Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3647~3661
e-ISSN: 2548-6861

Accuracy Analysis of the Capsule Network Method in Cataract Fundus Image Classification

Salshabila Adi Putri [1], Arnawan Hasibuan [2], Riski Suwanda [3]
* Teknik Informatika, Universitas Malikussaleh
salshabila.220170132@mhs.unimal.ac.id [1], arnawan@unimal.ac.id [2], riskisuwanda@unimal.ac.id [3]

Article Info

Article history:

Received 2026-07-03
Revised 2026-08-03
Accepted 2026-08-10

Keyword:

Capsule Network (CapsNet),
Cataract,
Deep Learning,
Fundus Image,
Image Classification.

ABSTRACT

Cataract is one of the leading causes of blindness worldwide and requires early detection to prevent vision deterioration. Advances in artificial intelligence, particularly deep learning, have enabled the development of automated systems for medical image classification. However, conventional Convolutional Neural Networks (CNNs) tend to lose important spatial information due to pooling operations, which may affect classification performance. This study aims to analyze the accuracy of the Capsule Network (CapsNet) method for cataract classification using fundus images. The proposed method employs image preprocessing techniques, including resizing, data augmentation, and normalization, before training the CapsNet model. The model was developed using Python and the PyTorch framework. Performance evaluation was conducted using a Confusion Matrix and several metrics, namely Accuracy, Precision, Recall, Specificity, F1-score, and Area Under the Curve (AUC). Experimental results showed that the model achieved a training accuracy of 86.29% and a best validation accuracy of 81.34%. Furthermore, testing on 1,813 fundus images resulted in an accuracy of 89.85%, precision of 94.56%, recall of 84.53%, specificity of 95.15%, F1-score of 89.26%, AUC-ROC of 95.27%, and Average Precision (AP) of 94.71%. These findings indicate that CapsNet is capable of effectively classifying cataract fundus images. Although the obtained performance is consistent with the theoretical capability of CapsNet to preserve spatial relationships among image features, this capability was not directly evaluated in the present study.

I. INTRODUCTION

Cataract is one of the leading causes of blindness worldwide and remains a significant public health concern. According to the World Health Organization (WHO), cataracts account for approximately 45% of all global blindness cases, and the number of affected individuals continues to increase as the world's population ages [1]. Furthermore, it is estimated that more than 15 million people will experience blindness due to cataracts by 2025 [2]. Data from the Global Burden of Disease (GBD) study in 2021 revealed that the burden of cataract-related disease increased from 3.42 million Years Lived with Disability (YLDs) in 1990 to 6.55 million YLDs in 2021, representing an increase of nearly 92% [3]. These findings highlight the importance of early detection and accurate diagnosis in reducing the prevalence of cataract-induced blindness.

In Indonesia, cataracts are also the primary cause of blindness among the elderly population. Based on the Rapid Assessment of Avoidable Blindness (RAAB) report published by the Ministry of Health of the Republic of Indonesia, the prevalence of blindness among individuals aged 50 years and older is approximately 3%, with cataracts accounting for nearly 81% of all blindness cases [4]. This high prevalence emphasizes the need for intelligent diagnostic systems capable of identifying cataracts quickly, accurately, and efficiently. One promising approach is the application of

Artificial Intelligence (AI), particularly Deep Learning techniques, in medical image analysis.

Fundus image analysis is a non-invasive method widely used to evaluate retinal conditions and internal eye structures. Although cataracts primarily affect the eye lens, the resulting lens opacity can significantly reduce the visibility of retinal structures captured in fundus photographs [5]. Therefore, the visual quality and characteristics of fundus images can serve as indirect indicators for cataract identification and classification. Moreover, fundus image-based computer-aided diagnosis systems are considered practical because they do not require invasive procedures and can be integrated with automated AI-based analysis systems.

Recent advances in Deep Learning have significantly improved the performance of medical image analysis. Numerous studies have demonstrated that Deep Learning models can automatically extract discriminative features and achieve high classification accuracy across various medical imaging tasks. In ophthalmology, Deep Learning approaches have been successfully applied to detect glaucoma, diabetic retinopathy, macular degeneration, and cataracts. The DeepOpacityNet study demonstrated that fundus images could be utilized for automatic cataract detection through an explainable Deep Learning framework with promising classification performance [6]. Similarly, the Fundus Photograph-Based Cataract Evaluation Network (DCEN) showed that fundus image-based cataract classification can improve the accuracy of cataract severity assessment [7].

Most existing medical image classification studies employ Convolutional Neural Networks (CNNs) and their variants, such as Convolutional Recurrent Neural Networks (CRNNs) [8]. CNNs are highly effective in extracting visual features through convolution operations and have become the dominant architecture for image classification tasks. However, CNNs suffer from a fundamental limitation associated with pooling operations, which may lead to the loss of important spatial information within images [9]. This limitation can negatively affect the model's ability to preserve part-whole relationships, particularly in complex medical images such as retinal fundus photographs [10].

To address these limitations, Sabour et al. introduced Capsule Network (CapsNet), a Deep Learning architecture designed to preserve spatial relationships between image features through a mechanism known as Dynamic Routing Between Capsules [11]. Unlike CNNs, which represent features as scalar values, CapsNet represents features as vectors, allowing the model to retain information regarding object position, orientation, and other spatial attributes. This capability enables CapsNet to capture visual patterns more comprehensively and improve classification performance on images with complex spatial variations.

Several previous studies have demonstrated the effectiveness of CapsNet in ophthalmic image analysis. In Optical Coherence Tomography (OCT) image classification, CapsNet achieved higher classification accuracy than conventional convolutional architectures [12]. Furthermore, Govindharaj et al. reported that a hybrid CapsNet and U-Net++ architecture achieved an accuracy of 98.23% for glaucoma classification [13]. Another study employing the InceptionCaps architecture obtained an Area Under the Curve (AUC) value of 0.9556 on the RIM-ONE v2 dataset [14].

These findings indicate the strong potential of CapsNet for various ophthalmological image classification tasks.

Recent studies have shown that deep learning methods, particularly Convolutional Neural Networks (CNNs), have achieved promising performance in cataract classification using retinal fundus images. Most previous studies have employed CNN-based architectures, such as conventional CNN, AlexNet, Siamese CNN (SCNN), DeepOpacityNet, and ResNet50, which have demonstrated high classification accuracy for fundus image analysis. However, these studies primarily focused on improving classification performance and paid limited attention to preserving the spatial relationships among retinal features, as pooling operations in CNN-based architectures may reduce important spatial information. Although Capsule Network (CapsNet) has demonstrated promising results in several ophthalmic image analysis tasks, including glaucoma and Optical Coherence Tomography (OCT) image classification, its application to cataract classification using retinal fundus images remains relatively limited. Therefore, there is still a research gap in understanding the effectiveness of CapsNet for cataract fundus image classification and evaluating whether its capsule-based representation can provide competitive performance compared with existing CNN-based approaches reported in the literature [15].

Another challenge in cataract classification is the quality of retinal fundus images obtained from cataract patients. Lens opacity frequently introduces blur, reduced contrast, and optical artifacts that degrade image quality, making feature extraction more difficult and potentially affecting the performance of deep learning models. Consequently, further investigation is required to evaluate the capability of Capsule Networks in classifying cataract fundus images under realistic imaging conditions.

