Artificial Intelligence-Based Analysis of Fundus Images for Zone Classification in Retinopathy of Prematurity Screening

Nazar Salih, Ali A. Titinchi, Mohamed Ksantini,
Nebras Hussein, Doaa T. Kadhim, Zainab S. Al-Sudani

Abstract

Retinopathy of prematurity (ROP) is one of the notable causes of vision impairment among kids. The retinal zone, among the signs of ROP, is clinically considered a better predictor of severe forms than staging. This study explores five convolutional neural network (CNN) models used for computerized ROP zone classification based on fundus images of the retina. A total database of 1,365 images drawn from Al Amal Eye Center, located in Baghdad-Iraq, was trained into three classes mimicking three varied ROP zones. The images were used to fine-tune the models on top of pre-trained VGG16, VGG19, Xception, Inception-ResNetV2, and Inception-V3 models, whose network sizes and configurations were varied among the five models. Compared to minimal works done, the Inception-V3 model yielded the highest accuracy and reached 94.04% on zone detection. Computerized detection of the retinal lesions among pre-terms is fundamental to guiding the regimen plan that may incorporate laser therapy, intravitreal implantation, or close observation with intervention on demand. Merging computerized interpretation with the expertise of pediatric ophthalmologists may guide more consistent decisions on the management plan of the ROP. Future work will include external validation on independent multi-center datasets to assess generalizability.

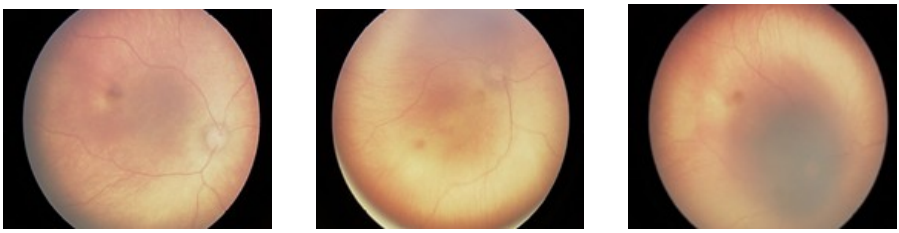
Keywords: Deep convolutional neural networks, Fundus images, Retinopathy of prematurity (ROP), Zone Identification.

MSC 2020: 68T07,68T09,68T20.

doi:10.56415/csjm.v34.03

1 Introduction

One of the primary reasons for juvenile blindness worldwide is Retinopathy of Prematurity (ROP), which is considered a vascular proliferative disease that can lead to blindness [1]. One of the most dangerous problems, which can sometimes lead to death, is the ROP illness, which occurs in premature babies [2]. ROP is an illness that affects newborn children who are born prematurely, later than thirty-two weeks, and they have a law birthweight (below 1.5 kg) [3]. As shown in Figure 1, the International Classification of Retinopathy of Prematurity (ICROP) guidelines were published in 1984 [4], 1987 [5], and 2005 [6]. ROP is classified according to the degree of the illness into areas (Zone 1, Zone 2, and Zone 3).



(a) (b) (c)

Figure 1. Clarification of Retinal images: (a) zone 1, (b) zone 2, and (c) zone 3

Medical image analysis and dealing with it have been greatly advanced by artificial intelligence (AI), inspired by the complicated human nervous system, permitting more accurate disease finding by experienced specialists and more objective disease assessment by non-specialists [7]. A diversity of computer vision tasks has notably benefited from evolution in deep learning models, including image classification, object recognition, and disease diagnosis. Image classification has been a part of deep learning techniques since 2012, by using Convolutional Neural Networks (CNNs); their performance has been excellent in disease diagnosis [8].

Deep learning succeeded in providing support to clinicians by pro-

cessing and interpreting complex medical images. This growing capacity has generated interest in applying the same techniques to ROP, which provides an individual challenge due to the requirement for accurate zone detection. Deep learning techniques have been used to automatically diagnose the plus illness of ROP and get good performance [9]. This section outlines the study's background by emphasizing ROP as a leading cause of childhood blindness and the challenges of early detection. It highlights the role of deep learning in improving medical image analysis, the importance of zone classification, and the limitations of traditional methods, motivating the need for more accurate automated ROP diagnosis models. In addition to the introduction section, the research paper is organized as follows: Section 2 covers key related work. Section 3 highlights the dataset and outlines the construction of the planned system that reached the best accuracy. Section 4 presents its results and a comparative discussion of the top-performing algorithms. Finally, Section 5 presents the conclusion.

