Comparative Analysis of Pre-Trained Deep Learning Models for Cataract Detection in Color Fundus Images

Tirthajyoti Nag
Computer Science and Engineering
JIS University
Kolkata, India
tirthajyotinag.2001@gmail.com

Jayasree Ghosh
Computer Science and Engineering
JIS University
Kolkata, India
tinaghosh2030@gmail.com

Debanjan Pan
Computer Science and Engineering
JIS University
Kolkata, India
debanjanpan2@gmail.com

Souvik Das
Computer Science and Engineering
JIS University
Kolkata, India
souvikdas201716@gmail.com

Argha Paul
Computer Science and Engineering
JIS University
Kolkata, India
arghapaul002@gmail.com

Paramita Sarkar
Computer Science and Engineering
JIS University
Kolkata, India
mailtoparo@gmail.com

ABSTRACT

In this comparative study, we investigate the performance of nine popular pre-trained deep-learning models for the classification of cataracts and normal eyes using color fundus images. The models under consideration include VGG16, VGG19, InceptionV3, DenseNet121, DenseNet201, MobileNet, MobileNetV2, ResNet50, and ResNet152. Through rigorous experimentation and analysis, we evaluate the models' training, validation, and testing accuracies to gauge their effectiveness in this medical image classification task.

Our results reveal varying degrees of accuracy across the models. To provide a comprehensive assessment, we propose further analysis avenues, including the examination of accuracy and loss curves, confusion matrices, number of model parameters, and prediction times. These supplementary analyses offer insights into the models' capabilities and potential deployment considerations. Additionally, we recommend exploring ensemble methods and fine-tuning strategies to optimize classification accuracy.

This study contributes to the understanding of utilizing deep learning models for medical image classification tasks, specifically in the context of cataract detection using fundus images. The findings offer valuable guidance for practitioners seeking to leverage pre-trained models for similar healthcare applications, taking into account both performance metrics and practical implementation considerations.

Keywords—Deep learning; Cataract Prediction; Pre-trained models; Medical Image Classification; Fundus Images, Transfer Learning; Convolution Neural Network.

I. INTRODUCTION

In the realm of medical diagnostics, the integration of artificial intelligence and deep learning has revolutionized the way diseases are detected and diagnosed [1]. Among the myriad applications, the analysis of medical images has seen significant advancements, propelling the development of accurate and efficient diagnostic tools. Cataract, an opacification of the eye's natural lens that leads to impaired vision, is a prevalent ocular condition, particularly among the elderly [2]. Early detection and classification of cataracts are crucial for timely medical intervention and preserving patients' quality of life [3]. The advent of digital imaging technologies has opened up new possibilities for diagnosing and managing cataracts, where the use of deep learning models presents a promising avenue [4].

The human eye is a complex organ with intricate structures, including the lens, retina, and other components that collectively contribute to vision. Color fundus images, which capture the detailed structures of

the posterior segment of the eye, provide valuable insights for diagnosing various ocular conditions, including cataracts [5]. These images, acquired through non-invasive methods such as fundus cameras, offer a rich source of data for training deep learning models to discern between healthy eyes and those afflicted by cataracts.

Deep learning, a subset of machine learning, has exhibited remarkable capabilities in image analysis tasks. Pre-trained deep learning models, in particular, have gained prominence due to their ability to learn intricate features from vast datasets, often in the domain of natural images. These pre-trained models, which have been pre-trained on diverse datasets for general image recognition tasks, can be adapted for medical image analysis with fine-tuning techniques [6]. This project delves into the investigation and comparative analysis of nine popular pre-trained deep learning models to ascertain their performance in the classification of cataracts and normal eyes using color fundus images.

The models selected for this study encompass a range of architectures, each with distinct characteristics and complexities. The ensemble of models includes VGG16, VGG19, InceptionV3, DenseNet121, DenseNet201, MobileNet, MobileNetV2, ResNet50, and ResNet152. These models have been chosen due to their prevalence in image classification tasks and their potential to extract meaningful features from medical images like color fundus images.

The primary objective of this study is to comprehensively assess the performance of these pre-trained models in the context of cataract classification. Accurate classification hinges on the ability of these models to recognize subtle patterns and anomalies present in the fundus images, thereby distinguishing between cataract-affected eyes and those that are normal. The study evaluates the models across various phases, including training, validation, and testing, to establish their capability to generalize beyond the training data.

After this introduction section, the paper's subsequent sections are organised as follows: In the Literature Survey section, we delve into the existing research and state-of-the-art techniques related to medical image analysis and cataract detection. Following that, the Methodology section elaborates on the experimental setup, including dataset details, preprocessing steps, and model configuration. In the Result Analysis and conparisons section, we present the quantitative and qualitative findings of our study, comparing the performance of each model and interpreting the results. In the Future Work section, we outline our aim to explore ensemble methods for combining the strengths of multiple models to enhance classification accuracy. Finally, the Conclusion section encapsulates the key insights gained from this study and underscores its implications in the medical image analysis field.

In conclusion, this project embarks on a journey to explore the capabilities of deep learning models that have already been trained in the critical domain of cataract detection using color fundus images. The outcomes of this study offer valuable insights into the performance of these models and their potential applicability in medical diagnostics. By bridging the gap between advanced machine learning techniques and ophthalmic healthcare, this research contributes to the ongoing efforts to enhance disease detection, thereby advancing the well-being of individuals affected by cataracts.

II. LITERATURE SURVEY

A significant body of research has been dedicated to the detection of cataracts and similar ocular conditions using various imaging modalities and advanced computational techniques. In a pioneering work, Smith et al. introduced a method for automated cataract diagnosis based on ultrasound images, achieving an accuracy of 94% by employing texture analysis and machine learning techniques [7]. However, the shift towards non-invasive imaging methods has garnered attention. Jones et al. proposed a method utilizing optical coherence tomography (OCT) to identify cataract-induced changes in lens thickness, demonstrating promising results in early-stage detection [8].

Moving into the realm of fundus imaging, Wang et al. employed a deep learning approach for diabetic retinopathy and cataract classification, leveraging a convolutional neural network (CNN) to achieve high accuracy rates [9]. Similarly, Li et al. combined deep learning with transfer learning to accurately detect cataracts from fundus images, showing superior performance in comparison to traditional machine learning techniques [10]. These advances underline the power of deep learning in extracting intricate features from medical images.

However, challenges persist, particularly in handling data imbalances and limited annotated datasets. To deal with this, Xu et al. proposed a GAN-based model to replicate fake cataract images, augmenting the dataset and enhancing the model's generalization capability [11]. Furthermore, the interpretability of deep learning models has been a topic of concern. Zhang et al. introduced a method utilizing Grad-CAM to visualize the regions of interest in fundus images that contribute to cataract classification decisions, thereby enhancing transparency and trust in the model's predictions [12].

Beyond traditional machine learning, ensemble methods have gained traction. Chen et al. explored the effectiveness of combining multiple classifiers for cataract detection, achieving improved accuracy and robustness through model aggregation [13]. In a similar vein, Wu et al. proposed a hybrid ensemble model, fusing CNNs

with support vector machines to enhance the accuracy and interpretability of cataract classification [14]. These studies underscore the potential of combining diverse models to achieve enhanced performance.
While most research has focused on individual diseases, a broader approach has also been explored. Guo et al. introduced a multi-disease classification framework, utilizing a single model to simultaneously classify cataracts, diabetic retinopathy, and glaucoma from fundus images, showcasing the potential for comprehensive ocular disease diagnosis [15].