Based on these challenges, this study aims to analyze the classification performance of the Capsule Network (CapsNet) method for cataract detection using retinal fundus images. The proposed model was developed using Python and the PyTorch framework. Prior to model training, image preprocessing techniques, including resizing, data augmentation, and normalization, were applied to improve data quality and model robustness. Model performance was comprehensively evaluated using a Confusion Matrix and several classification metrics, including Accuracy, Precision, Recall, Specificity, F1-score, and Area Under the Curve (AUC).

The novelty of this study lies in its specific investigation of Capsule Network for binary cataract classification using retinal fundus images, a research area that remains relatively limited compared with CapsNet applications for glaucoma, diabetic retinopathy, and Optical Coherence Tomography

(OCT) image analysis. In addition, this study employs a comprehensive evaluation using multiple medical image classification metrics and investigates the classification performance of CapsNet, whose architecture is theoretically designed to preserve spatial relationships among retinal feature. Furthermore, the proposed model is implemented in a Flask-based web application to demonstrate its potential applicability as a computer-aided cataract classification system.

The main contributions of this study are summarized as follows: (1) the implementation of a Capsule Network model for binary classification of cataract and normal retinal fundus images; (2) the development of a complete classification framework consisting of image preprocessing, data augmentation, model training, validation, and independent testing; (3) a comprehensive performance evaluation using Accuracy, Precision, Recall, Specificity, F1-score, and AUC; and (4) the implementation of the trained model in a Flask-based web application for classification testing and result visualization. The findings of this study are expected to contribute to the development of CapsNet-based medical image classification and provide a reference for future studies on automated cataract screening.

II. METHOD

This study employs an experimental research method using a Deep Learning approach to analyze the accuracy of the Capsule Network (CapsNet) method in classifying cataract fundus images. The study focuses on developing and evaluating a CapsNet model to distinguish between normal fundus images and cataract fundus images based on the visual characteristics contained within the images.

The research stages include dataset collection, image preprocessing, dataset splitting into training, validation, and testing data, designing the Capsule Network architecture, model training, performance evaluation, and web-based system implementation. Training data is used for the model learning process, validation data is used to monitor model performance during training, while testing data is used to evaluate the model's ability on unseen data.

Model performance evaluation is carried out using a Confusion Matrix, which produces values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on these values, several evaluation metrics are calculated, namely Accuracy, Precision, Recall, Specificity, F1-Score, Area Under Curve (AUC), and Average Precision (AP). These metrics are used to measure the model's ability to accurately classify cataract fundus images. The best-performing model is then implemented into a web-based system using the Flask framework to facilitate automatic fundus image classification.

A. Dataset Collection

The dataset used in this study consists of retinal fundus images belonging to two categories, namely cataract and normal. The images were obtained from a publicly available retinal fundus image dataset available on the Kaggle platform. Since this study utilized a secondary public dataset, the class labels provided by the original dataset publisher were adopted directly without any additional manual annotation or relabeling. The dataset was selected because it contains sufficient variations in retinal appearance and image quality, making it suitable for evaluating deep learning models for cataract classification.

To support the development and evaluation of the proposed Capsule Network (CapsNet), the dataset was divided into three subsets: training, validation, and testing datasets. The training dataset was used to learn model parameters and optimize network weights during the learning process. The validation dataset was employed to monitor model performance during training, assist in model selection, and reduce the risk of overfitting. Meanwhile, the testing dataset was reserved exclusively for the final evaluation and was not involved in either the training or validation process, ensuring an unbiased assessment of the model's generalization capability.

The dataset distribution used in this study is presented in Table I.

TABLE I
DATASET SIZE

Image Class	Training Data	Validation Data	Testing Data	Total
Cataract	4.752	298	905	5.955
Normal	5.428	313	908	6.649
Total	10.180	611	1.813	12.604

As shown in Table I, the training dataset consists of 10,180 retinal fundus images, including 4,752 cataract images and 5,428 normal images. The validation dataset contains 611 images, comprising 298 cataract images and 313 normal images. In addition, an independent testing dataset consisting of 1,813 retinal fundus images was used exclusively to evaluate the final performance of the trained model. The testing dataset includes 905 cataract images and 908 normal images, resulting in a relatively balanced class distribution that helps reduce classification bias during model evaluation.

The separation of the training, validation, and testing datasets was performed before the model development process to ensure that no testing images were used during training or hyperparameter tuning. This separation minimizes the possibility of data leakage and enables a more reliable evaluation of the proposed model when classifying previously unseen retinal fundus images. The testing dataset was completely isolated from the training and validation processes and was used exclusively for the final performance evaluation after model training had been completed. This protocol ensured that the reported testing performance reflected the model's generalization capability on unseen retinal fundus images without introducing data leakage.

Since this study utilized a publicly available Kaggle dataset, patient identifiers and image acquisition metadata

were not available. Therefore, patient-wise separation among the training, validation, and testing subsets could not be verified. Although the dataset was partitioned before model development to prevent image-level data leakage, the possibility that multiple images from the same patient may exist across different subsets cannot be completely excluded. This limitation should be considered when interpreting the reported performance. Future studies are encouraged to employ clinically collected retinal fundus datasets with patient-level identifiers to ensure strict patient-wise data partitioning and further minimize the risk of patient-level data leakage.

It should be noted that this study relied on a publicly available Kaggle dataset. Therefore, the quality of image annotations depends on the information provided by the original dataset publisher, and no independent clinical verification or relabeling by the authors was performed. Consequently, further validation using clinically annotated retinal fundus datasets is recommended to strengthen the clinical applicability of the proposed model.

B. Image Preprocessing

Image preprocessing is an essential stage in this study to ensure that retinal fundus images are transformed into a standardized format before being processed by the Capsule Network (CapsNet) model. This stage aims to improve image quality, reduce irrelevant variations, and ensure that all input data have a consistent representation. The preprocessing pipeline was implemented using the PyTorch framework and the torchvision library in the Google Colaboratory environment.

The preprocessing process consists of image resizing, data augmentation, normalization, conversion to numerical data, and batch data determination. These steps are performed sequentially before the images are fed into the Capsule Network model.

1) Image Resize: The retinal fundus images used in this study have varying dimensions and resolutions. Therefore, all images were resized to 128 × 128 pixel**s** using the Resize() function from the PyTorch *torchvision* library. This process standardizes the input size required by the Capsule Network (CapsNet) model, enabling consistent feature extraction across all images. Furthermore, image resizing reduces computational complexity and improves training efficiency while preserving the essential retinal features needed for cataract classification. Figure 1 presents an example of the image resizing result from the original fundus image to the standardized size of 128 × 128 pixels.

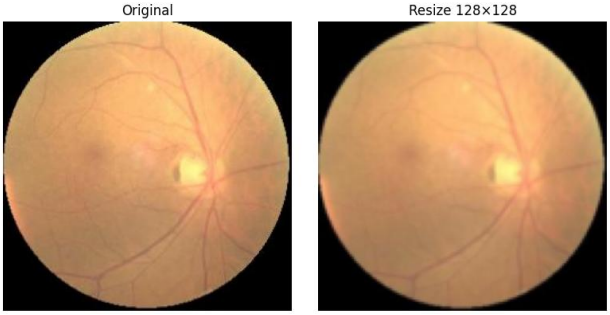


Figure 1. Resized Fundus Image (128 x128 pixels)

2) Data Augmentation**:** Data augmentation was applied to the training dataset to increase data diversity and improve the generalization capability of the Capsule Network (CapsNet) model. This technique generates new variations of retinal fundus images through several random transformations while preserving important retinal structures.

