A Deep Learning Approach for Cataract Detection Using VGG16

Ahmed M. Salaheldin[1]*, Ahmed El Bialy[1], Rahma Saad[1]

ABSTRACT

Cataracts are a leading cause of visual impairment globally, especially among the aging population. Early and accurate detection is essential to prevent irreversible vision loss. This study proposes a deep learning–based approach for automated cataract detection using retinal fundus images. The VGG16 convolutional neural network (CNN) was employed and fine-tuned for binary classification between normal and cataract-affected eyes. The dataset, sourced from publicly available repositories, was limited to two classes, Normal and Cataract—with approximately 1,000 images per class. Data preprocessing and augmentation techniques were applied to enhance model generalization and robustness. The model achieved an accuracy of 92%, indicating its reliability for clinical use. Comparative analysis with existing literature demonstrates the model's competitive performance, highlighting its potential as a practical and accessible tool for aiding ophthalmologists in early cataract diagnosis. Future work may include multi-disease detection and the integration of explainable AI for increased transparency in medical decision-making.

Keywords: Cataract, deep learning, VGG16, retinal images, CNN, image classification, medical imaging.

I. INTRODUCTION

Cataract is considered as one of the major causes of vision impairment and blindness worldwide, especially among the aging population. The condition leads to blurred vision, glare sensitivity, and difficulty seeing at night [1]. According to the World Health Organization, cataracts are responsible for over 50% of blindness cases in low- and middle-income countries. Although cataract surgery is widely available and highly effective, early diagnosis remains crucial for timely treatment and to prevent permanent visual damage [2].

Traditional methods for cataract detection rely on clinical examination and imaging techniques such as slit-lamp photography and fundus imaging [3]. However, these methods are often time-consuming, require specialist interpretation, and may not be accessible in remote or underserved areas. To overcome these challenges, artificial intelligence (AI), particularly deep learning, has emerged as a promising tool in the field of medical image analysis.

AI-based systems, especially those utilizing convolutional neural networks (CNNs) [4], have shown remarkable success in automating the detection of various eye diseases, including diabetic retinopathy, glaucoma, and cataract. These models can learn complex features from large datasets of retinal images, enabling high-accuracy classification and prediction. Deep learning not only reduces the burden on ophthalmologists but also provides a scalable solution for mass screening programs.

[1] Systems and Biomedical Engineering Department, Higher Institute of Engineering, EL Shorouk Academy, Cairo, Egypt.

* Corresponding Author

Ramakrishnan et al. [5] present a recent study proposed a hybrid architecture combining convolutional neural networks (CNNs) with support vector machines (SVMs) to enhance feature extraction and classification of fundus images for ocular disease detection. This approach addresses the limitations of manual image interpretation, which can be time-consuming and error prone. By leveraging automated techniques, the system aids in early and accurate cataract diagnosis. The models were optimized using Intel's oneDNN library within the oneAPI environment and evaluated on the publicly available Ocular Disease Intelligent Recognition (ODIR) dataset. Among the tested models, MobileNet achieved the highest performance with an accuracy of 98.36%.

Tashkandi [6] presents a deep learning–based approach for the early detection of multiple eye conditions, including cataracts, glaucoma, age-related macular degeneration (AMD), diabetic retinopathy, and high myopia. The researcher emphasizes the limitations of traditional diagnostic methods and propose a user-friendly system that simplifies and accelerates retinal image analysis. By leveraging a large dataset and advanced image processing techniques, the model was trained to accurately recognize disease patterns. Among the evaluated models, CNN, VggNet, MobileNet, and hybrid deep learning architectures achieved superior performance, with accuracy exceeding 98%, significantly outperforming traditional machine learning models like SVM and Random Forest.