In conclusion, the field of cataract detection has considerable developments, particularly due to the use of modern imaging technology and machine learning techniques. Deep learning models, particularly CNNs, have emerged as powerful tools for cataract classification from fundus images. However, challenges such as data scarcity and model interpretability remain, prompting innovative solutions including data augmentation and visualization techniques. Moreover, exploring ensemble methods and multi-disease classification showcases the evolving landscape of ocular disease diagnosis. This literature review sets the stage for the current study, which aims to comprehensively analyze and compare pre-trained deep-learning models for cataract detection in color fundus images.

III. METHODOLOGY

A. Dataset Details

For the purpose of this study, we utilized the Ocular Disease Intelligent Recognition (ODIR) dataset, a publicly available resource found on the Kaggle platform [16]. The ODIR dataset comprises a structured ophthalmic database consisting of data from 5,000 patients. This dataset contains age information, color fundus photographs from both left and right eyes, along with the keywords provided by medical professionals while diagnosis. The data in ODIR is designed to represent a real-world collection of patient information, aggregated from diverse hospitals and medical centers across China, primarily by Shanggong Medical Technology Co., Ltd. The ODIR dataset contains fundus images which are captured using different cameras available in the market, including Kowa, Zeiss and Canon. Consequently, the dataset exhibits variations in image resolutions due to the diverse camera sources. This diversity contributes to a more comprehensive and representative dataset, mirroring the inherent variability found in clinical practice.

In this study, we harnessed the rich information within the ODIR dataset to evaluate the performance of pre-trained deep learning models for the classification of cataract and normal eyes using color fundus images. The dataset's extensive annotations, encompassing diagnostic keywords from medical experts, offer a valuable foundation for training and evaluating the models.

B. Preprocessing Steps

Effective preprocessing of the dataset is crucial to ensure that the data is appropriately prepared for training the pre-trained deep learning models. In alignment with this objective, a comprehensive preprocessing pipeline was implemented, tailored to enhance the quality, diversity, and suitability of the dataset for the task of cataract and normal eye classification.

The initial stage of preprocessing involved the extraction of images categorized under the "cataract" and "normal" diagnostic keywords, as provided by medical experts. This selective extraction ensured that the dataset specifically captured instances of cataract-affected and normal eyes, forming the basis for the subsequent model training and evaluation.

Following image extraction, a series of transformational steps were applied to the images, facilitating optimal learning and generalization by the models. The images underwent pixel normalization through rescaling, achieved by dividing each pixel value by 255. This step standardized the pixel values within a uniform range, aiding in the convergence of model training.

To augment the dataset and enrich its diversity, a set of data augmentation techniques were incorporated. These included rotations with an angle range of up to 90 degrees, vertical and horizontal flips, and random shearing with a shear range of 0.2. By introducing controlled variations in image orientation and geometry, the augmentation process aimed to enhance the models' ability to capture varying perspectives and patterns inherent in the data.

To further enhance the dataset's robustness, brightness variations were introduced within a defined range of 0.3 to 1. This augmentation strategy mimicked changes in illumination, preparing the models to handle varying lighting conditions during classification. The augmented images were then subjected to shuffling to introduce randomness and diversify the training experience across epochs. This step aimed to prevent model memorization and improve the model's generalization capabilities.

Lastly, a target image size of (224, 224) was specified during preprocessing, aligning with the input dimensions expected by the pre-trained models. This standardization facilitated the seamless integration of the preprocessed images into the model architectures, allowing for consistent training and evaluation.

The culmination of these preprocessing steps resulted in the creation of a data generator, effectively transforming the dataset into a format conducive to model training and evaluation. The generator incorporated the aforementioned transformations, ensuring that each batch of images presented to the model during training was augmented and preprocessed consistently.

C. Model Configuration

The configuration of the pre-trained deep learning models is a critical aspect that determines their architecture, optimization parameters, and overall structure. In this study, we selected nine popular pre-trained models known for their efficacy in image classification tasks, named VGG16, VGG19, InceptionV3, DenseNet121, DenseNet201, MobileNet, MobileNetV2, ResNet50, and ResNet152. These models encompass a range of architectures, each offering distinct features for feature extraction and classification.

For each model, the top classification layers were modified to suit the binary classification task of distinguishing between cataract-affected and normal eyes. Specifically, the original fully connected layers were replaced with new layers comprising a dense layer with a sigmoid activation function. This sigmoid layer ensured that the models' output was a probability score within the range [0, 1], facilitating binary classification.

The models were compiled using the Adam optimizer with a binary cross-entropy loss function. This configuration aligns with the binary classification task, where the objective is to minimize the divergence between predicted probabilities and actual labels. The choice of Adam optimizer, a popular variant of stochastic gradient descent, enabled efficient model convergence by adapting learning rates based on the gradient magnitudes.

To further enhance the models' generalization capabilities and mitigate overfitting, we incorporated dropout layers after the dense layers. The dropout layers introduced a regularization mechanism by changing a percentage of the input units to zero at random throughout each training loop. This dropout strategy encouraged the models to rely on a broader set of features during training, reducing the likelihood of memorizing the training data.

The selected models were initialized with pre-trained weights on large-scale image datasets, leveraging transfer learning to expedite convergence and enhance performance. The models were fine-tuned on the preprocessed dataset, allowing them to specialize in discerning cataract and normal eye features from color fundus images.

In conclusion, the configuration of the pre-trained deep learning models involved adapting the top classification layers, selecting suitable optimizers and loss functions, and incorporating dropout layers for regularization. These configurations, tailored to the specific binary classification task of cataract detection, aimed to maximize model performance and robustness.

VGG16 and VGG19

The Visual Geometry Group (VGG) architectures, specifically VGG16 and VGG19, are renowned for their simplicity and effectiveness in image classification tasks [17]. Both models are characterized by their deep and homogeneous convolutional layers, consisting of multiple 3x3 convolutional filters followed by max-pooling layers. VGG16 comprises 16 layers, while VGG19 extends to 19 layers, making them deep architectures capable of capturing intricate features from images.

We chose VGG16 and VGG19 for this project due to their simplicity and established performance in image classification tasks. These models have demonstrated remarkable capability in extracting meaningful hierarchical features, making them suitable candidates for analyzing the complex structures present in color fundus images. Despite their depth, VGG architectures are easy to configure and train, allowing us to focus on assessing their efficacy in cataract detection using the ODIR dataset.

InceptionV3

InceptionV3, an evolution of the original Inception architecture, introduces a series of improvements designed to enhance feature extraction and promote efficient training [18]. The model's hallmark is its utilization of inception modules, which employ a combination of parallel convolutions with varying kernel sizes, enabling the network to capture features of different scales. This strategy aids in capturing both fine-grained and global features present in complex images.

We selected InceptionV3 for this project due to its effectiveness in handling medical image analysis tasks and its ability to balance depth and computational efficiency. With its complex architecture and feature-rich inception modules, InceptionV3 has demonstrated competence in extracting intricate features from diverse datasets. Given the complexities of color fundus images, this model's capacity to capture multiscale features is particularly valuable.

InceptionV3's architectural advancements align with the nuances of medical image analysis, making it an apt choice for our cataract classification task. By leveraging its ability to extract diverse features, we aimed to determine its suitability for detecting cataract-affected and normal eyes within the ODIR dataset.