In this study, data augmentation was implemented using the PyTorch *torchvision* library and included Random Horizontal Flip (50%), Random Vertical Flip (30%), Random Rotation (15°), Color Jitter, and Random Resized Crop (85%–100% scale). These transformations were used to simulate variations in image orientation, acquisition angle, illumination conditions, and image focus that commonly occur during fundus image acquisition.

The augmentation process helps reduce overfitting and improves the model's ability to recognize retinal features under different imaging conditions. Figure 2 presents examples of the augmentation results applied to retinal fundus images.

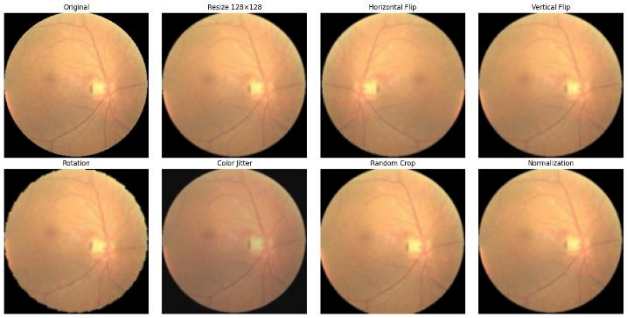


Figure 2. Data Augmentation Process

3) Normalization**:** Normalization was performed to standardize the distribution of pixel values across all retinal fundus images, thereby improving training stability and accelerating model convergence. In this study, image normalization was implemented using the Normalize() function from the PyTorch *torchvision* library with the ImageNet normalization parameters, namely mean = (0.485, 0.456, 0.406) and standard deviation = (0.229, 0.224, 0.225) for the RGB channels.

Before normalization, each image was converted into a tensor representation using the ToTensor() function. Subsequently, the pixel values were standardized according to:

$$x_{norm} = \frac{x - mean}{std}$$

where:

- x_{norm} = normalized pixel value
- x = original pixel value
- $mean$ = dataset mean
- std = dataset standard deviation

By standardizing pixel values, the Capsule Network model can learn more effectively and become less sensitive to variations in image intensity.

4) Conversion to Numerical Data: After the resizing, augmentation, and normalization stages, all retinal fundus images were converted into numerical tensor representations using the ToTensor() function from the PyTorch *torchvision* library. This step is essential because the Capsule Network (CapsNet) model can only process data in numerical form.

The ToTensor() function automatically converts image data into PyTorch tensors and scales pixel values from the range of 0–255 to 0–1 according to:

$$T(x) = \frac{x}{255}$$

where:

- $T(x)$= tensor value after conversion,
- x= original pixel value.

In addition to scaling pixel values, the tensor conversion process changes the image dimension format from $(H, W, C)$to (C, H, W), which is required by the PyTorch framework and the CapsNet architecture. Consequently, retinal fundus images with a size of $128 \times 128 \times 3$are transformed into tensors of size $3 \times 128 \times 128$. This transformation enables efficient numerical computation and feature extraction during the training and testing processes of the CapsNet model.

5) Batch Data Determination: After the dataset was loaded using the PyTorch DataLoader, batch data validation was performed to ensure that the tensor dimensions, data types, and batch size were compatible with the Capsule Network (CapsNet) architecture. This validation step is important to verify that the preprocessed images can be processed correctly during training.

In this study, a batch size of 32 was used for both training and validation. The validation results showed that the input tensor had the shape:

(B,C,H,W)

where:

- B= batch size,
- C= RGB color channels,
- H= image height,
- W= image width.

Accordingly, each batch was represented as a tensor of size:

$$32 \times 3 \times 128 \times 128$$

In addition, all image tensors were validated to ensure that they were stored in the float32 format, which is suitable for deep learning computations on both CPU and GPU platforms. The validation process also confirmed that the normalized pixel values were distributed within the expected range, indicating that the preprocessing pipeline had been successfully applied. These results demonstrate that the dataset was correctly loaded and prepared for the CapsNet training process.

Overall, each preprocessing stage contributes to improving the performance and robustness of the proposed Capsule Network model. Image resizing standardizes the input dimensions while reducing computational complexity, enabling consistent feature extraction across all retinal fundus images. Data augmentation increases image diversity by simulating variations in image orientation, illumination, and scale, thereby reducing overfitting and improving the model's generalization capability. Image normalization stabilizes the distribution of pixel values, facilitating faster convergence and more stable optimization during training. Finally, tensor conversion and batch organization enable efficient numerical computation and parallel processing within the PyTorch framework. Collectively, these preprocessing procedures ensure consistent data representation and contribute to the stability and classification performance of the proposed CapsNet model.

C. Capsule Network Architecture

Capsule Network (CapsNet) is a deep learning architecture introduced by [11]. to address the limitations of Convolutional Neural Networks (CNNs) in preserving spatial relationships between features. Unlike CNNs, which represent features as scalar values, CapsNet represents features as vectors, allowing the spatial relationships between objects to be better preserved.

In this study, the Capsule Network architecture consists of a Convolution Layer, Rectified Linear Unit (ReLU) activation function, Primary Capsule Layer, Digit Capsule Layer, Dynamic Routing mechanism, and Squashing Function. The overall architecture is designed to extract discriminative retinal features and classify fundus images into cataract and normal categories.

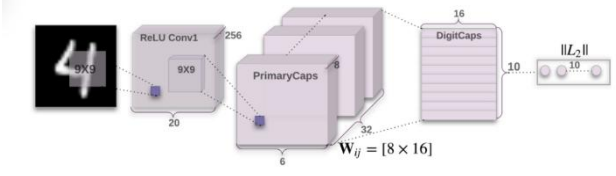


Figure 3. Capsule Network Architecture

1) Convolution Layer: The Convolution Layer is responsible for extracting initial features from retinal fundus images, such as edges, textures, and other visual patterns. The

result of the convolution process is a feature map, which is then used as input for the next layer..

$$(y * w)[i,j] = \sum_m \sum_n x\ [i+m, j\ +\ n]\ .w\ [m,n]$$

where $y(i,j)$is the output value at position (i, j), $x(i+m, j+n)$is the input pixel value, and $w(m,n)$is the convolution kernel or filter.

2) Activation Function (ReLU): After the convolution process, the Rectified Linear Unit (ReLU) activation function is applied to introduce non-linearity into the model, enabling it to learn more complex patterns..

$$f(x) = max\ (0,x)$$

where $f(x)$is the output of the activation function and $x$is the input value.

3) Primary Capsule Layer*:* The Primary Capsule Layer transforms the feature maps produced by the convolution layer into a set of capsule vectors. Each capsule outputs a vector whose length represents the probability of the presence of a specific feature, while the direction of the vector encodes the characteristics and spatial information of that feature.

$$u_i = w_i x_i + b_i$$

where $u_i$is the output of the i-th capsule, $W_i$is the weight matrix, $x_i$is the capsule input, and $b_i$is the bias term.

