 DOI: 10.35382/TVUJS.14.3.2024.9

CATARACT CLASSIFICATION USING MOBILENETV2-BASED MODEL

Khanh-Duy Nguyen[1,4*], Thai-Son Nguyen[2], Min-Te Sun[3]

Abstract – *Cataract occurs when the lens of the eyes, normally transparent, becomes cloudy. Clouded vision resulting from cataracts can pose challenges in activities such as reading, nighttime driving, and discerning facial expressions of acquaintances. Ensuring quality of vision now requires early detection of cataracts. This study aims to create a deep-learning classification system capable of distinguishing between healthy eyes and those affected by cataracts. To achieve this, modifications such as skip connections and channel-wise attention have been integrated into the pre-trained MobileNetV2 model to formulate the proposed model. Furthermore, augmentation technique and unsharp masking filter are implemented in the pre-processing dataset to augment the image count and improve image quality. The findings indicate that the model achieved an accuracy rate of 98,80% for ODIR-2019 in 2 categories: cataract and normal.*

Keywords: cataract classification, channel-wise attention, mobileNetV2, skip connection.

I. INTRODUCTION

Cataract is an eye disease that develops when fluid accumulates in the lens of the eye, leading to blurred vision [1]. Cataracts represent the prevailing treatable cause of acquired blindness on a global scale [2]. The National Eye Institute (NEI) reports that approximately 24.4 million Americans aged 40 and older are affected by cataracts [3]. Cataracts can be caused by various factors, such as exposure to sunlight and other ultraviolet radiation sources, diabetes, excessive smoking or drinking, obesity, genetic predisposition, prior eye surgeries, eye trauma, and prolonged use of corticosteroid medications [1]. Indicators of cataracts include blurred vision, difficulty seeing clearly, frequent alterations in lens prescription, distorted colors, double vision, and heightened sensitivity to glare.

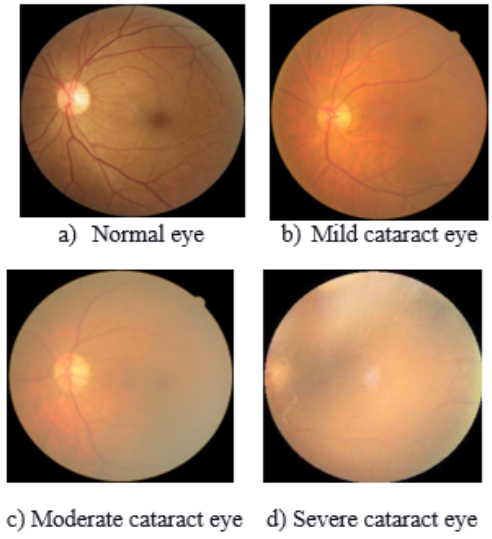

Fig. 1: Cataract classification

Figure 1 illustrates a comparison between a healthy eye and one afflicted with cataracts [4]. Figure 1 (a) shows the normal eye in which the tiny blood vessels can be seen very clearly. Figure 1 (b) and Figure 1 (c) show the mild and moderate stages where blood vessels can be seen. Figure 1(d) displays the severe stage of cataract because the blood vessels are invisible.

Detecting cataracts in their early stages can avoid several consequences, such as blindness

[1,2]Tra Vinh University, Vietnam
[3]National Central University, Taoyuan, Taiwan
[4]Postgraduate student, National Central University, Taoyuan, Taiwan
*Corresponding author: nkduy@tvu.edu.vn
Received date: 28th June 2024; Revised date: 09th September 2024; Accepted date: 16th September 2024

[5]. This manual process requires clinical skills and experience from ophthalmologists. In particular, cataract classification refers to recognizing highly complex blood vessel system. Therefore, there is a need for techniques that could easily diagnose cataract. In this study, an attempt is made to handle this issue by building a deep-learning model to help classify cataracts. In this work, a deep learning model was introduced, designed by incorporating skip connection and channel-wise attention using the MobileNetV2 model. The MobileNetV2 model was pre-trained by the ImageNet dataset.

The contributions of the article are presented as follows:

- The categorization of images that depict normal and cataract-afflicted eyes is accomplished through the utilization of a deep-learning framework trained on ImageNet datasets.

