Detection of Cataract Disease Using Convolution Neural Network with Autoencoder

Mrs S.B.Saleema Parvin[1], Dr.A.Devendran[2]

Abstract: Cataracts are one of the most prevalent visual conditions that people experience as they age. A cataract is when the lens of the eyes develops a fog. The major signs and symptoms of this disease are blurred vision, fading colors, and difficulties seeing in bright light. Having trouble doing a number of chores is typically the outcome of these symptoms. Therefore, early cataract identification and prevention may aid in lowering the prevalence of blindness. This study uses Convolutional Neural Networks (CNN) to categorize cataract disease using an image dataset which are freely accessible. The proposed CNN fuses the Autoencoder (AE) advantages of CNN as pre-trained to preserve better correlation between the patches. When comparing it with classical diagnosis technique, image classification by CNN is potential performance and cost efficient method. As a result, the goal of the current study is to create a model for predicting cataracts. A human grader may find it difficult to recognize the earliest small alterations in the optic disc. The Deep Learning (DL) encoder can learn subtle characteristics in fundus images of individuals with early-stage cataracts because we have access to cataract fundus images that have been classified based on a thorough ophthalmologic examination.

Keyword: Cataract detection, Fundus images, Convolution Neural Network, Autoencoder, Retinal diseases

Introduction

Light is refracted into the retina by the eye's lens, a transparent, circular part of the eye that extends beyond the iris [1]. An eye condition called a cataract causes the lens of the eye to become clouded. According to the World Health Organization, 65.2 million people worldwide have cataracts, which is a significant increase over the number of people who have glaucoma, corneal opacities, trachoma and diabetic retinopathy combined that have the same severity of vision loss or blindness. In 2025, there will be more than 40 million blind individuals globally, with cataracts accounting for more than 50% of all occurrences of blindness [2]. Smoking, heredity, alcoholism, dietary or metabolic issues, drugs, or prolonged exposure to sunlight can all cause cataracts [3] [4]. Additionally, it has links to a number of other disorders. As the condition worsens, fundus disease can lead to blindness and visual loss. Slit-lamp photography, which involves reflecting light within the lens structure and inferring the presence of the cataract from the non-uniform lighting of the refracted intensity, is a method for diagnosing cataracts. The American Cooperative Cataract Research Group protocol, the Oxford Clinical Cataract Classification, and the Lens Opacities Classification System III are well-known protocols for classifying the cataract disease. These protocols are based on

[1]Ph.D Research Scholar, Faculty of Computer Applications Dr.MGR Educational and Research Institute, Maduravoyal, Chennai, saleemasaffiaahil@gmail.com
[2]Professor, Faculty of Management Studies, Dr.MGR Educational and Research Institute, Maduravoyal, Chennai, devendran.mba@drmgrdu.ac.in

complicated procedures and require highly skilled ophthalmologists [5]. Additionally utilized for cataract screening include the ultrasound biomicroscope, optical coherence tomography, and the ultrasound backscattering signal. Their diagnostic procedures are expensive and dependent on difficult procedures. Through ophthalmoscopy, where the blurred appearance of the retinal components is equivalent to the quality of visual acuity, the cataract is simply identified. It is possible to grade cataracts using a retinal fundus picture with greater precision, even for untrained graders, which facilitates access to eye diagnostics [6].

A detailed study has been conducted on the detection of cataracts. They created an eight-layered Deep Convolutional Neural Network (DCNN) model and balanced lumination issues in the dataset to construct an automated cataract grading method [7]. The Deep convolutional Network was added to the similar technique to let each convolutional layer extract information more precisely [8]. They have trained a DCNN model similar to Res-Net50 and they have achieved a good degree of accuracy without pre-processing data [9]. State-of-the-art (SOTA) performance in tasks like image classification and object recognition has recently been attained by certain suggested CNN based models, including ResNet, EfcientNet, GoogLeNet and VGGNet [10]. Several of these methods have also been used to determine the cause of fundus diseases. Several researchers have sought to include Self-Attention (SA) processes into Computer Vision (CV) in light of the success of SA models like Transformer8 in natural language processing.

Recent research by Vision Transformer (ViT) has demonstrated that on ImageNet-1K, a nearly single vanilla Transformer layer is sufficient to produce respectable performance. When pretrained on the expansive private JFT-300M dataset, ViT specifically produced outcomes equivalent to SOTA CNNs, showing that the Transformer model has a larger model capacity than CNNs. Additionally, despite Transformer designs' growing popularity in vision tasks and their competitive performance against CNN architectures in a number of vision tasks, the great performance is dependent on the availability of a large amount of training data [11]. In low data volume circumstances, transformer topologies continue to lag behind CNNs. As a result, with a small sample size, Transformer-based models have yet to be applied to the field of fundus disease categorization.

The self-attention layer having a global receptive field possesses a higher capacity to model features than the convolutional layer, which can make up for the convolutional layer's lack of global modelling capabilities since it has a stronger inductive bias prior and faster convergence. As a result, they are regarded as being a part of the same backbone network for multistage feature extraction. Low-level local characteristics are extracted using convolution, while long-term dependencies are captured using the transformer. With the combination of CNN architecture and Transformer architecture, the model achieves better generalization performance and stronger learning ability, making it an excellent candidate for fundus image classification tasks due to its higher model capacity and stronger learning ability. Due to the fact that such a model is typically used on mobile devices, we only enable the worldwide receptive field once the feature map size has been down sampled to a reasonable level, which is comparable to the actual scenario. In this paper, a technique based on CNN for automated ophthalmological illness diagnosis from fundus pictures was provided. The ODIR dataset was used to test the suggested strategy, which performed better than previous approaches utilizing a similar dataset in numerous criteria. The proposed approach's benefits and limitations can be stated.

