

Redefining Cataract Diagnosis: Leveraging MobileNetV2-Based CNN for Automated Detection from Fundus Images

A. Sri Nagesh*, Popuri Keerthika¹, Raguthu Dheeraj¹, Vetcha Sai Manikanta Saketh¹

¹Department of Computer Science and Engineering, RVR & JC College of Engineering, Guntur, A.P, India.

Article Type:

Research Article

Received on

18/03/2024

Revised on

25/04/2024

Accepted on

29/05/2024

Published on

30/06/2024

*Corresponding author email:

asrinagesh@gmail.com

ISSN:

DOI Number (Article)

<https://doi.org/10.55306/CJDTEs.2024.1104>

Article Citation:

A. Sri Nagesh, Popuri Keerthika, Raguthu Dheeraj, Vetcha Sai Manikanta Saketh, **Redefining Cataract Diagnosis: Leveraging MobileNetV2-Based CNN for Automated Detection from Fundus Images** Ci-STEM Journal of Digital Technologies and Expert Systems, 1(1), 2024, pp34-45

Copyright: © 2024 by the authors.

Licensee Ci-STEM Global Services Foundation, Kolkata, India.

This article is an open-access article. Distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

ABSTRACT

Cataracts remain a prevalent cause of global blindness, emphasizing the urgent need for timely and accurate diagnostic solutions. This study introduces an innovative automated cataract detection method using fundus images, employing three distinct convolutional neural network architectures: MobileNetV2, EfficientNetB0, and ResNet50. Through transfer learning, these types of models were developed using a dataset containing both cataract and non-cataract fundus images. The outcomes show how successful the suggested models were in precisely classifying fundus images, with notable performance indicators including F1-score, recall, accuracy, and precision. The MobileNetV2-based model achieved 98% accuracy, EfficientNetB0 attained 68% accuracy, while the ResNet50 model, enhanced with additional dense layers, attained 69% accuracy. Furthermore, comprehensive evaluation including classification reports and confusion matrices validates the robustness and generalization capabilities of the models, confirmed through cross-validation on independent test sets. This convolutional neural network-based approach holds promise for scalable, cost-effective automated cataract detection in clinical settings, with the potential for further advancements in model interpretability and dataset diversification to enhance its applicability in diverse populations.

Keywords:

Cataracts, blindness, diagnostic solutions, fundus images, convolutional neural network, confusion matrices, detection.

1. INTRODUCTION

Cataracts represent a significant global health concern, contributing substantially to blindness worldwide. Detecting cataracts early is crucial for effective intervention and prevention of vision loss. This study introduces an innovative technique for automated cataract detection utilizing fundus images, employing advanced intense learning models: MobileNetV2, EfficientNetB0, and ResNet50. Accurate and effective diagnosis methods are desperately needed as the frequency of cataracts rises in order to expedite treatment and enhance patient outcomes.

Cataracts, marked by a clouding of the lens of the eye, obstruct the passage of light to the retina, leading to impaired vision. While cataracts typically progress slowly and may not initially cause significant vision problems, they can eventually result in blindness if left untreated, particularly in individuals over 40 years old. Early detection of cataracts is essential for timely intervention and reducing the risk of irreversible vision loss, underscoring the importance of automated screening methods.

Recent advancements in artificial intelligence have opened up new possibilities for automating cataract detection, offering a promising solution for scalable and cost-effective screening programs. Fundus images, which provide detailed views of the eye's interior, are particularly well-suited for this purpose due to their accessibility and diagnostic value. However, existing

automated detection systems often face challenges such as suboptimal accuracy and high computational complexity, highlighting the need for improved detection algorithms.

In this work, we provide a novel method for cataract identification based on deep-learning models designed especially for fundus image analysis. By leveraging the capabilities of MobileNetV2, EfficientNetB0, and ResNet50, we aim to enhance detection accuracy while minimizing computational overhead. Through rigorous investigation and evaluation, the effectiveness of proposed models have demonstrated effectiveness in accurately classifying fundus images and identifying cataracts with high precision and recall.

Our contributions include the construction of sophisticated deep-learning algorithms for cataract detection and the establishment of an extensive dataset of fundus images. By harnessing the power of innovative technology, we seek to enhance cataract screening's effectiveness and accessibility, eventually leading to better global outcomes for eye health.

The subsequent sections of this paper provide detailed descriptions of our suggested deep learning architectures and an overview of relevant field research, describe the experimental setup and methodology, present the experimental results and analysis, and conclude with consequences and future research directions. Through our efforts, we aim to address the critical need for accurate and efficient diagnostic solutions for cataracts, ultimately enhancing patient care and reducing the global burden of vision impairment.

