Received: September 12, 2024. Revised: October 30, 2024. 184

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

DAUCD: Deep Attention U-Net for Cataract Detection Leveraging CNN Frameworks

Maneesha Vadduri[1] Kuppusamy P[1]*

[1]*School of Computer Science and Engineering, VIT-AP University, Amaravathi, Andhra Pradesh 522237, India*
* Corresponding author's Email: drpkscse@gmail.com

Abstract: Cataract remains the leading cause of blindness globally, accounting for nearly half of all cases. This study presents the Deep Attention U-Net for Cataract Diagnosis (DAUCD) model, leveraging advanced deep learning techniques to improve both the segmentation and classification of cataract in retinal images. The proposed DAUCD method integrates Attention U-Net architectures with pre-trained backbones (ResNet50, Inception-v3, and VGG19) for precise blood vessel segmentation, followed by the classification of segmented outputs using VGG16. The model achieved a classification accuracy of 98.24%, with a sensitivity of 99.77%, a specificity of 97.83%, and an AUC of 99.24%, particularly excelling with the ResNet50-based backbone. The dataset, curated from multiple sources including the cataract dataset, ODIR5K, eye-diseases-classification, and cataract eyes datasets, comprises a total of 10,444 fundus images. It was designed to support both segmentation and classification tasks, with images evenly distributed across cataract and non-cataract classes. This comprehensive dataset provided a strong foundation for validating the effectiveness and generalizability of the proposed DAUCD model. The findings of this research underscore the robustness and efficiency of the DAUCD model in medical image analysis, offering promising advancements in early detection and treatment outcomes for cataract patients.

Keywords: Cataract, Fundus images, Classification, Deep learning, Attention U-Net.

1. Introduction

Cataract ranks among the primary causes of vision loss in developed countries, accounting for nearly half of the cases [1]. By 2024, the global population of visually impaired individuals is estimated to reach around 150 million [2]. Cataracts are often divided into three main types: Nuclear Cataract, Cortical Cataract, and Posterior Subcapsular. Several causes may lead to the deterioration of the lens protein. Due to these causes, the metabolism of the lens was disrupted, resulting in the development of cataracts. Cataract occurs when the presence of a cloudy lens obstructs the passage of light onto the retina, leading to a loss of visual clarity. Cataracts are more prevalent in those above 40 years of age, and the probability of developing cataracts increases as one ages. Treatment for cataracts has become a concern due to the steadily rising frequency of the condition and its growing impact [3]. It is recommended to be promptly recognized and promptly treated.

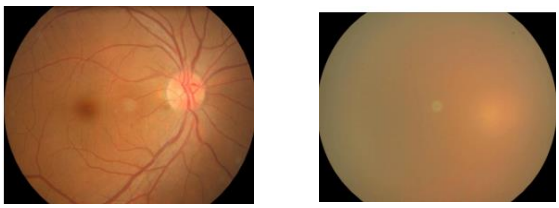
There is continuing research on cataract risk factors [4-6]. Age-related nuclear and cortical cataracts are often encountered in older individuals [7]. UV-B radiation is a risk factor. Contact and smoking may contribute to visual function changes, although the likelihood of them causing significant alterations is low. To address cataracts, it is important to recognize and treat them early. If discovered early, some actions may be performed to mitigate the progression by taking preventive measures, such as using sunglasses that reduce glare [8]. Surgical procedures frequently provide effective solutions for severe cataracts that considerably affect a patient's daily life. Currently, four main procedures are used for the identification and grading of cataracts. The initial approach involves the light-focus technique. The second approach uses iris image projection. The third method involves the slit lamp assessment, while

the fourth employs ophthalmoscopic transillumination Nevertheless, physical evaluation may be prone to subjectivity, requires a significant amount of time, and incur high costs [9].

Hence, considering the social and economic aspects, it is very logical to get automated cataract diagnosis by the use of artificial intelligence. The proposed DAUCD model introduces a novel integration of Attention U-Net with pre-trained backbones, such as ResNet50, Inception-v3, and VGG19. This integration allows the model to enhance blood vessel segmentation by focusing on essential features and filtering out irrelevant ones through attention mechanisms. Furthermore, the segmented images are classified using VGG16, ensuring precise identification of cataract presence. The emphasis on eye care is crucial since it offers an opportunity to positively impact the lives of individuals globally [10].

The DAUCD model demonstrates several advantages over existing methods. It enhances robustness and efficiency in distinguishing fine retinal structures through its multi-stage approach. Additionally, the integration of multiple pre-trained backbones improves the model's ability to generalize across diverse retinal images, addressing the limitations of traditional single-architecture methods. By employing attention gates, the model effectively captures detailed retinal features, thereby achieving higher diagnostic precision. Cataracts are classified into two categories: Non-cataract and Cataract. Fig. 1 presents two distinct retinal fundus images. In Fig. 1 (a), a normal retinal fundus is displayed, with capillaries and vascular cells clearly visible. Conversely, Fig. 1 (b) shows a cataract-affected image, where blurriness conceals the capillaries and vascular structures. At this point, most individuals have substantial vision loss.

The paper is organized as follows: Section 2 reviews related work on cataract detection techniques. Section 3 elaborates on the model architecture and implementation. Section 4 discusses the experimental results and discussion. Section 5 concludes the paper by summarizing the key findings and offering directions for future research.



(a) (b)

Figure. 1 Retinal Fundus images: (a) Non-Cataract and (b) Cataract

2. Related work

In recent years, the field of medical image analysis has seen remarkable progress in developing automated diagnostic methods for eye diseases. The rise of deep learning techniques has enabled significant advancements in the detection and grading of cataracts. This literature review aims to provide a comprehensive overview of existing research on automated approaches for cataract diagnosis, highlighting key innovations, challenges, and advancements in leveraging deep learning methods for detection and classification tasks.

Several studies have explored various automated frameworks to enhance cataract detection and grading using retinal fundus images, demonstrating notable improvements in diagnostic accuracy, robustness, and efficiency.

Existing Methods and their Limitations

A Deep CNN (DCNN) was presented for cataract identification and grading, which made use of feature maps from the architecture's pooling layers. This approach was time-efficient, with cataract identification and grading accuracies of 93.52% and 86.69%, respectively. However, relying on pooling layers may result in the loss of crucial spatial details, affecting the model's accuracy in distinguishing similar retinal features, especially in complex cases [11].

DCNN and Random Forests were used to grade cataracts at six levels. The suggested DCNN extracted fundus image characteristics at various levels using three modules. DCNN provided a feature dataset that RF utilized to construct the more complex six-level cataract grading. On average, this approach was 90.69% accurate. The six-level grading system may help doctors better comprehend patients. However, the reliance on manual tuning of RF parameters and separate feature extraction may limit automation and scalability [12].

A computer-assisted procedure for assessing the cataract severity, ranging from mild to severe, using the fundus pictures was provided. This technique employed previously trained CNNs through transfer learning to automate the classification of cataracts. The final classification was performed using the feature extraction process and the SVM classifier, which had a four-stage CCR of 92.91%. The use of separate CNNs for feature extraction and SVM for classification can lead to information loss and reduced adaptability due to the lack of direct integration between these stages [13].

A cataract grading system using a Tournament-Based Ranking CNN was introduced, which employs a tournament framework to improve accuracy in

identifying underrepresented classes, achieving an exact match accuracy of 68.36%. The manual setup of the tournament structure may limit adaptability to different datasets, requiring frequent reconfiguration [14].