- The deep-learning model utilizing MobileNetV2 is crafted to integrate skip connections and channel-wise attention blocks, resulting in superior performance when compared to the original pre-trained MobileNetV2.

The rest of this study is structured as follows. In Section II, a review of the literature is presented. The technical details are given in Section III. Section IV depicts the dataset, evaluation metrics, and experimental setup. The results are analyzed in Section V. Finally, Section VI provides the conclusion.

II. LITERATURE REVIEW

Research on cataracts has been ongoing for many years. Studies investigate the implementation of machine learning into cataract classification. Fan et al. [6] first proposed using principal component analysis (PCA) to reduce the dimensionality of the input dataset. Subsequently, they employ two distinct methods, namely wavelet feature extraction and the sketch method, to extract image features. The researchers then opted for support vector machines (SVM), bagging, random forest, and gradient-boosted decision trees to classify 455 fundus images. The findings indicate that the highest accuracy, reaching 84.77%, can be achieved using the sketch feature with SVM. Dong et al. [7] select a range of comprehensive pre-processing techniques, including the maximum entropy contrast, the Krisch template filter, and the adaptive weighted median filter. Subsequently, they employ SVM for classification, achieving an accuracy of 94.91% using wavelet features.

The rise of deep learning has led to increased popularity of neural network models for categorizing cataracts. The research carried out by Kaur et al. [8] employs the ResNet-18 model to categorize images of eye illness into four distinct groups. Their study employed a dataset containing 4,217 photographs and attained an accuracy rate of 94%. The approach of Zhang et al. [9] propose employing the deep convolutional neural network for cataract detection, achieving a detection accuracy of 93.52%. Neural networks are increasingly becoming deeper, and there is a growing trend towards utilizing pre-trained models in computer vision tasks. Singh et al. [10] introduce a pre-trained visual geometry group (VGG) 16 model designed to detect cataract disease, achieving an accuracy rate of 96.10%. Also, the study of Funde et al. [5] applies the VGG19 model to detect cataracts.

III. RESEARCH METHODS

A. Preprocessing data

Data augmentation Data augmentation has become a necessary part of the successful application of deep learning models on image data to improve the sufficiency and diversity of training data [11]. In this research, the data shortage was overcome by employing geometric image augmentation, as discussed in the study of Awaluddin et al. [12] and Maharana et al. [13]. New transformed images are obtained from the original images using Keras ImageDataGenerator, developed as part of 'keras.preprocessing.image' module and provided many parameters to augment the data. The study set seven parameters in ImageDataGenerator, in particular:

- Rotation range = 20: The images can be rotated by up to 20 degrees.

- Width shift range = 0.01: The images can be shifted horizontally up to 1% of their total.
- Height shift range: The images can be vertically shifted up to 1% of their total height.
- Horizontal flip: The images can be randomly flipped horizontally.
- Vertical flip: The images can be randomly flipped vertically.
- Shear range = 0.1: The images can be shared by up to 1%.
- Zoom range = 0.1: The images can be zoomed in or out by up to 1%.

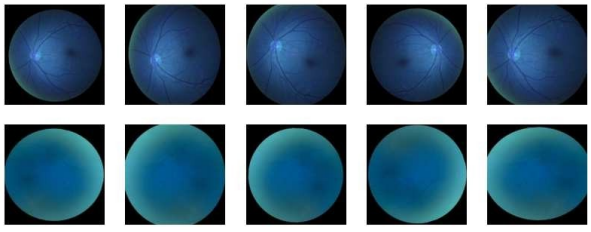

Fig. 2: Augmentation images

Figure 2 shows that creating five distinct images from each single original image has resulted in a five-fold increase in the number of images within the input dataset.