2. LITERATURE REVIEW:

V Harini, V Bhanumathi's [4] study proposes a cost-effective auxiliary diagnosis system leveraging fundus image analysis for grading and classification of cataracts.

Preprocessing techniques, such as contrast improvement and image resampling, and elimination of noise, are employed to enhance image quality and facilitate dataset manipulation. A mean filter is utilized for Gaussian noise reduction, considering the high-quality nature of fundus images captured by fundus cameras.

Wavelet decomposition, chosen for its simultaneous time and frequency analysis capabilities, decomposes fundus images into high and low-frequency components, enabling better distinction between vessels and background. The Haar wavelet transform is selected for its ease of implementation, particularly in distinguishing high-frequency components indicative of vessel structures.

In order to distinguish edges, clever edge detection is used after preprocessing, and then it is enlarged for better visibility. Higher levels of cataract severity are suggested by fewer components in the edge-detected image. The number of connected components in the image serves as an indicator of cataract severity.

The study presents a comprehensive methodology for automated grading and classification of cataracts, integrating preprocessing, splinter decomposition, and edge detection techniques. MATLAB is employed for implementation, with various cataract grades tested to evaluate system performance.

Linglin Zhang et al. [5] endeavour to explore the effectiveness and efficiency of employing Deep Convolutional Neural Nets (DCNNs) for automated cataract classification and identification. Additionally, it visualizes feature maps available at the pool5 layer, offering insights into the semantic meaning of extracted features, thereby explaining the depiction of features by DCNNs.

Many population-based clinical retinal fundus image datasets, including up to 5620 pictures, are used to thoroughly validate the suggested DCNN-based classification approach. The study comes to two important conclusions: firstly, the method effectively overcomes challenges posed by local uneven illumination and eye reflections, improving the accuracy of DCNN classification. Secondly, as the sample size increases, the DCNN classification accuracies improve and exhibit more stable fluctuations. Notably, the proposed method achieves high accuracies of 93.52% and 86.69% for cataract detection and grading tasks, respectively, outperforming contemporary state-

of-the-art approaches. This underscores the significance of addressing the challenges associated with cataract detection and grading, particularly in leveraging advanced techniques like DCNNs to improve accuracy and efficiency in automated systems.

Juyel Rana, Syed Md. Galib [6] study proposes a methodology centered on developing a smartphone application for self-screening cataract detection. The lack of ophthalmologists and specialized equipment in underdeveloped nations like Bangladesh makes it difficult to diagnose and cure cataracts promptly. To address these issues, researchers have explored automated smartphone-based cataract detection techniques, offering a convenient and accessible solution for early screening.

Leveraging the ubiquity of smartphones with well-focused front cameras, the application allows individuals to perform self-assessment anywhere, anytime. By utilizing advanced features of Android Studio, SDK, NDK, and OpenCV, the application facilitates real-time image processing and analysis.

The proposed model involves launching the application, which activates the front camera and initiates face and eye detection using OpenCV. Subsequently, the application identifies and isolates the pupils using the Cascade Classifier, storing the pupil images for further analysis. A pseudo-code algorithm outlines the cataract detection process, involving color determination of the pupils and classification into various stages based on color matching with an existing database. Smartphone-based solutions offer promising avenues for early intervention and improved healthcare access, potentially revolutionizing cataract screening practices.

Meimei Yang et al. [2] study suggest using a neural network classifier based on retinal image classification for automated cataract detection, aiming to enhance diagnostic efficiency and alleviate the physical and economic burdens on patients and society.

Previous studies have highlighted the significance of preprocessing, classifier construction and feature extraction in developing effective automated detection systems. Preprocessing techniques are essential for enhancing image quality and reducing noise, thus improving subsequent analysis accuracy. The proposed methodology incorporates an enhanced trilateral filter to reduce noise and an upgraded Top-bottom hat transformation to boost image contrast.

Feature extraction plays a pivotal role in capturing relevant information for classification. Luminescence, representing image clarity, and gray co-occurrence matrices are used to extract 40 features, such as entropy, contrast, and homogeneity measurements. Additionally, gray-gradient co-occurrence matrices provide edge information, enabling the computation of features such as gradient dominance and energy.

The classification process employs a backpropagation (BP) neural network with two layers, consisting of 40 input neurons representing extracted features and 4 output neurons corresponding to cataract severity levels. The network is trained using a subset of retinal images, with training data accounting for 70% and validation and test data each representing 15% of the dataset. Training utilizes the conjugate gradient descent method, with validation data preventing overfitting.