An automated cataract detection method employing DCNNs and a trained Res-Net classifier model with a 95.77% accuracy was suggested. The approach lacks pre-processing to address image quality variations, making it less robust and accurate with noisy or low-quality images [15].

A hybrid CRNN model was introduced for cataract classification, combining CNNs (AlexNet, GoogLeNet, ResNet, and VGGNet) with RNNs to capture spatial correlations in image patches, achieving 96.39% accuracy. The model's reliance on patch-based analysis may overlook global image context, and its validation on a limited dataset raises concerns about generalizability [16].

Transfer learning with Inception-V4 was used to tackle CNN issues like overfitting, high computation costs, and fading gradients, achieving 96% accuracy in cataract classification. However, the method relies on standard CNN architectures and lacks multi-scale feature extraction, potentially limiting its ability to capture both fine details and broader patterns in complex images [17].

A stacking approach graded cataracts into six stages using ResNet-18 for high-level feature extraction and GLCM for texture features. Two SVM classifiers processed these features, with a fully connected neural network achieving 94.75% accuracy. However, the reliance on manually engineered GLCM features may limit flexibility and generalizability to diverse datasets [18].

A deep learning model was proposed for cataract detection and grading, utilizing a flexible ResNet-based architecture with 18 and 50 layers for different tasks. The model processes fundus images through the G channel and outputs predictions with heatmaps to localize key areas. It achieved state-of-the-art accuracy of 97.2% for detection and 87.7% for grading, with interpretability validated by ophthalmologists. The model's reliance on the G channel may overlook valuable features in the R and B channels, potentially affecting classification accuracy [19].

VGG-19 was applied to automate eye disease detection using fundus images, achieving 95% accuracy in classifying normal versus cataract cases within the ODIR-5K dataset. However, the absence of segmentation might limit precision in complex cases, where detailed region-specific analysis is essential [20].

An EfficientNet and ML-Decoder (Multi-Label Decoder) based deep learning model achieved 95.7% accuracy in detecting multi-label retinal diseases from fundus images. The model uses SAM optimizer, image transformations, and pixel-level fusion of eye images, outperforming state-of-the-art methods on the ODIR dataset with fewer parameters. However, class imbalance remains unresolved as GAN-based solutions were proposed but not implemented [21].

EfficientNetB3 achieved 96.94% accuracy in multi-label classification of fundus images using the ODIR-5K dataset. The model showed strong potential for early diagnosis on edge devices, reducing computational costs. The study primarily focused on accuracy and F1-score, but did not fully explore other important metrics like precision, recall [22].

CataractNetDetect, a multi-label deep learning classification system, integrates features from paired fundus images (left and right eyes) using models like ResNet-50, DenseNet-121, and Inception-V3. Trained on the ODIR-5K dataset, it achieved 94% accuracy, along with an F1-score of 98.0% and AUC of 97.9%, outperforming conventional models in diagnosing cataracts and other ocular diseases. CataractNetDetect's reliance on paired images may limit its effectiveness when only single-eye data is available or when there are significant differences between eyes [23].

A deep learning-based system using ResNet50 achieved 92% accuracy on the ODIR-5K dataset for diagnosing eye diseases like diabetic retinopathy, glaucoma, and cataracts. The approach employs transfer learning, image preprocessing, and oversampling but lacks attention mechanisms, potentially affecting accuracy in detecting subtle or overlapping features [24].

CataractEyeNet, a deep learning-based system using an enhanced VGG-19 model, achieved 96.78% accuracy in detecting cataract disorders. However, relying solely on VGG-19 may limit its adaptability to diverse datasets and variations in lens images, potentially affecting performance in different imaging conditions [25].

An ensemble deep learning approach was developed to detect and classify eight eye diseases from fundus images, achieving 90.24% accuracy for cataract detection using CLAHE, Gaussian filtering, and augmentation. However, the model lacks a dedicated segmentation step, potentially limiting its accuracy in detecting subtle abnormalities or distinguishing overlapping features in complex cases [26].

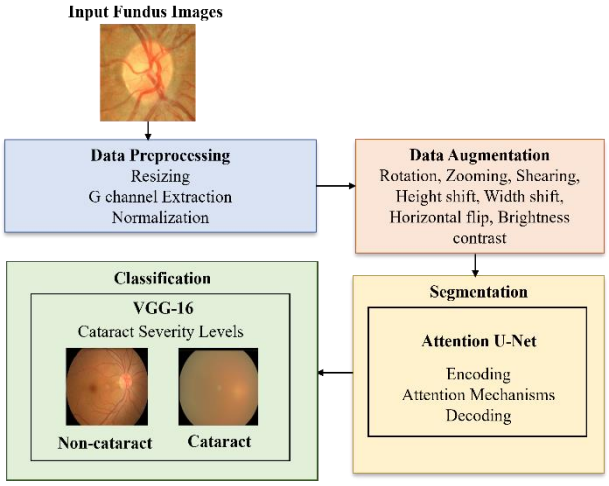


Figure. 2 Work Flow Diagram

3. Model architecture and implementation

The overview of the proposed methodology, depicted in Fig. 2, encompasses several key components: Data Preprocessing, Data Augmentation, Deep Learning Attention U-Net frameworks for segmentation, and the identification of Cataract Severity Levels by using VGG-16 as a classifier.

3.1. Dataset

A comprehensive collection of fundus cataract images is curated from multiple databases and open-source datasets available on Kaggle, addressing the lack of standardized benchmark datasets. This research employs a selection of images sourced from the cataract dataset [27], odir5k [28], eye-diseases-classification [29], cataract eyes Kaggle [30], and cataract dataset [31], as illustrated in the Table 1.

The dataset comprises 10,444 fundus images, evenly distributed across two classes, with the cataract class containing 5,244 images and the non-cataract class containing 5,200 images after preprocessing, as detailed in the upcoming section.

The dataset is categorized into training, testing, and validation subsets in a 70:20:10 ratio. This results in 7,311 designated for training, 2,089 images are allocated for testing, and 1,044 images are reserved for validation. The distribution of images in each set is as follows:

Training Set: 3,671 images with cataract and 3,640 are non-cataract images.

Testing Set: 1,049 images with cataract and 1,040 are non-cataract images.

Validation Set: 524 images with cataract and 520 are non-cataract images.

Table 1. Overview of Dataset: Train, Test, and Validation Splits

Dataset Split	Total Images	Cataract Images	Non-Cataract Images
Training Set	7,311	3,671	3,640
Testing Set	2,089	1,049	1,040
Validation Set	1,044	524	520

3.2. Data preprocessing

To make fundus images more effective for analysis, several preprocessing steps are crucial. First, we resize the images using bicubic interpolation, ensuring they fit the necessary parameters for further examination. We then perform green channel extraction from the RGB images to combat uneven lighting issues. The green channel stands out for its ability to highlight important details while maintaining the essential features of the original images as illustrated in Fig. 3, and it also speeds up processing time by about a third. Next, we normalize the images by standardizing the intensity values, which involves subtracting the average pixel intensity and normalizing by the standard deviation.

We also use data augmentation to expand the dataset to enhance model performance, reducing the risk of overfitting. This involves generating new images by rotating them at different angles (0°, 45°, 90°, 135°, 180°), flipping them horizontally, cropping them at the corners, and shifting them within a specific frame. These techniques ensure that the model can generalize better across different datasets. The result is a collection of high-quality, resilient images prepared for additional analysis, as illustrated in Fig. 4.

