CATARACT DETECTION USING DIGITAL CAMERA IMAGES

Abhijith J
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
Kottayam,Kerala, India
abhijithj2024@cs.ajce.in

Athul Krishna S
Department Of Computer Science And Engineering
Amal Jyothi College of Engineering
Kottayam,Kerala,India
athulkrishnas2024@cs.ajce.in

Amarthyag P
Department Of Computer Science And Engineering
Amal Jyothi College of Engineering
Kottayam,Kerala, India
amarthyagp2024@cs.ajce.in

Angela Rose Baby
Department Of Computer Science And Engineering
Amal Jyothi College of Engineering
Kottayam,Kerala,India
angelarosebaby2024@cs.ajce.in

Ms.Mekha Jose
Department Of Computer Science And Engineering
Amal Jyothi College of Engineering
Kottayam,Kerala,India
mekhajose@amaljyothi.ac.in

***Abstract*—Cataracts, a common eye condition, are a major cause of vision problems worldwide. Finding cataracts early is important so that they can be treated right away.Slit lamps and fundus cameras are two common instruments used to detect cataracts; both are very expensive and require domain expertise.Therefore Cataract may remain undetected at early stages, and when detected at later stages it need expensive medical intervention. In this paper we propose a novel approach for cataract detection from digital images which is a solution to the above mentioned problem. Here we utilize a Convolutional Neural Network (CNN) model for image classification. Use of smartphones for capturing images and detection of cataract lead to simple and easily accessible solution to cataract detection to common people.**

***Index Terms*—Cataract,CNN,VGG-16,ResNet,Inception V3**

I. Introduction

Cataracts, a condition where the eye's natural lens becomes clouded, continue to be a major contributor to vision impairment and blindness on a global scale. While the elderly are most at risk, there are other factors such as injury, medication, or genetics that can also lead to cataracts in younger individuals. As the world's population ages at a rapid pace, the number of cases is predicted to rise, creating significant challenges in the field of public health. Detecting cataracts early and seeking prompt treatment is essential in halting their development and maintaining clear eyesight.

For cataracts to be adequately managed and treated, early detection is essential. As the condition advances, it can greatly impede vision and make daily tasks challenging, ultimately diminishing one's quality of life. Initial indicators may include hazy sight, heightened sensitivity to bright light, and trouble seeing in dimly lit environments. Yet, cataracts often progress gradually, causing people to delay seeking medical assistance until their vision is significantly impacted.

The importance of detecting visual impairments early cannot be emphasized enough, especially given the immense impact they have globally. [1] Over 2.2 billion people worldwide are estimated to be blind or visually impaired by the World Health Organization, with cataracts accounting for a large portion of this figure. It is shocking to see that over half of blind cases globally are caused by cataracts, hence prevention efforts must be prioritized.

Conventional strategies for detecting cataracts typically involve clinical evaluations and specialized imaging methods, like fundus photography. Although these methods can be successful, they are often constrained by factors like availability, expense, and the specialized knowledge needed for analysis. Furthermore, in many areas, especially in resource-limited settings, there is a dearth of skilled eye doctors and facilities for comprehensive vision care

Utilizing age-old techniques for detecting cataracts as a foundation, the latest technological developments, specifically the widespread usage of advanced smartphones with top-notch cameras, offer a hopeful pathway towards early diagnosis. Our objective is to develop a dependable and efficient model for diagnosing cataracts from digital photos taken with cellphones by utilizing the capabilities of deep learning techniques, particularly [2] Convolutional Neural Networks (CNNs). This approach boasts multiple benefits over traditional techniques, such as increased accessibility, cost-efficiency, and scalability. The utilization of smartphone-based screening programs offers a promising avenue for connecting with marginalized communities living in isolated or low-resource regions. This approach eliminates the reliance on specialized eye care facilities and equipment, thus minimizing the strain on healthcare infrastructure. With our proposed CNN model, we capitalize on the abundance of visual information captured by digital images to quickly and precisely detect key markers of cataract pathology, making it an effective tool for large-scale screening endeavors.