The proposed approach offers a systematic framework for automated cataract detection, leveraging advanced computational techniques to analyze retinal images efficiently. By automating the classification process, this methodology promises to enhance diagnostic accuracy, reducing reliance on manual assessments, and improving overall healthcare outcomes.

A. B. Jagadale et al. [7] Paper proposes a computer-aided system utilizing slit lamp images to detect cataracts at an early stage. Lens detection, segmentation, feature extraction, and categorization are all part of the process. Minimizing human error and inter and intra-grade variability is one of the main concerns. The three primary types of cataracts are nuclear, cortical, and post-subcapsular.

Traditionally, ophthalmologists rely on slit lamp observation aided by the lens Opacity Classification System (LOCS-III) is a tool used to diagnose and detect cataracts. One popular

treatment option for cataract-related vision impairment is lens replacement surgery. Previous studies emphasize the efficacy of computer-aided discovery and grading, with accurate lens localization being critical for successful diagnosis.

Lens localization is achieved using the Hough circle detection transform, leveraging the circular nature of the lens. Measures of statistical significance like mean, contrast, and energy are extracted to characterize the lens, and a support vector machine (SVM) is employed for categorization.

The suggested method makes use of a database of slit lamp photos that it obtained from an eye facility, with lens detection accuracy validated using the Hough circle detection transform. Statistical feature extraction ensures robustness to variations in illumination sources. Features that have been retrieved from photos and classified into various degrees of cataract severity are used to train the SVM.

Overall, this methodology emphasizes the importance of accurate lens localization and the role of advanced computational techniques, such as SVM, in improving cataract detection efficiency.

While traditional methods such as slit lamp observation with retro-illumination images are effective, they require expert interpretation and may be subject to variability.

Previous research has identified distinctive characteristics in retro-illumination images of cataracts, wherein opacities manifest as dark regions within the lens. Cortical cataracts typically exhibit spoke-like structures, originating from the equator and extending towards the central lens region, while posterior subcapsular (PSC) cataracts appear as patches closer to the central part. These images are captured in two modes, anterior and posterior, with the former offering sharper details, particularly for cortical cataracts.

However, challenges arise due to uneven illumination, artifacts, and noise, complicating traditional image processing techniques such as thresholding and edge detection. To address these issues, researchers propose leveraging texture analysis, as cataract regions often exhibit richer texture compared to non-cataract regions. This paper uses local entropy to quantify texture, indicating the image's local roughness.

To enhance the effectiveness of texture-based detection, X. Gao et al. [1] introduced enhanced texture measurement. This involves weighting the texture based on the inverse intensity of the image, thereby amplifying darker regions indicative of cataract opacities. By partitioning the lens image into central and outer regions, differentiating between cortical and PSC cataracts becomes feasible, as their locations and severity levels vary.

In feature extraction, many statistics are computed to characterize the texture and improve texture measurements. These statistics include mean, standard deviation, skewness, and kurtosis. Linear discriminant analysis (LDA) is then employed to train a classifier using these features. This proposed methodology demonstrates promising results, getting an accuracy of 84.8% on a clinical database comprising a substantial number of image pairs.

Overall, this method highlights the potential of CAD systems in cataract detection, offering advantages in scalability, consistency, and efficiency. By utilizing cutting-edge methods for image processing and machine learning algorithms, such systems show promise for both mass screening initiatives and assisting clinicians in grading cataracts accurately.

In their pioneering research, Gao et al. [3] introduced a groundbreaking approach to automatically grade nuclear cataracts using slit-lamp images. Slit-lamp imaging is a cornerstone diagnostic tool in ophthalmology, renowned for its ability to provide high-resolution images showing the anterior part of the eye, including the lens. These images offer clinicians a detailed understanding of the structural changes associated with cataracts, facilitating accurate diagnosis and monitoring of disease progression. In their study, Gao et al. harness the potential of slit-lamp images as a rich source of data for cataract grading. Unlike traditional grading methods that often rely on manual assessment or predefined image features, their system integrates deep learning algorithms to autonomously learn discriminative features directly from these images. By using a variety of

datasets of slit-lamp photos to train recursive neural networks (RNN) and convolutional neural networks (CNN), the proposed system effectively captures subtle variations indicative of cataract severity. By using this novel method, cataract grading becomes more efficient and there are new opportunities to use modern imaging modalities in ocular diagnosis. The successful integration of slit-lamp imaging with deep learning techniques underscores the transformative potential of interdisciplinary research in advancing clinical practices and improving patient care in ophthalmology.