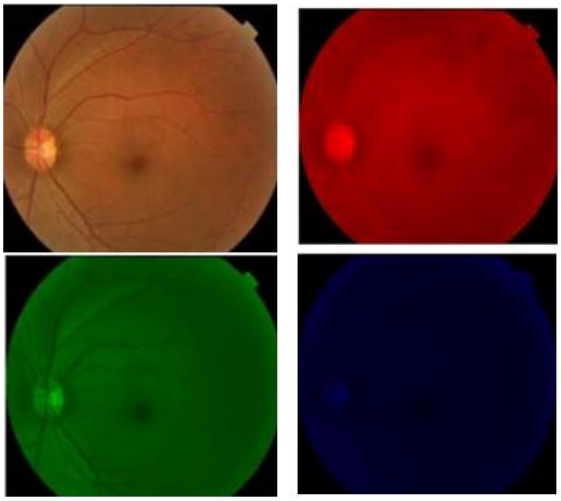


Figure. 3 RGB image in the first column, Red channel in the second column, Green channel in the third column, Blue channel in the fourth column

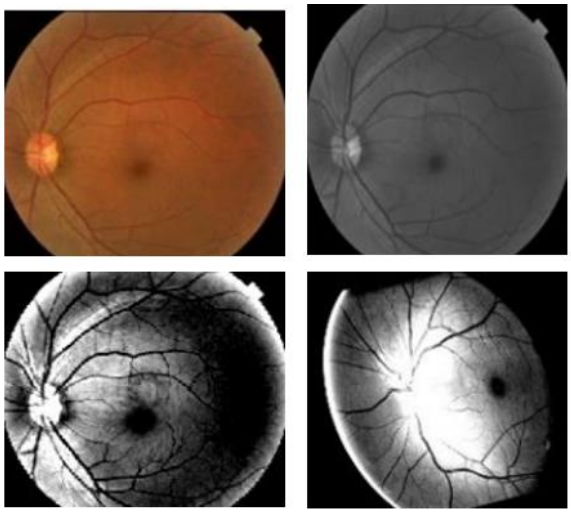


Figure. 4 The resulting fundus images after the preprocessing steps: Resized image in the first column, Green channel of the image in the second column, Normalized image in the third column, Augmented images in the fourth column

3.3. Attention U-Net architecture

Fully Convolutional Networks (FCNs) have gained considerable interest in image segmentation research, particularly with the U-Net design. The U-Net structure is highly effective in binary segmentation, making it widely applicable in biomedical image segmentation tasks [32]. The U-Net design comprises two main components: (1) the encoder, referred to as the contracting path, and (2) the decoder, also called the expansion path. In the encoder, there are 5 stages, each utilizing a sequence of convolutional layers followed by max-pooling layers to progressively reduce the spatial dimensions of the image and extract localized features. The decoder path also contains 5 stages, where transposed convolutional layers are used for upsampling to restore spatial resolution while refining the feature maps.

Unlike architectures that rely on dense layers, the U-Net uses only convolutional layers, making it capable of handling images of varying sizes [33]. Additionally, skip connections are utilized to pass feature maps from the encoder to the corresponding layers in the decoder, helping retain spatial details that might otherwise be lost during downsampling. These skip connections are s in our Attention U-Net model by integrating attention gates at each of the 5 skip connections, these connections enable the model to concentrate on the most important spatial information This addresses a limitation in the original U-Net design, where early-stage feature maps are often less informative due to a lack of focus on important details [34].

The DAUCD model further improves performance by replacing the standard U-Net encoder with pre-trained backbones (ResNet50, Inception-v3, and VGG19), which provide more powerful and efficient feature extraction. The combination of these enhancements allows our model to deliver more accurate segmentation results in biomedical image analysis.

3.4. Proposed DAUCD model

The U-Net model has been improved to produce optimum outcomes, resulting in the development of Attention U-Net. This design has an encoder, the decoder, and the attention gates included in each level's skip connection. Our proposed Attention U-Net integrates attention gates at each skip connection. For evaluating the optimal segmentation performance, replaced the original encoder of the standard U-Net with pre-trained networks ResNet50, Inception-v3, and VGG19 serving as backbones in contracting path. These three models share identical decoders, which include convolutional, upsampling, and concatenation layers [35]. The concatenation layers merge the upsampled output with the feature maps from the encoder. Additionally, each convolutional layer is succeeded by batch normalization and the Rectified Linear Unit (ReLU) activation function. The ReLU activation function, denoted as (1), is known as the input $\mathcal{X}$ of a neuron, which enhances the model's performance. The main distinction among the three networks is found in encoder. A prevalent method for training the deep neural networks, batch normalization improves efficiency and the consistency.

$$eLU(\mathcal{X}) = max(0,1) \tag{1}$$

In the Attention U-Net framework, each level incorporates an attention gate that receives the signal inputs: G and X, G being a deeper layer gating signal, providing the spatial information. The signal X comes from a skip connection which contains richer feature representations. Combining signals X and G ensures the retention of both attributes and spatial details. X represents the encoder's feature map, while

Figure. 5 AG Gate Design Overview

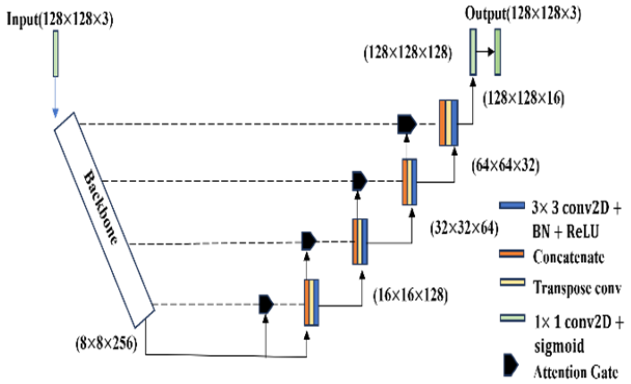


Figure. 6 Attention U-Net Model Framework

G represents the gating signal in the decoder path's attention gate, as illustrated in Fig. 5. The attention gate's output, X_1 is represented as $X \cdot \alpha$. The vector of attention coefficients α is defined in (2), with $\sigma_1$and σ_2 corresponding to the ReLU and sigmoid activation functions, correspondingly. w_X and w_G represents linear transformations, while B_G and $B_\theta$indicate the biases.

$$\alpha = \sigma_2(w_\theta^t(\sigma_1(w_X^t X + w_G^t G + B_G)) + B_\theta) \quad (2)$$

The proposed attention U-Net architecture is illustrated in Fig. 6. This model replaces the backbone with pre-trained models, specifically ResNet50, Inception-v3, and VGG19. The first architecture utilizes Inception-v3 as the encoder, blending layers of convolution with max-pooling. Inception modules are used instead of the original U-Net's convolutional layers, applying 3 × 3 and 1 × 1 convolutions alongside 3 × 3 max-pooling. The alternative design utilizes VGG19 as the encoder, with convolutional and max-pooling layers, increasing the filter size after each pooling operation. The third architecture features ResNet50 utilizes residual blocks with skip connections to address the issue of vanishing gradients.

3.5. Process of classification

After performing blood vessel segmentation with the Attention U-Net model, which utilizes pre-trained backbones like ResNet50, Inception-v3, and VGG19, the segmented images are then fed into a VGG16 model for classification.

The classification process begins by passing the segmented retinal images through a series of convolutional layers in VGG16. These convolutional layers apply multiple filters to the input images to extract hierarchical features. The ReLU activation function is applied after each convolutional layer to introduce non-linearity, enhancing the model's capacity to capture intricate patterns. Max-pooling layers follow each convolution block, downsampling the feature maps and reducing their spatial dimensions while retaining key information. As illustrated in Fig. 7, the VGG16 architecture consists of multiple convolutional layers, each followed by max-pooling operations that systematically decrease the feature map size while preserving essential information for classification.