Aida Jones [7] highlights the effectiveness of Support Vector Machines (SVM) in grading canine cataracts, particularly in addressing challenges related to age and genetic factors. The proposed AI-based model utilizes standard mobile phone images to classify cataract severity, offering a practical solution for veterinarians. The model underwent binary and multi-class classification tests, with performance validated using K-fold cross-validation. The system achieved accuracy rates of 83% without cross-validation and 81% with 10-fold cross-validation, demonstrating the model's reliability and potential in veterinary ophthalmology.

Triwijoyo [8] presents a study that focuses on developing an automated cataract detection system using a deep neural network architecture applied to fundus images. Cataracts, characterized by age-related lens clouding, are a leading cause of visual impairment, with over 120,000 new cases reported annually. The proposed method involves four key steps: data acquisition, preprocessing, model training using a convolutional neural network (CNN), and testing. The model was trained on 410 fundus images (329 for training and 81 for validation) and evaluated using a confusion matrix and the approach achieved accuracy of 99%.

El Harti [9] presents a study to address the growing need for automated diagnostic tools in ophthalmology due to the rising prevalence of eye disorders and limited specialist availability. The researchers propose a hybrid model that integrates a classical neural network with four machine learning algorithms to classify eye diseases, including glaucoma, diabetic retinopathy, and cataracts. The VGG16 model, pre-trained on the ImageNet dataset, is used for feature extraction, with the resulting features fed into the machine learning classifiers. The model was trained and evaluated using a diverse set of retinal images from publicly available sources, including the Indian Diabetic Retinopathy Image Dataset and other retinal databases.

The aim of the study is to implement an AI model for the detection of cataract of from fundus images which can be used as a decision support system to reduce human error and save time.

The rest of the paper is organized as follows: Section 2 describes the materials and methodology used in the research. Section 3 reports the main results, which are further examined in detail in Section 4. Section 5 provides the conclusion and suggests possible avenues for future investigation.

II. MATERIALS AND METHODS

This section outlines the systematic approach employed to develop a deep learning-based model for cataract detection using retinal images. The methodology encompasses all stages of the project, including dataset selection and preparation, image preprocessing, model architecture design, training strategy, and performance evaluation. A combination of transfer learning and convolutional neural network techniques was utilized to optimize classification accuracy while ensuring computational efficiency. The materials used in this study include a curated dataset of fundus images, image augmentation tools, and pre-trained neural network models. The following subsections provide detailed descriptions of each step in the process as shown in Figure. 1.

