8*, 75264–75278. DOI: https://doi.org/10.1109/ACCESS.2020.2988510

Cordes, A. H., Couceiro, I. B., Alvarenga, A. D., Malinovski, I., Dominguez, C. T., Andrade, C. V. d., ... Von Der Weid, J. P. (2021). Practical considerations for OCT applications. *Journal of Physics: Conference Series*, *1826*(1), 012064. DOI: https://doi.org/10.1088/1742-6596/1826/1/012064

Coşkun, M., Yildirim, Ö., Uçar, A., & Demir, Y. (2017). An overview of popular deep learning methods. *European Journal of Technic*, *7*(2), 165–176.

Das, V., Dandapat, S., & Bora, P. K. (2019). Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images. *Biomedical Signal Processing and Control*, *54*, 101605. DOI: https://doi.org/10.1016/J.BSPC.2019.101605

Devi, M. S. S., Ramkumar, S., Kumar, S. V., & Sasi, G. (2021). Detection of diabetic retinopathy using OCT image. *Materials Today: Proceedings*, *47*(1), 185-190. DOI: https://doi.org/10.1016/J.MATPR.2021.04.070

Dong, Y., Hou, J., Zhang, N., & Zhang, M. (2020). Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. *Complexity*, *2020*(1680845). DOI: https://doi.org/10.1155/2020/1680845

Fadel, C., Holmes, W., & Bialik, M. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Boston, MA: The Center for Curriculum Redesign.

Farsiu, S., Chiu, S. J., O'Connell, R. V., Folgar, F. A., Yuan, E., Izatt, J. A. ... Toth, C. A. (2014). Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology*, *121*(1), 162–172. DOI: https://doi.org/10.1016/j.ophtha.2013.07.013

Forte, R., Cennamo, G. L., Finelli, M. L., & De Crecchio, G. (2009). Comparison of time domain Stratus OCT and spectral domain SLO/OCT for assessment of macular thickness and volume. *Eye, 23*(11), 2071–2078. DOI: https://doi.org/10.1038/eye.2008.363

Gołębiewska, J., Brydak-Godowska, J., Moneta-Wielgoś, J., Turczyńska, M., Kęcik, D., & Hautz, W. (2017). Correlation between choroidal neovascularization shown by OCT angiography and choroidal thickness in patients with chronic central serous chorioretinopathy. *Journal of Ophthalmology*, *2017*, 3048013. DOI: https://doi.org/10.1155/2017/3048013

Gupta, D. (2011). Fundus fluorescein angiography and OCT. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *249*(3), 463). DOI: https://doi.org/10.1007/s00417-010-1467-x

Horowitz, M. C., Allen, G. C., Saravalle, E., Cho, A., Frederick, K., & Scharre, P. (2018). *Artificial intelligence and international security*. Washington, DC: CNAS.ORG.

Jancy, P. L., S, Duela, J. S., Devi, D. M. R., & Lakshmi, C. P. (2021). Diabetic retinopathy diagnosis in OCT images using convolutional neural network. In *2nd International Conference on Smart Electronics and Communication* [ICOSEC] (p. 1043-1047). DOI: https://doi.org/10.1109/ICOSEC51865.2021.9591885

Kamble, R. M., Chan, G. C., Perdomo, O., Kokare, M., Gonzalez, F. A., Müller, H., & Mériaudeau, F. (2018, December). Automated diabetic macular edema (DME) analysis using fine tuning with inception-resnet-v2 on OCT images. *In 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (p. 442-446). Sarawak, Malaysia: IEEE. DOI: https://doi.org/10.1109/IECBES.2018.8626616

Kaymak, S., & Serener, A. (2018). Automated age-related macular degeneration and diabetic macular edema detection on OCT images using deep learning. In *IEEE 14th International Conference on Intelligent Computer Communication and Processing* [ICCP] (p. 265–269). DOI: https://doi.org/10.1109/ICCP.2018.8516635

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., ... Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*(5), 1122-1131. DOI: https://doi.org/10.1016/j.cell.2018.02.010

Kouziokas, G. N. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transportation Research Procedia*, *24C*, 467–473. DOI: https://doi.org/10.1016/J.TRPRO.2017.05.083

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *25th International Conference on Neural Information Processing Systems* [NIPS' 12] (p. 1097–1105). DOI: https://doi.org/10.1145/3065386