As the input passes through successive convolutional and pooling layers, VGG16 learns high-level features that represent complex attributes of the segmented images. At the conclusion of the convolutional stages, the feature maps are flattened into a one-dimensional vector, which is then passed into a sequence of fully connected layers. These fully connected layers, analogous to the decision-making part of the network, analyze the learned features and determine the relationship between them.

Finally, the output from the last fully connected layer passes through a softmax activation function, which outputs the probabilities of the image belonging to two distinct classes: "cataract" or "non-cataract." The class with the highest probability is chosen as the final classification label. This classification task effectively builds on the precise blood vessel segmentation by the Attention U-Net, ensuring that the system combines segmentation accuracy with robust diagnostic capabilities. In this way, the model segments the retinal images and accurately classifies them into their respective categories, facilitating an end-to-end approach for cataract detection.

4. Experimental results and discussion

All the experiments were carried out within a Google Colab environment, on a system with a 2.30 GHz Intel Core i7 processor, and 16 GB of RAM. The dataset was split into three parts: 70% for training, 10% for validation, and 20% for testing.

The Attention U-Net model was enhanced with three distinct backbones ResNet50, Inception-v3, and VGG19, and undergone training with 150 epochs while a learning rate set to 1×10^(-3) using a dataset of fundus images. For optimization, Inception-v3 utilized the Adam algorithm, while Stochastic Gradient Descent (SGD) was used in training ResNet50, VGG19. Model performance on the test set was measured through metrics such as accuracy, sensitivity, specificity, and precision.

Following segmentation, the VGG16 model was utilized for classification, where the segmented images were classified into "cataract" or "non-

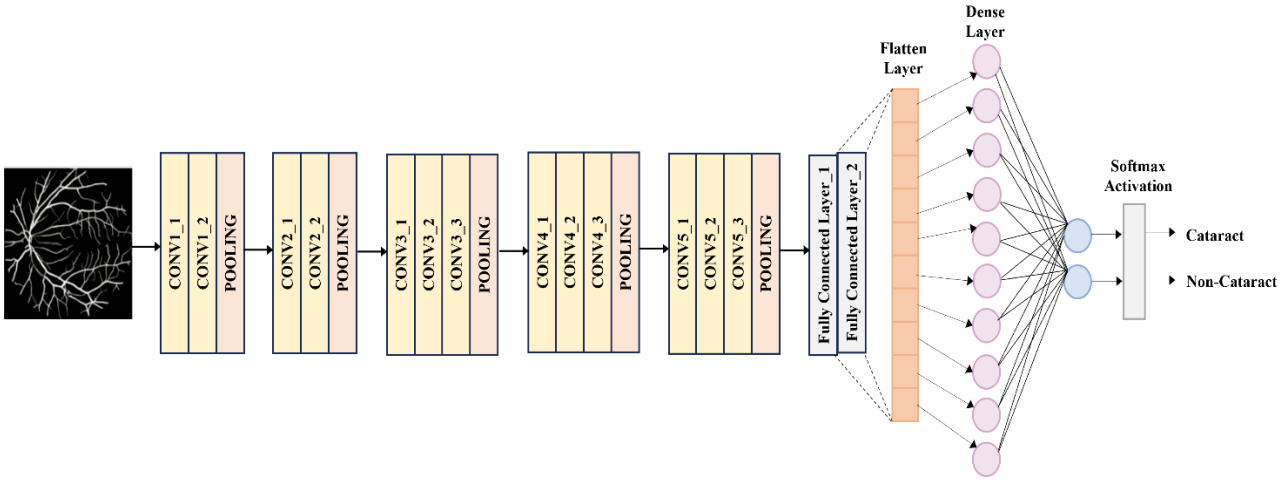


Figure. 7 VGG16 architecture for binary classification

cataract." The VGG16 classifier was also trained with the Adam optimizer, and its performance was assessed using the same metrics and AUC to ensure robust diagnostic capability across both segmentation and classification tasks.

4.1. Evaluation criteria

This study evaluates each architecture's performance using multiple metrics: Accuracy (Acc), sensitivity (Sen), specificity (Spe), precision (Pre), and the Area Under the Curve (AUC). Equations (3) through (6) represent the corresponding calculations. Acc is the count of correct predictions generated by a framework in these equations. Pre represents the proportion of correct positive predictions out of all positive predictions generated by the model. Sen assesses the accuracy of the model by measuring the ratio of properly detected positive examples to the total number of real positive instances. Spe, in contrast, measures the accuracy of identifying negative cases and indicates the model's capability to accurately detect non-positive examples. The AUC metric quantifies the discriminatory power of a classifier in distinguishing between many classes. It offers a concise overview of the Receiver Operating Characteristic (ROC) curve.

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$Sen = \frac{TP}{TP+FN} \quad (4)$$

$$Spe = \frac{TN}{TN+FP} \quad (5)$$

$$Pre = \frac{TP}{TP+FP} \quad (6)$$

4.2. Performance evaluation

Table 2 presents the segmentation performance metrics for each architecture, where Models K, L, and M correspond to U-Net frameworks using ResNet50, Inception-v3, and VGG19 as backbones, respectively. Model K (ResNet50) demonstrates the highest performance, with an accuracy (Acc) of 97.50%, sensitivity (Sen) of 98.10%, specificity (Spe) of 96.23%, and precision (Pre) of 97.20%. Model M (VGG19) and Model L (Inception-v3) show slightly lower segmentation performance, with Model M achieving an Acc of 96.10%, Sen of 98.50%, Spe of 90.86%, and Pre of 94.21%. Model L follows with an Acc of 95.80%, Sen of 97.50%, Spe of 93.63%, and Pre of 97.50%. These results suggest that Model K provides superior segmentation performance, outperforming the other models in nearly all metrics.

Table 2. Segmentation Performance Metrics for Models K, L, and M

Metrics	U-Net Backbones		
	K	L	M
Acc	97.50%	95.80%	96.10%
Sen	98.10%	97.50%	98.50%
Spe	96.23%	93.63%	90.86%
Pre	97.20%	97.50%	94.21%

Table 3. Classification Performance Metrics for Models K, L, and M

Metrics	U-Net Backbones		
	K	L	M
Acc	98.24%	96.36%	97.33%
Sen	99.77%	98.28%	99.10%
Spe	97.83%	94.34%	92.68%
Pre	98.42%	98.24%	95.12%
AUC	99.24%	97.36%	98.33%

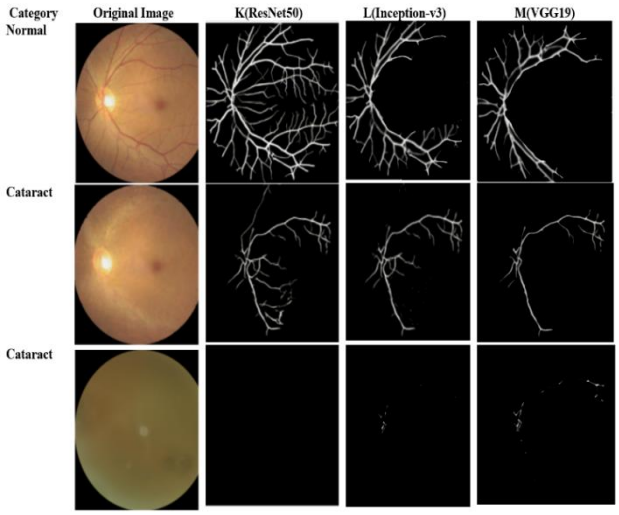


Figure. 8 Comparison of Retinal Blood Vessel Segmentation Across Different U-Net Backbones for Normal and Cataract-Affected Images

For classification, Table 3 summarizes the performance metrics where the same backbones are used in the U-Net model for segmentation, and the segmented images are classified into "cataract" or "non-cataract" using the VGG16 model. Model K again leads the classification task, achieving an Acc of 98.24%, Sen of 99.77%, Spe of 97.83%, Pre of 98.42%, and an AUC of 99.24%. Model M (VGG19) and Model L (Inception-v3) exhibit slightly lower classification performance, with Model M reaching an Acc of 97.33%, Sen of 99.10%, Spe of 92.68%, Pre of 95.12%, and an AUC of 98.33%. Model L shows the lowest classification performance, with an Acc of 96.36%, Sen of 98.28%, Spe of 94.34%, Pre of 98.24%, and an AUC of 97.36%.