4) Digit Capsule Layer: The Digit Capsule Layer is a higher-level capsule layer that receives outputs from the Primary Capsule Layer. In this study, the layer consists of two output capsules representing the normal and cataract classes.

$$\hat{u}_{j|i} = W_{ij} u_i$$

where $\hat{u}_{j|i}$is the predicted vector from capsule i to capsule j, $W_{ij}$is the transformation matrix, and $u_i$is the output of the previous capsule layer.

5) Dynamic Routing: Dynamic Routing is used to determine the relationships between capsules based on the level of agreement among detected features. This mechanism allows the model to preserve spatial relationships between features, ensuring that important information in retinal fundus images is maintained during classification.

$$s_j = \sum_i c_{ij}\ \hat{u}_{j|i}$$

where $s_j$is the input to the j-th higher-level capsule, $\hat{u}_{j|i}$is the predicted output vector from capsule $i$to capsule j, and $c_{ij}$is the coupling coefficient representing the probability of connection between capsule $i$and capsule j.

6) Squashing Function: The squashing function is used to constrain the length of the output vector to the range between 0 and 1 without changing its direction. The longer the vector, the higher the probability of the class represented by the capsule.

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|}\ .$$

where $v_j$is the output vector of capsule j, $s_j$is the input to capsule j, and $\|\ s_j\ \|$is the magnitude of vector s_j.

D. Model Training Configuration

The Capsule Network model was trained using the Python programming language and the PyTorch framework on the Google Colaboratory platform, utilizing GPU acceleration to improve computational efficiency. The dataset that had undergone the preprocessing stage was divided into training, validation, and testing sets to support the model training and evaluation processes.

During training, the Adam (*Adaptive Moment Estimation*) optimizer was employed to update model weights based on the gradients generated through the backpropagation process. In addition, a *ReduceLROnPlateau* learning rate scheduler was implemented to automatically adjust the learning rate when the model performance ceased to improve. An *early stopping* mechanism was also applied to reduce the risk of overfitting by terminating the training process when no improvement was observed after a predefined number of epochs.

Table II
OPTIMIZER CONFIGURATION

No	Parameter	Value
1	Optimizer	Adam
2	Learning rate	0.001
3	Weight decay	0.0001
4	Scheduler	ReduceLROnPlateau
5	Mode	min
6	Factor	0.5
7	Patience	3

In addition to the optimizer parameters, the hyperparameters used during the training process are presented in Table III.

TABLE III
MODEL TRAINING HYPERPARAMETERS

No	Hyperparameter	Value
1	Number of Class	2
2	Batch SIZE	32
3	Epoch	50
4	Learning rate	0.001
5	Optimizer	Adam
6	Weight decay	0.0001
7	Routing Iteration	3
8	Early Stopping Patience	10
9	Mixed Precision Training	True

As shown in Table III, the model was trained for 50 epochs using a batch size of 32 and a learning rate of 0.001. Model optimization was performed using the Adam optimizer with a weight decay value of 0.0001 to improve the model's generalization capability. The Capsule Network architecture employed three routing iterations to optimize the dynamic routing process between capsules. Furthermore, an early stopping mechanism with a patience value of 10 was applied to prevent overfitting and reduce unnecessary training time. Mixed precision training was also utilized to improve memory efficiency and accelerate GPU computation during the training process.

Throughout the training phase, model performance was monitored using the validation dataset by observing the loss and accuracy values at each epoch. The best-performing model was subsequently selected and evaluated using the testing dataset to assess its capability in classifying normal and cataract fundus images.

E. Model Evaluation

Model evaluation was conducted to measure the capability of the Capsule Network in classifying fundus retinal images into two categories, namely normal and cataract. The evaluation process was performed using the testing dataset, which was not involved in either the training or validation stages, thereby providing an unbiased assessment of the model's generalization ability on unseen data.

In this study, model performance was evaluated using a Confusion Matrix, which produces four classification outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values were subsequently used to calculate several evaluation metrics, including Accuracy, Precision, Recall, Specificity, F1-Score, Area Under the Curve (AUC), and Average Precision (AP).

TABLE IV
CONFUSION MATRIX

Actual Class	Predicted Cataract	Predicted Normal
Cataract	*True Positive* (TP)	*False Negative* (FN)
Normal	*False Positive* (FP)	*True Negative* (TN)

In the confusion matrix, *True Positive* (TP) represents the number of cataract images correctly classified as cataract. *True Negative* (TN) represents the number of normal images correctly classified as normal. *False Positive* (FP) refers to the number of normal images incorrectly classified as cataract, while *False Negative* (FN) represents the number of cataract images incorrectly classified as normal.

1) Accuracy: Accuracy measures the overall proportion of correctly classified samples among all testing data.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

2) Precision: Precision measures the proportion of correctly predicted positive samples among all positive predictions generated by the model.

$$Precision = \frac{TP}{TP + FP}$$

3) Recall: Recall measures the model's ability to identify all actual positive samples correctly.

$$Recall = \frac{TP}{TP + FN}$$

4) Specificity: Specificity measures the model's ability to correctly identify negative samples.

$$Specificity = \frac{TN}{FP + TN}$$

5) F1-Score: F1-Score is the harmonic mean of Precision and Recall and is used to evaluate the balance between both metrics.

$$F1 - Score = 2x \frac{Precision\ x\ Recall}{Precision + Recall}$$

6) Area Under the Curve (AUC): The *Area Under the Curve* (AUC) is used to evaluate the model's ability to distinguish between positive and negative classes based on the *Receiver Operating Characteristic* (ROC) curve. AUC values range from 0 to 1, where values closer to 1 indicate better classification performance.

7) Average Precision (AP): *Average Precision* (AP) is used to evaluate model performance based on the *Precision-Recall* curve. This metric represents the average precision value across different recall levels and provides a more comprehensive assessment of classification performance, particularly for imbalanced datasets.

III. RESULTS AND DISCUSSION

This section presents the results obtained from the implementation and evaluation of the proposed Capsule Network (CapsNet) model for cataract fundus image classification. The discussion begins with the image preprocessing stage, followed by an analysis of the training performance of the CapsNet model. Subsequently, the classification results are evaluated using various performance metrics, including Accuracy, Precision, Recall, Specificity, F1-Score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and Average Precision (AP). Furthermore, confusion matrix analysis and graphical evaluations through ROC and Precision–Recall curves are presented to provide a comprehensive assessment of the model's performance. Finally, the best-performing model is implemented in a web-based application for automated cataract fundus image classification.

A. Training Performance of CapsNet

The training process was conducted for 50 epochs using the training dataset, while the validation dataset was used to evaluate the model's ability to generalize to unseen data. Model performance was monitored through training and validation loss, training and validation accuracy, and learning

rate scheduling. These metrics were used to analyze the learning behavior and convergence of the proposed Capsule Network (CapsNet) model.

1) Training and Validation Loss:

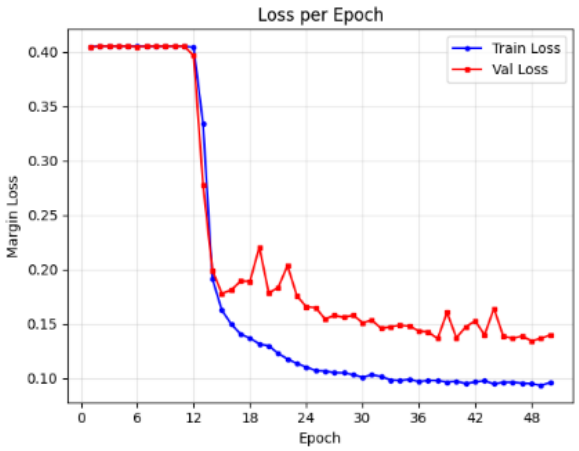


Figure 4. Loss Per Epoch Graph

Figure 4 presents the training and validation loss obtained during the training process. At the beginning of training, both training and validation loss were relatively high, with values around 0.40. This condition indicates that the model was still in the early learning stage and had not yet learned representative features from retinal fundus images.