Unsharp masking filter

The unsharp masking filter (UMF) improves image contrast by adding a version of the image that highlights edges to the original input [14]. Let a color input image be given as I(i,j), where (i,j) is the pixel position of the image. The main idea is that the original image takes away the blurred version to obtain the detailed part [15]. In this paper, the Gaussian smoothing filter is employed to seek blurred images. The UMF can be calculated as Formula (1).

$$UMF(x,y) = I(x,y) - \sum_i \sum_i K(i,j)I(x-i,y-j) \quad (1)$$

Where:

UMF (x,y): The output image at position (x, y)

I (x,y): The original image at position (x, y)

K (i,j): The Gaussian kernel at position (i, j)

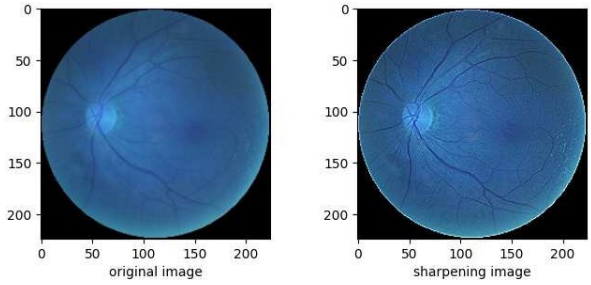

Fig. 3: Sharpening images

Figure 3 illustrates the outcomes of image post-application of an unsharp masking filter. The output image clearly reveals retinal blood vessels, including veins and arteries, along with the optic nerve and macula, which are also notably discernible. This enhancement has potential benefits for advancing the learning trajectory of deep learning models.

B. Global pooling block

Global pooling blocks take the multidimensional feature map and convert it to a one-dimensional feature vector using global pooling [16]. Global pooling layer transforms (H x W x C) feature map to (1 x 1 x C) where (H x W) is the input image size and C is the channel. There are two kinds of global pooling: global max pooling (GMP) and global average pooling (GAP). In this study, GAP was utilized. Figure 4 illustrates an example of global average pooling.

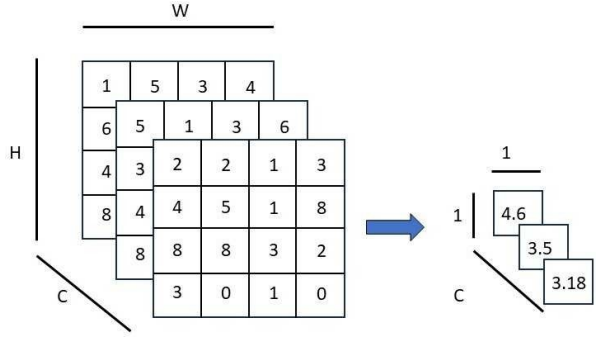

Fig. 4: Global average pooling

Let z be the feature map, for each channel C, the GPA operation can be represented as Formula (2).

$$z_C = \frac{1}{H \times W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} x_{ij} C \qquad (2)$$

Where H and W are the height and width of the feature map, and x_{ijC} is the value of the C-th channel at location (i,j).

C. Long skip connection

The revolution of the convolutional neural network (CNN), especially, the deep convolutional neural network which has many layers depicted exemplary performance in computer vision and image processing [17]. When the network becomes deeper and starts converging, adding more layers leads the deep learning model to a higher training error and causes over-fitting, known as the degradation problem [18]. In Shafiq et al. [18], skip connections, particularly in the context of residual networks (ResNet) and similar architectures, were added to the deep learning model to help mitigate the degradation problem.

Skip connections are links that bypass one or more layers in a CNN and connect the output of an earlier layer to the input of a later layer. By using this method, performance degradation due to too many stacked convolution structures can be avoided, and such a quick connection mechanism can perform identity mapping in multi-layer structures.

$$y = F(x, W_i) + x \qquad (3)$$

Where x and y are the input and output vectors.

D. Channel-wise attention

Channel-wise attention is a technique used in neural networks to enhance model performance by focusing on the most informative features across the channel dimension of the input data. To fully capture channel dependencies with the channel- wise descriptor, a simple gating mechanism can be applied. With z_C as global information of input x at C-th channel from Formula (2), the simple gating mechanism with sigmoid function is calculated by Hu et al. [19].

$$s_C = \sigma\big(W2\big(\delta(W1 z_c + b_1)\big) + b_2\big) \qquad (4)$$

where δ refers to the rectified linear unit (ReLU) activation function as Formula (5).

$$\delta(x) = \max(0, x) \qquad (5)$$

And σ refers to the sigmoid activation function as Formula (6).