These results confirm that the U-Net framework with ResNet50 as the backbone is the most effective model for both segmentation and classification tasks, achieving the highest performance across all key metrics.

Fig. 8 illustrates the segmentation results of blood vessels from retinal images using three different backbones ResNet50 (Model K), Inception-v3 (Model L), and VGG19 (Model M). The original fundus images are presented on the left, showing both normal and cataract-affected retinas. The segmentation outputs of the three models are compared for both normal and cataract images. ResNet50 captures more detailed and continuous blood vessel structures for the normal image, followed by Inception-v3 and VGG19. However, in the cataract-affected images, the models show diminished ability to capture the vessels due to the cloudiness introduced by cataracts, with the ResNet50 backbone still providing the most robust segmentation, while Inception-v3 and VGG19 exhibit less accurate and incomplete vessel segmentation, particularly in images with severe cataracts. The effectiveness of ResNet50 in segmenting vessels is most apparent in clearer images, making it the best-performing model.

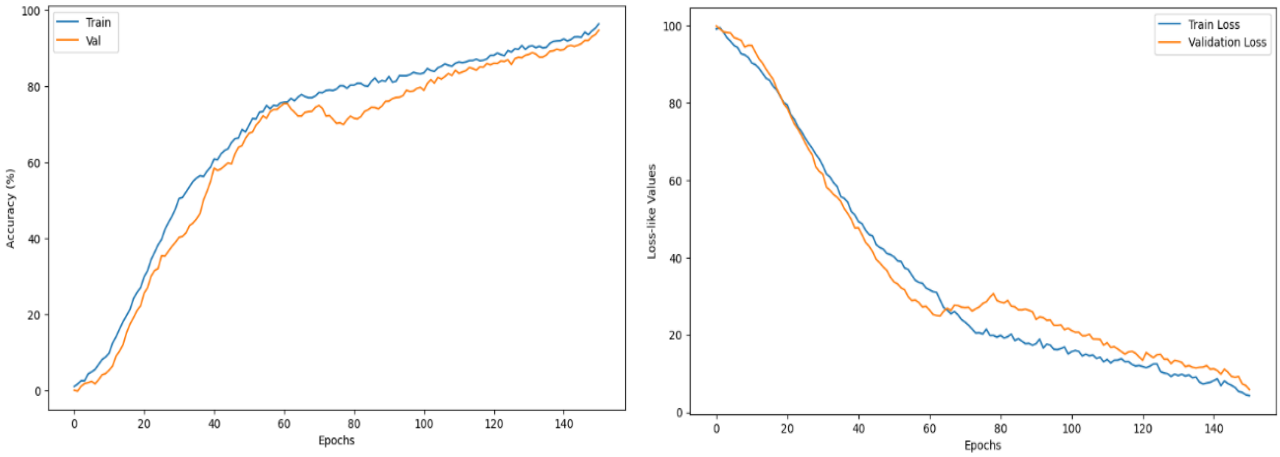
In Fig. 9, the accuracy and loss curves for the classification task using VGG16 after the segmentation stage are shown. Models K, L, and M represent ResNet50, Inception-v3, and VGG19, respectively, as the backbones used during the segmentation process. These curves depict the classification accuracy of the VGG16 model as it processes the segmented images and classifies them into "cataract" or "non-cataract" categories. The close alignment of training and validation accuracy, along with the steady decrease in loss, indicates stable training and suggests that the model is not overfitting. This highlights the model's strong generalization capability and consistent learning performance.

Fig. 10 displays the confusion matrix for the classification predictions made using VGG16 after segmentation, across three different U-Net backbones. The matrix is divided into four sections: the top-left quadrant shows True Positives (TP), representing correctly classified cataract cases, and the top-right indicates False Positives (FP), where non-cataract images are incorrectly classified as cataract. The bottom-left represents False Negatives (FN), where cataract images are misclassified as non-cataract, and the bottom-right shows True Negatives (TN), representing correctly classified non-cataract images. The results indicate that the classification model performs well, with high true positive and true negative values across all backbones, reflecting strong accuracy in both cataract and non-cataract classifications. Low values in the false positive and false negative sections suggest minimal misclassification, further highlighting the robustness of the classification technique across different backbones.

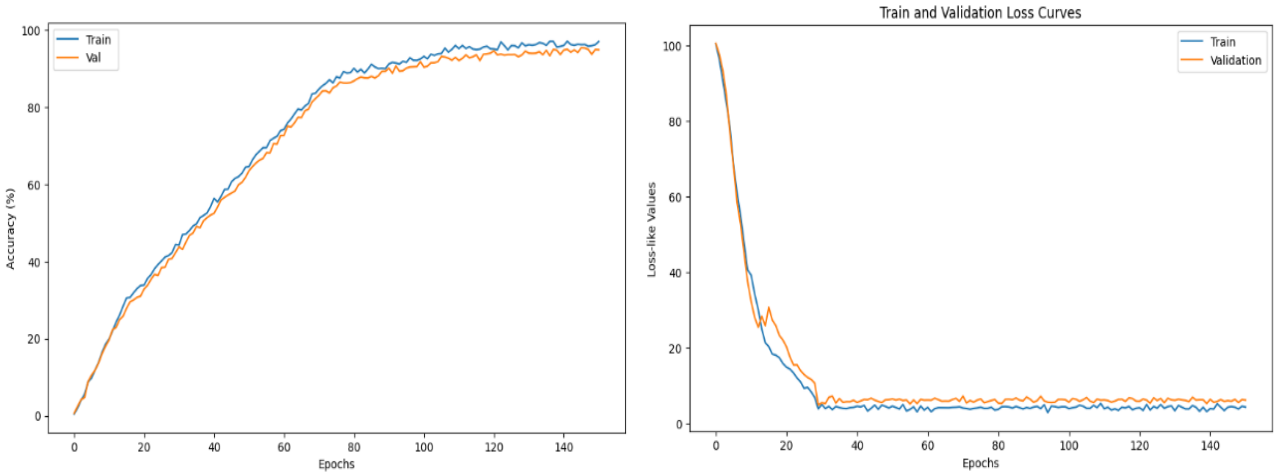
The ROC curve offers a complete assessment of the classification model’s efficacy at various thresholds, emphasizing the balance between sensitivity and specificity.

In Fig. 11, the ROC curves for the classification task using VGG16 after segmentation with ResNet50, Inception-v3, and VGG19 backbones are shown. These curves are positioned close to the top-left corner, indicating high classifier performance and accuracy.