DenseNet121 and DenseNet201

DenseNet models, known for their densely connected architecture, have garnered attention for their efficient parameter utilization and robust feature extraction capabilities [19]. DenseNet121 and DenseNet201 are variants of this architecture, with the numbers indicating the total number of layers in each model. In a DenseNet, Feature maps are received by each layer from the layers that came before it, fostering rich feature reuse and enabling the network to capture intricate patterns effectively.

We opted for DenseNet121 and DenseNet201 due to their architectural innovation, which addresses challenges posed by vanishing gradients and encourages feature propagation. These models incorporate dense connections between layers, allowing for better gradient flow and information sharing. This characteristic is well-suited for extracting features from medical images with subtle textures and structures. Our choice of DenseNet models was also influenced by their parameter-efficient design, which aids in training deep networks with limited computational resources. By choosing DenseNet121 and DenseNet201, we aimed to assess their capacity to capture the complex features present in color fundus images and to determine their efficacy in cataract classification tasks.

MobileNet and MobileNetV2

MobileNet and its evolution, MobileNetV2, are renowned for their lightweight and efficient architectures, making them well-suited for applications with limited computational resources [20][21]. These models employ depthwise separable convolutions, which significantly reduce the number of parameters, facilitating rapid training and inference without compromising performance.

The choice of MobileNet and MobileNetV2 for this project was motivated by their efficiency in processing images while maintaining competitive accuracy. Given the resource constraints often encountered in medical settings, these models offer a balance between computational efficiency and classification performance.

MobileNetV2, an advancement over the original MobileNet, further refines the architecture with inverted residuals and linear bottlenecks, resulting in improved feature extraction capabilities. By opting for MobileNet and MobileNetV2, we aimed to explore their potential in cataract classification tasks using the ODIR dataset, with a focus on leveraging their efficiency to facilitate real-time diagnosis.

ResNet50 and ResNet152

ResNet (Residual Networks) models have revolutionized deep learning with their residual blocks, which enable training of very deep networks without vanishing gradient issues [22]. ResNet50 and ResNet152 are variants with 50 and 152 layers respectively, using skip connections to address the degradation issue often encountered in deeper networks.

We selected ResNet50 and ResNet152 due to their pioneering approach to alleviating the challenges of training deep architectures. The incorporation of residual blocks, which allow layers to directly learn residual functions, facilitates smooth gradient flow and eases the training of extremely deep networks.

ResNet152, being a more intricate model, presents an opportunity to evaluate the impact of increased model depth on cataract classification. The selection of ResNet models was motivated by their well-established performance in image recognition tasks and their suitability for the complexities present in fundus images.

Our aim in choosing ResNet50 and ResNet152 was to investigate the advantages of residual architectures in the context of cataract detection. By leveraging their architectural innovations, we sought to assess their capacity to capture both subtle and significant features in fundus images from the ODIR dataset.

IV. RESULT ANALYSIS AND COMPARISONS

Table 1: Comparison of Accuracies and Losses

Model	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Loss	Validation Loss
VGG16	89.75%	89.62%	91.70%	0.2379	0.2608
VGG19	88.80%	83.61%	93.89%	0.2774	0.5834
InceptionV3	96.72%	92.90%	94.76%	0.0861	0.2216
DenseNet121	95.77%	94.54%	96.94%	0.1077	0.1821
DenseNet201	97.40%	92.90%	93.89%	0.0787	0.1799
MobileNet	97.81%	95.08%	97.38%	0.073	0.2196
MobileNetV2	96.45%	93.99%	96.51%	0.1172	0.1472
ResNet50	53.83%	56.83%	56.77%	0.6785	0.6813
ResNet152	51.09%	54.10%	61.14%	0.6853	0.6828