A significant decrease in loss can be observed between epochs 12 and 14. This reduction indicates that the CapsNet model began to successfully learn important discriminative features from the retinal fundus images, resulting in more accurate predictions. After this stage, the training loss continued to decrease gradually and reached approximately 0.094 at the end of training.

Similarly, the validation loss exhibited an overall downward trend throughout the training process. Although several fluctuations were observed, the validation loss eventually converged to approximately 0.138. These fluctuations are expected because validation data were not used during model optimization. Nevertheless, the overall decreasing trend indicates that the model maintained stable learning behavior and good generalization capability.

Furthermore, the difference between training loss and validation loss remained relatively moderate throughout the training process. This result suggests that the CapsNet model did not experience significant overfitting and was able to maintain consistent performance on unseen validation data.

2) Training and Validation Accuracy

Figure 5 illustrates the progression of training and validation accuracy during the training process. At the initial epochs, both accuracy values remained relatively low, ranging from approximately 48% to 52%. This behavior indicates that the model was still learning the basic characteristics of retinal fundus images and had not yet developed an effective classification capability.

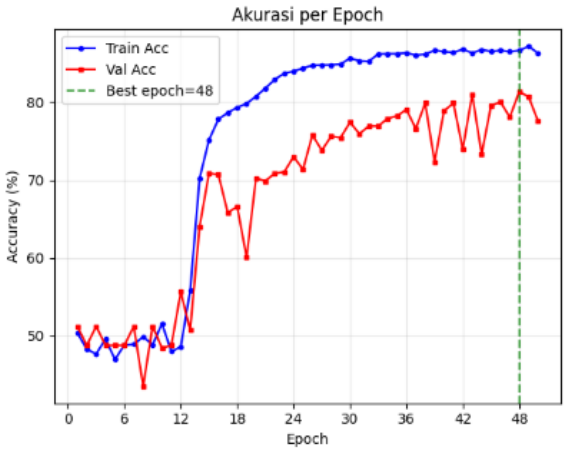


Figure 5. Accuracy per Epoch Graph

A substantial increase in accuracy was observed around epoch 13, corresponding to the period where the loss values decreased significantly. This result indicates that the CapsNet model successfully learned discriminative retinal features associated with cataract and normal classes. After this stage, both training and validation accuracy continued to improve and gradually stabilized.

At the end of training, the model achieved a training accuracy of 86.29% and a validation accuracy of 81.34%. The highest validation accuracy was obtained at epoch 48, which was selected as the best-performing model during the training process. The relatively stable gap between training and validation accuracy indicates that the model achieved good generalization performance without exhibiting severe overfitting.

Overall, the accuracy curves demonstrate that the proposed CapsNet model was capable of learning meaningful retinal fundus features and achieved satisfactory classification performance for cataract detection.

3) Learning Rate Schedule

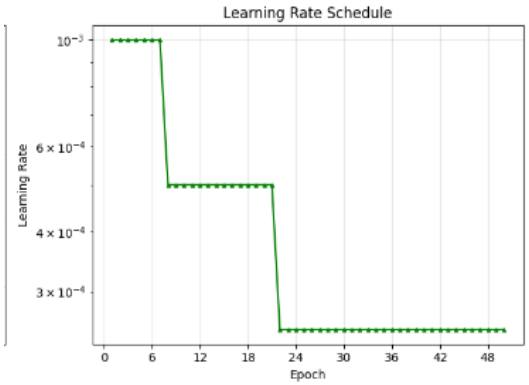


Figure 6. Learning Schedule Graph

Figure 6 shows the learning rate adjustments applied throughout the training process. The initial learning rate was set to 0.001 and was automatically updated using the ReduceLROnPlateau scheduler. This strategy reduces the learning rate whenever the validation accuracy fails to improve for several consecutive epochs.

Based on the graph, the learning rate was first reduced from 0.001 to 0.0005 and subsequently decreased to 0.00025. These adjustments occurred when the improvement in validation accuracy became stagnant, allowing the optimizer to perform finer parameter updates and preventing large oscillations near the optimum solution.

The gradual reduction of the learning rate contributed to the stabilization of both accuracy and loss values during the later stages of training. As a result, the model was able to converge more effectively and achieve its best validation performance at epoch 48. The combination of learning rate scheduling and model checkpointing therefore played an important role in improving training stability and overall classification performance.

TABLE V
TRAINING PERFORMANCE SUMMARY

Metric	Value
Best Training Accuracy	86.29%
Best Validation Accuracy	81.34%
Final Training Loss	0.094
Final Validation Loss	0.138
Best Epoch	48

Based on the training results summarized in Table V, the proposed CapsNet model achieved a best training accuracy of 86.29% and a best validation accuracy of 81.34%. The final training loss and validation loss were 0.094 and 0.138, respectively, indicating that the model successfully converged during the learning process. Furthermore, the highest validation performance was obtained at epoch 48, which was selected as the final model for testing and deployment. The relatively small gap between training and validation performance suggests that the model achieved good generalization capability and did not suffer from severe overfitting.

B. Classification Performance Evaluation

The performance of the proposed Capsule Network (CapsNet) model was evaluated using the independent testing dataset consisting of 1,813 retinal fundus images, including 905 cataract images and 908 normal images. Model evaluation was conducted using several classification metrics derived from the confusion matrix, namely Accuracy, Precision, Recall (Sensitivity), Specificity, F1-Score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and Average Precision (AP).

1) Confusion Matrix Analysis

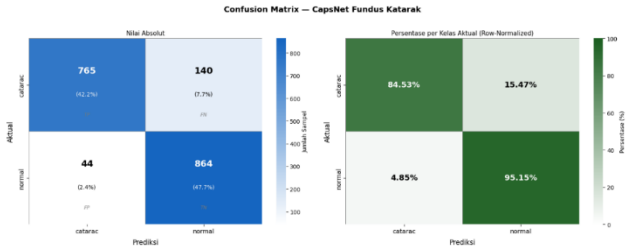


Figure 7. Confusion Matrix

The confusion matrix was used to compare the predicted labels with the ground-truth labels of the testing dataset. In this study, the cataract class was considered the positive class, while the normal class was considered the negative class.

Based on the testing results, the model correctly classified 765 cataract images as cataract (True Positive) and 864 normal images as normal (True Negative). Meanwhile, 44 normal images were incorrectly classified as cataract (False Positive), and 140 cataract images were incorrectly classified as normal (False Negative).

The confusion matrix indicates that the proposed CapsNet model was able to correctly identify the majority of cataract and normal fundus images. Although several false negative cases were observed, the overall classification performance demonstrates good discrimination capability between the two classes.