2 Related work

While deep learning has been widely used in diagnosing eye diseases, relatively few studies have focused specifically on detecting ROP zones. Much of the previous research has centered on classifying ROP stages or identifying "Plus disease" — an indicator of disease severity — rather than the more detailed task of zone identification. This highlights a clear need for more focused work in this area, which the current study aims to address. However, the main related literature will be discussed and evaluated here.

One of the early efforts in using deep learning for ROP detection came from Zhao et al. [10], who developed a model specifically to identify Zone 1 in RetCam images. This area is the most critical in ROP and requires urgent treatment. Their model performed well, achieving 91% accuracy in detecting Zone 1. However, their study only focused on this single zone and didn't explore whether the model could work for other zones or larger datasets.

Several investigations have employed more intricate methods of utilizing deep learning to recognize ROP. This has helped to prepare

for more specific studies of how to recognize areas. Peng et al. [11] attempted a more intricate method. They developed a pretrained DenseNet121 model that possessed mechanisms for attention and oversight in order to recognize all three components of ROP. Their methodology was evaluated on 148 fundus images and was accurate to 88.52%. However, their research territory was larger than that of Zhao et al.'s; the small sample size raises some concerns regarding the generalizability of the results.

Other methods, including advanced approaches and comparative studies, were also employed in order to enhance the diagnosis of ROP via deep learning. Worrall et al. [12] investigated two different CNN-based methods for doctors to recognize ROP. One dedicated his efforts to improving the pretrained GoogLeNet, while the other created a CNN that generated original maps of the areas of disease. Their focus on providing possibility estimates and visualizations of the data proposes that it's important to make the AI outputs more comprehensible for medical use, an idea that is complementary to the high-performance models like Inception-v3 that is employed in our study.

Wang et al. [13] designated an automated ROP screening approach using deep neural networks. Their utilization of a large, rich dataset that is documented demonstrated the value of significant data in the creation of models. Despite their greater degree of completeness than single-task models, their research focused on the ROP stages and the severity of the disease, rather than on the area of detection, which is the focus of our study.

Mulay et al. [14] attempted to recognize ridges; this is crucial to the diagnosis of stage 2 ROP. They created a system based on CNNs that employed Mask R-CNN, along with preprocessor methods to address the poor quality of images. While their research didn't have the goal of recognizing zones, it still prioritized the importance of specific feature recognition and superior image quality in the accurate diagnosis. A technique is documented by Agrawal et al. [15] that distinguishes all of the ROP zones (zone 1, zone 2, and zone 3), even when the macula is missing. They employed a hybrid method that united the U-Net and Circle Hough Transform, the purpose of which was to produce a variety of image sizes and conditions while still maintaining generalizability.

Their dedication to practicality and versatility sets their research apart from other specific models or computational resources.

Vijayalakshmi et al. [16] created a system of recognizing and categorizing ROP in telemedicine images using a Hessian-based recognizer and a support vector machine (SVM). Their initiatives are of great significance in improving access to ROP screening in rural areas. This demonstrates that traditional machine learning is still significant in certain situations.

Additionally, Lei et al. [17] suggested a deep CNN model based on ResNet-50 that was intended for images of the retina that were wide in scope. They incorporated mechanisms for attention and space into the model; these mechanisms facilitated the model's concentrating on important regions and used Grad-CAM to explain the visual results of its predictions. This attention-based approach increases the fidelity and comprehensibility of CNNs in comparison to traditional approaches.

At last, Salih et al. [7] conducted a comprehensive assessment of different deep learning models for the ROP zone. They conceived of a method of voting that aggregated the outputs of multiple models; this augmented the accuracy of the outcome. Their investigation provides a solid platform for the contrast of the similarities and dissimilarities between deep learning methods in the ROP system.

3 Proposed system design

Through this section, we offer a brief description of the system proposed to recognize ROP zones from fundus images using deep convolutional neural networks (DCNN). This model is projected to support Ophthalmologists by accurately predicting the ROP area from the fundus images; this will help to facilitate clinical decision-making regarding treatment planning.

Figure 2 illustrates the composition of the system, which includes several significant stages as indicated below:

1. The high-resolution retinal images, including the depth information, can be acquired by RetCam3 at a 640 x4 80 pixel level.