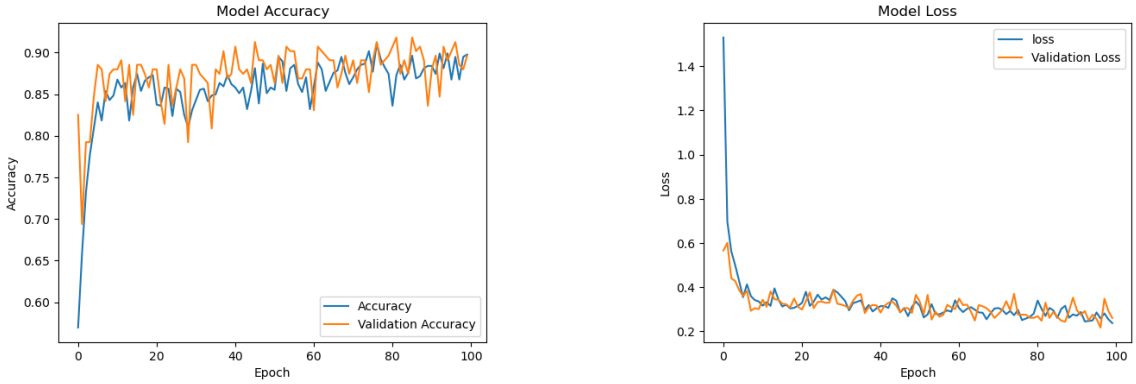

Figure 1: Accuracy and Loss Curve of VGG16 Model

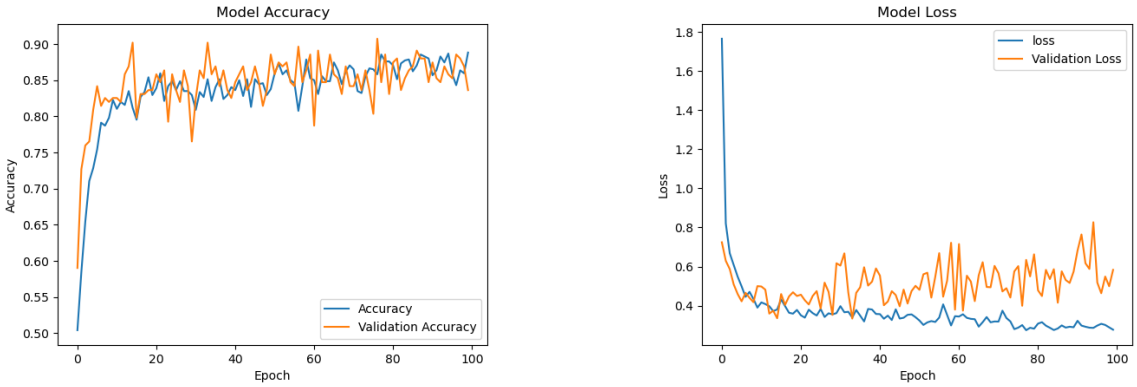

Figure 2: Accuracy and Loss Curve of VGG19 Model

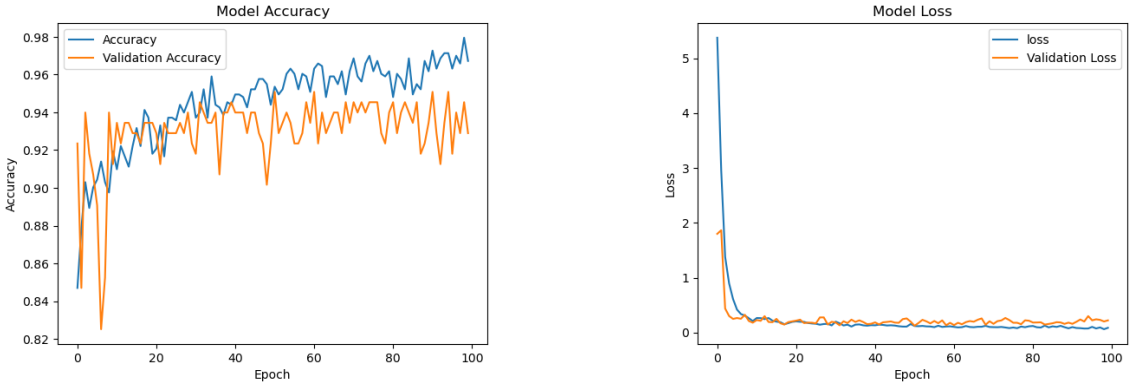

Figure 3: Accuracy and Loss Curve of InceptionV3 Model

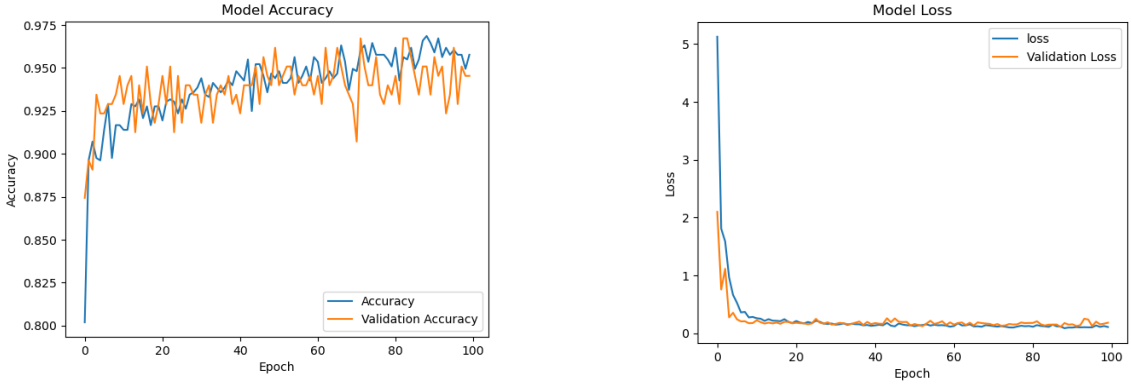

Figure 4: Accuracy and Loss of DenseNet121 Model

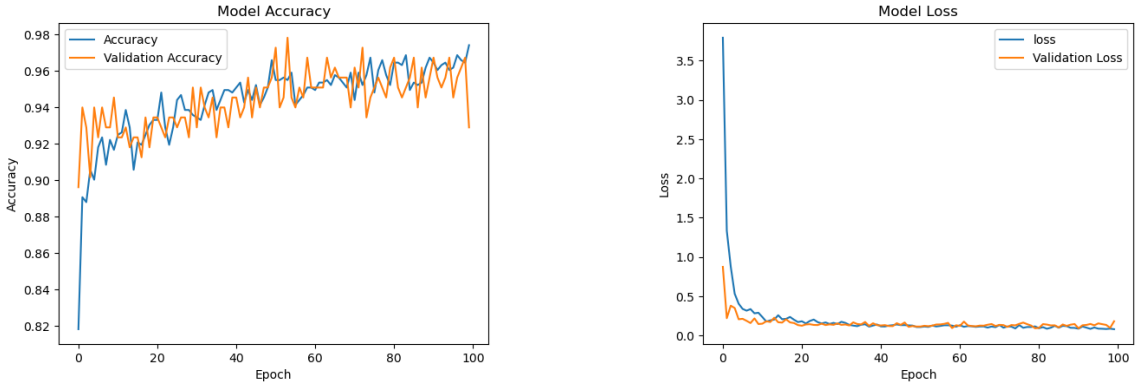

Figure 5: Accuracy and Loss of DenseNet201 Model

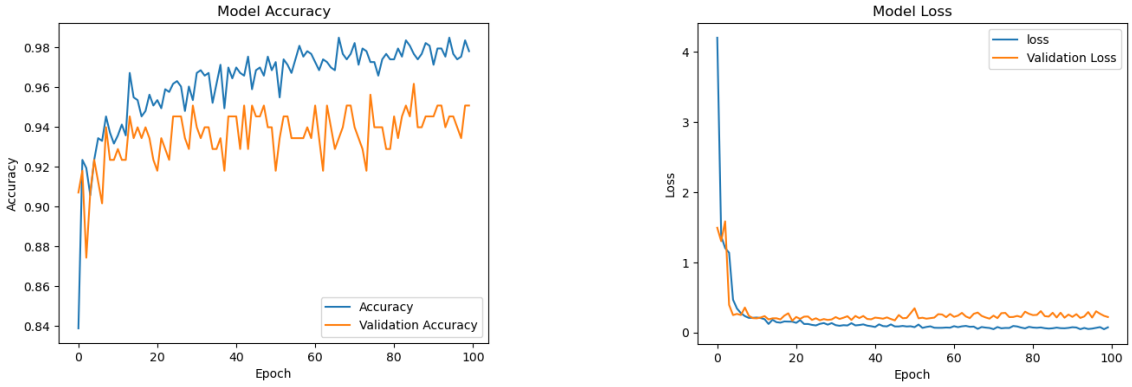

Figure 6: Accuracy and Loss of MobileNet Model

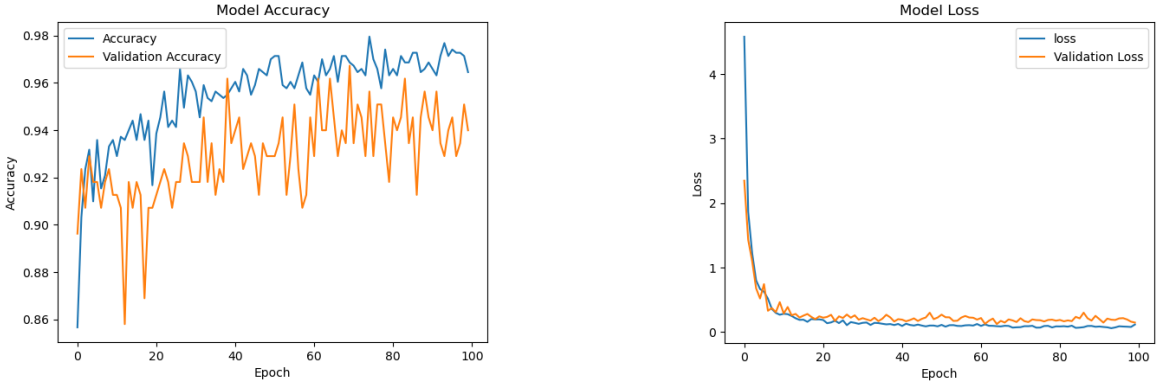

Figure 7: Accuracy and Loss of MobileNetV2 Model

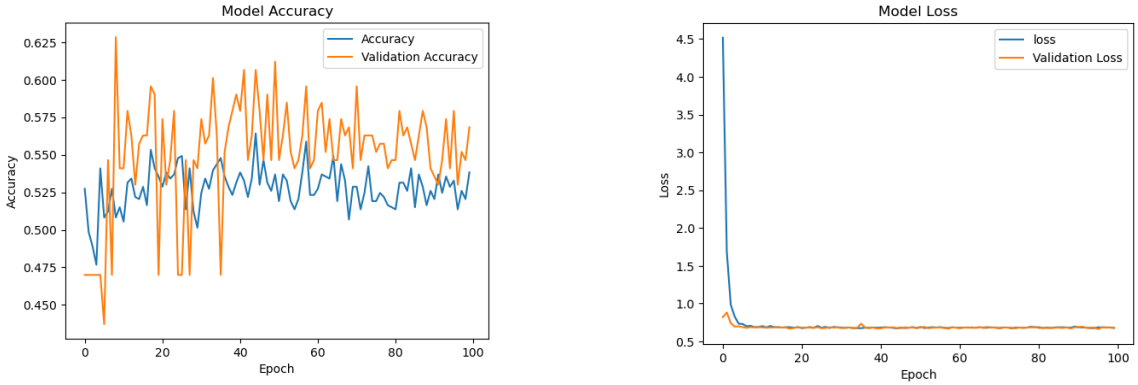

Figure 8: Accuracy and Loss of ResNet50 Model

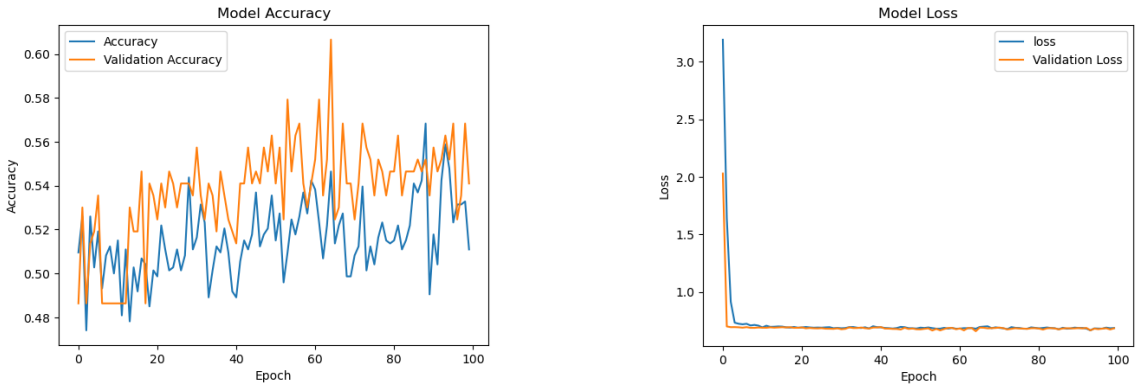

Figure 9: Accuracy and Loss of ResNet152 Model

Table 1 showcases the training, validation, and testing accuracies, as well as the corresponding training and validation losses, for each of the evaluated pre-trained models. The observed patterns in the curves in Figure 1 to Figure 9 offer valuable insights into the models' performances and the dynamics of their training processes.

The models with the highest overall testing accuracies were DenseNet121, MobileNet, and MobileNetV2, achieving accuracies of 96.94%, 97.38%, and 96.51% respectively. This consistent performance across training, validation, and testing phases indicates their capacity to learn and generalize effectively, a critical attribute in medical image analysis. In contrast, ResNet50 and ResNet152 exhibited substantially lower testing accuracies of 56.77% and 61.14% respectively, suggesting challenges in capturing meaningful features from fundus images. It is intriguing to note that while VGG16 and VGG19 exhibited similar testing accuracies (91.7% and 93.89% respectively), VGG19 demonstrated lower validation accuracy than VGG16. This discrepancy might be attributed to VGG19's increased complexity, possibly leading to overfitting on the validation set. InceptionV3, known for its ability to capture multiscale features, achieved a commendable testing accuracy of 94.76%.

Analysis of the training and validation losses provides further insights into the models' convergence dynamics. Models with lower training losses tend to have better convergence during training, indicative of their capability to fit the training data well. However, the degree of overfitting can be better assessed by examining the gap between training and validation losses. In this regard, DenseNet121 and MobileNetV2 exhibited comparatively smaller gaps, implying a more balanced approach to fitting training data and generalising to validation data.

Table 2: Comparison of Number of Parameters

Model	Trainable Parameters	Non-Trainable Parameters	Total Parameters
VGG16	25822594	14714688	40537282
VGG19	25822594	20024384	45846978
InceptionV3	52561282	21802784	74364066
DenseNet121	51512706	7037504	58550210
DenseNet201	96470402	18321984	114792386