II. RELATED WORKS

It has long been the practice to evaluate the severity of cataracts using ophthalmic medical images, and it is anticipated that these images will enhance the ability to detect and diagnose cataract sickness. Ultrasonic, slit-lamp, fundus, retroillumination, and digital camera images are among the various types of ophthalmic images that are generally used to diagnose cataracts. The two most used techniques for identifying cataracts are the slit-lamp image and the fundus imaging. On the other hand, individuals who reside in rural regions may find it difficult to obtain equipment for taking fundus photos. Digital camera images are more practical than ophthalmology imaging when it comes to cataract identification. Therefore, a cataract detection method based on photos from smartphones' cameras is needed for the initial detection of cataracts.

[3] offers an automated technique for cataract detection and grading as immature, mature, normal, and hyper mature based on fundus images. The five hidden layers of the CNN model utilized in this study include filter sizes of 8, 16, 32, 64, and 128 channel outputs. It makes use of a completely connected layer and SoftMax activation mechanism to ascertain the cataract class. The Adam Optimizer yields a training accuracy of 0.93 and a learning rate of 0.001.

[4] uses fundus images for cataract classification.Images are normalized and shrunk to 512 by 512 pixels in order to eliminate image defects.It is done to supplement data.When the machine learning model is being trained, it helps to lessen overfitting. The model employed is the ResNet architecture, which is typically made up of 152 convulsion levels plus one fully connected layer.

[5] In this case, the original RGB fundus photos are transformed to the green channel. Eight layers make up the model; the first five are convolutional layers and the final three are fully connected layers. The model is a DCNN. A four-way softmax receives the output from the final fully connected layer and generates a distribution across the four class labels.Each layer is made up of numerous 2D feature maps, each of which often captures a specific aspect of the image, like color, object category, or characteristics, while also preserving the spatial data at coarse resolution.

[6] When users take ocular images using the smartphone-based slit-lamp, several photographs of each individual would be taken. Tenengard algorithm is applied in order to resolve the blurry image issue. The author developed a novel fusion technique that combines the ShuffleNet model, the Gray Level Co-occurrence Matrix-based support vector machine, and the YOLO v3 model. The stacking method uses the model output values that need to be combined as the input values in the next step and the final classification label as the output value in order for the logistic regression approach to achieve the coefficient.

[7] Preprocessing is carried out and fundus pictures are obtained.The DCNN feature extractor, which uses a residual network with 17 convolutional layers to gain additional knowledge about abstract fundus images, is used for feature extraction. Shallow, residual, and pooling modules make up the entire DCNN. After the shallow and residual modules have extracted the features at the shallow, medium, and deep levels, the pooling module outputs the last feature vectors for the RF grader. A more thorough six-level cataract grading is achieved using RF. This approach achieves an average detection accuracy of 97.4%, according to experiments.

[8] comprises 16 layers arranged into 4 blocks, with 2 layers in each block. RGB pictures of 224 by 224 pixels are the input for the first block. A 3x3 kernel and "valid" padding are used in each convolutional layer. To decrease the spatial dimensions, Max-Pooling is used with a stride of 2. Rectified Linear Unit (ReLU) activation functiongenerate non-linear behavior. Feature extraction is handled by the network's first four blocks. Each block has 32, 32, 64, and 128 filters, increasing in number gradually. These blocks' outputs are aggregated into a feature map for additional processing. Fully connected layers get the feature map as input for categorization. The layers that are entirely connected consist of dropout, dense, and flatten layers. Dropout layers are employed to stop overfitting in denser layers, which contain 64, 128 and 256 neurons. A sigmoid activation mechanism is used in the last dense layer to detect binary cataracts (cataracts vs. normal eyes).A 99.13% accuracy rate on average is attained.

[9] a method for detecting cataracts using digital camera photos that uses CNN is proposed.