Kumar, A., & Sodhi, S. S. (2020). Comparative analysis of gaussian filter, median filter and denoise autoenocoder. In *7th International Conference on Computing for Sustainable Global Development* [INDIACom] (p. 45–51). DOI: https://doi.org/10.23919/INDIACOM49435.2020.9083712

Law, M., Seah, J., & Shih, G. (2021). Artificial intelligence and medical imaging: applications, challenges and solutions. *The Medical Journal of Australia*, *214*(10), 450-452. DOI: https://doi.org/10.5694/MJA2.51077

Leitgeb, R. A. (2019). En face optical coherence tomography: a technology review . *Biomedical Optics Express, 10*(5), 2177-2201. DOI: https://doi.org/10.1364/BOE.10.002177

Li, Feng, Chen, H., Liu, Z., Zhang, X., & Wu, Z. (2019). Fully automated detection of retinal disorders by image-based deep learning. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *257*, 495–505. DOI: https://doi.org/10.1007/s00417-018-04224-8

Li, F. F., Karpathy, A., & Johnson, J. (2016). Stanford University CS231n: Convolutional neural networks for visual recognition. *Neural networks*, *1*. Retrieved December 17, 2021, from https://cs231n.github.io/convolutional-networks/

Li, J.-P. O., Liu, H., Ting, D. S. J., Jeon, S., Chan, R. V. P., Kim, J. E., ... Ting, D. S. W. (2021). Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. *Progress in Retinal and Eye Research*, *82*, 100900. DOI: https://doi.org/10.1016/J.PRETEYERES.2020.100900

Li, Z., Cheng, K., Qin, P., Dong, Y., Yang, C., & Jiang, X. (2021). Retinal OCT image classification based on domain adaptation convolutional neural networks. *14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* [CISP-BMEI], (p. 1–5). DOI: https://doi.org/10.1109/CISP-BMEI53629.2021.9624429

Miranda, M., & Romero, F. J. (2019). Antioxidants and retinal diseases. *Antioxidants*, *8*(12). DOI: https://doi.org/10.3390/ANTIOX8120604

Mezni, I., Slama, A. B., Mbarki, Z., Seddik, H., & Trabelsi, H. (2021). Automated identification of SD-optical coherence tomography derived macular diseases by combining 3D-block-matching and deep learning techniques. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *9*(6), 660–669. DOI: https://doi.org/10.1080/21681163.2021.1926329

Mishra, S. S., Mandal, B., & Puhan, N. B. (2019). Multi-level dual-attention based CNN for macular optical coherence tomography classification. *IEEE Signal Processing Letters*, *26*(12), 1793–1797. DOI: https://doi.org/10.1109/LSP.2019.2949388

Motozawa, N., An, G., Takagi, S., Kitahata, S., Mandai, M., Hirami, Y., ... Kurimoto, Y. (2019). Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes. *Ophthalmology and Therapy*, *8*, 566–583. DOI: https://doi.org/10.1007/s40123-019-00207-y

Najeeb, S., Sharmile, N., Khan, M. S., Sahin, I., Islam, M. T., & Hassan Bhuiyan, M. I. (2019). Classification of retinal diseases from OCT scans using convolutional neural networks. *10th International Conference on Electrical and Computer Engineering* [ICECE], (p. 465–468). DOI: https://doi.org/10.1109/ICECE.2018.8636699

OpenCV: *Denoising*. (2021, December 18). Retrieved December 18, 2021, from https://docs.opencv.org/4.x/d1/d79/group__photo__denoise.html#ga4c6b0031f56ea3f98f768881279ffe93

OpenCV: *Image Thresholding*. (2021, December 12). Retrieved December 12, 2021, from https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html

Padmavathi, K., & Thangadurai, K. (2016). Implementation of RGB and grayscale images in plant leaves disease detection–comparative study. *Indian Journal of Science and Technology*, *9*(6), 1-6.

Rajagopalan, N., Venkateswaran, N., Josephraj, A. N., & Srithaladevi, E. (2021). Diagnosis of retinal disorders from optical coherence tomography images using CNN. *PLoS ONE*, *16*(7), e0254180. DOI: https://doi.org/10.1371/JOURNAL.PONE.0254180

Rasti, R., Rabbani, H., Mehridehnavi, A., & Hajizadeh, F. (2018). Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Transactions on Medical Imaging*, *37*(4), 1024–1034. DOI: https://doi.org/10.1109/TMI.2017.2780115