$$\sigma(x) = \frac{1}{1 + e^{-x}} \qquad (6)$$

The final output of the channel-wise attention block is [19] as Formula (7).

$$\widetilde{x_c} = s_c \cdot x_c \qquad (7)$$

E. MobileNetV2 Model

MobileNet, a CNN architecture introduced by Howard et al. [20], addresses the challenge of high computing demands. Crafted with constrained resources in mind, it achieves optimal accuracy levels [21]. MobileNet employs depthwise separable convolutions to create compact deep convolutional neural network, making it a proficient choice for mobile and embedded vision tasks [20]. The Inception models later employed depthwise separable convolutions to lower computational expenses [22]. There are two parts in depthwise separable convolution: depthwise convolution and pointwise convolution. The depthwise convolution refers to convolution, in this context, pertaining to operations that remain within individual channels, thereby implying that each channel within the feature map utilizes its own convolution kernel exclusively for processing [20]. The calculation of the depthwise

convolution proposed by Howard et al. [20] is presented as Formula (8).

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} F_{k+i-1,l+j-1,m} \quad (8)$$

where K refers to the convolution kernel, and F refers to the feature map. Although depthwise convolution is very efficient, it cannot create new features [20]. It is necessary to combine the output of depthwise convolution through 1 x 1 convolution (pointwise). Figure 5 illustrates the MobileNet architecture. MobileNetV2, developed by

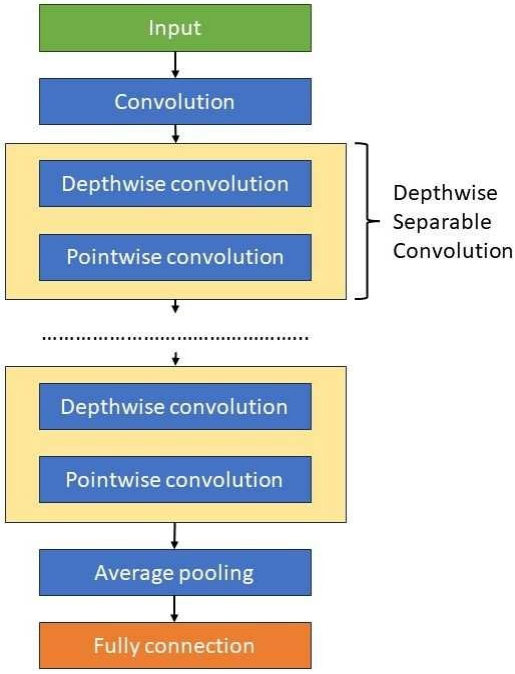

Fig. 5: MobileNet architecture

Sandler et al. [23], builds upon this by introducing inverted residuals and linear bottlenecks, providing improved performance and efficiency. Figure 6 shows the MobileNetV2 architecture. There are a few differences in the structure of depthwise separable convolution. One other pointwise convolution was added. Additionally, the residual connection which gives the architecture a better accuracy compared to non-residual architecture is incorporated. The MobileNetV2

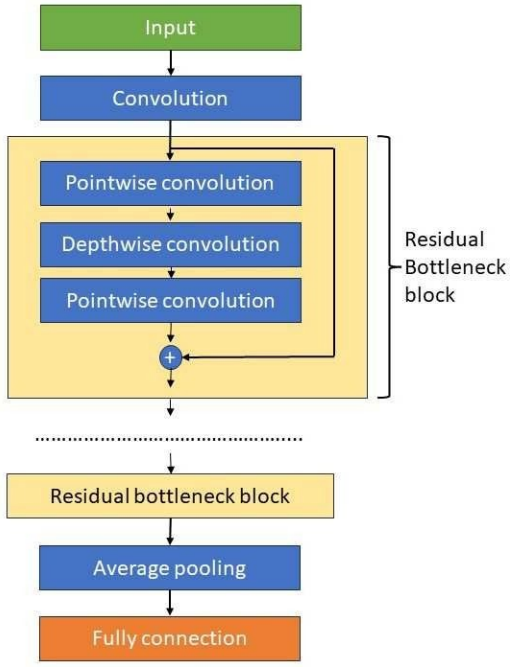

Fig. 6: MobileNetV2 architecture

was selected as the backbone model for this study due to its simple structure, which facilitates rapid training.