2) Quantitative Performance Evaluation

TABLE VI
PERFORMANCE EVALUATION OF THE PROPOSED CAPSNET MODEL

No	Evaluation Metric	Result
1	*Accuracy*	89,85%
2	*Precision*	94,56%
3	*Recall* (*Sensitivity*)	84,53%
4	*Specificity*	95,15%
5	*F1-score*	89,26%
6	AUC-ROC	95,27%
7	AP (AUC-PR)	94,71%

The model achieved an overall accuracy of 89.85%, indicating that most retinal fundus images were classified correctly. The precision value of 94.56% demonstrates that the majority of images predicted as cataract were indeed cataract cases, reflecting a low false-positive rate.

The recall value of 84.53% indicates that the model successfully detected most cataract cases in the testing dataset. In medical image classification, recall is particularly important because it reflects the ability of the system to identify diseased patients. Furthermore, the specificity value of 95.15% shows that the model was highly effective in recognizing normal fundus images and minimizing incorrect diagnoses of healthy cases.

The obtained F1-Score of 89.26% demonstrates a good balance between precision and recall, indicating stable

classification performance across both classes. Overall, these results suggest that the proposed CapsNet model provides reliable and robust performance for cataract fundus image classification.

Although the testing accuracy (89.85%) was slightly higher than the best training accuracy (86.29%), this observation does not necessarily indicate data leakage. During the training stage, various data augmentation techniques, including random rotation, horizontal and vertical flipping, color jitter, and random resized cropping, were applied to increase image variability and improve the model's robustness. Consequently, the training process became more challenging because the model was required to learn from more diverse image variations. In contrast, the testing dataset underwent only image resizing and normalization without augmentation, resulting in more consistent image characteristics. Furthermore, the training, validation, and testing datasets were separated before model development, and no testing images were involved in either the training process or hyperparameter tuning. Therefore, the higher testing accuracy is more likely attributable to differences in preprocessing and data characteristics rather than data leakage. Nevertheless, further evaluation using independent clinically annotated datasets is recommended to confirm the generalization capability of the proposed model under real-world clinical conditions.

Based on the confusion matrix, the proposed CapsNet model correctly classified the majority of retinal fundus images; however, a number of false-positive and false-negative predictions were still observed. False-positive predictions indicate that several normal retinal fundus images were incorrectly classified as cataract, whereas false-negative predictions correspond to cataract images that were incorrectly identified as normal. In medical screening applications, false-negative cases are particularly important because they may delay the identification of patients who require further clinical examination.

The observed misclassifications may be associated with variations in retinal fundus image quality, including non-uniform illumination, low contrast, image blur, and natural anatomical variations among patients. In addition, mild cataract cases may exhibit subtle retinal characteristics that are visually similar to normal fundus images, making the classification task more challenging. Since this study did not include a detailed visual inspection of the misclassified images, the exact causes of these prediction errors could not be determined. Future studies should perform qualitative analyses of false-positive and false-negative cases to better understand the characteristics of difficult retinal fundus images and further improve the classification performance of the proposed model.

3) ROC and Precision–Recall Curve Analysis

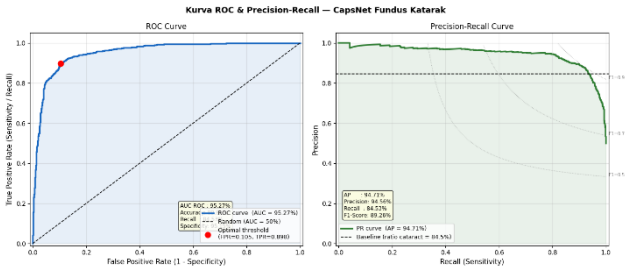


Figure 8. ROC Curve and Precision-Recall Curve CapsNet Model

Figure 8 presents the Receiver Operating Characteristic (ROC) curve and Precision–Recall (PR) curve of the proposed Capsule Network (CapsNet) model. These graphical evaluations were used to assess the model's classification capability across various decision thresholds and provide a more comprehensive performance analysis beyond confusion matrix-based metrics.

The proposed CapsNet model achieved an AUC-ROC value of 95.27%, indicating excellent discriminative capability in distinguishing cataract fundus images from normal fundus images. The ROC curve is positioned close to the upper-left corner of the graph, demonstrating that the model is capable of maintaining a high True Positive Rate (TPR) while keeping the False Positive Rate (FPR) relatively low. This result confirms that the model can effectively separate positive and negative classes under different classification thresholds.

In addition, the Precision–Recall curve produced an Average Precision (AP) value of 94.71%, indicating that the model consistently maintained high precision across different recall levels. The high AP score demonstrates that most images predicted as cataract were indeed cataract cases, while the number of incorrect positive predictions remained relatively low. This characteristic is particularly important in medical image classification, where accurate disease detection is essential for supporting clinical decision-making.

Overall, the high AUC-ROC and AP values indicate that the proposed Capsule Network model achieved strong classification performance, excellent class-separation capability, and good generalization on unseen testing data. These findings demonstrate that the proposed CapsNet model achieved strong classification performance on retinal fundus images. Although the obtained results are consistent with the theoretical advantages of Capsule Networks in preserving spatial feature relationships, this capability was not directly evaluated in the present study and therefore requires further empirical investigation.

Although the proposed Capsule Network (CapsNet) achieved strong classification performance, the capability of preserving spatial relationships among retinal features was not directly evaluated in this study through feature visualization or comparative experiments with CNN-based architectures. Therefore, the effectiveness of CapsNet in

maintaining spatial information cannot be empirically confirmed based solely on the experimental results presented.

Nevertheless, the satisfactory classification performance obtained in this study is consistent with the theoretical design of CapsNet. Unlike conventional CNNs that represent image features as scalar activations, CapsNet represents image features as vectors and employs dynamic routing between capsules to model part-whole relationships among visual structures. This mechanism enables the network to encode not only the existence of retinal features but also their spatial attributes, such as orientation and relative position, which may be beneficial for retinal fundus image analysis.

In this study, the proposed model achieved an accuracy of 89.85%, an AUC-ROC of 95.27%, and an Average Precision of 94.71%, indicating that the model successfully learned discriminative retinal representations for cataract classification. However, these performance metrics alone do not constitute direct evidence that the preserved spatial relationships contributed to the observed classification performance.

Therefore, further studies should incorporate explainable artificial intelligence (XAI) techniques, feature visualization methods, capsule activation analysis, or direct comparisons with modern CNN architectures trained under identical experimental settings. Such analyses would provide stronger empirical evidence regarding the contribution of capsule representations and dynamic routing to retinal feature learning and cataract fundus image classification.

In addition, this study evaluated the proposed CapsNet model using a single train-validation-test split on a publicly available Kaggle dataset and did not employ k-fold cross-validation or external validation using clinically collected retinal fundus datasets. Therefore, although the obtained results demonstrate promising classification performance, the reported metrics may not fully reflect the model's generalization capability across different patient populations, imaging devices, or clinical environments. Future research should perform k-fold cross-validation and evaluate the proposed model on independent multicenter clinical datasets to provide stronger evidence regarding its robustness, reliability, and clinical applicability.

Furthermore, although multiple evaluation metrics, including Accuracy, Precision, Recall, Specificity, F1-score, AUC-ROC, and Average Precision, were employed to assess the proposed CapsNet model, the present study did not include statistical analyses such as confidence interval estimation or statistical significance testing. Therefore, the stability of the reported performance and the statistical significance of the observed differences relative to previous studies cannot be quantitatively confirmed. This limitation should be considered when interpreting the comparative performance of the proposed model. Future research should incorporate statistical validation methods, such as confidence interval estimation, bootstrap analysis, or appropriate statistical significance tests, to provide a more rigorous evaluation of model reliability and enable more objective comparisons with other deep learning architectures.