Literature Review

Cataract identification and categorization are made possible by the fundus image. The majority of conventional cataract diagnosis techniques use optical coherence tomography and human-engineered characteristics. In order to grade and categorize cataracts, Fan et al. [12] collected wavelet- and sketch-based features, decreased the features dimensionality via Principal Component Analysis (PCA), and used Machine Learning (ML) techniques. For the identification of cataract and conjunctivitis, Manchalwar et al. [13]

employed a Histogram of Oriented Gradient (HOG) component with minimum distance classifier. Qiao et al. [14] suggested a different method for analysing fundus images that involved manually extracting and weighting three feature sets (wavelet, colour and texture) before using SVM to classify the images. The decision tree approach was used by Xiong et al. [15] to classify the visible structure's pixel number, standard deviation, and mean into five stages of cataract blurred vision with vitreous opacity. Dong et al. [16] provided a method for classifying cataracts based on information about the vessels, in which vessel information was obtained using a Kirsh template filter, wavelet and texture characteristics were retrieved, and then SVM was used to grade the cataracts. These techniques only extract features in one direction that is inadequate to fully capture the fundus image's intricacies. Cao et al. [17] introduced an enhanced Haar wavelet feature that included a detail component for displaying texture information in vertical, horizontal, and diagonal dimensions. The majority vote approach used for cataract classification has proven to be more accurate at detecting and grading cataracts. Xiong et al. [18] suggested a novel method for automatically classifying cataracts and it is based on a multifeatured set that fuses textural characteristics from the Grey-Level Co-Occurrence Matrix (GLCM) with more advanced characteristics from a previously trained ResNet model.

Huang, Jonathan, et al. [19] used four distinct models in the example of ensemble learning. To improve the model's accuracy, they appliedResNet50, Vgg12 and AclSincNet. The creation of each model was aided by pre-training on audio set data. Ensemble learning has been accomplished using all of these models, and the validation set of trials was utilized to get the findings. The highest outcome after ensemble averaging every model was 83.01%. Kumar et al., [20] proposed the medical image categorization may be carried out using an array of modified CNN methods. This approach is applicable to biological research, instruction, and diagnostics. The training pictures were 6776 photos, whereas the test images were 4166 images. The CNN architectures AlexNet and Google-Net were each utilized for an image classification task. The investigations have been carried out using both ensemble and individual models. The strategy produced an overall accuracy of 96.59%. Finally, Mahmoud Smaida and Serhii Yaroshchak. [21] utilised three distinct models in their study. To increase accuracy, they applied the VGG-16 model, CNN model and Inception-v3 model. To increase the performance of their models, stacking, bagging and boosting were used in combination with datasets on eye diseases. Combining all the various structures into one bagging ensemble yields an accuracy of 86.43%.

The purpose of this research, as stated by Parampal et al. [22], is to help ophthalmology by applying DL technology to identify cataract along with other retinal diseases. 75137 fundus images are included in the data sets. CNNs are mostly used in DL. Slit-lamp images of paediatric cataracts have also been subjected to DL. When it came to grade, density, and location of paediatric cataracts, a CNN algorithm demonstrated great sensitivity and specificity. Patient empowerment, early diagnosis, and the identification of curable eye diseases would all be made possible with DL tools. When using single-label image classification methods to train DL networks for diagnosis, the network may incorrectly link some information to the existence of a disease. The research objective of Julia et al., [23] is to assist ophthalmology using ai to detect cataract and other retinal disease. The data set is 1475 fundus images are used applications focus on ROP or congenital cataracts. The Inception architecture, which incorporates transfer learning through pretraining on ImageNet, was adopted by all CNN-based systems. They used 3 types of CNN methods there are normal CNN, pre-CNN and plus-CNN with accuracy (76.42%-97.69%). Almost all of these algorithms focus solely on identifying illnesses at a particular moment in time, without considering longitudinal imaging. Its high reliability and have good sensitivity and specialty. They used deep rop and rop di architecture. In the work of Wei Lu et al., [24] solemnly introduces AI algorithms which are used build up an AI model which is used to detect any anomaly in an eye CNN, which operates on the concept of DL, is one such technique that is frequently employed for detection.

It consists of multiple convolutional layers that considers several extracts of the input images and combines them to produce a final image which is than compared to the predetermined scale to detect any anomaly. Also, this paper lays down different steps which are used to build up a successful AI model.

Research Methodology

This research focuses on detecting cataract patients through image processing using left and right eye fundus photograph. The images of retina is collected, image preprocessing, image processing through AE as Transfer Learning (TL) in which image is processed and classified precisely for providing better detection of cataract because it is one of the dangerous ocular disease that may cause blindness. There are various CNN architecture available as existing but the additional information is collected through channel from other features which is not available in the pre-training original channel. There are various type of ocular disease fundus images but this research focuses on cataract and the cataract associated ocular diseased patients. The image of normal patients and cataract patients shown in figure 1a and 1b who may be possible of affecting through cataract due to aging in nature which is one of the most criteria for early detection of cataract. This architecture illustrates the CNN-AE with Inception-V2 for determining the cataract detection using fundus image is shown in figure 2. In this experimental research, the AE act as TL which used to reconstruct and de-noise the image has comprising into two different process namely encoder process and decoder process.