3. PROPOSED MODELS

The proposed architecture for cataract detection makes use of MobileNetV2, a convolutional neural network (CNN) architecture that is effective and lightweight and designed for mobile and embedded applications. MobileNetV2 is recognized for its depth-wise separable convolutions and linear bottlenecks, which enable it to uncover a balance between computational and classification performance. By utilizing MobileNetV2 as the backbone of the model, we make use of its capacity to minimize computational overhead and extract significant information from input photos, making it well-suited for analyzing fundus images for cataract discovery.

MobileNetV2 is a deep learning architecture designed to facilitate efficient inference on mobile and embedded devices. It is a successor to the original Mobile Net, developed by Google researchers. MobileNetV2 builds upon the principles of its predecessor, aiming to achieve high accuracy while minimizing computational cost and model size.

One key feature of MobileNetV2 is its use of inverted residual blocks. These blocks consist of a lightweight bottleneck layer followed by an expansion layer and a linear projection layer. This design helps to reduce the number of parameters and computational complexity while maintaining expressive power in the network. The inverted residual blocks enable MobileNetV2 to achieve better performance compared to traditional CNNs with similar model sizes.

Another important aspect of MobileNetV2 is its emphasis on improving the efficiency of depth-wise separable convolutions. Depth-wise separable convolutions split the conventional convolution process into distinct pointwise and depth-wise convolutions, significantly reducing the computational cost. MobileNetV2 introduces additional optimizations to this scheme, such as using a linear bottleneck between the depth-wise and pointwise convolutions, further enhancing efficiency without compromising accuracy.

Furthermore, MobileNetV2 incorporates techniques like batch normalization, ReLU activations, and global average pooling to improve training stability and performance. These elements enhance the network architecture's overall efficacy and allow it to accomplish remarkable outcomes on a range of tasks, such as semantic segmentation, object identification, and picture classification. MobileNetV2 has been widely adopted in the ML (Machine Learning) community due to its versatility and efficiency. It is especially well-suited for implementation on resource-constrained devices like edge computing platforms, mobile phones, and Internet of Things devices due to its lightweight design.

By providing a balance between accuracy and computational efficiency, MobileNetV2 continues to be a valuable tool for developers seeking to deploy deep learning models in real-world applications.

In this architecture, MobileNetV2 serves as the feature extractor, processing fundus images to extract relevant features indicative of cataracts. The initial layers of MobileNetV2 are responsible for capturing low-level elements like textures and edges, gradually progressing to higher-level features representing more abstract patterns present in the input images. These features are crucial for distinguishing between fundus images with and without cataracts.

To tailor MobileNetV2 for the cataract detection task, additional layers are introduced after the feature extraction stage. These layers are made up of convolutional processes designed to improve the model's capability to distinguish between fundus pictures with and without cataracts by further

honing the retrieved characteristics. By incorporating these additional convolutional layers, the model gains the capacity to capture subtle variations and intricate spatial patterns specific to cataracts, thereby improving its classification performance.

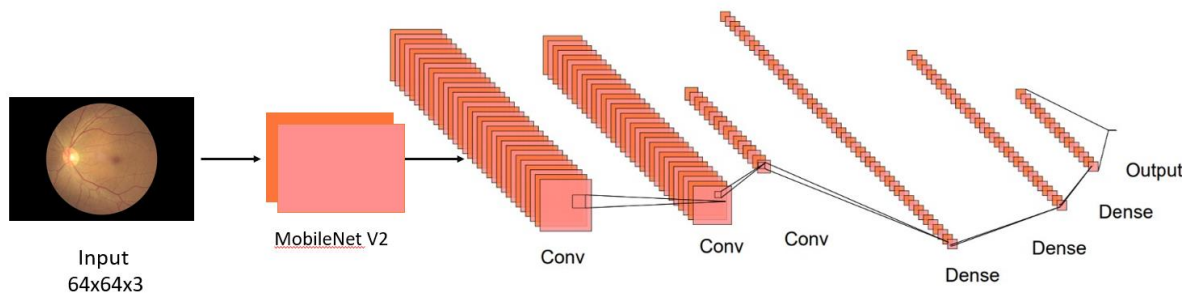


Figure 1. MobileNet V2 based CNN Architecture

In our proposed architecture, dropout regularization is essential for improving the model's resilience and capacity for generalization. Following the first dense layer with 128 neurons, a dropout rate of 30% is applied, meaning that during each training iteration, 30% of the neurons in this layer are randomly deactivated. This strategic dropout percentage is meticulously selected through empirical experimentation to strike a delicate balance between preventing overfitting and preserving essential information required for accurate classification.