F. Proposal system

The proposal system is depicted in Figure 7. After loading a dataset (sized 244 x 244 x 3), images are enhanced by applying a UMF to sharpen their details. The original dataset contains 594 cataract images and 500 normal images. Through augmentation, the dataset expands to include 5,440 images for both cataract and normal categories. These augmented images are then scaled by dividing each pixel value by 255. Subsequently, the dataset is divided into training and testing sets. Within the training set, images are further divided into training data (3,481 images sized 244 x 244 x 3) and validation data (871 images sized 244 x 244 x 3), while the test data consists of 1,088 images sized 244 x 244 x 3. The training and validation images are inputted into a classification model, utilizing MobileNetV2 which was pre-trained with the ImageNet dataset

as the backbone model. The output shape of MobileNetV2 is (7 x 7 x 1280). Following MobileNetV2, a GAP layer, which is applied to convert the output shape of MobileNetV2 into (1 x 1 x 1280) and a channel-wise attention block are integrated. Another GAP layer processes the input to match the shape of the channel-wise attention output and then combines them. Model performance is assessed using predicted and test images, with metrics such as accuracy, precision, recall, and F1-score being utilized for evaluation.

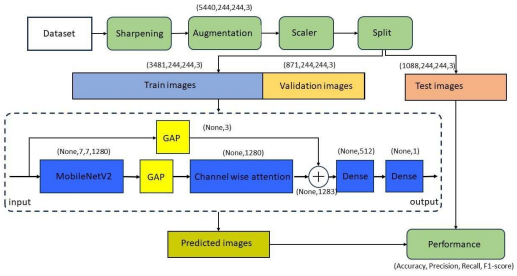

Fig. 7: Proposed system flow for cataract classification

IV. EXPERIMENTS

A. Dataset

The ODIR-2019 dataset (ODIR-5K) is used in the experiments. The dataset is assembled from 5,000 patients of different ages and color fundus photographs from both left and right eyes. Also, doctor's diagnostic keywords are attached to this dataset. The diversity labels include normal (N), glaucoma(G), cataract (C), AMD (A), hypertension (H), myopia (M), and other abnormalities (O). In this research, cataract (C) and normal (N) data are used as datasets to train and test the proposed model. The dataset can be easily found at ODIR-2019 Grand Challenge [24] or ODIR-5K Kaggel dataset [25].

B. Evaluation metrics

To evaluate the performance of the classification model, a confusion matrix was utilized. In the confusion matrix, there are four metrics including:

- True positive (TP): An instance for which both predicted and actual values are positive.
- False positive (FP): An instance for which the predicted value is positive but the actual value is negative.
- False negative (FN): An instance for which the predicted value is negative but the actual value is positive.
- True negative (TN): An instance for which both predicted and actual values are negative.

In classification task, Accuracy, Recall, Precision, and F1 Score are regular evaluation metrics that are often used to assess the performance of deep learning model as Formula (9), (10), (11) and (12).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$Recall = \frac{TP}{TP + TN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

C. Experimental setup

The study used Python on TWCC Taiwan AI Cloud with an NVIDIA Tesla V100 32GB GPU. The parameters for the experimental setup are detailed in Table 1.

Table 1: Experimental setup parameters

Parameter	Value
Batch size	64
Learning rate	1×10^{-4}
Optimizer	Adam
Loss function	Binary Cross-Entropy
Early stopping	Enabled
Training data size	3,481 images
Validation data size	871 images
Testing data size	1,088 images

Furthermore, binary cross-entropy was used to train the model as Formula (13).

$$L_{BCE} = \Sigma[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (13)$$

During model training, early stopping technique was employed to avoid overfitting.

V. RESULTS AND DISCUSSION

A. *The performance of the proposal model*

Figure 8 illustrates the fluctuations observed in both train and validation accuracy throughout each epoch. The training accuracy curve gradually increases from below 92% and reaches 100% by epoch 23, then maintains on the chart. Concurrently, the validation accuracy curve follows a similar pattern, rising from near 92% until achieving 98% at epoch 25, then stabilizing.