The seven layers of the model suggested includes two layers of convolution, two Max-pooling layers, a flatten layer, and two dense layers. Three primary colors—red, green, and blue—across 64×64 sizes were present in the input photos. ReLUs, or rectified linear units, were used to activate convolutional layers. Two max-pooling layers were implemented with a pool size of 2×2. In the first dense layer, 128 neurons and the ReLU activation function were employed. In the last dense layer, a single neuron and the Sigmoid activation function were used as a binary classifier. There was no dropout rate in place.Binary cross-entropy is used by the learning algorithm as a loss function.

III. METHODOLOGY

A. Dataset

Digital camera photos acquired by Krishnabojha [10] are included in the dataset used in this study, which can be accessed on github. It includes both normal eye photos and cataract images. Of the 9668 photos, 80% are utilized for training and the remaining 20% are used to assess the performance of the model.

B. Model Description

Convolutional Neural Networks (CNNs) are deep neural networks that construct a complex feature hierarchy through the use of convolutional processes, pooling layers, and non-linear activation functions.In contrast to the manual feature extraction approach, which separates these two processes, deep learning-based methods combine them during the feature extraction as well as classification phases.

TABLE I
DESCRIPTION OF VARIOUS LAYERS OF THE SUGGESTED MODEL

Layer	Filter	Config.	Stride	Output
Conv2D	32	KS: 3x3; ReLU	-	200 x 200 x 32
Max-pooling2D	-	KS: 2x2	2	-
Conv2D	32	KS: 3x3; ReLU	-	-
Max-pooling2D	-	KS: 2x2	2	-
Conv2D	64	KS: 3x3; ReLU	-	-
Max-pooling2D	-	KS: 2x2	2	-
Conv2D	128	KS: 3x3; ReLU	-	-
Max-pooling2D	-	KS: 2x2	2	-
Flatten	-	-	-	-
Dense	-	64; ReLU	-	-
Dropout	-	0.4	-	-
Dense	-	128; ReLU	-	-
Dropout	-	0.4	-	-
Dense	-	256; ReLU	-	-
Dropout	-	0.5	-	-
Dense	-	Sigmoid	-	-

Table 1 provides a description of model's sixteen levels. Half of these layers are included in four blocks, each with two levels; the other half is utilized for classification.

The first block uses 200 × 200 RGB (three input channel) photos as inputs. There are 32 filters used, all with proper padding and a kernel size (KS) of 3 x 3.Next, a Max-Pooling (MP) layer with a stride of 2 is applied to reduce the data representation's spatial dimensions (width and height). More pixels essentially result in fewer pixels since more pixels mean more parameters and more data. In the end, this block is activated by the ReLU activation function, which means that the matrix negative values are taken to be 0 and the positive values stay unchanged.

The second block is the same block with the same parameter values. Next, the third block is the same as the second, but this time it has 64 filters. In the fourth block, the number of filters is raised to 128.A feature map is fed into the fully connected layers from the aggregate outputs of the four blocks. The purpose of these layers—the dropout, dense, and flatten layers—is to detect cataracts. Three sets of dense and dropout layers are built, with the dense layers comprising 64, 128 and 256 attending neurons, in order to gather the changing properties of the cataract.Additionally, three dropout layers with values of 0.4, 0.4, and 0.5 are set up to set 40%, 40%, and 50% of the neurons in the hidden layers to 0 at each training phase update in order to prevent the model from overfitting. Since cataract detection is a binary classification, the sigmoid activation function which is given below is used in the last dense layer.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

The last activation function is the sigmoid function, whose input is represented by the symbol x. There are N classes in the sigmoid layer, and each class is associated with a single neuron. The two main groups in our system are those who have cataracts and those who have normal eye health. In binary classification, the CNN design generates output at two neurons. Given a cataract image, the contribution of the surviving and second neurons would be either 1 or 0, or vice versa.