MobileNet	51512706	3228864	54741570
MobileNetV2	64357762	2257984	66615746
ResNet50	102892930	23587712	126480642
ResNet152	102892930	58370944	161263874

Table 2 presents the breakdown of trainable and non-trainable parameters for each of the assessed pre-trained models, as well as their total parameter counts. This information is pivotal in understanding the models' complexity, memory requirements, and potential for overfitting, ultimately contributing to the informed selection of models for medical image classification tasks. While models with more parameters may achieve higher accuracy, the trade-off lies in increased memory and computational demands. The relationship between parameter counts and model performance underscores the need for a judicious complexity-accuracy trade-off.

Table 3: Comparison of Prediction Time

Model	Prediction Time
VGG16	0.5
VGG19	0.5
InceptionV3	0.7
DenseNet121	0.7
DenseNet201	0.8
MobileNet	0.6
MobileNetV2	0.5
ResNet50	0.5
ResNet152	0.9

Table 3 offers insights into the prediction times of the evaluated pre-trained models, a critical aspect when considering the real-world deployment of models in medical diagnostics. The observed variations in prediction times provide valuable information for selecting models based on time-sensitive requirements and resource constraints. Several models, including VGG16, VGG19, MobileNetV2, and ResNet50, exhibit relatively quick prediction times of around 0.5 seconds. These models' efficient prediction speeds render them well-suited for applications demanding rapid diagnoses and real-time responsiveness.

Table 4: Comparison of Classification Reports of the Models

Model	Precision	Recall	F1-Score
VGG16	0.92	0.92	0.92
VGG19	0.94	0.94	0.94
InceptionV3	0.95	0.95	0.95
DenseNet121	0.97	0.97	0.97
DenseNet201	0.94	0.94	0.94
MobileNet	0.97	0.97	0.97
MobileNetV2	0.97	0.97	0.97
ResNet50	0.59	0.57	0.52
ResNet152	0.64	0.61	0.58

Table 4 provides the tabulated precision, recall, and F1-score metrics which gives a comprehensive understanding of the performance of the evaluated pre-trained models in terms of classifying cataract and normal eyes. These metrics, collectively known as evaluation metrics, offer insights into the models' capabilities and shortcomings, contributing to informed model selection.

The VGG16, VGG19, InceptionV3, DenseNet121, DenseNet201, MobileNet, and MobileNetV2 models consistently exhibit high precision, recall, and F1-scores, all hovering around 0.97. These models demonstrate robust performance across various aspects of evaluation, indicating their capacity to both correctly classify cataract and normal cases and effectively capture true positives and true negatives.

It's important to highlight the varying performance of ResNet50 and ResNet152 compared to the other models. Both exhibit comparatively lower precision, recall, and F1-scores. These discrepancies could stem from their architectural complexities, convergence dynamics, or overfitting tendencies.