1a

1b

Figure 1a. Cataract disease identified in fundus image. 1b. Normal patient identified in fundus image

Input dataset as fundus image of eye

Original images path with its original dimensions

Generate images from original image path for TL using AE

Reconstructed Path

De-noised images as pre-trained images are reconstructed

Original image path $F_X = f(x)$

TL using AE as $F_{TR} = f(X_{ReCon})$

Integrating original image pass and pre-trained AE method

Fine tuning of pre-trained AE with Insception-V2 as classification layers

High accurate in predicting cataract disease from the fundus images

Figure 2 Proposed architecture of CNN-AE with Inception-V2 for predicting cataract disease

Data collection

The fundus image involves for 5000 patients with 6392 images available with left as well as right eyes fundus picture along with patient details like patient ID, sex and the age. This dataset consist of various ocular disease images but the images considered for experimental is normal, glaucoma and its associates as well as DR images. The dataset is shown in figure 3 has been collected by M/S. Shanggong Medical Technology Co., Ltd. from different hospitals/medical centers in China.

ID	Patient Age	Patient Sex	Left-Fundus	Right-Fundus	Left-Diagnostic Keywords	Right-Diagnostic Keywords	N	D	G	C	A	H	M	O	filepath	labels	target	filename
0	69	Female	0_left.jpg	0_right.jpg	cataract	normal fundus	0	0	0	1	0	0	0	0	../input/o	['N']	[1, 0, 0, 0,	0_right.jpg
1	57	Male	1_left.jpg	1_right.jpg	normal fund	normal fundus	1	0	0	0	0	0	0	0	../input/o	['N']	[1, 0, 0, 0,	1_right.jpg
2	42	Male	2_left.jpg	2_right.jpg	laser spoti¼	moderate non p	0	1	0	0	0	0	0	1	../input/o	['D']	[0, 1, 0, 0,	2_right.jpg
4	53	Male	4_left.jpg	4_right.jpg	macular epi	mild nonprolife	0	1	0	0	0	0	0	1	../input/o	['D']	[0, 1, 0, 0,	4_right.jpg
5	50	Female	5_left.jpg	5_right.jpg	moderate n	moderate non p	0	1	0	0	0	0	0	0	../input/o	['D']	[0, 1, 0, 0,	5_right.jpg
6	60	Male	6_left.jpg	6_right.jpg	macular epi	moderate non p	0	1	0	0	0	0	0	1	../input/o	['D']	[0, 1, 0, 0,	6_right.jpg
7	60	Female	7_left.jpg	7_right.jpg	drusen	mild nonprolife	0	1	0	0	0	0	0	1	../input/o	['D']	[0, 1, 0, 0,	7_right.jpg
8	59	Male	8_left.jpg	8_right.jpg	normal fund	normal fundus	1	0	0	0	0	0	0	0	../input/o	['N']	[1, 0, 0, 0,	8_right.jpg
9	54	Male	9_left.jpg	9_right.jpg	normal fund	vitreous degen	0	0	0	0	0	0	0	1	../input/o	['O']	[0, 0, 0, 0,	9_right.jpg
10	70	Male	10_left.jpg	10_right.jpg	epiretinal m	normal fundus	0	0	0	0	0	0	0	1	../input/o	['N']	[1, 0, 0, 0,	10_right.jpg
11	60	Female	11_left.jp	11_right.jpg	moderate n	moderate non p	0	1	0	0	0	1	0	0	../input/o	['D']	[0, 1, 0, 0,	11_right.jpg
13	60	Female	13_left.jp	13_right.jpg	pathologica	pathological my	0	0	0	0	0	0	1	0	../input/o	['M']	[0, 0, 0, 0,	13_right.jpg
14	55	Male	14_left.jp	14_right.jpg	normal fund	macular epireti	0	0	0	0	0	0	0	1	../input/o	['O']	[0, 0, 0, 0,	14_right.jpg
15	50	Male	15_left.jp	15_right.jpg	normal fund	myelinated ner	0	0	0	0	0	0	0	1	../input/o	['O']	[0, 0, 0, 0,	15_right.jpg
16	54	Female	16_left.jp	16_right.jpg	normal fund	pathological my	0	0	0	0	0	0	1	0	../input/o	['M']	[0, 0, 0, 0,	16_right.jpg
17	57	Male	17_left.jp	17_right.jpg	drusen	drusen	0	0	0	0	0	0	0	1	../input/o	['O']	[0, 0, 0, 0,	17_right.jpg
18	58	Male	18_left.jp	18_right.jpg	pathologica	pathological my	0	0	0	0	0	0	1	0	../input/o	['M']	[0, 0, 0, 0,	18_right.jpg
19	45	Male	19_left.jp	19_right.jpg	mild nonpro	mild nonprolife	0	1	0	0	0	0	0	0	../input/o	['D']	[0, 1, 0, 0,	19_right.jpg
21	76	Female	21_left.jp	21_right.jpg	epiretinal m	epiretinal mem	0	0	0	0	0	0	0	1	../input/o	['O']	[0, 0, 0, 0,	21_right.jpg
23	47	Male	23_left.jp	23_right.jpg	hypertensiv	hypertensive re	0	0	0	0	0	1	0	0	../input/o	['H']	[0, 0, 0, 0,	23_right.jpg
24	75	Female	24_left.jp	24_right.jpg	normal fund	cataract	0	0	0	1	0	0	0	0	../input/o	['C']	[0, 0, 0, 1,	24_right.jpg