TABLE VII
COMPARISON OF THE PROPOSED CAPSNET WITH PREVIOUS STUDIES

Reference	**Model**	**Dataset**	**Main Objective**	**Performance**
Elsawy et al. (2023)	DeepOpacityNet (CNN-based)	Color fundus photographs	Cataract detection with explainable visualization	High classification performance with heatmap visualization
Gao et al. (2023)	Dual-Stream ResNet50	1,340 retinal fundus images	Cataract type classification and severity grading	Accuracy: 97.62%, Sensitivity: 98.20%, F1-score: 94.01%
Mulyasari et al. (2024)	AlexNet (CNN)	4,217 Kaggle fundus images	Multi-disease retinal classification	Precision: 87%, Recall: 88%, F1-score: 88%
Rahmadwati et al. (2024)	Siamese CNN (SCNN)	Retinal fundus images	Binary cataract classification	Accuracy: 91.25%, Precision: 91%
Firdaus et al. (2022)	CNN	512 Kaggle eye images	Binary cataract classification	Accuracy: 99.74%
This Study	Capsule Network (CapsNet)	12,604 Kaggle retinal fundus images	Binary cataract classification	Accuracy: 89.85%, Precision: 94.56%, Recall: 84.53%, Specificity: 95.15%, F1-score: 89.26%, AUC: 95.27%

Table VII presents a comparison between the proposed Capsule Network model and several previous studies on cataract classification using retinal fundus images. Previous studies predominantly employed convolutional neural network (CNN)-based architectures, including DeepOpacityNet, ResNet50, AlexNet, Siamese CNN, and conventional CNN models. These approaches demonstrated promising classification performance; however, most of them relied on convolution and pooling operations for feature extraction.

In contrast, this study investigates the application of Capsule Network (CapsNet), whose architecture is theoretically designed to preserve spatial relationships among image features through vector representations and dynamic routing. The proposed model achieved a testing accuracy of 89.85% on the selected retinal fundus image dataset. However, the results presented in Table VII should not be interpreted as a direct performance comparison because the previous studies were conducted using different datasets, image acquisition conditions, preprocessing pipelines, data augmentation strategies, classification objectives, and evaluation protocols. Consequently, differences in the reported performance may be influenced by experimental settings rather than by the network architecture itself.

Therefore, the comparison in Table VII is intended to position the proposed method within the existing literature and to demonstrate that Capsule Network has been relatively less explored for cataract fundus image classification compared with conventional CNN-based architectures. The objective of this study was to evaluate the feasibility and classification performance of CapsNet rather than to establish its superiority over other deep learning models. A direct and objective comparison would require all competing architectures, such as ResNet, DenseNet, EfficientNet, and Vision Transformer, to be trained and evaluated using the same dataset, preprocessing procedures, hyperparameter settings, and evaluation protocol. Such an experimental comparison was beyond the scope of the present study and is recommended as future work to provide a fair assessment of the relative strengths and limitations of CapsNet.

TABLE VIII
COMPUTATIONAL CONFIGURATION AND EFFICIENCY CONSIDERATIONS OF THE PROPOSED CAPSNET MODEL

Metric	**Value**
Training Environment	Google Colaboratory
Deep Learning Framework	PyTorch
Total Training Time	152.3 minutes
Best Validation Accuracy	81.34%
Testing Dataset Size	1,813 retinal fundus images
Total Inference Time	119.6 seconds
Throughput	15 images/second

Table VIII summarizes the computational performance of the proposed Capsule Network (CapsNet) model during both the training and testing stages. The model was implemented using the PyTorch deep learning framework in the Google Colaboratory environment. During training, the proposed model required a total training time of 152.3 minutes to achieve the best validation accuracy of 81.34%. This training duration reflects the computational cost required for optimizing the network parameters through multiple learning iterations while maintaining satisfactory classification performance on the validation dataset.

During the testing stage, the trained CapsNet model processed 1,813 retinal fundus images in 119.6 seconds, corresponding to an average throughput of approximately 15 images per second. These results indicate that the proposed model is capable of performing retinal fundus image classification within a relatively short inference time, suggesting that CapsNet has the potential to support offline computer-aided diagnosis systems for cataract screening. Although the inference speed is promising, practical deployment in real-world clinical environments should also consider hardware specifications, computational resources, and system scalability.

Despite the satisfactory computational performance observed in this study, several computational efficiency metrics were not systematically evaluated, including the number of trainable parameters and memory consumption. These metrics are important because they influence the feasibility of deploying deep learning models in healthcare facilities with limited computational resources or mobile screening devices. Therefore, future research should include a comprehensive computational efficiency analysis and compare the proposed CapsNet model with other state-of-the-art deep learning architectures, such as ResNet, DenseNet, EfficientNet, and Vision Transformer, to provide a more objective assessment of the trade-off between classification performance and computational cost.

C. Web-Based Implementation Results

To demonstrate the practical applicability of the proposed Capsule Network (CapsNet) model, the trained model was deployed into a web-based application developed using the Flask framework. The application integrates the complete classification pipeline, including image upload, preprocessing, model inference, and result visualization. Through this implementation, users can classify retinal fundus images automatically without requiring direct interaction with the underlying deep learning model.

1) Web Application Interface

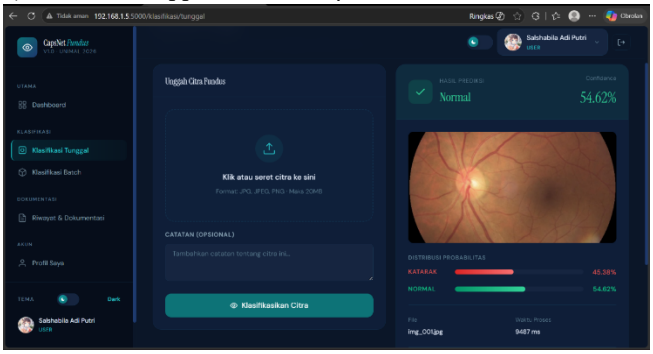


Figure 9. Single Classification Page

The developed web application provides several functional modules, including user authentication, dashboard management, single-image classification, batch-image classification, and classification history. After logging into the system, users can upload retinal fundus images through the provided interface. The uploaded images are

automatically processed using the same preprocessing procedures employed during model development, ensuring consistency between the training and deployment environments.

For single-image classification, the system predicts whether the uploaded fundus image belongs to the Cataract or Normal class and displays the corresponding confidence score. In addition, the batch classification feature allows multiple fundus images to be processed simultaneously, improving efficiency when handling larger datasets.