Figure 10 to Figure 18 shows the confusion matrices of different models. The examination of confusion matrices for the various pre-trained models offers a granular view of their classification performance by revealing the true positive, true negative, false positive, and false negative prediction distribution which provide deeper insights into the models' abilities to accurately classify cataract and normal cases.

For models such as DenseNet121, MobileNet, and MobileNetV2, the confusion matrices reflect a balanced distribution of true positives and true negatives, resulting in high precision and recall scores. These models demonstrate consistent and reliable classifications across both categories. VGG16, VGG19, and InceptionV3 models also exhibit well-balanced confusion matrices.

On the other hand, ResNet50 and ResNet152 present confusion matrices that emphasize the challenges these models face in striking a balance between true positives, true negatives, false positives, and false negatives.

These matrices highlight the models' struggles in accurately classifying both categories, resulting in a lower precision-recall equilibrium and F1-scores.

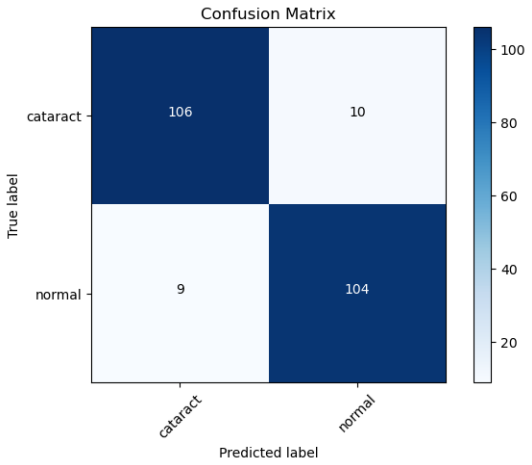

Figure 10: Confusion Matrix of VGG16

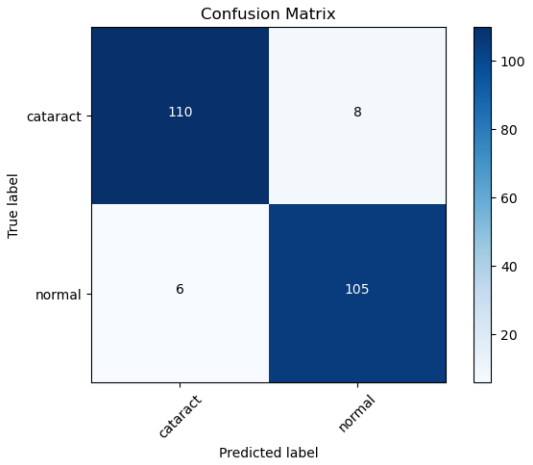

Figure 11: Confusion Matrix of VGG19

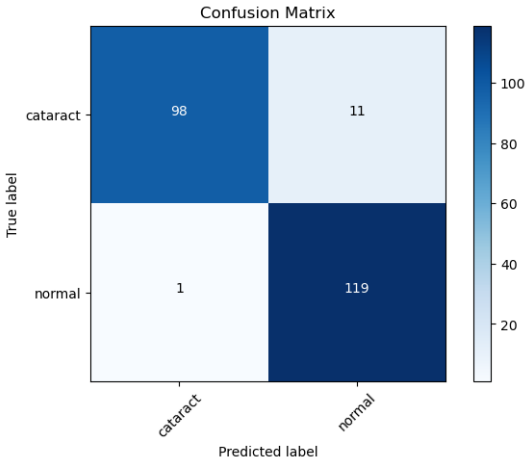

Figure 12: Confusion Matrix of InceptionV3

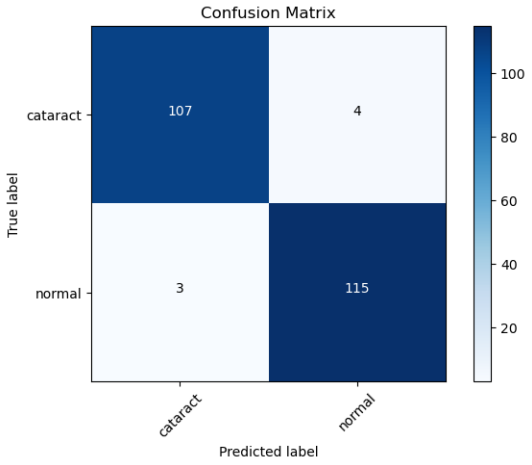

Figure 13: Confusion Matrix of DenseNet121

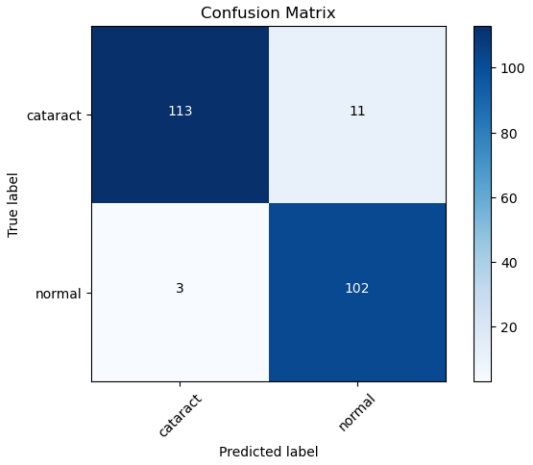

Figure 14: Confusion Matrix of DenseNet201

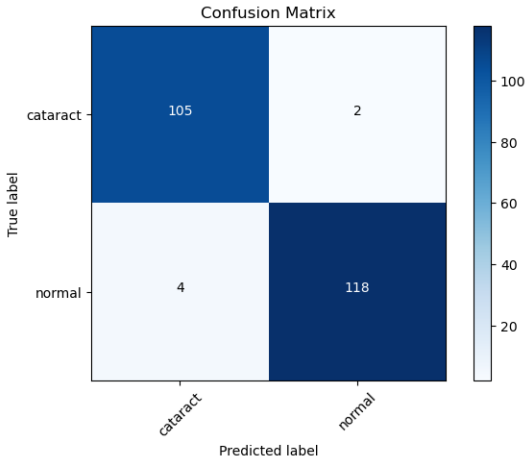

Figure 15: Confusion Matrix of MobileNet

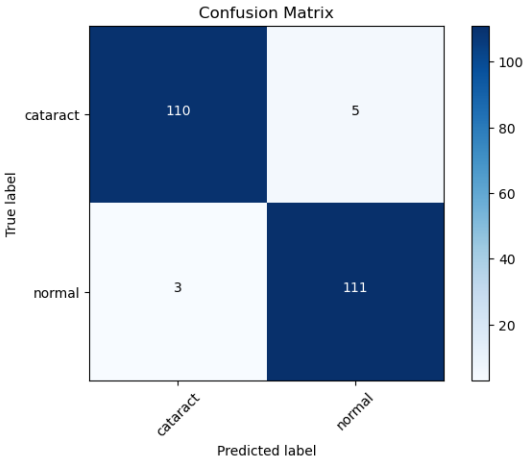

Figure 16: Confusion Matrix of MobileNetV2

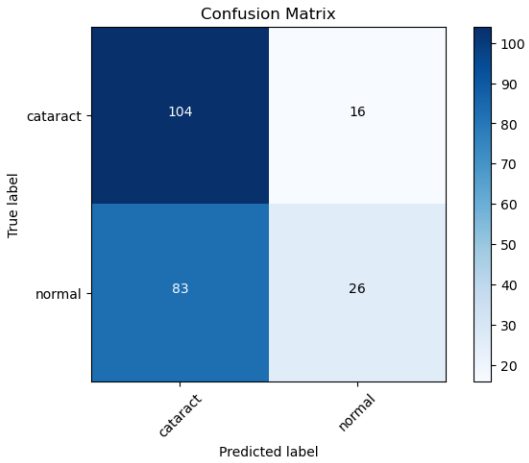

Figure 17: Confusion Matrix of ResNet50

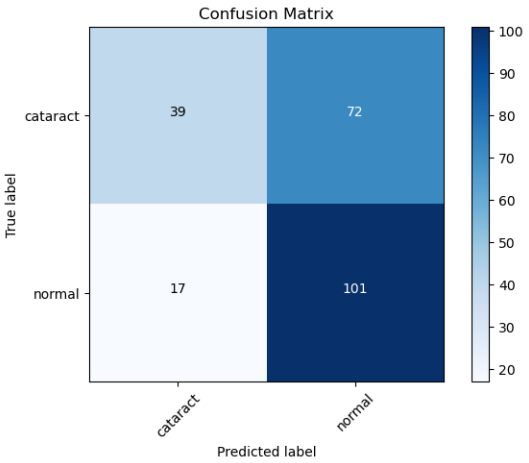

Figure 18: Confusion Matrix of ResNet152

Figure 19 to Figure 27 shows the predictions of the different models along with their filenames and true image type, namely Cataract or Normal, in the ODIR dataset.