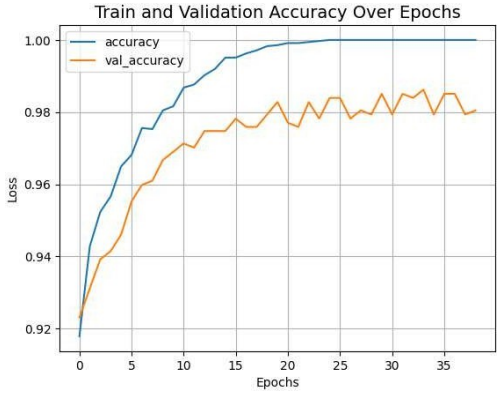

Fig. 8: Training accuracy curves of the proposal model

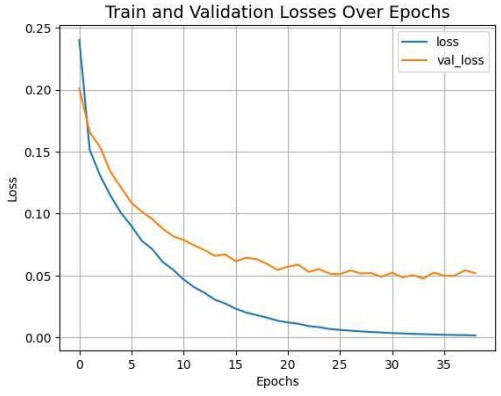

Fig. 9: Training loss curves of the proposal model

Figure 9 illustrates the loss during both training and validation for each epoch. The training loss initiates close to 0.25 and gradually decreases until hitting 0.0016 by epoch 39. Similarly, the validation loss goes down from 0.20 and reaches 0.052 by epoch 31, then remains steady.

B. *Comparison with other models*

Figure 10 and Figure 11 illustrate the confusion matrix of both the original MobileNetV2 and the proposed model. It is evident that there is a higher count of true positives in Figure 11 compared to Figure 10. Conversely, the count of false positives in Figure 11 is observed to be lower than in Figure 11.

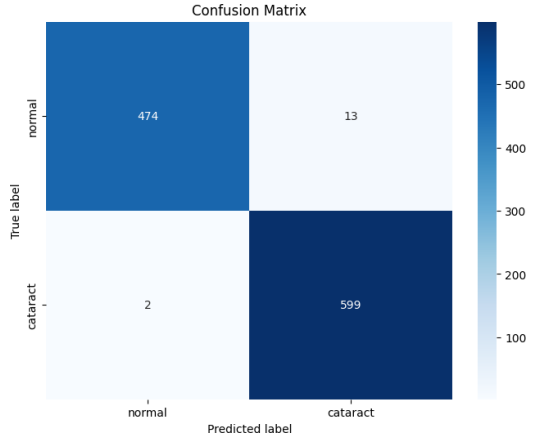

Fig. 10: Confusion matrix of original MobileNetV2

Table 2 summarizes the performance comparison of various models, bringing together metrics such as accuracy, precision, recall, and F1-score for DCNN, pre-trained VGG16, VGG19, MobileNetV2, and the proposed model. The proposed model excels with an accuracy of 98.80%, significantly outperforming DCNN (93.52%), VGG19 (94%), and pre-trained VGG16 (96.10%). Most notably, it shows a marked improvement over MobileNetV2, which achieved 98.62% accuracy. This enhancement underscores the effectiveness of the modifications, such as incorporating skip connections and channel-wise attention. These enhancements