Furthermore, after the subsequent dense layers comprising 64 and 32 neurons, dropout percentages of 20% and 30% are applied, respectively. Through repeated experimentation targeted at refining the model's performance on the validation set, these dropout rates are likewise experimentally calculated.

By systematically adjusting these dropout percentages, we aim to mitigate the probability of overfitting while ensuring that the model retains valuable features and patterns crucial for effective cataract classification.

One of the key benefits of MobileNetV2-based architecture is its adaptability to varying input image sizes. Fundus images, obtained through different imaging devices or techniques, often come in different resolutions and aspect ratios. To accommodate this variability, the model accepts input images of fixed dimensions, typically resized to a square shape for uniformity. By resizing the input images to a predetermined size, such as 64x64 pixels, the model can process fundus images of diverse resolutions without sacrificing performance. This flexibility enhances the model's versatility and applicability in real-world scenarios, where fundus images may exhibit varying resolutions and aspect ratios.

After the convolutional layers, global average pooling is applied to aggregate spatial information across the entire feature map. This pooling operation reduces the dimensionality of the features extracted while preserving essential information relevant to cataract detection. Subsequently, the final classification is carried out by introducing completely linked dense layers with dropout regularization. The model may learn intricate correlations between the retrieved characteristics and the existence or absence of cataracts in the input pictures thanks to these layers. Dropout regularization promotes improved generalization to unknown data by randomly deactivating neurons during training, preventing overfitting.

The model's output layer, which consists of a single neuron with a sigmoid activation function, generates a binary classification output that indicates the probability that the input picture has a cataract. The model is optimized with the Adam optimizer and binary cross-entropy loss during training, iteratively adjusting its parameters to minimize the discrepancy between predicted and

ground truth cataract labels. Fine-tuning may be performed on the base MobileNetV2 layers during training to adapt this model to the specific characteristics of the cataract detection task.

Table 1 Description of different layers of the MobileNetV2-based model

Layer	Filters	Config	Stride	Output
mobilenetv2_1.00_224	-	-	-	(2, 2, 1280)
conv2d_82	64	3×3, relu	(1,1)	(2,2,64)
conv2d_83	64	3×3, relu	(1,1)	(2,2,64)
conv2d_84	64	3×3, relu	(1,1)	(2,2,64)
conv2d_85	64	3×3, relu	(1,1)	(2,2,64)
global_average_pooling2	-	-	-	(32,)
flatten_9	-	-	-	(32,)
dense_24	128	relu	-	(128,)
dropout_17	-	-	-	(128,)
dense_25	64	relu	-	(64,)
dropout_18	-	-	-	(64,)
dense_26	32	relu	-	(32,)
dense_19	-	-	-	(32,)
dense_27	1	sigmoid	-	(1,)

Table 2. Description of different layers of the ResNet50 based model

Layer	Filters	Config	Stride	Output
ResNet50(base model)	N / A	N / A	N / A	(64, 64, 3)
Conv2D	64	(3, 3) ReLU	Same	(64, 64, 64)
Dropout	N / A	p = 0.3	N/A	(64, 64, 64)
Conv2D	32	(3, 3) ReLU	Same	(64, 64, 32)
(64, 64, 32)	N / A	p = 0.2	N / A	(64, 64, 32)
GlobalAveragePooling2D	N / A	N / A	N / A	(32,)
Flatten	N / A	N / A	N / A	(32,)
Dense	128	ReLU, L2(0.01)	N / A	(128,)
Dense	64	ReLU, L2(0.01)	N / A	(64,)
Dense	1	Sigmoid	N / A	(1,)

In this architecture, the Base model ResNet50 is followed by two convolutional layers, each with a dropout layer to avoid overfitting. To preserve the input's spatial dimensions, the convolutional layers have 64 and 32 filters, respectively, along with padding and ReLU activation. Dropout is implemented with 0.3 and 0.2 probability, respectively, following the first and second convolutional layers. After reducing the output's spatial dimensions using global average pooling and flattening, densely linked layers with L2 regularization and ReLU activation are used. The output of the last dense layer is a binary classification using sigmoid activation.