Figure 3 Dataset for ocular disease

In this experimental research, the AE act as TL which used to reconstruct and de-noise the image has comprising into two different process namely encoder process and decoder process shown in figure 4. In the encoder process, the

input has been read out into compressed form whereas the input layer has considered original features that get connected to the smaller neurons of intermediate layer which is efficient for dimensionality reduction. The compressed representation of image act as the intermediate layer which is furthermore connected to

output layer as decoder process that performs as a wise versa of encoder process. While training the AE, the similar data is considered as output with reconstructed input. The major aim is to reduce the loss among the reconstructed data as well as the original.
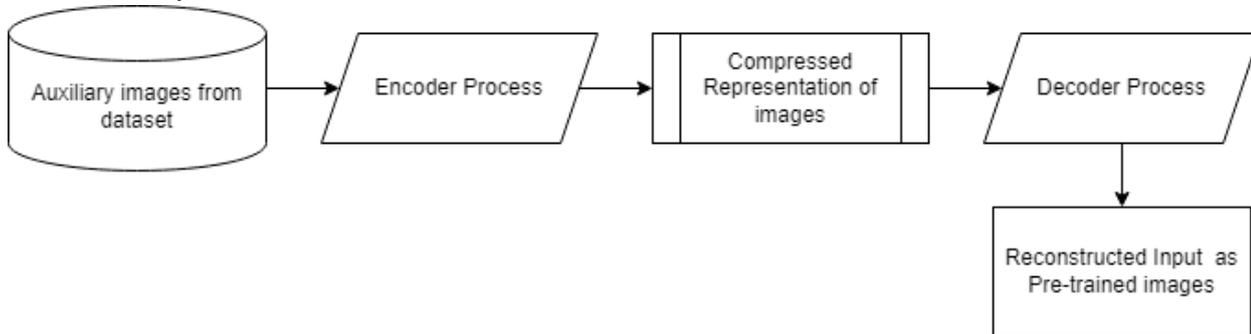
Figure 4 Working of AE method

This proposed methodology has focused on improving the accuracy of detecting the cataract disease using TL as AE with CNN model by Inception-V2 for the exploitation of an input. The architecture strength is AE and inception blocks of CNN that convolute the similar data on various scales in parallel through input split into a multi-scale pathway and merged it finally in term of single concatenated output. The major merits in managing data at various spatial resolutions and parallel process has minimize the computational cost through network sparsity. However, the proposed model has reported for providing lower error rate by minimizing the use of parameter amount with no loss in accuracy. Hence, the training of Inception-V2 with Mini-Batch size as 20, epochs as 1000, learning rate is 0.0001 and Adam as an optimizer that assist in providing the first-order gradient-based optimization of the stochastic objective function. Thus, the Adam optimizer is provided as an objective functions considered in the equation 1 and 2.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)g_1$$
(1)

$$n_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2$$
(2)

Where,

m_t and n_t = First order and second order moments

g_t = Gradient of error surface

β_1 and β_2 = Estimating Decaying rate for first order moment and second order moment

According to this CNN-AE with Inception-V2 architecture, the layer of Inception-V2 with 1586 layer of CNN is considered which consists of an alternate

convolution layer and pooling layer as sequence and ended with full connected layer. The fine-tuned model considered in this experiment is CNN in which all convolution layers have utilized 3x3 size masked with the stride of 1 or 2 followed by pooling layers for minimizing the spatial resolution for providing a form of invariance to translation. With the use of masking size 3x3 as well as 8x8 by 2 strides whereas all convolution layer inputs have padded 0 in preserving the grid size. In Inception module, three type of filter size such as 1x1, 5x5 and 3x3 have been unified using pooling layer. Final output of Inception block is the resultant feature maps concatenation for four sequence path. The four sequence path as follows

- First path contains 1x1 convolutions that perform as a selective highway network utilized for passing the selected data forwarded with no transformations.
- In the subsequent two path such as second and third, the 1x1 convolutions have attended by a multiscale transformation of 5x5 and 3x3 convolutions for providing various features.
- In the final path, 1x1 convolution has continued by 3x3 pooling to the translation-invariant features extraction.

Hence, the general module is frequent in the network, and finally the last layer has connected to a full connected layer that associated with a softmax classifier which estimates the network's global loss in the training phase.

In this experimental research, TL has utilized two stages initially utilized in the generation of an auxiliary images from original through AEs and subsequently with the painstaking process of network training have been minimized using CNN-AE using Inception-V2. This TL is utilized for improving the feature space in which both source and target domain are the same but are limited for

distinct tasks. The TL is used for transferring the knowledge from auxiliary learners to deep CNN with Inception-V2 that solves the learning issues by classification. It improves the feature space over target domain that help in improving the target learning potential for predictive function in the target domain.

Experimental Results and Discussion

The eye disease dataset performances with and without TL is used to assess the proposed SAE-based CNN-AE architecture's learning capability. Data from eye images and statistical data is converted to 2D images in order to take advantage of the 2D convolution operators' capacity for learning in the CNN model on information about eye diseases like DR, and glaucoma in this research. This study focuses on early prediction of glaucoma patient through fundus images and statistical dataset. Each pixel's brightness in the image varies according to the attributes of the patients. In the eye disease dataset, the AE is initially trained from scratch in which backpropogation process is used to optimize the parameters after the weights have been initialized randomly. Datasets from 20% validation have been utilized in evaluating parameters. Figure 5 illustrate the fine tuning of proposed AE with CNN-AE model in which the sequential layer as inspection-V2 is utilized and dense layer illustrates the eight different target classes in which earlier cataract is the research key target.

```
_____
Layer (type)                Output Shape              Param #
================================================================
inception_resnet_v2 (Functi (None, 9, 9, 1536)        54336736
onal)

batch_normalization_203 (Ba (None, 9, 9, 1536)        10752
tchNormalization)

global_average_pooling2d (G (None, 1536)              0
lobalAveragePooling2D)

dense (Dense)               (None, 512)               786944

dense_1 (Dense)             (None, 256)               131328

dense_2 (Dense)             (None, 128)               32896

dense_3 (Dense)             (None, 8)                 1032

================================================================
Total params: 55,299,688
Trainable params: 955,272
Non-trainable params: 54,344,416
_____
```