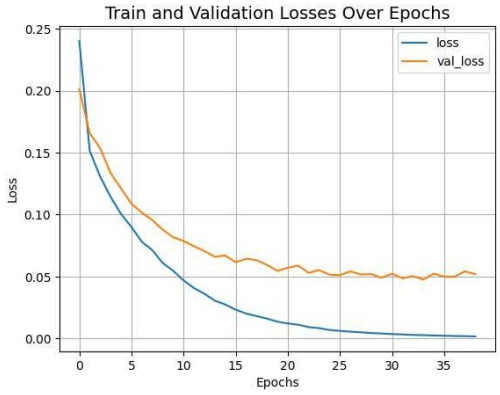

Fig. 11: Confusion matrix of the proposal model

Table 2: Performance comparison of different models

Model	Accuracy	Precision	Recall	F1-score
Diffusion-convolutional neural networks (DCNN) [9]	93.52%	—	—	—
Pre-trained VGG16 [10]	96.10%	—	—	—
VGG19 [26]	94%	—	—	—
MobileNetV2	98.62%	0.9733	0.9958	0.9844
Proposal model	98.80%	0.9774	0.9958	0.9865

address the degradation problem often seen in deep networks and enable the model to focus more effectively on the most informative features. Consequently, the proposed model not only achieves higher precision (0.9774 vs. 0.9733) but also boasts a slightly better F1-score (0.9865 vs. 0.9844), while maintaining the same recall (0.9958) as MobileNetV2. Additionally, the use of data augmentation and unsharp masking filters in preprocessing significantly enhanced the quality and diversity of the training data. Overall, these specific improvements enable the proposed model to outperform both the baseline MobileNetV2 and general purpose architectures like VGG16 and VGG19, demonstrating its potential effectiveness and reliability for cataract classification.

VI. CONCLUSION

In conclusion, the study has successfully developed and validated a novel deep-learning model based on MobileNetV2, which has demonstrated exceptional performance in classifying cataract-affected and normal eyes. By integrating advanced techniques such as skip connections and channel-wise attention, along with rigorous data augmentation and image preprocessing methods, the proposed model achieved an impressive accuracy of 98.80% on the ODIR-2019 dataset. This performance surpasses several existing models, highlighting the potential of the proposed approach in enhancing cataract detection capabilities, which could significantly aid in early diagnosis and potentially reduce the global burden of cataract-induced vision impairment. In the future, the study could be extended to classify a wider range of eye diseases using the model, which would enhance its usefulness in clinical settings. This improvement would allow for more detailed diagnostic evaluations and help in the better management of different eye conditions.

REFERENCES

[1] Gayathri S, Saran S, Kumar PS, Singh IP. Cataract disease classification using convolutional neural network architectures. In: *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*. 02^{nd}–04^{th} March 2023; Tuticorin, India. IEEE; 2023. p.992–998. http://dx.doi.org/10.1109/ICEARS56392.2023.10085502.

[2] Steinmetz JD, Bourne RR, Briant PS, Flaxman SR, Taylor HR, Jonas JB, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. *The Lancet Global Health*. 2021;9(2): 144–160.

[3] Gao Z, Gorenflo M, Kaelber DC, Monnier VM, Xu R. Drug repurposing for reducing the risk of cataract extraction in patients with diabetes mellitus: integration of artificial intelligence-based drug prediction and clinical corroboration. *Frontiers in Pharmacology*. 2023;14: 1181711. https://doi.org/10.3389/fphar.2023.1181711.

[4] Yadav S, Yadav JKPS. Automatic cataract severity detection and grading using deep learning. *Journal of Sensors*. 2023;2023(1): 2973836. https://doi.org/10.1155/2023/2973836.

[5] Funde K, Joshi J, Castelino K, Patil P, Mulla N. Cataract detection by leveraging VGG-19 classification model on retinal images. In: *13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 03rd–05th October 2022; Kharagpur, India. IEEE; 2022. p.1–6. https://doi.org/10.1109/ICCCNT54827.2022.9984433.

[6] Fan W, Shen R, Zhang Q, Yang JJ, Li J. Principal component analysis based cataract grading and classification. In: *17th International Conference on E-health Networking, Application & Services (HealthCom)*. 4th–17th October 2015; Boston, MA. IEEE; 2015. p.459–462. https://doi.org/10.1109/HealthCom.2015.7454545.

[7] Dong Y, Wang Q, Zhang Q, Yang J. Classification of cataract fundus image based on retinal vascular information. In: *Smart Health: International Conference (ICSH 2016)*. 24th–25th December 2016. Springer; 2017. p.166–173. https://doi.org/10.1007/978-3-319-59858-1_16.