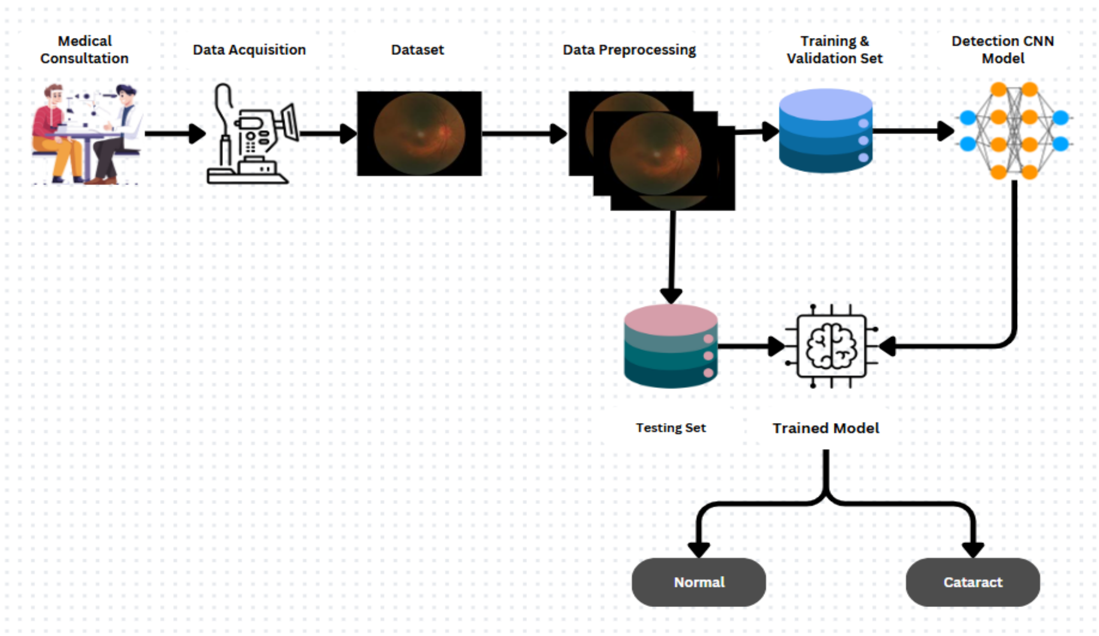


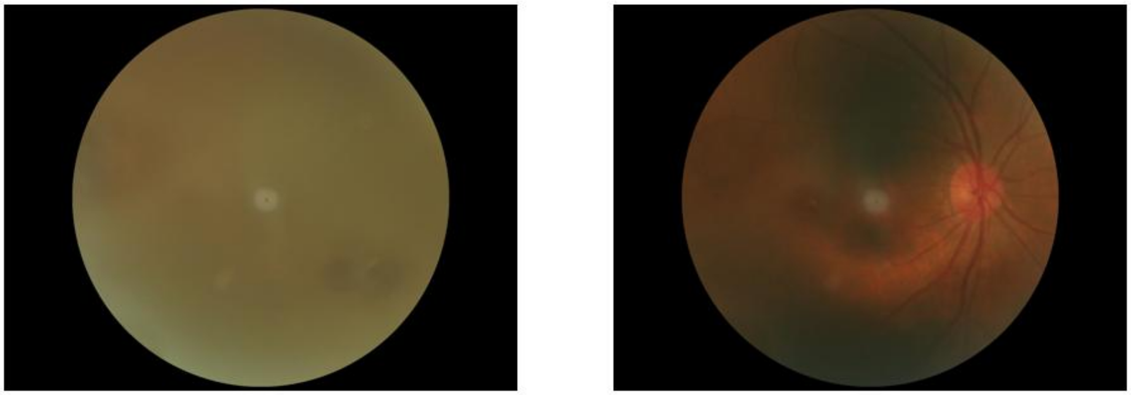
Figure 1 Block diagram of the proposed model for cataract detection

A. Dataset

The dataset utilized in this research includes retinal images grouped into four categories: Normal, Diabetic Retinopathy, Cataract, and Glaucoma. For the purpose of this research, we focused exclusively on the Normal and Cataract classes to enable binary classification. Each of these two categories includes approximately 1,000 high-resolution retinal images, ensuring a balanced dataset suitable for training and evaluation as shown in Figure .2.

The images were sourced from a combination of publicly available and diverse databases, including the Indian Diabetic Retinopathy Image Dataset (IDRiD) [10], the Ocular Recognition dataset [11], and the High-Resolution Fundus (HRF) dataset [12]. These datasets provide a wide range of retinal image variations in terms of quality, contrast, and pathological features, thereby enhancing the robustness and generalizability of the model.

To facilitate proper model training and fair performance assessment, the dataset was divided into three portions. The training set comprised 80% of the images and was used to train the deep learning model. Another 10% formed the validation set, which helped in tuning hyperparameters and tracking performance throughout training. The final 10% was allocated as the testing set, used exclusively to evaluate the model's accuracy on previously unseen data. This systematic division enables a thorough and reliable assessment of the model's capability to distinguish between cataract and normal retinal images.



a b

Figure 2 Sample of the utilized dataset a) cataract case b) normal case

B. Preprocessing

To ensure optimal performance of the deep learning model and enhance its ability to generalize across diverse image inputs, a comprehensive data preprocessing and augmentation pipeline was applied. This step was essential to reduce the risk of overfitting, improve robustness, and enable the model to learn meaningful patterns from varied representations of the same class.

Preprocessing involved standardizing the input images by rescaling pixel intensity values from the original range of [0, 255] to a normalized range of [0, 1]. This was achieved by dividing all pixel values by 255, which facilitates faster convergence during training and ensures compatibility with the pre-trained model input format.