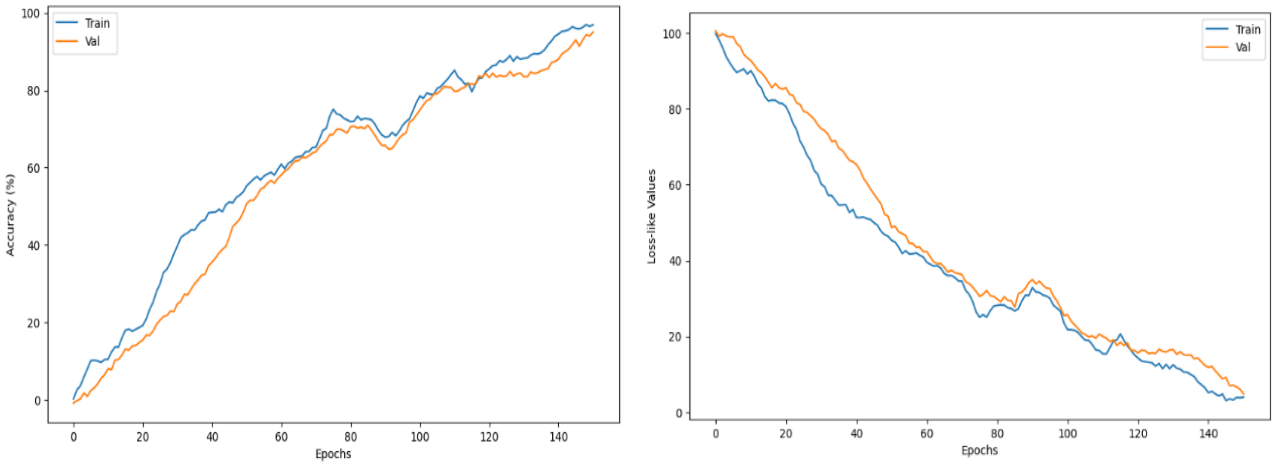
Among the models evaluated, the classification accuracy using the ResNet50-based U-Net exhibits the best performance, as its ROC curve is closer to



(a)



(b)



(c)

Figure. 9 Accuracy and Loss curves for the: (a) Inception-v3, (b) VGG-19, and (c) ResNet50 backbones

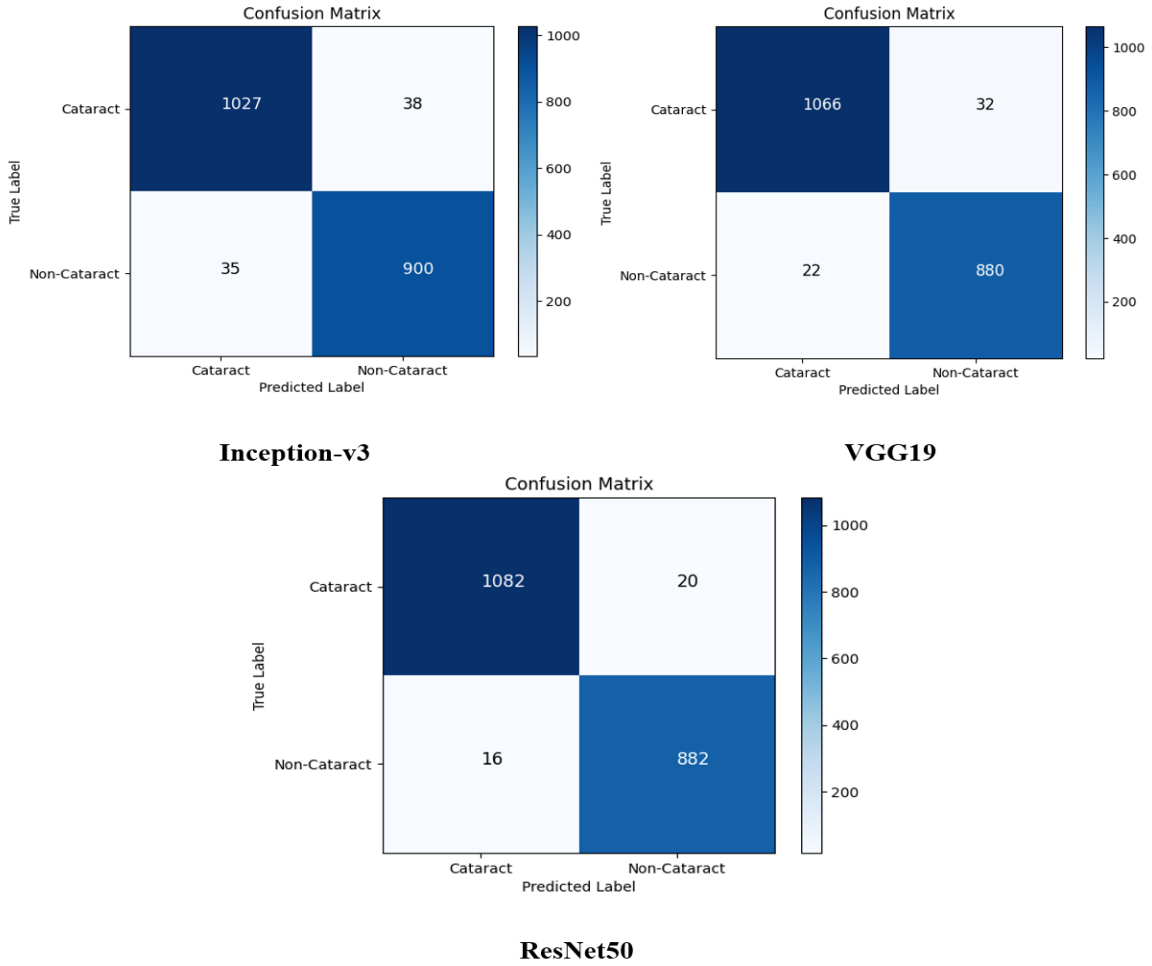


Figure. 10 Confusion Matrix Comparison Across Diverse Backbones

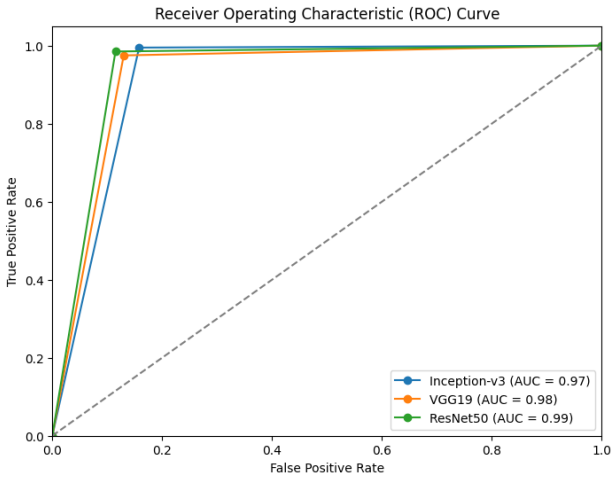


Figure. 11 ROC Curves for U-Net Models Across Different Backbones

the top-left corner compared to the Inception-v3 and VGG19-based models. This highlights superior sensitivity and specificity for the ResNet50 backbone in the classification of cataract and non-cataract images. The ResNet50-based model demonstrates exceptional classification performance, further emphasizing its robustness and reliability across the dataset.