This CNN model is developed using Python and a Flutter application is employed alongside for user interaction. The application connects to the model through Flask to deliver prediction outputs to users. Users can input digital images of the eye through the application, which then provides a prediction indicating whether the eye shows signs of cataracts or not.

IV. RESULT AND DISCUSSION

We go over the performance of the suggested model in this section. Using the same dataset, the cataract detection capabilities of our model and three other pre-trained algorithms are evaluated.

The confusion matrix indicates that a model that classifies images into cataract and non-cataract categories performs well. Photos with cataracts are represented by 0 in this context, while photos without cataracts are represented by 1.The frequency of accurate predictions along the diagonal (853 for non-cataract and 712 for cataract) relative to the off-diagonal values (incorrect predictions) indicates the model's ability to differentiate between cataract and non-cataract images. Figure 1 displays the suggested model's confusion matrix.

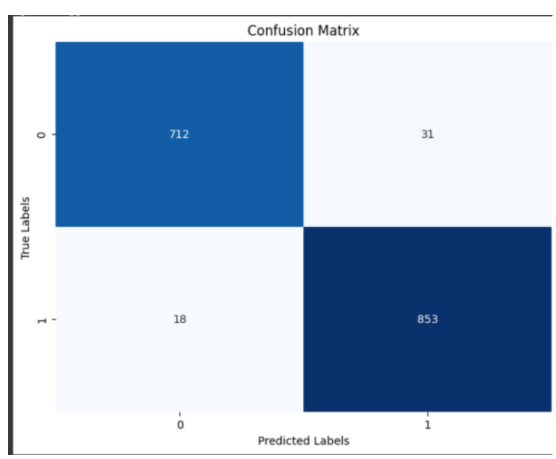

Fig. 1. Confusion Matrix

The performance of a classification model over training epochs is represented graphically by the fig 2 graph.

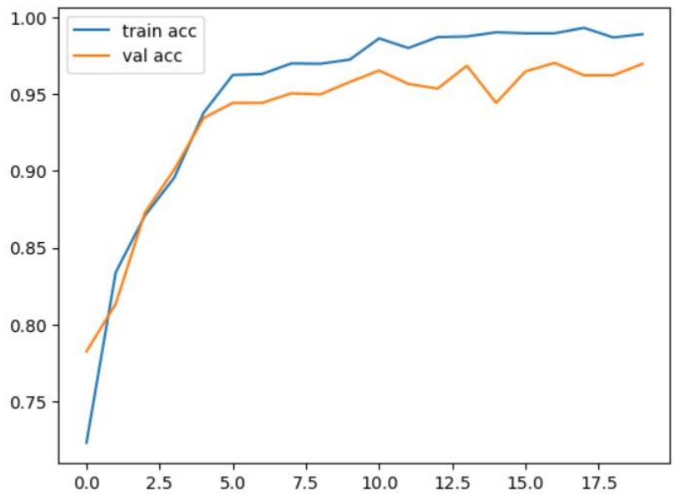

Fig. 2. Training And Validation Graph

The total number of epochs, or full cycles of training the model on the whole training dataset, is tracked on the X-axis. The Y-axis, which illustrates how effectively the model classifies instances, represents accuracy. Two plotted lines are present,

Training Accuracy (train acc): This line represents the model's prediction performance using the training data that it is currently being trained on. It is typically shown in blue.

Correctness of Validation (val acc): This line, which is usually displayed in orange, indicates the model's accuracy on an alternative validation dataset.

This validation set is essential for determining how well the model prevents overfitting and generalizes to new data.The steady increase in training and validation accuracy with increasing epochs is the graph's encouraging finding. This suggests that the model is assiduously gaining knowledge from the training set. Furthermore, it appears that as epochs pass, the separation between the two lines is increasing smaller. Although a better training accuracy is initially expected, a narrowing gap shows the model has learned complicated patterns without overfitting to its training dataset. . Overall, the model's performance is indicated by this visualization, which shows that it can learn from new data and generalize to previously unseen data.