Figure 5 Sequential layer with CNN-AE as fine tuning model

The training mode is made to iterate for 25 epochs to obtain the high and better understating of model information and it assist in producing high accuracy in earlier prediction of glaucoma diseases. Figure 6 illustrates the first 17 epochs and finally obtained accuracy in 25th epoch of this model is 0.9815 and validation accuracy is 0.9286.

```
Epoch 1/25
155/155 [==============================] - 95s 531ms/step - loss: 1.3262 - accuracy: 0.4681 - val_loss: 2.0902 - val_accuracy: 0.2541
Epoch 2/25
155/155 [==============================] - 80s 519ms/step - loss: 1.1800 - accuracy: 0.5392 - val_loss: 1.8619 - val_accuracy: 0.3026
Epoch 3/25
155/155 [==============================] - 82s 526ms/step - loss: 1.0526 - accuracy: 0.5872 - val_loss: 1.4909 - val_accuracy: 0.3528
Epoch 4/25
155/155 [==============================] - 82s 532ms/step - loss: 0.9321 - accuracy: 0.6508 - val_loss: 1.3496 - val_accuracy: 0.4219
Epoch 5/25
155/155 [==============================] - 82s 530ms/step - loss: 0.8169 - accuracy: 0.6940 - val_loss: 1.3478 - val_accuracy: 0.4400
Epoch 6/25
155/155 [==============================] - 82s 531ms/step - loss: 0.6663 - accuracy: 0.7551 - val_loss: 0.8785 - val_accuracy: 0.6365
Epoch 7/25
155/155 [==============================] - 82s 531ms/step - loss: 0.5339 - accuracy: 0.8139 - val_loss: 0.7694 - val_accuracy: 0.6933
Epoch 8/25
155/155 [==============================] - 83s 536ms/step - loss: 0.4556 - accuracy: 0.8361 - val_loss: 0.5945 - val_accuracy: 0.7804
Epoch 9/25
155/155 [==============================] - 83s 533ms/step - loss: 0.3261 - accuracy: 0.8833 - val_loss: 0.4963 - val_accuracy: 0.8158
Epoch 10/25
155/155 [==============================] - 82s 531ms/step - loss: 0.2507 - accuracy: 0.9176 - val_loss: 0.3404 - val_accuracy: 0.8701
Epoch 11/25
155/155 [==============================] - 82s 529ms/step - loss: 0.1972 - accuracy: 0.9334 - val_loss: 0.2931 - val_accuracy: 0.9071
Epoch 12/25
155/155 [==============================] - 82s 532ms/step - loss: 0.1792 - accuracy: 0.9368 - val_loss: 0.2068 - val_accuracy: 0.9317
Epoch 13/25
155/155 [==============================] - 82s 532ms/step - loss: 0.1350 - accuracy: 0.9546 - val_loss: 0.2133 - val_accuracy: 0.9268
Epoch 14/25
155/155 [==============================] - 84s 542ms/step - loss: 0.1200 - accuracy: 0.9598 - val_loss: 0.1382 - val_accuracy: 0.9548
Epoch 15/25
155/155 [==============================] - 86s 556ms/step - loss: 0.1016 - accuracy: 0.9659 - val_loss: 0.1243 - val_accuracy: 0.9613
Epoch 16/25
155/155 [==============================] - 84s 541ms/step - loss: 0.1051 - accuracy: 0.9653 - val_loss: 0.1548 - val_accuracy: 0.9441
Epoch 17/25
155/155 [==============================] - 85s 550ms/step - loss: 0.0776 - accuracy: 0.9736 - val_loss: 0.1085 - val_accuracy: 0.9671
```