Table 3 Description of different layers of the EfficientB0 based model

Layer	Filters	Configuration	Output Shape
Efficient NetB0 (base)	-	-	(64, 64, 3)
Conv2D	64	(3 * 3) ReLU, Same	(64, 64, 64)
GlobalAveragePooling2D	-	-	(64,)
Flatten	-	-	(64,)
Dense	128	ReLU, L2(0.01)	(128,)
Dropout	-	p = 0.3	(128,)
Dense	64	ReLU, L2(0.01)	(64,)
Dropout	-	p = 0.2	(64 ,)
Dense	1	Sigmoid	(1,)

Another design that has been suggested is an altered form of EfficientNetB0, which begins with an input shape of (64, 64, and 3). After applying ReLU activation and 64 3x3 filters as the first convolutional layer, global average pooling is used to compress spatial data into a (64,) shape. After that, the output is flattened, becoming a 64-size one-dimensional array. Two tightly linked layers are then used to lessen overfitting. The first layer has 128 neurons, ReLU activation, and L2 regularization with a coefficient of 0.01; it is followed by dropout with a likelihood of 0.3. The second dense layer is composed of 64 neurons that exhibit L2 regularization and ReLU activation. A second dropout layer with a likelihood of 0.2 follows.

4. IMPLEMENTATION DETAILS

Dataset

Our dataset consists of 860 fundus images from Ocular Disease Recognition and Kaggle datasets for our dataset. Three subsets are created from these photos: 10% (86 images) are used for validation, 10% (6688 images) are used for training, and the remaining 10% (86 images) are used for testing. Fundus pictures, taken with a specialized fundus camera, show the rear eye part, including important components such as the macula, optic disc, retina, and blood vessels. The model gains the ability to recognize features and patterns in fundus pictures throughout the training phase by learning from the training data. To avoid overfitting and guarantee that the model performs well on previously unknown data, the weights of the model are normalized during the validation step. Finally, during testing, the model is evaluated to assess its accuracy and loss, crucial metrics for gauging its performance in diagnosing ocular diseases.

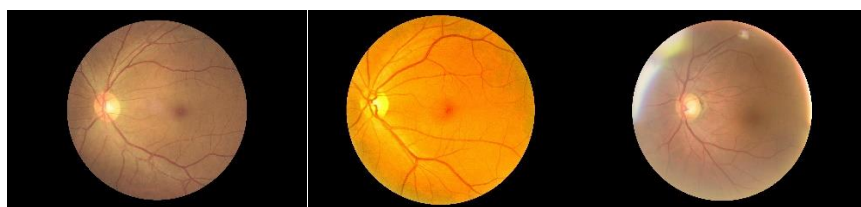


Figure 2 Sample Images in the dataset

In our experiment, we used Google Colab and had a strong computer setup with a Tesla T4 GPU, an Intel Core i9-10850K CPU, and 64GB of RAM. We built a custom CNN model using TensorFlow and Keras, which we adjusted with a fixed learning rate of 0.0001 using the Adam optimizer. We trained the model for 55 epochs, each time processing small groups of 32 images.

This helped our model learn better and improve its performance over time. After training, our model didn't just give simple predictions; it provided probabilities for different categories, which can be helpful for doctors in making decisions about patient care. We have also developed a user-friendly website to assist individuals in uploading images and obtaining predictions using FastAPI. FastAPI is a powerful Python framework for creating APIs. Its asynchronous capabilities and automatic documentation generation make it a popular choice for rapid development and scalability. With intuitive design and seamless functionality, users can easily upload their images and receive accurate predictions. This platform intends to give a convenient and accessible solution for image analysis, enhancing user experience and facilitating informed decision-making.