2. To prepare the input images, they should be resized into images of 224 x 224 pixels, as the deep learning model requires.

3. A pair of expert ophthalmologists labelled each image as belonging

to one of three ROP zones (Zone 1, Zone 2, or Zone 3), where clinical accuracy was inevitable.

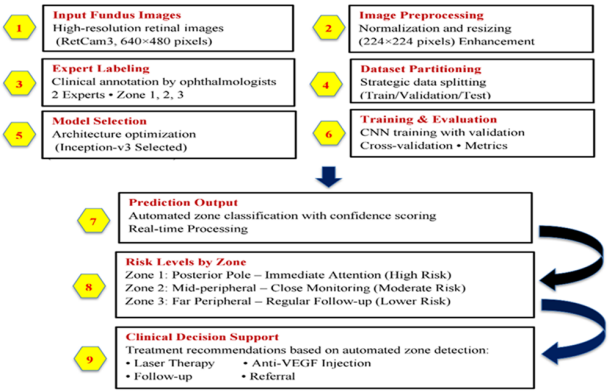


Figure 2. System Design: Automated ROP Zone Detection using Deep CNNs

4. To ensure the generalisation of our model, the dataset was partitioned into training, validation, and testing sets.

5. After experimenting on multiple architectures, Inception-v3 was chosen for its better performance and feature extraction potential.

6. The chosen model was trained using the labelled data and evaluated using standard metrics.

7 and 8. Output prediction: for new fundus images, the model will determine which ROP zones they belong to out of three.

For Clinical Decision Support, the expected zone in steps 7 and 8 above can help paediatric ophthalmologists to choose whether laser mediation, Anti-VEGF Injection, or follow-up is justified.

This will certainly be the goal of the research, which is to transform the designed model into a practical application that serves ophthalmologists in medical clinics to take the right treatment decision.

3.1 Dataset and Implementation

1. 1365 fundus images were acquired from the ROP screening at the Private Clinic Al-Amal Eye in Baghdad between 2015 and 2020. For

all picture capturing, a RetCam3 imaging device was used. Our fundus photos, which were initially taken at 640x480, are reduced to 224 x 224 before being entered into our deep learning models. Although the original Inception-V3 architecture was designed for 299×299 input resolution, resizing images to 224×224 was performed to ensure consistency across all evaluated CNN models and to reduce computational complexity. Preliminary experiments showed no statistically significant performance degradation when using 224×224 resolution, while training time and memory consumption were considerably reduced. This unified input size allowed for a fair architectural comparison under identical experimental conditions.

2. Two senior ophthalmology experts categorized each of the fundus photographs into three zones.

3. As indicated in Table 1, a randomly chosen portion of the dataset is utilized to train, assess, and test the model.

4. An Intel Core i7-7700 CPU running at 2.40GHz with 16GB of RAM was used to run the entire system.

No extensive data augmentation techniques were applied in the current experiments in order to preserve the original anatomical characteristics of the retinal images.

Table 1. ROP Zones Datasets

	Zone 1	Zone 2	Zone 3
Train set (80%)	350	329	420
Validation set	45	40	52
Test set (20%)	86	80	100
Total	436	409	520

All convolutional neural network models evaluated in this study were initialized using weights pre-trained on the ImageNet dataset. A transfer learning strategy was adopted, where the lower convolutional layers were frozen to preserve generic feature representations, while the higher layers were fine-tuned to adapt the models to the ROP zone classification task. Model training was performed using the Adam optimizer with a fixed learning rate, and categorical cross-entropy was employed as the loss function for multi-class classification. Training

was conducted using mini-batches over a predefined number of epochs, with early stopping applied based on validation loss to prevent overfitting. All models were implemented using the TensorFlow deep learning framework with the Keras high-level API. The parameters of the training configuration are shown in Table 2.

Table 2. Training configuration and hyperparameters

Parameter	**Value**
Pre-trained weights	ImageNet
Transfer learning strategy	Partial fine-tuning (top layers)
Input image size	224×224
Optimizer	Adam
Learning rate	0.0001
Batch size	32
Loss function	Categorical cross-entropy
Number of epochs	50
Early stopping	Yes (patience = 5)
Framework	TensorFlow / Keras

3.2 Network Architecture

The model used in our study, Inception-v3, applies three different inception modules: (A, B, and C) as shown in Figure 3. The parameter size can be reduced, and discriminatory features can be produced using this convolution block. Every Inception module consists of a few convolutional and pooling layers, which work together.