Figure 6 Epochs of CNN-AE as fine tuning model

Figure 7 has illustrated the right eye of the patients in which the training accuracy is 98.70% and validation accuracy is 97.36% at 25 epochs. The training accuracy initiated from 49.02% at 1 epoch and progressively increased to 98.70% and similarly the validation accuracy has initiated from 16.53% to 97.36%. As the epoch increases the CNN-AE has fine tuning that assist in improving the validation accuracy till 97.36%.
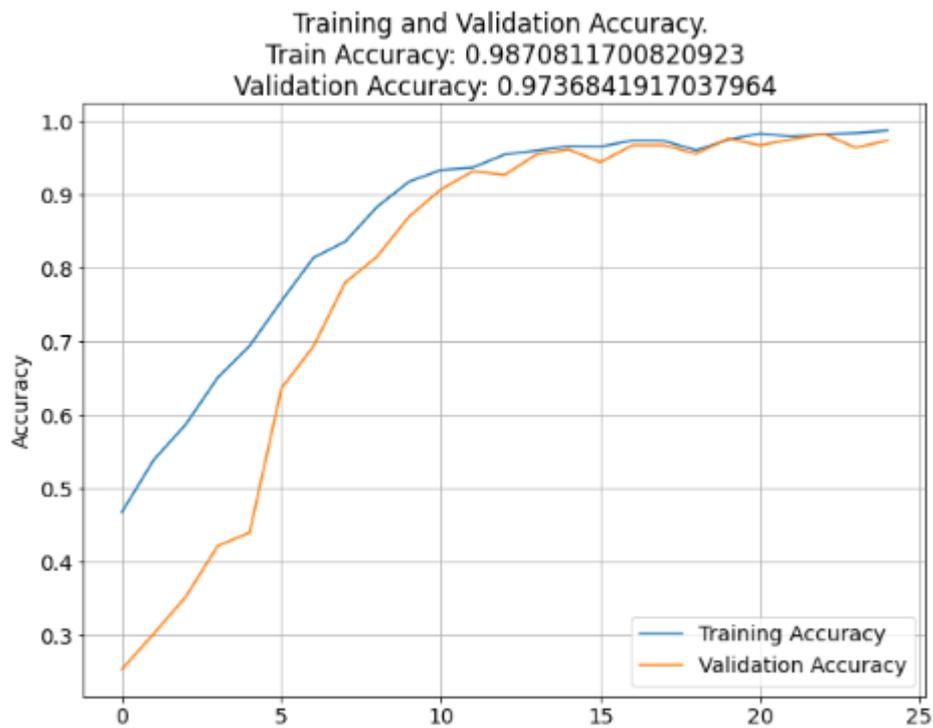
Figure 7 Accuracy for proposed CNN-AE model in right eye

Figure 8 has illustrated the loss for training and validation of proposed CNN-AE model for right eye patient records. The loss of CNN-AE model initiated with 1.323 and progressively decreases while the epoch count increases and finally at epoch 25 as 0.063. Similarly in the case of validation loss, the loss get decreased from 2.176 to 0.080.

Figure 8 Loss for proposed CNN-AE model in right eye
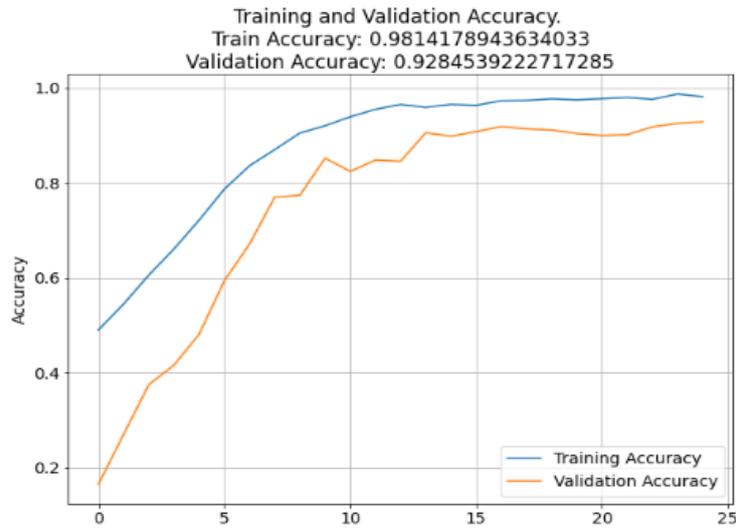
Figure 9 Accuracy for proposed CNN-AE model in left eye

Figure 9 has illustrated the left eye of the patients in which the training accuracy is 98.14% and validation accuracy is 92.84% at 25 epochs. The training accuracy initiated from 50.13% at 1 epoch and progressively increased to 98.14% and similarly the validation accuracy has initiated from 17.12% to 92.84%. As the epoch increases the CNN-AE fine tuning assist in improving the validation accuracy till 92.85%.

Figure 10 Loss for proposed CNN-AE model for left eye

Figure 10 has illustrated the loss for training and validation of proposed CNN-AE model for right eye patient records. The loss of CNN-AE model initiated with 1.415 and progressively decreases while the epoch count increases and finally at epoch 25 as 0.063. Similarly in the case of validation loss, the loss get decreased from 2.247 to 0.522.

According to table 1 and figure 11, the accuracy of training and testing of CNN-AE model in right eye is 98.71% and 97.36% correspondingly is higher than left eye accuracy which is 98.14% and 92.85% respectively. The model has performed better in both eyes but generate high result in right eye.