4.3. Ablation study

In this ablation study, we examine the impact of removing attention gates from the DAUCD model, which uses three U-Net backbones: ResNet50 (Model K), Inception-v3 (Model L), and VGG19 (Model M). Attention gates, which help focus on key spatial features during segmentation, were removed, and the model performance was evaluated across segmentation and classification tasks. All other components, including the pre-trained backbones and the VGG16 classifier, remained unchanged. The performance of the models without attention gates is summarized below in Tables 4 and 5 respectively.

The ablation results demonstrate that the removal of attention gates led to a consistent drop in performance across all models. In terms of segmentation accuracy, Model K (ResNet50) saw a reduction from 97.50% to 95.50%, Model L

(Inception-v3) decreased from 95.80% to 93.10%, and Model M (VGG19) dropped from 96.10% to 94.50%.

A similar pattern was observed in classification accuracy, with Model K falling from 98.24% to 96.10%, Model L from 96.36% to 94.50%, and Model M from 97.33% to 95.80%.

The models also experienced significant declines in specificity, particularly Model K, which dropped from 97.83% to 94.10%, indicating a reduced ability to correctly identify non-cataract cases. Similarly, Model L decreased from 94.34% to 92.23%, and Model M from 92.68% to 89.63%. The AUC values followed this downward trend, with Model K dropping from 99.24% to 97.50%, Model L from 97.36% to 95.36%, and Model M from 98.33% to 96.33%. These results clearly confirm that the attention gates play a crucial role in improving the model's ability to extract meaningful features during the segmentation process, which directly contributes to more accurate classification outcomes.

This ablation study reaffirms that attention gates are essential for achieving optimal performance in the DAUCD model. The absence of attention gates led to a decline in both segmentation and classification performance across all models, with the most notable reductions in accuracy, specificity, and AUC. For instance, Model K's classification accuracy decreased by 2.14%, while its specificity dropped by 3.73%. These findings emphasize the importance of attention mechanisms in enhancing feature extraction and ensuring accurate diagnostic outcomes in cataract detection.

Table 4. Segmentation Performance Metrics for Models K, L, and M (Without Attention Gates)

Metrics	U-Net Backbones		
	K	L	M
Acc	95.50%	93.10%	94.50%
Sen	96.10%	95.30%	96.63%
Spe	93.23%	91.32%	88.29%
Pre	94.63%	95.21%	91.12%

Table 5. Classification Performance Metrics for Models K, L, and M (Without Attention Gates)

Metrics	U-Net Backbones		
	K	L	M
Acc	96.10%	94.50%	95.80%
Sen	97.30%	96.10%	97.50%
Spe	94.10%	92.23%	89.63%
Pre	95.12%	96.20%	95.12%
AUC	97.50%	95.36%	96.33%

4.4. Comparative analysis and discussion

The accuracy and loss curves shown in Fig. 9 indicate that the DAUCD method effectively avoids both underfitting and overfitting, highlighting its robustness in managing training and testing of image data. This performance surpasses the existing methods, as detailed in Table 6 and illustrated in Fig.12. The proposed DAUCD model surpassed other existing approaches with an accuracy of 98.24%, as clearly detailed in Table 6.

The proposed DAUCD model demonstrates significant advancements in cataract detection accuracy, achieving a remarkable 98.24%. This performance surpasses several state-of-the-art methods using the same ODIR-5K dataset for consistency in comparison. For instance, VGG-19, as reported by [20], achieved an accuracy of 95.27%. The CataractNetDetect model reached an accuracy of 94.21% [23]. Similarly, a ResNet-50-based architecture yielded an accuracy of 92.00% [24]. Another method, the DNN-based ensemble model, achieved a lower accuracy of 90.24% [26]. Despite these strong results, the DAUCD model outperformed them all, demonstrating its superior capability in handling the complexities of cataract detection through the innovative use of Attention U-Net for segmentation and VGG-16 for classification.

Table 6. Comparison of the proposed DAUCD Model with existing approaches.

Model	Acc (%)
VGG-19 [20]	95.27
CataractNetDetect [23]	94.21
ResNet-50 [24]	92.00
DNN-based Ensemble Model [26]	90.24
DAUCD (Proposed)	98.24

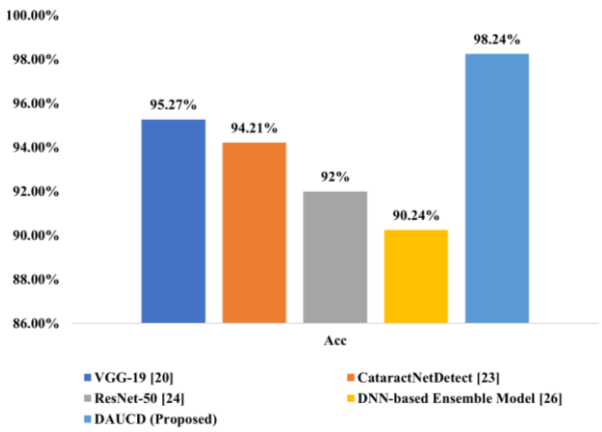


Figure. 12 Performance Comparison of the proposed DAUCD model with existing approaches in terms of Accuracy

The improved accuracy indicates that the DAUCD model effectively leverages attention mechanisms to enhance feature extraction and classification, providing more precise and reliable diagnostic results.

5. Conclusion

In conclusion, the proposed DAUCD model significantly advances cataract detection through its combination of segmentation and classification tasks, demonstrating superior accuracy and robust performance. By integrating attention mechanisms within the U-Net architecture for blood vessel segmentation and leveraging pre-trained CNN backbones such as ResNet50, Inception-v3, and VGG19, the model achieves impressive segmentation results. Furthermore, the classified outputs from the VGG16 model, based on these segmented images, attain a high classification accuracy of 98.24%. The DAUCD model effectively addresses the challenges of underfitting and overfitting, ensuring reliable diagnostic results in both segmentation and classification tasks. Its high sensitivity and specificity validate its clinical effectiveness, potentially improving early detection and treatment outcomes for cataract patients. This work underscores the promise of deep learning and attention mechanisms in medical image analysis, setting the stage for further research and advancements in this critical area of healthcare.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Maneesha Vadduri contributed to the problem analysis and writing of the article. Kuppusamy P as the co-author, formulated the problem statement and organized the manuscript, ensuring accurate interpretations. All authors have reviewed and approved the final version of the manuscript for submission.

Acknowledgments

The authors express their gratitude to the editors and reviewers.

Notations:

Variable	Notation
$\mathcal{X}$	Input to the ReLU activation function.
X	Feature map from the encoder (skip connection).
G	Gating signal from the decoder path.
X_1	Attention gate output $\mathrm{X} \cdot \alpha$.
α	Attention coefficients.
$\mathrm{w}_{\mathrm{X}}^{t}$	Transposed weight matrix for X.
w_{G}^{t}	Transposed weight matrix for *G*.
B_G	Bias for the gating signal *G*.
B_θ	Bias for the attention gate.
σ_1	ReLU activation function.
σ_2	Sigmoid activation function.
w_θ	Weight matrix in the attention gate.
w_θ^t	Transposed weight matrix for the attention gate.

References

[1] M. Vadduri and P. Kuppusamy, "Diabetic Eye Diseases Detection and Classification Using Deep Learning Techniques—A Survey", In: *Proc. of International Conference on Information and Communication Technology for Competitive Strategies*, Singapore, pp. 443-454, 2022.

[2] M. Vadduri and P. Kuppusamy, "Enhancing Ocular Healthcare: Deep Learning-Based Multi-Class Diabetic Eye Disease Segmentation and Classification", *IEEE Access*, Vol. 11, pp. 137881-137898, 2023.