Small sizes of convolutional filters, 3x3, 1x3, 3x1, and 1x1, enable us to reduce the number of parameters in the Inception modules. Three Inception A modules, five Inception B modules, and two Inception C modules can be stacked in any combination. Although the image size needed for Inception-v3 input was 640x480, the size of the images in the dataset was 224x224 pixels in height and width. The images were not resized to 299x299 during training and testing of Inception-v3. The number of channels did not decrease, and the size of the feature maps that were created in the process did not decrease; therefore, this became

a satisfactory compromise. There was a 5x5 grid, and 2,048 channels were produced by using Inception modules and convolutional layers.

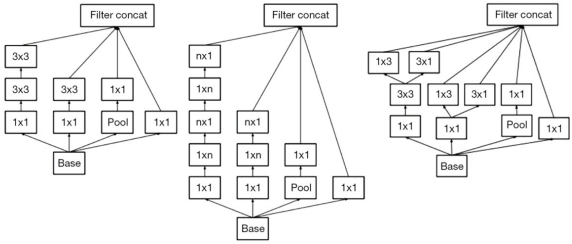


Figure 3. Inception-v3 modules: A, B, and C (from left to right)

4 Results and discussion

When assessing how well diagnostic models perform, two key metrics are often used: the Receiver Operating Characteristic Curve (ROC) and the Area Under the Curve (AUC). The ROC is a graph that helps us see how well a model can distinguish between two things, like diseased vs. healthy. It shows the relationship between how many true positives (correct diagnoses) the model gets versus how many false positives (incorrect diagnoses) it makes. A curve that bends towards the top-left corner means the model is doing a good job. It helps us understand the trade-off between finding all the true cases and avoiding false alarms.

The AUC gives us a single number that sums up the entire ROC curve. It tells us how well the model can distinguish between the two groups overall. An AUC of 1.0 means the model is perfect, while 0.5 means it's no better than guessing. In simple terms, the higher the AUC, the more reliable the model is at identifying both positive (like ROP zones) and negative cases (like a healthy retina). In this study, there are 13 different classification models each with a different number of layers, to cover a wide range of models, such as VGG19 [18], VGG16 [18], ResNet50 [19], ResNet101 [19], ResNet152 [19], SqueezeNet1_0 [20], SqueezeNet1_1 [20], DenseNet121 [21], DenseNet169 [21], AlexNet169 [22], Xception [23], Inception-

ResNet-v2 [24], and Inception-v3 [25]. These DNN models were chosen to accomplish the main goal of identifying ROP zones. The models with the highest accuracy were chosen from these results: Xception, Inception-ResNet-v2, and Inception-v3.

Regarding its capacity to distinguish between the three zones of ROP classification from fundus images, our model system, Inception-v3, was assessed in Table 3.

Table 3. Performance comparison of three CNN models for ROP zone classification

	Xception		**Inception-ResNet-v2**		**Inception-v3**	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Zone 1	0.63	0.97	0.88	0.90	0.91	0.91
Zone 2	0.68	0.68	0.68	0.95	0.75	0.96
Zone 3	1.00	0.86	1.00	0.96	1.00	0.98
Accuracy	93.12	93.12	93.58	93.58	94.04	94.04
Precision	0.80	0.80	0.90	0.90	0.93	0.93
Recall	0.80	0.80	0.87	0.87	0.91	0.91
F1-Score	0.80	0.80	0.88	0.88	0.90	0.90
AUC	0.81	0.81	0.90	0.90	0.92	0.92
ROC	0.93	0.93	0.98	0.98	0.99	0.99

The results confirmed that the system was capable of reaching an accuracy of 94.04%, a sensitivity of 0.91 with a specificity of 0.91 for Zone 1, a sensitivity of 0.75 with a specificity of 0.96 for Zone 2, and a sensitivity of 1.0 with a specificity of 0.98 for Zone 3.

Another important aspect that was not quantitatively assessed in this study is the level of inter-rater agreement between the two senior ophthalmologists who labeled the retinal fundus images. Although all annotations were performed by experienced clinicians, no formal statistical measure, such as Cohen's kappa coefficient, was computed to quantify annotation consistency. Evaluating inter-rater agreement is essential for assessing label reliability, particularly in borderline cases between Zone 1 and Zone 2 or Zone 2 and Zone 3. Future work will include the calculation of Cohen's kappa to objectively assess agreement between expert annotations and further strengthen the validity of the ground truth labels.