Table 1 Comparison of accuracy performed in right eye and left eye

Evaluation Metrics for the DL models	CNN-AE method		CNN method	
	Accuracy in Right eye (%)	Accuracy in Left eye (%)	Accuracy in Right eye (%)	Accuracy in Left eye (%)
Training	98.70	98.14	93.82	93.15
Testing	97.38	92.84	93.56	91.47
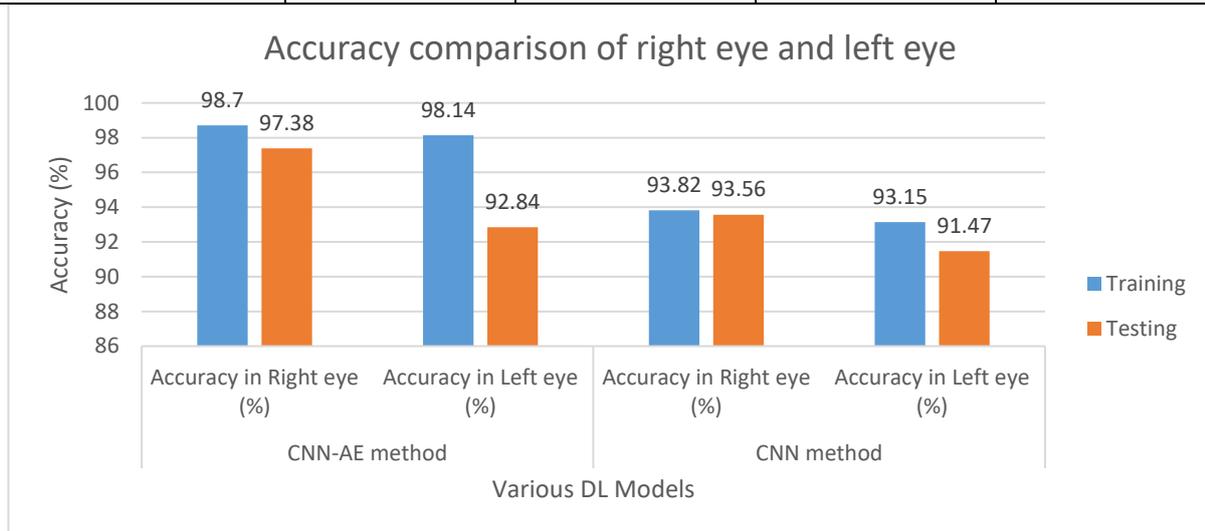
Figure 11 Accuracy performance of right eye and left eye

Table 2 Comparison of loss performed in right eye and left eye

Evaluation Metrics for the DL models	CNN-AE method		CNN method	
	Loss in Right eye	Loss in Left eye	Loss in Right eye	Loss in Left eye
Training	0.063	0.063	0.983	0.985
Testing	0.080	0.522	1.175	1.964

According to table 2 and figure 12, the loss of training and testing of CNN-AE model in right eye is 0.063 and 0.080 correspondingly is equal with left eye loss value in training but lesser than left eye loss value in testing are 0.063 and 0.522 respectively. The model has performed better in both eyes but generate low loss result in right eye.
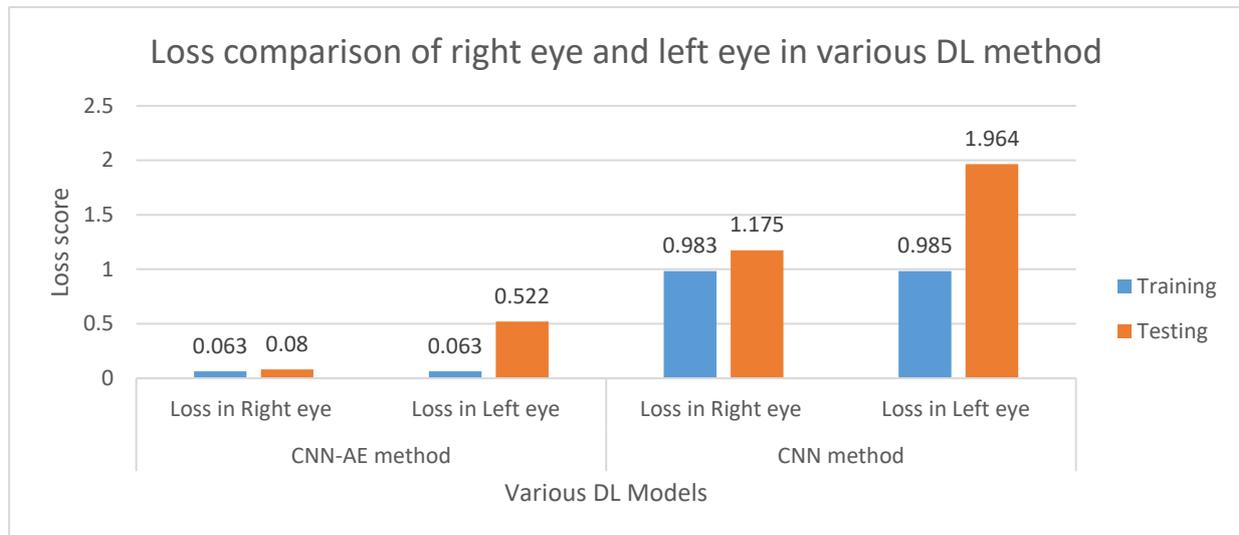
Figure 12 Loss performance of right eye and left eye

Based on these obtained result, the proposed CNN-AE produce high accuracy in predicting glaucoma disease with right eye is better while compared to left eye. The validation accuracy of right eye is 97.38% higher than left eye is 92.85%.

Conclusion

This research has proposed a unique approach named AE to enhance the functional strength of CNN. The usage of AE has assisted CNN in performing significantly. However, the architecture of CNN has made use of the AE concept and in addition, TL has a better generative model that has been employed to improve input representation. Hence, the AE method serves as an auxiliary learner in segregating the dispersion and variation of the input data. The performance of TL has assisted in generating channels from auxiliary learners accessible with the original CNN model's feature set which help in the creation of a complex features hierarchy. In order to get a further advantage of reducing computing time, TL is also utilized for the fine-tuning the proposed model with CNN-AE training. The proposed model has outperformed with high prediction results in identifying cataract. The training accuracy and testing accuracy is obtained for both the right and left eye in which data sample collected from the right eye has shown high accuracy in predicting cataract disease is 97.38% while compared to left eye which is 92.84%. Moreover, the intent to use different auxiliary learners both in a supervised and unsupervised manner.

Reference

[1] S. Hu et al., 'Unified Diagnosis Framework for Automated Nuclear Cataract Grading Based on Smartphone Slit-Lamp Images', IEEE Access, vol. 8, pp. 174169–174178, 2020, doi: 10.1109/ACCESS.2020.3025346.