Data augmentation techniques were then applied to artificially expand the training dataset and introduce variability. These included:

- Random rotation of images by up to ±20 degrees to simulate different viewing angles.
- Width and height shifts of up to 20% to account for positional variation of the retina within the frame.
- Shearing transformations mimic slight distortions in imaging conditions.
- Zoom operations to allow the model to focus on features at varying scales.
- Horizontal flipping to introduce symmetry-based variations and improve spatial learning.

These augmentations were applied randomly during each training iteration, effectively creating a dynamic and diverse set of training samples. As a result, the model was better equipped to handle inconsistencies in image orientation, size, and lighting—conditions commonly encountered in real-world clinical settings.

C. Model Architecture

In this study, the VGG16 model was adopted as the base architecture for cataract classification. VGG16, pre-trained on the ImageNet dataset, is well-regarded for its depth, simplicity, and strong performance in image recognition tasks. Its use of small convolutional filters and deep layered structure makes it effective for learning rich feature representations from images.

To adapt VGG16 for binary classification of retinal images (Normal vs. Cataract), several modifications were made to its architecture:

The original fully connected (top) layers of the VGG16 model were removed, allowing for a custom classification head to be attached.

- A Flatten layer was introduced to convert the output feature maps into a one-dimensional vector.
- A Dense layer containing 128 units with ReLU activation was incorporated to capture advanced feature representations from the extracted data.
- To reduce overfitting, a Dropout layer with a dropout rate of 0.5 was included, randomly deactivating half of the neurons during training.
- The architecture concludes with a Dense layer utilizing a sigmoid activation function, which generates a probability value to distinguish between cataract and non-cataract cases.

During initial training, the convolutional base (feature extraction layers) of VGG16 was frozen to retain the learned weights from ImageNet, which helps in transferring general image features to the medical domain without overfitting on a relatively smaller dataset.

D. Model Training and Validation

The modified VGG16 model was compiled using the Adam optimizer, which is known for its efficiency and adaptive learning rate capabilities. The binary cross-entropy loss function was employed as it is well-suited for binary classification tasks. Accuracy was selected as the primary performance metric to evaluate the model's ability to correctly distinguish between Normal and Cataract images.

Training was carried out over 20 epochs with a batch size of 32, ensuring a balance between convergence speed and stability. To monitor model performance and prevent overfitting, a Model Checkpoint callback was implemented, which automatically saved the model with the highest validation accuracy during training. This structured training procedure allowed the model to progressively learn discriminative features and generalize effectively to unseen data, ultimately leading to robust classification performance.

E. Model Evaluation

To comprehensively assess the performance of the cataract detection model, a range of evaluation metrics and visualization techniques were employed. These metrics provided insights into both the accuracy and reliability of the model across different aspects of classification:

- Binary Cross-Entropy Loss: This metric measures how much the predicted probability values deviate from the true class labels. It is the main loss function used during training and helps the model reduce prediction errors.
- Accuracy: This metric reflects the percentage of correctly classified images out of all samples. While it is a straightforward indicator of overall performance, it may not fully represent effectiveness when the dataset is imbalanced.
- Precision: This evaluates how many of the images predicted as positive (cataract) are actually correct, indicating how dependable the model is at identifying cataract cases.
- Recall: This metric measures the model's effectiveness in detecting all true cataract instances, showing how well it captures actual positive cases.

- F1-Score: This provides a balanced measure by combining precision and recall through their harmonic mean, making it especially valuable when dealing with uneven class distributions.

These metrics, along with confusion matrix visualizations and ROC curves, collectively demonstrated the effectiveness and robustness of the proposed model in accurately classifying retinal images into Normal and Cataract categories.

III. RESULTS

The effectiveness of the developed deep learning model for cataract detection was assessed using common classification metrics such as accuracy, precision, recall, and F1-score. The summarized results in the table highlight the model's strong capability to accurately differentiate between normal and cataract-affected retinal images.