[3] X.-Q. Zhang, Y. Hu, Z.-J. Xiao, J.-S. Fang, R. Higashita, and J. Liu, "Machine learning for cataract classification/grading on ophthalmic imaging modalities: a survey", *Machine Intelligence Research*, Vol. 19, No. 3, pp. 184-208, 2022.

[4] L. Cao, H. Li, Y. Zhang, L. Zhang, and L. Xu, "Hierarchical method for cataract grading based on retinal images using improved Haar wavelet", *Information Fusion*, Vol. 53, pp. 196-208, 2020.

[5] N. Nur, S. Cokrowibowo, and R. Konde, "Cataract detection in retinal fundus image using gray level co-occurrence matrix and k-nearest neighbor", In: *Proc. of International Joint Conference on Science and Engineering 2021 (IJCSE 2021)*, pp. 268-271, Atlantis Press, 2021.

[6] I. Weni, P. E. P. Utomo, B. F. Hutabarat, and M. Alfalah, "Detection of cataract based on image features using convolutional neural networks", *Indonesian Journal of Computing and Cybernetics Systems*, Vol. 15, No. 1, pp. 75-86, 2021.

[7] R. B. J. Simanjuntak, Y. Fuâ, R. Magdalena, S. Saidah, A. B. Wiratama, and I. Daâ, "Cataract classification based on fundus images using convolutional neural network", *JOIV: International Journal on Informatics Visualization*, Vol. 6, No. 1, pp. 33-38, 2022.

[8] S. Yadav and J. K. P. S. Yadav, "Enhancing Cataract Detection Precision: A Deep Learning Approach", *Traitement du Signal*, Vol. 40, No. 4, 2023.

[9] M. A. Syarifah, A. Bustamam, and P. P. Tampubolon, "Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer", In: *Proc. of AIP Conference Proceedings*, Vol. 2296, No. 1, AIP Publishing, pp. 1-5, 2020.

[10] K. Y. Son, J. Ko, E. Kim, S. Y. Lee, M.-J. Kim, J. Han, E. Shin, T.-Y. Chung, and D. H. Lim, "Deep learning-based cataract detection and grading from slit-lamp and retro-illumination photographs: Model development and validation study", *Ophthalmology Science*, Vol. 2, No. 2, pp. 100147, 2022.

[11] 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. of IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, pp. 60-65, 2017.

[12] J. Ran, K. Niu, Z. He, H. Zhang, and H. Song, "Cataract detection and grading based on combination of deep convolutional neural network and random forests", In: *Proc. of International Conference on Networking Infrastructure and Digital Content (ICNIDC)*, pp. 155-159, 2018.

[13] T. Pratap and P. Kokil, "Computer-aided diagnosis of cataract using deep transfer learning", *Biomed. Signal Process. Control*, Vol. 53, Art. No. 101533, 2019.

[14] D. Kim, T. J. Jun, Y. Eom, C. Kim, and D. Kim, "Tournament based ranking CNN for the cataract grading", In: *Proc. of 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1630-1636, 2019.

[15] M. R. Hossain, S. Afroze, N. Siddique, and M. M. Hoque, "Automatic detection of eye cataract using deep convolution neural networks (DCNNs)", In: *Proc. of IEEE Region 10 Symposium (TENSYMP)*, pp. 1333-1338, 2020.

[16] A. Imran, J. Li, Y. Pei, F. Akhtar, T. Mahmood, and L. Zhang, "Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network", *Visual Computer*, Vol. 37, No. 8, pp. 2407-2417, 2021.

[17] M. U. Raza, Z. Saeed, S. Samer, A. Mobeen, and A. Samer, "Classification of eye diseases and detection of cataract using digital fundus imaging (DFI) and inception-V4 deep learning model", In: *Proc. of International Conference on Frontiers of Information Technology (FIT)*, pp. 137-142, 2021.

[18] H. Zhang, K. Niu, Y. Xiong, W. Yang, Z. He, and H. Song, "Automatic cataract grading methods based on deep learning", *Computer Methods and Programs in Biomedicine*, Vol. 182, Art. No. 104978, 2019.

[19] J. Li, X. Xu, Y. Guan, A. Imran, B. Liu, L. Zhang, J.-J. Yang, Q. Wang, and L. Xie, "Automatic cataract diagnosis by image-based interpretability", In: *Proc. of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3964-3969, 2018.

[20] S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, "A new technique for cataract eye disease diagnosis in deep learning", *Periodicals of Engineering and Natural Sciences*, Vol. 11, No. 6, pp. 14-26, 2023.

[21] O. Sivaz and M. Aykut, "Combining EfficientNet with ML-Decoder classification head for multi-label retinal disease classification", *Neural Computing and Applications*, pp. 1-11, 2024.

[22] M. Pektaş, "Performance analysis of efficient deep learning models for multi-label classification of fundus image", *Artificial Intelligence Theory and Applications*, Vol. 3, No. 2, pp. 105-112, 2023.

[23] W. N. Ismail and H. A. Alsalamah, "A novel CatractNetDetect deep learning model for effective cataract classification through data fusion of fundus images", *Discover Artificial Intelligence*, Vol. 4, No. 1, Art. No. 54, 2024.

[24] M. A. Mohamed, M. A. Zakaria, E. Hamdi, R. E. Tawfek, T. M. Taha, Y. M. Afify, R. W. Elshinawy, and M. H. Ahmed, "Multi-Class Eye Disease Classification Using Deep Learning", In: *Proc. of 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 489-494, 2023.

[25] A. Sohail, H. Qayyum, F. Hassan, and A. U. Rahman, "CataractEyeNet: A novel deep learning approach to detect eye cataract disorder", In: *Proc. of International Conference on Information Technology and Applications (ICITA 2022)*, Singapore, pp. 63-75, 2023.

[26] A. A. Jeny, M. S. Junayed, and M. B. Islam, "Deep neural network-based ensemble model for eye diseases detection and classification", *Image Analysis and Stereology*, Vol. 42, No. 2, pp. 77-91, 2023.

[27] J. R. Ng, "Cataract Dataset", *Kaggle*, 2019. [Online]. Available:

https://www.kaggle.com/datasets/jr2ngb/cataractdataset/data
[28] T. Mahamed, "ODIR 5K Classification", *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/tanjemahamed/odir5k-classification
[29] G. Doddi, "Eye Diseases Classification", *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification
[30] T. S. Borges, "Cataract Eyes Kaggle", *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/thiagosantosborges/cataracteyeskaggle
[31] S. K. Prabhakar, "Cataract, DR, Normal, Glaucoma Fundus Images Dataset", *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/drskprabhakar/cataract-dr-normal-glaucoma-fundus-images-dataset
[32] X. Zhao, S. Wang, J. Zhao, H. Wei, M. Xiao, and N. Ta, "Application of an attention U-Net incorporating transfer learning for optic disc and cup segmentation", *Signal, Image and Video Processing*, Vol. 15, pp. 913-921, 2021.
[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", In: *Proc. of Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Munich, Germany, Oct. 5-9, 2015, Part III, Vol. 18, pp. 234-241, 2015.
[34] J. Kim, L. Tran, E. Y. Chew, and S. Antani, "Optic disc and cup segmentation for glaucoma characterization using deep learning", In: *Proc. of 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 489-494, 2019.
[35] T. Shyamalee and D. Meedeniya, "CNN based fundus images classification for glaucoma identification", In: *Proc. of 2022 2nd International Conference on Advanced Research in Computing (ICARC)*, pp. 200-205, 2022.