[2] W. Song, Y. Cao, Z. Qiao, Q. Wang, and J. Yang, 'An Improved Semi-Supervised Learning Method on Cataract Fundus Image Classification', in 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), Jul. 2019, vol. 2, pp. 362–367, doi: 10.1109/COMPSAC.2019.10233.

[3] R. Sigit, M. Kom, M. B. Satmoko, D. K. Basuki, S. Si, and M. Kom, 'Classification of Cataract Slit-Lamp Image Based on Machine Learning', in 2018 International Seminar on Application for Technology of Information and Communication, Sep. 2018, pp. 597–602, doi: 10.1109/ISEMANTIC.2018.8549701.

[4] V. Agarwal, V. Gupta, V. M. Vashisht, K. Sharma, and N. Sharma, 'Mobile Application Based Cataract Detection System', in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Apr. 2019, pp. 780– 787, doi: 10.1109/ICOEI.2019.8862774.

[5] L. Cao, H. Li, Y. Zhang, L. Xu, and L. Zhang, 'Hierarchical method for cataract grading based on retinal images using improved Haar wavelet', Apr. 2019, arXiv:1904.01261.

[6] Y. Zhou, G. Li, and H. Li, 'Automatic Cataract Classification Using Deep Neural Network With

Discrete State Transition', IEEE Transactions on Medical Imaging, vol. 39, no. 2, pp. 436–446, Feb. 2020, doi: 10.1109/TMI.2019.2928229.

[7] Linglin Zhang, Jianqiang Li, i Zhang, He Han, Bo Liu, Yang, J., & Qing Wang. (2017). Automatic cataract detection and grading using Deep Convolutional Neural Network. 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC).

[8] Hossain, M. R., Afroze, S., Siddique, N., & Hoque, M. M. (2020). Automatic Detection of Eye Cataract using Deep Convolution Neural Networks (DCNNs). 2020 IEEE Region 10 Symposium (TENSYMP). Published. https://doi.org/10.1109/tensymp50017.2020.9231045

[9] Doshi, D., Shenoy, A., Sidhpura, D., & Gharpure, P. (2016). Diabetic retinopathy detection using deep convolutional neural networks. 2016 International Conference on Computing, Analytics and Security Trends (CAST). Published. https://doi.org/10.1109/cast.2016.7914977

[10] . Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).

[11] Yuan, L., Hou, Q., Jiang, Z., Feng, J. & Yan, S. Volo: Vision outlooker for visual recognition. arXiv preprint arXiv:2106.13112 (2021).

[12] Fan, W., Shen, R., Zhang, Q., Yang, J.J., Li, J.: Principal component analysis based cataract grading and classification. In: 2015 17th International Conference on E-health Networking, Application and Services (HealthCom), IEEE, pp. 459–462 (2015)

[13] Manchalwar, M., Warhade, K.: Detection of cataract and conjunctivitis disease using histogram of oriented gradient. Int. J. Eng. Technol. (IJET) (2017)

[14] Qiao, Z., Zhang, Q., Dong, Y., Yang, J.J.: Application of SVM based on genetic algorithm in classification of cataract fundus images. In: 2017 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, pp. 1–5 (2017)

[15] Xiong, L., Li, H., Xu, L.: An approach to evaluate blurriness in retinal images with vitreous opacity for cataract diagnosis. J. Healthc. Eng. 2017 (2017).

[16] Dong, Y., Wang, Q., Zhang, Q., Yang, J.: Classification of cataract fundus image based on retinal vascular information. In: International Conference on Smart Health, Springer, pp. 166–173 (2016)

[17] Cao, L., Li, H., Zhang, Y., Zhang, L., Xu, L.: Hierarchical method for cataract grading based on retinal images using improved haar wavelet. Inf. Fusion 53, 196–208 (2020)

[18] Xiong, Y., He, Z., Niu, K., Zhang, H., Song, H.: Automatic cataract classification based on multi-feature fusion and SVM. In: 2018 IEEE 4th International Conference on Computer and Communications (ICCC), IEEE, pp. 1557–1561 (2018)

[19] Huang, J., Lu, H., Lopez Meyer, P., Cordourier, H., & del Hoyo Ontiveros, J. (2019). Acoustic Scene Classification Using Deep Learning-based Ensemble Averaging. Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019). https://doi.org/10.33682/8rd2-g787

[20] Kumar, A., Kim, J., Lyndon, D., Fulham, M., & Feng, D. (2017). An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification. IEEE Journal of Biomedical and Health Informatics, 21(1), 31–40. https://doi.org/10.1109/jbhi.2016.2635663

[21] Smaida, M., & Yaroshchak, S. (2020). Using Ensemble Learning for Diagnostics of Eye Diseases. International Journal of Scientific and Research Publications (IJSRP), 10(10), 273–279. https://doi.org/10.29322/ijsrp.10.10.2020.p10639

[22] Parampal S.Grewal MD, Faraz Oloumi PhD, Uriel Rubin MD, Matthew T.S.Tennant MD, FRCSC., Deep learning in ophthalmology: a review, Canadian Journal of Ophthalmology, 2018.

[23] Julia E. Reid, MD, Eric Eaton, PhD., Artificial Intelligence for Paediatric Ophthalmology,arXiv:1904.08796v1,2019.

[24] Wei Lu, Yan Tong, Yue Yu, Yiqiao Xing, Changzheng Chen, Yin Shen., Applications of Artificial Intelligence, Hindawi Journal of Ophthalmology,2018.