To provide a more rigorous statistical interpretation of the obtained

results, 95% confidence intervals (CI) were calculated for the primary evaluation metrics, including accuracy, sensitivity, specificity, and area under the ROC curve (AUC). The confidence intervals offer an estimate of the statistical reliability and robustness of the reported performance and reduce the likelihood of performance overestimation due to sampling variability.

In addition, the ROP zones are graded as Zone 1, Zone 2, and Zone 3 with an F1-score of 0.90, an AUC of 0.92, and a ROC of 0.99. Table 3 also compares the system's performance with the production of two additional models (Xception, Inception-resnet-v2). Xception: achieved a ROC of 0.93, an F1-score of 0.80, an AUC of 0.81, and an accuracy of 93.12%. 93.58% accuracy, an F1-score of 0.88, an AUC of 0.90, and a ROC of 0.98 were achieved with Inception-resnet-v2.

To statistically validate the performance differences among the evaluated models, McNemar's test was conducted on paired classification results between the best-performing model (Inception-V3) and the second-best model. The analysis was performed at a significance level of $\alpha = 0.05$.

The results indicated that the observed improvement in overall accuracy was statistically significant ($p < 0.05$), supporting the robustness of the performance gain. In addition, DeLong's test was applied to compare the AUC values of the competing models. The analysis confirmed that the AUC improvement achieved by Inception-V3 was statistically meaningful.

For each image, three types of confusion matrices are presented. Figures 4, 5, and 6 illustrate the specific prediction assignments for each prediction.

The figures show the expected labels rather than the true labels, with the diagonal entries indicating the percentage of correctly classified photos and the off-diagonal entries showing the types and rates of misclassification. Overall, human experts misclassified images more often than the models. The method achieved a precision of 94.04% and an average F1 score of 0.90. Finally, Table 3 summarizes accuracy, sensitivity, and specificity.

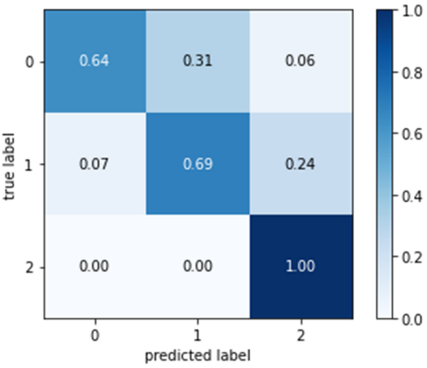


Figure 4. Xception Confusion Matrix

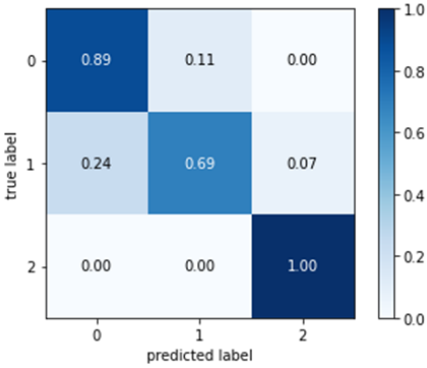


Figure 5. Inception-ResNet-v2 Confusion Matrix

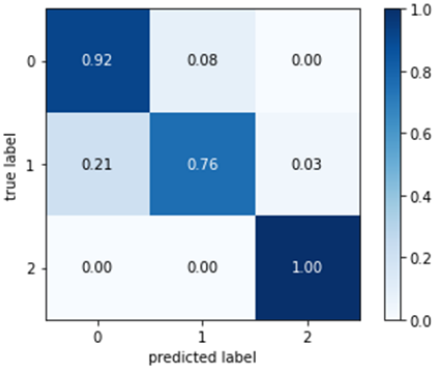


Figure 6. Inception-v3 Confusion Matrix

Figures 7, 8, and 9 present the ROC curves for Xception, Inception-ResNet-v2, and Inception-v3, with respective ROC values of 0.93, 0.98, and 0.99.