Rawat, C. S., & Gaikwad, V. S. (2014). Signal analysis and image simulation for optical coherence tomography (OCT) systems. *Ieeexplore.* In *International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT),* (p. 626-631). DOI: https://doi.org/10.1109/ICCICCT.2014.6993037

Rosenman, J., Roe, C. A., Cromartie, R., Muller, K. E., & Pizer, S. M. (1993). Portal film enhancement: technique and clinical utility. *International Journal of Radiation Oncology, Biology, Physics*, *25*(2), 333–338. DOI: https://doi.org/10.1016/0360-3016(93)90357-2

Saha, S., Nassisi, M., Wang, M., Lindenberg, S., kanagasingam, Y., Sadda, S., & Hu, Z. J. (2019). Automated detection and classification of early AMD biomarkers using deep learning. *Scientific Reports, 9, 10990*. DOI: https://doi.org/10.1038/s41598-019-47390-3

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. *BMC Medical Informatics and Decision Making, 21*(125), 1–23. DOI: https://doi.org/10.1186/S12911-021-01488-9

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *13*(1), 146. DOI: https://doi.org/10.1117/1.1631315

Sharma, A., Khanna, A. V., & Bhargava, M. (2021). Multi-label classification of retinal disorders in optical coherence tomography using deep learning.In *2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021* (p. 1750-1757). DOI: https://doi.org/10.1109/ICESC51422.2021.9532711

Son, L. H., Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE Access*, *7*, 23319–23328. DOI: https://doi.org/10.1109/ACCESS.2019.2899260

Srinivasan, P. P., Kim, L. A., Mettu, P. S., Cousins, S. W., Comer, G. M., Izatt, J. A., & Farsiu, S. (2014). Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical Optics Express, 5(10), 3568-3577*. DOI: https://doi.org/10.1364/BOE.5.003568

Tan, Z., Scheetz, J., & He, M. (2019). Artificial intelligence in ophthalmology: accuracy, challenges, and clinical application. *Asia-Pacific Journal of Ophthalmology*, *8*(3), 197–199. DOI: https://doi.org/10.22608/APO.2019122

Taş, S. P., Barin, S., & Güraksin, G. E. (2021). Deep learning for ophthalmological images. In *Deep Learning for Biomedical Applications* (p. 101-119). CRC Press.https://doi.org/10.1201/9780367855611-6

Tayal, A., Gupta, J., Solanki, A., Bisht, K., Nayyar, A., & Masud, M. (2021). DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases. *Multimedia Systems*. DOI: https://doi.org/10.1007/S00530-021-00769-7

Thomas, A., Harikrishnan, P. M., Ramachandran, Ramachandran, S., Manoj, R., Palanisamy, P., & Gopi, V. P. (2021). A novel multiscale and multipath convolutional neural network based age-related macular degeneration detection using OCT images. *Computer Methods and Programs in Biomedicine*, *209*, 106294. DOI: https://doi.org/10.1016/J.CMPB.2021.106294

Uysal, E., & Güraksin, G. E. (2021). Computer-aided retinal vessel segmentation in retinal images: convolutional neural networks. *Multimedia Tools and Applications*, *80*(3), 3505–3528. DOI: https://doi.org/10.1007/s11042-020-09372-w

Wang, J., Hormel, T. T., Gao, L., Zang, P., Guo, Y., Wang, X., ... Jia, Y. (2020). Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning. *Biomedical Optics Express*, *11*(2), 927-944. DOI: https://doi.org/10.1364/boe.379977

Wang, W., Xu, Z., Yu, W., Zhao, J., Yang, J., He, F., ... Li, X. (2019). Two-stream CNN with loose pair training for multi-modal AMD categorization. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. *Lecture Notes in Computer Science* (Vol. 11764, p. 156–164). Cham, SW: Springer.

Wüthrich, M., Trimpe, S., Cifuentes, C. G., Kappler, D., & Schaal, S. (2017). A new perspective and extension of the Gaussian Filter. *The International Journal of Robotics Research, 35*(14), 1731–1749. DOI: https://doi.org/10.1177/0278364916684019

Yadav, G., Maheshwari, S., & Agarwal, A. (2014). Contrast limited adaptive histogram equalization based enhancement for real time video system. In *International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014* (p. 2392–2397). DOI: https://doi.org/10.1109/ICACCI.2014.6968381

Ye, H., Gao, F., Yin, Y., Guo, D., Zhao, P., Lu, Y., ... Xia, J. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *European Radiology*, *29*, 6191–6201. DOI: https://doi.org/10.1007/s00330-019-06163-2