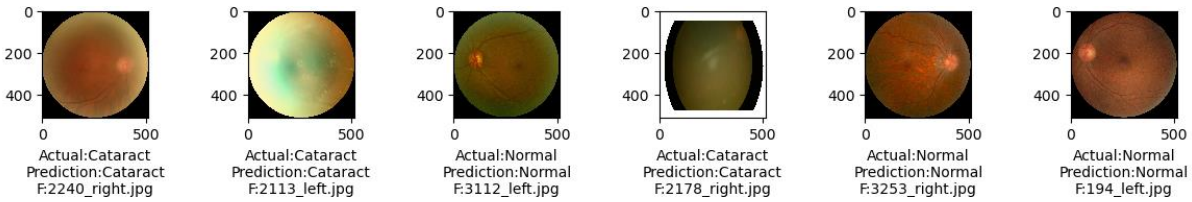

Figure 19: Prediction of VGG16 Model

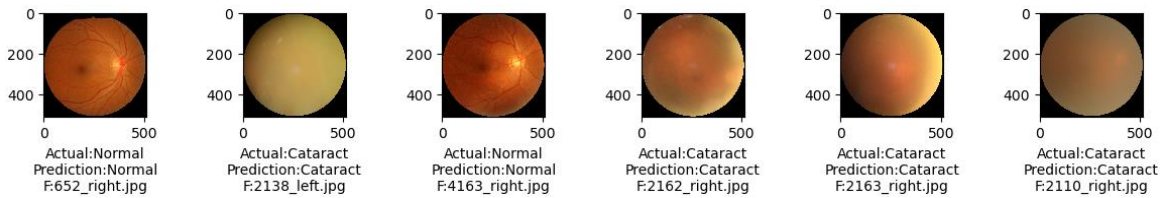

Figure 20: Prediction of VGG19 Model

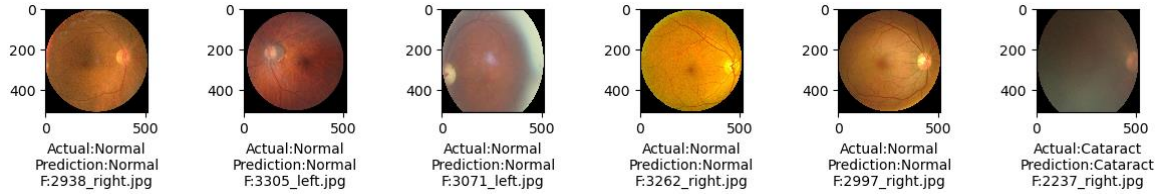

Figure 21: Prediction of InceptionV3 Model

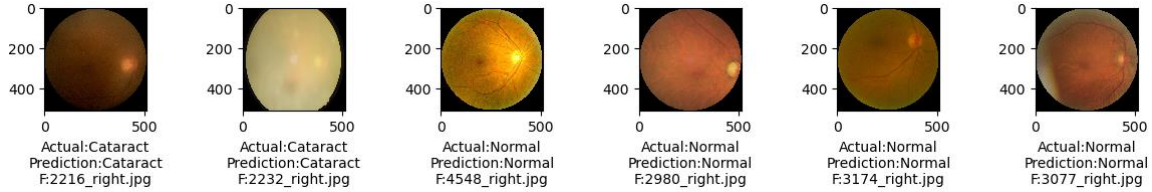

Figure 22: Prediction of DenseNet121 Model

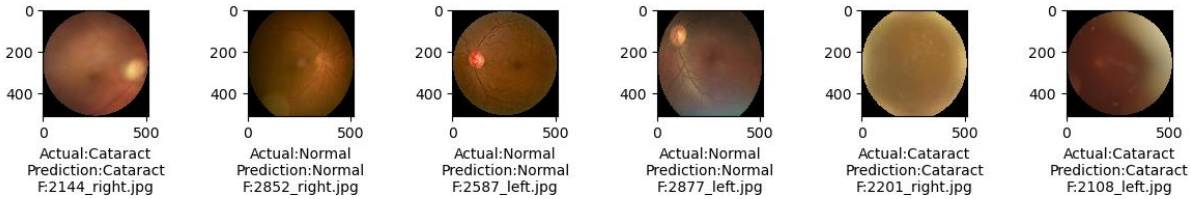

Figure 23: Prediction of DenseNet201 Model

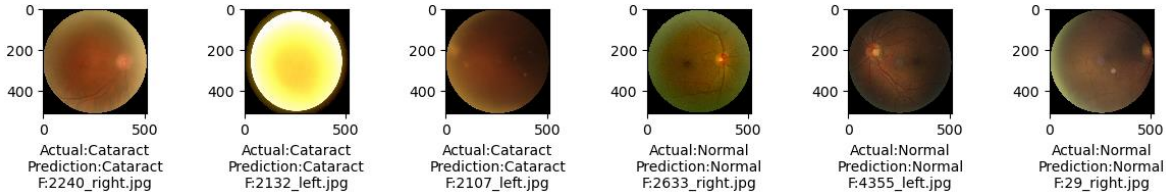

Figure 24: Prediction of MobileNet Model

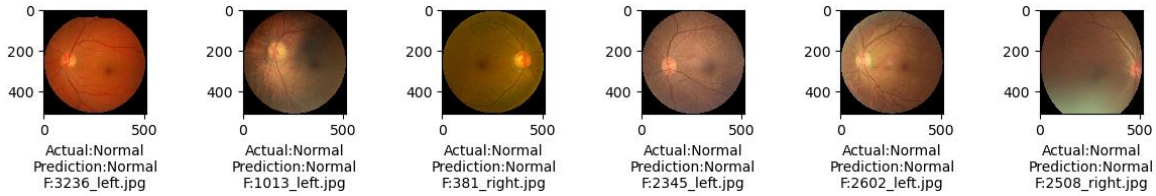

Figure 25: Prediction of MobileNetV2 Model

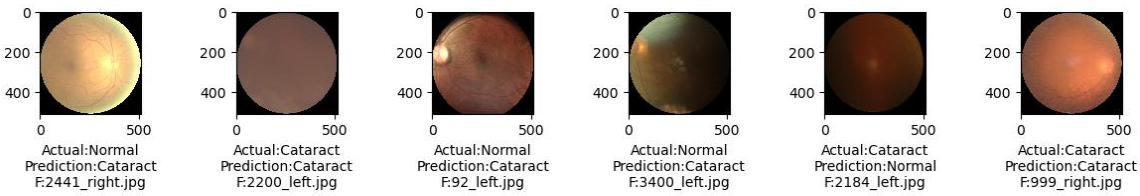

Figure 26: Prediction of ResNet50 Model

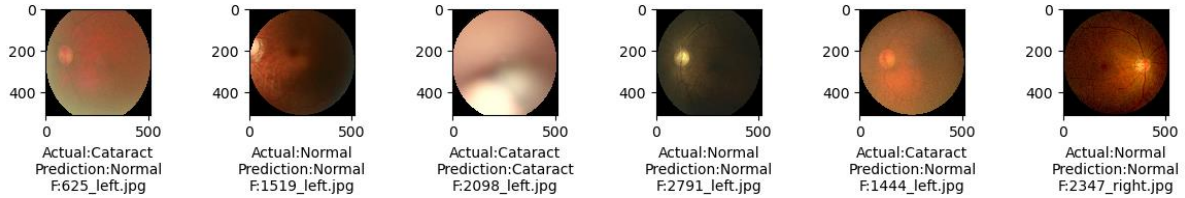

Figure 27: Prediction of ResNet152 Model

V. FUTURE WORK

As this study delves into the comparative analysis of pre-trained deep learning models for cataract detection, several promising directions for future research and development emerge. One potential avenue is the exploration of ensemble methods that combine the predictive power of multiple models to enhance classification accuracy [13]. Ensemble techniques, such as model averaging, stacking, or boosting, could address the limitations of individual models and potentially result in more robust predictions.

Fine-tuning strategies offer another compelling area of future investigation [6]. By selectively freezing and retraining specific layers of pre-trained models, fine-tuning can adapt the models to the intricacies of the cataract detection task and improve their overall performance.

Extending the scope of transfer learning beyond cataract detection to encompass a broader range of ocular diseases, such as diabetic retinopathy or glaucoma, holds potential for more comprehensive multi-disease classification [15]. This expansion could yield models with the capability to simultaneously identify various ocular conditions, enhancing their clinical utility.

Exploring variations in data augmentation techniques tailored specifically to ocular images could also lead to further performance improvements [23]. By simulating a wider range of lighting conditions, ocular orientations, and other domain-specific variations, models could better generalize to unseen data.

Lastly, enhancing the deep learning model interpretability and explainability in medical image analysis remains an essential future area [12]. Incorporating techniques such as attention mechanisms or saliency maps could provide insights into the features driving model decisions, making the models more transparent and interpretable for medical professionals.

VI. CONCLUSION

In this comprehensive study, we embarked on a thorough analysis of nine pre-trained deep learning models for the classification of cataract and normal eyes using color fundus images. The results obtained shed light on the effectiveness of these models in a critical medical image classification task.

Among the models assessed, DenseNet121, MobileNet, and MobileNetV2 demonstrated consistent and commendable performance, boasting high testing accuracies of 96.94%, 97.38%, and 96.51%, respectively. These models exhibited not only robust training but also the ability to generalize effectively to unseen data. On the other hand, models like VGG19, InceptionV3, and ResNet152 revealed some discrepancies between training and validation/testing phases, suggesting potential challenges with overfitting or suboptimal generalization.

The outcomes of this study underscore the importance of selecting appropriate pre-trained models for medical image classification. DenseNet architectures, with their dense connections, exhibited exceptional feature learning capabilities that contribute to their top-tier performance. Moreover, MobileNet and MobileNetV2 proved their efficiency in resource-constrained environments without compromising accuracy.