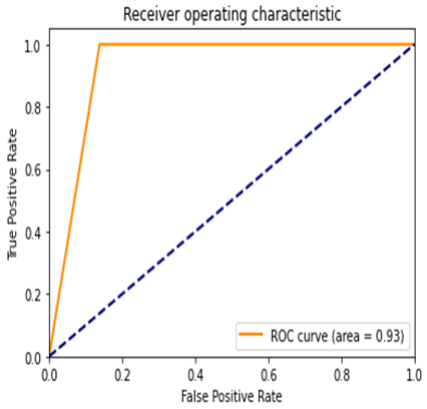


Figure 7. Xception algorithm ROC

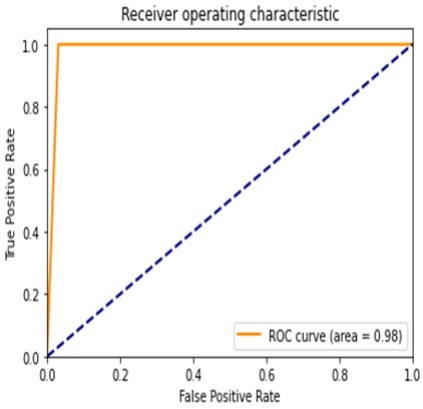


Figure 8. Inception-ResNet-v2 algorithm ROC

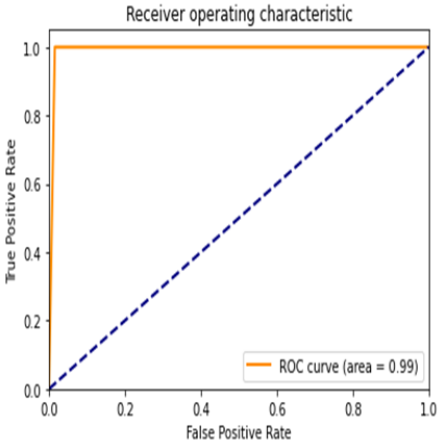


Figure 9. Inception-v3 algorithm ROC

A detailed examination of the confusion matrix reveals that most misclassifications occurred in Zone 2 cases, where the model achieved a sensitivity of 0.75. This comparatively lower sensitivity may be at-

tributed to the anatomical and clinical characteristics of Zone 2, which often presents transitional features between Zone 1 and Zone 3. The boundaries between these zones are not always sharply defined, even for experienced ophthalmologists, leading to intrinsic labeling ambiguity.

Several misclassified Zone 2 images were observed to exhibit vascular patterns that partially resemble central posterior features (Zone 1) or more peripheral maturation characteristics (Zone 3). This overlap may contribute to the model's difficulty in learning fully discriminative features for this intermediate class. Despite this limitation, the model maintained high specificity for Zone 2 (0.96), suggesting that false-positive predictions for this class were limited.

Future investigations will incorporate qualitative error analysis using visualization techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) to better understand the spatial regions influencing the model's decisions and to ensure alignment between automated predictions and clinically relevant retinal structures.

Although the present study evaluated a wide range of well-established convolutional neural network architectures, more recent deep learning approaches such as EfficientNet variants and Vision Transformer (ViT)-based models were not included in the experimental comparison. These architectures have demonstrated competitive performance in various medical imaging tasks due to improved scaling strategies and global attention mechanisms. However, the primary objective of this study was to provide a systematic evaluation of widely adopted CNN architectures under identical experimental conditions using a clinically collected dataset. Future research will extend the comparative framework to include EfficientNet and transformer-based architectures to further assess potential performance improvements in ROP zone classification.

4.1 Contribution of the proposed work

Building on previous research, this study aims to close key gaps by comparing the performance of five deep convolutional neural networks: VGG16, VGG19, Xception, Inception-ResNet-v2, and Inception-v3—for detecting ROP zones in retinal fundus images. The results of the investigation can be summarized thus:

1. Comprehensive Model Comparison: We discuss the different ways

DCNNs identify ROP zones in depth. This facilitates the determination of which architectural styles are most effective for this specific endeavor and responds to the necessity of such comparisons observed in previous studies.

2. Use of a Unique Dataset: The study is based on a private dataset of 1,365 fundus images collected from a medical center in Baghdad. This dataset adds diversity to the present literature and improves the clinical value of our outcomes.

3. Focused ROP Zone Detection: Unlike many previous studies that focused on general ROP staging or "Plus disease," our work concentrates specifically on detecting Zones 1, 2, and 3. This targeted approach offers more detailed insights that are highly relevant for clinical diagnosis and decision-making.