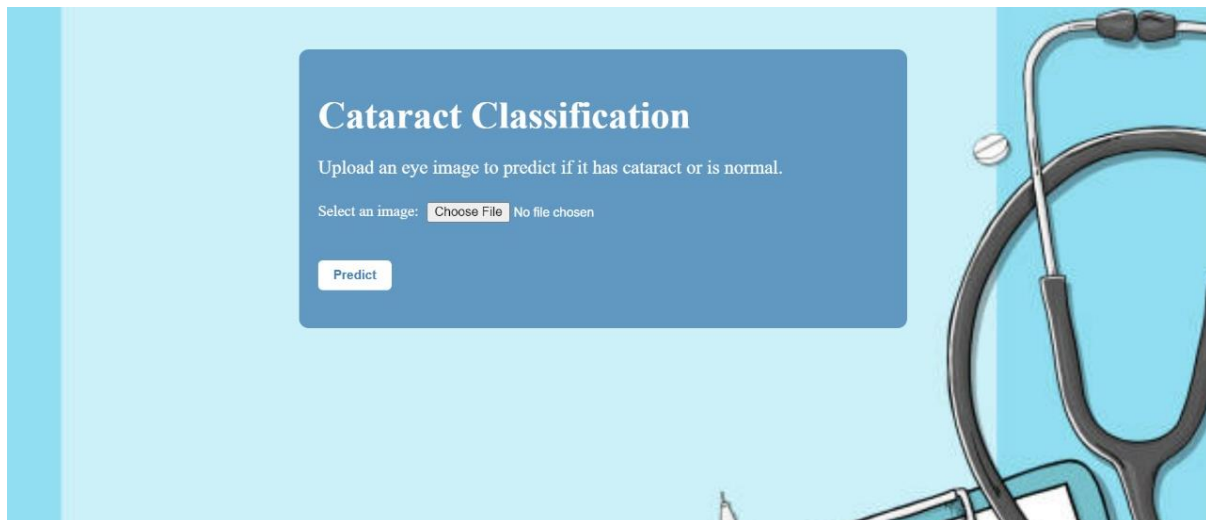


Figure 3 Picture of website

5. EXPERIMENTAL RESULTS

The experimentation results offer intricate insights into the performance of every implemented model, shedding light on their strengths and weaknesses in the circumstances of cataract detection. Beginning with the MobileNetV2-based model, it demonstrated exceptional accuracy, achieving a commendable training accuracy of 98.11% and a validation accuracy of 74.71%. This high accuracy indicates that the MobileNetV2 architecture effectively captured and learned discriminative features relevant to cataract classification. The MobileNetV2 model showcases the promising potential for robust cataract detection, benefiting from its efficient architecture and pre-trained weights.



Figure 4. Accuracy of the MobileNetv2-based Model

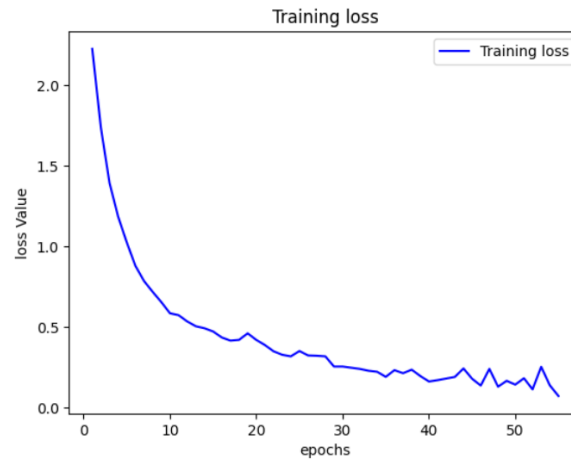


Figure 5 Loss of MobileNetv2-based Model

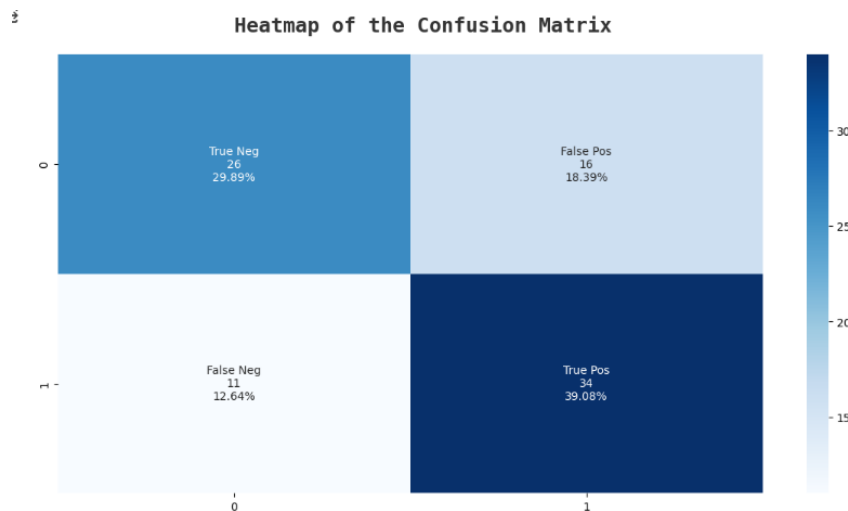


Figure 6 Confusion matrix of MobileNetv2 based model

Conversely, the ResNet50-based model exhibited a slightly lower training accuracy of 69% and displayed a validation accuracy of 59%. This model's performance suggests a better generalization capability compared to MobileNetV2, as indicated by the closer alignment between training and validation accuracies. The ResNet50 architecture, renowned for its deep residual connections, likely facilitated more effective feature extraction and representation, contributing to its improved generalization performance. Furthermore, the ResNet50 model's training loss of 0.1493 underscores its ability to minimize classification errors, leading to robust performance in cataract detection tasks.