Table 1 Summary of the obtained results for the cataract detection

Metric	Cataract	Normal	Overall
Accuracy	0.92	0.92	0.92
Precision	0.88	0.96	0.92
Recall	0.92	0.93	0.925
F1 Score	0.9	0.95	0.925

The overall classification accuracy achieved by the model is 92%, indicating a strong capability to correctly classify both normal and cataract cases. Notably, the precision for normal images (0.96) is higher than that for cataract images (0.88), suggesting the model is more confident in correctly identifying normal cases. On the other hand, the recall for cataract cases (0.92) reflects the model's effectiveness in correctly detecting the presence of cataracts, which is critical in a medical diagnosis context. The F1-score, which balances precision and recall, is 0.90 for cataract and 0.95 for normal, with an overall score of 0.925. These values confirm that the model performs reliably across both classes, making it a robust tool for aiding in the early detection of cataract disease. Overall, the evaluation metrics indicate that the trained model achieves a high level of diagnostic accuracy, with consistent performance across both categories, validating its suitability for clinical decision support.

IV. DISCUSSION

The results of this study demonstrate the effectiveness of a VGG16-based convolutional neural network model in accurately classifying retinal images into normal and cataract categories. The model achieved an overall accuracy of 92%, with a precision of 0.88 for cataract detection and 0.96 for normal images. The recall and F1-scores were 0.92 and 0.90 for cataracts, and 0.93 and 0.95 for normal cases, respectively. These metrics highlight the model's balanced performance across both classes and its reliability in clinical decision-making contexts, particularly in aiding early detection of cataracts. When compared with existing studies in the literature, our model shows strong performance, though some approaches report higher accuracy. For instance, Ramakrishnan et al. [5] proposed a hybrid CNN-SVM model optimized using Intel's oneDNN library, achieving a notably higher accuracy of 98.36% on the ODIR dataset. This suggests that hybrid models, especially those benefiting from hardware-level optimization and efficient feature classifiers like SVM, can outperform traditional CNN pipelines.

Similarly, Tashkandi [6] developed a multi-disease classification system using CNN, VGGNet, MobileNet, and hybrid models, all achieving over 98% accuracy. These results were made possible by using large, diverse datasets and advanced image preprocessing techniques. While our model performs slightly lower in accuracy, it maintains strong generalizability and was developed using a streamlined binary classification pipeline, which simplifies deployment in clinical settings. Aida Jones [7] focused on grading canine cataracts using SVMs and mobile-acquired images. Their study achieved 83% accuracy without cross-validation and 81% with 10-fold cross-validation. While not directly comparable to human cataract detection, it demonstrates the versatility of machine learning models across different domains. In contrast, our model's higher accuracy of 92% in a human clinical context reinforces the suitability of CNN-based approaches for medical imaging. Triwijoyo [8] achieved a remarkable 99% accuracy using a CNN-based cataract detection system trained on a smaller dataset of 410 fundus images. While impressive, the smaller dataset may limit the model's scalability and generalizability. Our model, trained on a larger and more diverse dataset, shows consistent and balanced performance, indicating better robustness in real-world applications.

Lastly, El Harti [9] proposed a hybrid framework combining classical neural networks with multiple machine learning algorithms. Their use of VGG16 as a feature extractor aligns with our approach, validating the strength of VGG16 in retinal image analysis. However, their focus extended beyond cataract detection to multiple eye diseases, which may dilute performance on any single condition. In contrast, our targeted binary classification framework offers a focused solution optimized for cataract diagnosis. In summary, while some existing studies report higher performance metrics, particularly those utilizing hybrid or ensemble methods, our model demonstrates a strong balance of accuracy, simplicity, and clinical applicability. Future work could explore integrating hybrid classifiers or incorporating multi-disease detection capabilities to enhance the model's versatility and diagnostic power.