We are now comparing the acquired result with the three prior pre-trained models based on their accuracy, precision, recall, and F1 score.The equations for above parameters are given below.

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \quad (2)$$

$$Precision = \frac{T_p}{T_p + F_p} \quad (3)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (4)$$

$$F1 - Score = \frac{2 * (Precision * Recall)}{(Precision * Recall)} \quad (5)$$

TABLE II
COMPARISON BETWEEN PROPOSED MODEL AND OTHER THREE PRE-TRAINED MODELS

Model	Accuracy	Precision	Recall	F1-Score
VGG-16	95.23	95.5	94.82	95.18
ResNet	69.76	67.00	77.13	71.71
Inception V3	95.54	95.50	97.63	96.5
Proposed CNN Model	96.88	96.49	97.93	97.20

Table 2 displays each model's performance results for recall, accuracy, precision, and f1-score. It is clear that the recommended CNN model, which has a 96.88% training accuracy, produces the best accuracy when the Adam optimizer is applied at a 0.001 learning rate. This is because the model's architecture is acceptable and not unduly complex for the dataset. The equivalent f1-score, precision, and recall for this model are 97.20%, 97.93%, and 96.49%, respectively.

V. CONCLUSION

The classification of cataracts from digital photographs was the subject of this work, which compared multiple CNN architectures: ResNet, Inception V3, VGG 16, and the CNN model that is recommended.To compare the four architectures, the AdamOptimizer is continually run with an average rate of learning of 0.001. With an accuracy of 96.88% on RGB input, the suggested CNN model produced the best system performance. The classification accuracy found in this study is an acceptable accuracy performance, based on the performance data. It is anticipated that this study will assist medical professionals in detecting cataracts early in order to prevent the condition's harmful effects and to provide the necessary medical care. Moving forward, our goal is to add more datasets in order to raise the cataract detection system's classification accuracy.

REFERENCES

[1] "Statistical Analysis By WHO." :https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] S. Gayathri, S. Saran, P. S. Kumar, and I. P. Singh, "Cataract disease classification using convolutional neural network architectures," in *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pp. 992–998, 2023.

[3] R. Simanjuntak, Y. Fu'adah, R. Magdalena, S. Saidah, A. Wiratama, and I. Ubaidah, "Cataract classification based on fundus images using convolutional neural network," *JOIV : International Journal on Informatics Visualization*, vol. 6, p. 33, 03 2022.

[4] A. Parulekar, A. Martins, M. Fernandes, P. Bhat, P. R. Shetgaonkar, and D. S. Aswale, "Advance cataract onset detection using deep learning," 2021.

[5] L. Zhang, J. Li, i Zhang, H. Han, B. Liu, J. Yang, and Q. Wang, "Automatic cataract detection and grading using deep convolutional neural network," in *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, pp. 60–65, May 2017.

[6] S. Hu, X. Wang, H. Wu, X. Luan, P. Qi, Y. Lin, X. He, and W. He, "Unified diagnosis framework for automated nuclear cataract grading based on smartphone slit-lamp images," *IEEE Access*, vol. 8, pp. 174169–174178, 2020.

[7] J. Ran, K. Niu, Z. He, H. Zhang, and H. Song, "Cataract detection and grading based on combination of deep convolutional neural network and random forests," in *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 155–159, Aug 2018.

[8] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "Cataract-net: An automated cataract detection system using deep learning for fundus images," *IEEE Access*, vol. 9, pp. 128799–128808, 2021.

[9] C.-J. Lai, P.-F. Pai, M. Marvin, H.-H. Hung, S.-H. Wang, and D.-N. Chen, "The use of convolutional neural networks and digital camera images in cataract detection," *Electronics*, vol. 11, no. 6, 2022.

[10] Krishnabojha, "Cataract Detection,–Using-CNN." :https://github.com/krishnabojha/Cataract Detection-using CNN. [Accessed 05-03-2022].