[8] Kaur G, Sharma N, Chauhan R, Kukreti S, Gupta R. Eye disease classification using resnet-18 deep learning architecture. In: *2nd International Conference on Futuristic Technologies (INCOFT)*. 24th–26th November 2023; Belagavi, Karnataka, India. IEEE; 2023. p.1–5. https://doi.org/10.1109/INCOFT60753.2023.10425690.

[9] Zhang L, Li J, Zhang, Han H, Liu B, Yang J, Wang Q. Automatic cataract detection and grading using deep convolutional neural network. In: *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*. 03rd August 2017; Calabria. IEEE; 2017. p.60–65. https://doi.org/10.1109/ICNSC.2017.8000068.

[10] Singh G, Guleria K, Sharma S. A pre-trained vgg16 model for cataract eye disease prediction. In: *3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. 29th–31st December 2023; Bangalore, India. IEEE; 2023. p.1–6. https://doi.org/10.1109/SMARTGENCON60755.2023.10442647.

[11] Yang S, Xiao W, Zhang M, Guo S, Zhao J, Shen F. Image data augmentation for deep learning: A survey. *arXiv* [Preprint] 2022. https://doi.org/10.48550/arXiv.2204.08610.

[12] Awaluddin BA, Chao CT, Chiou JS. Investigating effective geometric transformation for image augmentation to improve static hand gestures with a pre-trained convolutional neural network. *Mathematics*. 2023;11(23): 4783. https://doi.org/10.3390/math11234783.

[13] Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*. 2022;3(1): 91–99.

[14] Kwok N, Shi H. Design of unsharp masking filter kernel and gain using particle swarm optimization. In: *7th International Congress on Image and Signal Processing*. 14th–16th October 2014; Dalian, China. IEEE; 2014. p.217–222. https://doi.org/10.1109/CISP.2014.7003780.

[15] Song Y, Li C, Xiao S, Xiao H, Guo B. Unsharp masking image enhancement the parallel algorithm based on cross-platform. *Scientific Reports*. 2022;12(1): 20175. https://doi.org/10.1038/s41598-022-21745-9.

[16] Venkateswarlu IB, Kakarla J, Prakash S. Face mask detection using MobileNet and global pooling block. In: *IEEE 4th Conference on Information Communication Technology (CICT)*. 03rd–05th December 2020; Chennai, India. IEEE; 2020. p.1–5. https://doi.org/10.1109/CICT51604.2020.9312083.

[17] Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*. 2020;53: 5455–5516.

[18] Shafiq M, Gu Z. Deep residual learning for image recognition: A survey. *Applied Sciences*. 2022;12(18): 8972. https://doi.org/10.3390/app12188972.

[19] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 18th–23rd June 2018; Salt Lake City, UT, USA. IEEE; 2018. p.7132–7141.

[20] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* [Preprint] 2017. https://doi.org/10.48550/arXiv.1704.04861.

[21] Ubaidah IDWS, Fu'Adah Y, Sa'Idah S, Magdalena R, Wiratama AB, Simanjuntak RBJ. Classification of glaucoma in fundus images using convolutional neural network with mobilenet architecture. In: *1st International Conference on Information System Information Technology (ICISIT)*. 27th–28th July 2022; Yogyakarta, Indonesia. IEEE; 2022. p.198–203. https://doi.org/10.1109/ICISIT54091.2022.9872945.

[22] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *ICML15: Proceedings of the 32nd International Conference on International Conference on Machine Learning*. 06th–11th July 2015; Lille, France. ICML; 2015. p.448–456.

[23] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 18th–22nd June 2018; Salt Lake City, Utah, USA. IEEE; 2018. p.4510–4520.

[24] National Institute of Health Data Science at Peking University (NIHDS-PKU). *ODIR-2019 Challenge*. https://odir2019.grand-challenge.org/dataset/ [Accessed 1st May 2024].

[25] Larxel. *Ocular disease recognition*. https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k [Accessed 2nd May 2024]

[26] Mahmood SS, Chaabouni S, Fakhfakh A. A new technique for cataract eye disease diagnosis in deep learning. *Periodicals of Engineering and Natural Sciences*. 2023;11(6): 14–26.