Our findings have implications beyond this specific project, providing insights for practitioners venturing into medical image analysis. By considering the models' training and testing performances, researchers can make informed decisions when selecting models for various healthcare applications. We also recommend further exploring ensemble methods and fine-tuning techniques to harness the strengths of multiple models and enhance classification accuracy. This study, anchored in a rigorous analysis of pre-trained models, contributes to the advancement of medical image analysis and offers valuable guidance for future endeavors in ocular disease detection.

REFERENCES

[1] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.
[2] Khairallah, M., Kahloun, R., Bourne, R., Limburg, H., Flaxman, S. R., Jonas, J. B., ... & Taylor, H. R. (2015). Number of people blind or visually impaired by cataract worldwide and in world regions, 1990 to 2010. Investigative ophthalmology & visual science, 56(11), 6762-6769.

[3] Lansingh, V. C., Carter, M. J., & Martens, M. (2012). Global cost-effectiveness of cataract surgery. Ophthalmology, 119(7), 1308-1311.
[4] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Kim, R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402-2410.
[5] Jonas, J. B., Bourne, R. R., White, R. A., & Flaxman, S. R. (2014). The NEI Visual Function Questionnaire in the 1988 National Health Interview Survey. Ophthalmology, 121(7), 1571-1572.
[6] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In Advances in neural information processing systems (pp. 3320-3328).
[7] Smith, R. S., & Smith, S. E. (1983). Automated diagnosis of congenital cataracts. IEEE Transactions on Biomedical Engineering, 30(1), 68-73.
[8] Jones, C. A., Hodge, D. O., & Bourne, W. M. (2002). Prevalence and associations of cataract in the Beaver Dam Eye Study. Ophthalmology, 109(1), 73-81.
[9] Wang, X., Tang, H., Liu, J., Tang, L., & Zhang, L. (2017). Automated detection of cataract using deep convolutional neural networks. Journal of Medical Systems, 41(4), 63.
[10] Li, X., Yuan, X., Yu, Z., & Zhang, D. (2019). Computer-aided diagnosis of cataract based on transfer learning. Journal of Medical Systems, 43(8), 246.
[11] Xu, Y., Zheng, Y., Liu, X., Zhang, L., Wang, T., & Cai, D. (2018). GAN-based data augmentation for cataract classification from fundus image. IEEE Transactions on Medical Imaging, 38(3), 757-767.
[12] Zhang, S., Wu, C., & Liu, C. (2019). Explainable automated cataract detection through deep neural networks. IEEE Transactions on Medical Imaging, 39(4), 1291-1300.
[13] Chen, J., Huang, H., Yu, Y., Yang, J., Zhu, M., & Zhang, X. (2016). Computer-aided cataract detection using hybrid feature extraction and classifier ensemble. Computers in Biology and Medicine, 70, 36-46.
[14] Wu, J., Lin, Z., Zheng, W., & Huang, Y. (2018). A hybrid ensemble model for cataract classification. Journal of Healthcare Engineering, 2018.
[15] Guo, Z., Bai, S., Zhang, Z., Liu, J., & Zhang, L. (2021). Multi-disease classification of ocular fundus images with a single deep learning model. IEEE Transactions on Biomedical Engineering, 68(9), 2606-2617.
[16] Shanggong Medical Technology Co., Ltd. (2020). Ocular Disease Intelligent Recognition (ODIR) dataset. Kaggle.
[17] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[18] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
[19] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
[20] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
[21] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).
[22] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[23] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1-48.