V. CONCLUSION

This study presents a deep learning–based approach for the automated detection of cataracts using retinal fundus images, leveraging the VGG16 convolutional neural network architecture. Through careful dataset preparation, preprocessing, and model optimization, the proposed system achieved a high classification accuracy of 92%, along with strong precision, recall, and F1-scores across both normal and cataract classes. These results demonstrate the model's reliability and effectiveness in distinguishing cataract-affected eyes from healthy ones. The integration of transfer learning and data augmentation techniques played a key role in enhancing model generalization and robustness. Compared to traditional diagnostic methods, the proposed system offers a faster, more accessible, and less error-prone solution that can assist ophthalmologists in early cataract detection and clinical decision-making. While the model's performance is competitive, future research can focus on improving accuracy by incorporating hybrid models, expanding the dataset to include a wider variety of ocular diseases, and integrating explainable AI techniques to increase transparency in predictions. Additionally, deploying the model within a user-friendly interface would further support its adoption in real-world medical environments. The findings of this study reinforce the potential of deep learning as a powerful tool in ophthalmic diagnostics and lay the groundwork for more advanced, multi-condition diagnostic systems in the future.

REFERENCES

[1] "Global estimates on the number of people blind or visually impaired by cataract: a meta-analysis from 2000 to 2020," *Eye*, vol. 38, no. 11, pp. 2156–2172, 2024.

[2] "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.," *Lancet. Glob. Heal.*, vol. 9, no. 2, pp. e144–e160, Feb.

2021, doi: 10.1016/S2214-109X(20)30489-7.

[3] K. Y. Son *et al.*, “Deep learning-based cataract detection and grading from slit-lamp and retro-illumination photographs: Model development and validation study,” *Ophthalmol. Sci.*, vol. 2, no. 2, p. 100147, 2022.

[4] A. M. Salaheldin, M. Abdel Wahed, and N. Saleh, “A hybrid model for the detection of retinal disorders using artificial intelligence techniques,” *Biomed. Phys. Eng. Express*, vol. 10, no. 5, p. 55005, 2024, doi: 10.1088/2057-1976/ad5db2.

[5] A. B. Ramakrishnan, M. Madavan, R. Manikandan, and A. H. Gandomi, “A Hybrid Deep Learning Paradigm for Robust Feature Extraction and Classification for Cataracts,” *Appl. AI Lett.*, vol. 6, no. 2, p. e113, Apr. 2025, doi: https://doi.org/10.1002/ail2.113.

[6] A. Tashkandi, “Eye Care: Predicting Eye Diseases Using Deep Learning Based on Retinal Images,” *Computation*, vol. 13, no. 4. 2025. doi: 10.3390/computation13040091.

[7] A. Jones and T. B. and Vijayan, “AI-driven canine cataract detection: a machine learning approach using support vector machine,” *J. Chinese Inst. Eng.*, vol. 48, no. 1, pp. 94–102, Jan. 2025, doi: 10.1080/02533839.2024.2420100.

[8] B. K. Triwijoyo and G. H. Martono, “Automatic Cataract Detection and Classification Using Deep Neural Network,” in *2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2024, pp. 43–47. doi: 10.1109/COMNETSAT63286.2024.10862357.

[9] M. El Harti, S. Zaamoun, S. J. Andaloussi, and O. Ouchetto, “Classification of Eye Disorders Using Deep Learning and Machine Learning Models BT - Proceedings of the Second International Conference on Advances in Computing Research (ACR’24),” K. Daimi and A. Al Sadoon, Eds., Cham: Springer Nature Switzerland, 2024, pp. 184–194.

[10] P. Porwal *et al.*, “Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research,” *Data*, vol. 3, no. 3, p. 25, 2018.

[11] “Ocular Disease Recognition,” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k

[12] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, “Robust Vessel Segmentation in Fundus Images,” *Int. J. Biomed. Imaging*, vol. 2013, no. 1, p. 154860, Jan. 2013, doi: https://doi.org/10.1155/2013/154860.