In contrast, the EfficientNetB0-based model delivered the least satisfactory results among the three architectures, showcasing a training accuracy of 58% and a validation accuracy of 57%. This considerable performance gap between training and validation suggests significant overfitting, potentially attributable to the model's complexity relative to the dataset size. Despite its lower accuracy, the EfficientNetB0 model's training loss of 0.5681 implies a substantial effort to minimize classification errors during training. Nonetheless, the model's limited generalization capability, as evidenced by its poor performance on unseen validation data, highlights the importance of carefully balancing model complexity with dataset characteristics to achieve optimal performance in cataract detection applications.

6. CONCLUSION AND FUTURE WORK

So, we looked at the effectiveness of three different CNN designs for detecting cataracts in

pictures of the retinal fundus. However, MobileNetV2, which achieved remarkable training accuracy, appeared as the most promising design.

ResNet50, while demonstrating balanced performance and better generalization, achieved a lower overall accuracy compared to MobileNetV2. EfficientNetB0, on the other hand, showcased the pitfalls of using a complex model on a limited dataset, suffering from significant overfitting.

These findings suggest that MobileNetV2's efficient architecture and pre-trained weights hold significant promise for cataract detection. Future work should focus on mitigating overfitting in the MobileNetV2 model through techniques like dropout regularization or data augmentation. Hyperparameter optimization can further refine the model to achieve optimal performance. Additionally, incorporating a more diverse dataset that encompasses a wider range of image characteristics is crucial for enhancing the model's generalizability. By addressing these limitations, MobileNetV2 can be established as a robust and generalizable solution for cataract detection in real-world settings, potentially aiding in early diagnosis and improved patient care.

REFERENCES

- [1]. X. Gao, H. Li, J. H. Lim, and T. Y. Wong, "Computer-aided cataract detection using enhanced texture features on retro-illumination lens images," in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 1565–1568.
- [2]. M. Yang, J.-J. Yang, Q. Zhang, Y. Niu, and J. Li, "Classification of retinal image for automatic cataract detection," in Proc. IEEE 15th Int. Conf. e-Health Netw., Appl. Services (Healthcom), Oct. 2013, pp. 674–679.
- [3]. X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," IEEE Trans. Biomed. Eng., vol. 62, no. 11, pp. 2693–2701, Nov. 2015.
- [4]. V. Harini and V. Bhanumathi, "Automatic cataract classification system," in Proc. Int. Conf. Commun. Signal Process. (ICCSP), Apr. 2016, pp. 0815–0819.
- [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 Proc. IEEE 14th Int. Conf. Netw., Sens. Control (ICNSC), May 2017, pp. 60–65.
- [6]. J. Rana and S. M. Galib, "Cataract detection using a smartphone," in Proc. 3rd Int. Conf. Electr. Inf. Commun. Technol. (EICT), Dec. 2017, pp. 1–4.
- [7]. A.B. Jagadale, S. S. Sonavane, and D. V. Jadav, "Computer-aided system for early detection of nuclear cataract using circle Hough transform," in Proc. 3rd Int. Conf. Trends Electron. Information. (ICOEI), Apr. 2019, pp. 1009–1012.
- [8]. M. S. Junayed, A. A. Jeny, S. T. Atik, N. Neehal, A. Karim, S. Azam, and B. Shanmugam, "AcneNet—A deep CNN based classification approach for acne classes," in Proc. 12th Int. Conf. Inf. Commun. Technol. Syst. (ICTS), Jul. 2019, pp. 203–208.
- [9]. D. Kim, T. J. Jun, Y. Eom, C. Kim, and D. Kim, "Tournament based ranking CNN for the cataract grading," in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2019, pp. 1630–1636.
- [10]. T. Pratap and P. Kokil, "Computer-aided diagnosis of cataract using deep transfer learning," Biomed. Signal Process. Control, vol. 53, Aug. 2019, Art. no. 101533.
- [11]. R. Sigit, E. Triyana, and M. Rochmad, "Cataract detection using single layer perceptron based on a smartphone," in Proc. 3rd Int. Conf. Information. Comput. Sci. (ICICoS), Oct. 2019, pp. 1–6.

- [12]. X. Zhang, J. Lv, H. Zheng, and Y. Sang, "Attention-based multi-model ensemble for automatic cataract detection in B-scan eye ultrasound images," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–10.
- [13]. Kaggle. (2021). Ocular Disease Recognition. Accessed: Feb. 11, 2021.
- [14]. T. Pratap and P. Kokil, "Efficient network selection for computer-aided cataract diagnosis under noisy environment," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105927.
- [15]. M. S. M. Khan, M. Ahmed, R. Z. Rasel, and M. M. Khan, "Cataract detection using a convolutional neural network with VGG-19 model," in Proc. IEEE World AI IoT Congr. (AIIoT), May 2021, pp. 0209–0212.