2) Functional Testing Result

TABLE IX
FUNDUS IMAGE TESTING RESULTS

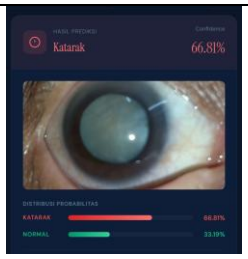
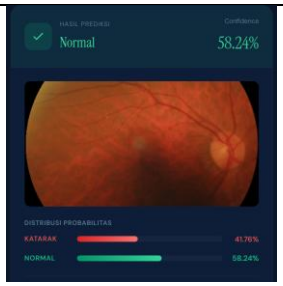
Classification Results	The Real Class	Information	Classification Accuracy Level
	Cataract	Correct	66.81%
	Normal	Correct	58.24%
	Diabetic Fundus Image	Wrong	61,94%
	Not detected	Correct	28%

Table IX presents several classification examples generated by the developed web-based application using retinal fundus images. The results demonstrate that the system was able to successfully perform automatic classification and provide prediction confidence scores for each uploaded image. Cataract fundus images were correctly identified as cataract, while normal fundus images were also classified appropriately, indicating that the trained CapsNet model maintained consistent performance when deployed in a real-world application environment. These results confirm that the integration between the image preprocessing pipeline, the trained model, and the web interface functioned properly during the inference process.

The classification confidence scores varied across different images, reflecting the diversity of retinal characteristics and image quality encountered during testing. For example, the cataract image achieved a confidence score of 66.81%, whereas the normal image obtained a confidence score of 58.24%. These results indicate that the model was able to recognize visual patterns associated with each class and generate corresponding predictions based on the learned retinal features. The variation in confidence values is expected because fundus images may differ in illumination, contrast, retinal visibility, and the severity of lens opacity.

In addition, an external diabetic fundus image was tested to observe the model's behavior when presented with retinal images outside the classes used during training. The model incorrectly classified this image with a confidence score of 61.94%, which is reasonable because diabetic retinal features were not included in the training dataset. This result highlights a limitation of the current binary classification model, which was specifically developed to distinguish only between cataract and normal fundus images. Therefore, the model should not be expected to accurately classify retinal diseases that were not represented during training.

Furthermore, the application was able to handle images that did not contain recognizable retinal fundus characteristics by returning a low-confidence prediction result. This functionality demonstrates the robustness of the deployed system when processing different types of image inputs. Overall, the testing results indicate that the proposed CapsNet model can be effectively integrated into a web-based platform and provide reliable classification outcomes for cataract screening applications. The successful deployment of the model into a Flask-based system further supports its potential implementation as a computer-aided diagnostic tool for assisting ophthalmologists in the early detection of cataracts.

IV. CONCLUSION

This research presented a Capsule Network (CapsNet)-based approach for the classification of cataract fundus images and evaluated its classification performance using retinal fundus images. The proposed architecture employs capsule representations and dynamic routing, which are theoretically designed to preserve spatial relationships among

image features. However, the capability of preserving spatial relationships was not directly evaluated in the present study.

Experimental evaluation demonstrated that the proposed CapsNet model achieved strong and reliable performance, yielding an accuracy of 89.85%, precision of 94.56%, recall of 84.53%, specificity of 95.15%, F1-score of 89.26%, AUC-ROC of 95.27%, and Average Precision (AP) of 94.71% on the independent testing dataset. These results indicate that the proposed model is capable of effectively distinguishing cataract fundus images from normal fundus images and demonstrates good generalization performance.

The study also showed that comprehensive preprocessing procedures, including image resizing, data augmentation, normalization, and tensor transformation, contributed to stable model training and satisfactory classification performance. Furthermore, the successful implementation of the trained model in a Flask-based web application demonstrates the feasibility of deploying the proposed approach as a computer-aided cataract screening system to support ophthalmologists and healthcare providers in early cataract detection.

Overall, the findings demonstrate that Capsule Network is a promising deep learning approach for cataract fundus image classification. Although the obtained classification performance is consistent with the theoretical advantages of CapsNet, the contribution of capsule representations and dynamic routing to preserving spatial relationships was not empirically validated in this study. Therefore, future research should incorporate explainable artificial intelligence (XAI) techniques, feature visualization, capsule activation analysis, and direct comparisons with modern CNN-based architectures trained under identical experimental settings. In addition, future studies should evaluate the proposed model using larger multicenter and clinically annotated retinal fundus datasets to further assess its robustness, generalization capability, and clinical applicability.

REFERENCES

[1] World Health Organization, "Cataract," World Health Organization. Accessed: Oct. 21, 2025. [Online]. Available: https://www.who.int/health-topics/cataract

[2] J. Dreyer, "Cataract Statistics Worldwide For 2025," Contact Lenses Plus. Accessed: Oct. 21, 2025. [Online]. Available: https://www.contactlensesplus.com/education/cataract-stats

[3] L. Lin *et al.*, "Global, regional, and national burden of cataract: A comprehensive analysis and projections from 1990 to 2021," *PLoS One*, vol. 20, no. 6 June, pp. 1–17, 2025, doi: 10.1371/journal.pone.0326263.

[4] V. P. R. W. Amd. Kep, "Katarak: Penyakit Mata yang Sering Tidak Disadari Hingga Terlambat." Accessed: Nov. 17, 2025. [Online]. Available: https://keslan.kemkes.go.id/view_artikel/3721/katarak-penyakit-mata-yang-sering-tidak-disadari-hingga-terlambat

[5] I. Santoso, A. M. Manurung, and E. R. Subhiyakto, "Comparison of ResNet-50, EfficientNet-B1, and VGG-16 Algorithms for Cataract Eye Image Classification," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 284–294, 2025, doi: 10.30871/jaic.v9i2.8968.

[6] M. A. S. Yudono, F. F. Ridha, D. Mardiyana, F. Al-Ghozi, and A. Maulana, "Klasifikasi Katarak Berdasarkan Optic Disc Citra Fundus Smartphone: Perbandingan Ekstraksi Ciri Tekstur Dan Metode Neural Network," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 1, pp. 203–212, 2025, doi: 10.25126/jtiik.2025129254.

[7] A. Elsawy *et al.*, "A deep network DeepOpacityNet for detection of cataracts from color fundus photographs," *Communications Medicine*, vol. 3, no. 1, pp. 1–11, 2023, doi: 10.1038/s43856-023-00410-w.

[8] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," *Information (Switzerland)*, vol. 16, no. 3, pp. 1–28, 2025, doi: 10.3390/info16030195.

[9] R. Nirthika, S. Manivannan, A. Ramanan, and R. Wang, "Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study," *Neural Comput. Appl.*, vol. 34, no. 7, pp. 5321–5347, 2022, doi: 10.1007/s00521-022-06953-8.

[10] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Ann. Transl. Med.*, vol. 8, no. 11, pp. 713–713, 2020, doi: 10.21037/atm.2020.02.44.

[11] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 3857–3867, 2017.

[12] T. Tsuji *et al.*, "Classification of optical coherence tomography images using a capsule network," *BMC Ophthalmol.*, vol. 20, no. 1, pp. 1–9, 2020, doi: 10.1186/s12886-020-01382-4.

[13] Govindharaj I, Karthick G, and Micheal G, "International Journal of Intelligent Systems And Applications In Engineering Accurate Segmentation and Classification of Glaucoma Disease Utilising Grey Wolf Based U-Net++ with Capsule Network," *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2024, no. 18s, pp. 445–455, 2024, [Online]. Available: www.ijisae.org

[14] G. Manohar and R. O'Reilly, "InceptionCaps: A Performant Glaucoma Classification Model for Data-Scarce Environment," *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science, AICS 2023*, 2023, doi: 10.1109/AICS60730.2023.10470926.

[15] M. U. Haq, M. A. J. Sethi, and A. U. Rehman, "Capsule Network with Its Limitation, Modification, and Applications—A Survey," Sep. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/make5030047.