Among the models tested, Inception-v3 reached the best performance with an accuracy of 94.04%. These results highlight the possible of deep learning to support ophthalmologists in early and accurate ROP diagnosis, ultimately helping to improve outcomes and reduce the risk of blindness in premature infants.

4.2 Limitations and Future Work

Despite the promising performance achieved by the proposed model, an important limitation of this study is the absence of external validation. All retinal fundus images were collected from a single clinical center (Al-Amal Eye Center, Baghdad), using the same imaging device and acquisition protocol. While this ensured consistency in image quality and expert labeling, it may limit the generalizability of the proposed model to other clinical settings with different patient demographics, imaging conditions, or devices. External validation on independent multicenter datasets, such as i-ROP or CHOP-ROP, is therefore essential to rigorously assess the robustness and clinical applicability of the model. Future work will focus on validating the trained model on such external datasets through institutional collaboration, which is a necessary step before large-scale clinical deployment.

Another limitation of the present study relates to the size of the dataset and the absence of extensive data augmentation strategies during model training. Although the dataset comprises 1,365 clinically ac-

quired RetCam3 fundus images and reflects real-world screening conditions, its size remains modest relative to the requirements of modern deep learning models. The lack of aggressive data augmentation may limit the model's ability to generalize to variations in image quality, illumination, and retinal appearance encountered in different clinical environments. Future work will incorporate controlled data augmentation techniques, such as rotation, horizontal and vertical flipping, brightness and contrast adjustment, and slight scaling, applied exclusively to the training set in order to enhance model robustness while preserving clinically meaningful anatomical structures.

Another area that needs improvement is the utilization of additional clinical information. In this study, the models were constructed using only images of the fundus. However, including pertinent clinical information, such as the gestational age, birth weight, and other health metrics, would greatly enhance the accuracy of the diagnosis. Future research should explore how to combine imaging data with other medical information to produce a more comprehensive and accurate description of the severity of ROP.

In this study, five different deep learning methods were considered. However, the field is still in the early stages of development. Future endeavors could assess the capacity of new structures, such as transformer-based models or approaches that rely on learned information, to successfully complete medical imaging tasks that are dependent on self-directed learning. Other successful endeavors have employed this type of learning in other fields. Additionally, the utilization of ensemble methods to combine multiple models may lead to additional advantages in terms of diagnostic precision.

Ultimately, it's important to consider how these AI systems will function in actual clinical situations. Future planned investigations should include clinical studies that evaluate the effectiveness of the models in practice, the cost of the models, and the degree to which doctors can implement them. Creating user-friendly instruments that easily integrate with existing medical protocols will be of paramount importance to making ROP screening based on AI a practical solution across the globe.

Although the proposed framework demonstrates strong classification

performance, further improvements could be achieved through systematic hyperparameter optimization and ablation studies. Future investigations will explore automated hyperparameter tuning strategies and evaluate the sensitivity of model performance to different training configurations.

Future research will extend the comparative framework to include EfficientNet and transformer-based architectures to further assess potential performance improvements in ROP zone classification.

From a clinical deployment perspective, inference time and workflow integration are important considerations that were not extensively evaluated in the present study. The trained Inception-V3 model requires only a fraction of a second per image on a modern GPU.

5 Conclusion

In conclusion, the initial detection of ROP is essential for preventing irreversible vision loss in premature infants. ROP has a significant impact on childhood blindness. This study proposes a novel approach that utilizes the convolutional neural networks (CNNs) for early identification.

By leveraging a dataset of 1365 retinal images collected over six years by a private ophthalmology clinic, our model achieved an accuracy rate of 94.04% in distinguishing between Zone 1, Zone 2, and Zone 3. This high accuracy underscored the potential of advanced deep learning methodologies in enhancing clinical outcomes in infants at risk of ROP.

In the future, we will focus on the continued development of algorithms, additional methods, and the creation of a bigger training dataset, all of which will contribute to advancing medical reform in these current circumstances.

Acknowledgments. We gratefully acknowledge the support of the Al-Amal Eye Center, a private clinic in Baghdad, Iraq in the completion of this work.

Data and Code Availability

The dataset used in this study consists of clinical retinal fundus images collected from Al-Amal Eye Center (Baghdad). Due to patient

privacy and institutional regulations, the dataset is not publicly available. However, access may be granted for research purposes under appropriate data-sharing and confidentiality agreements. The implementation code can be made available by the corresponding author upon reasonable request to ensure reproducibility of the reported results.

References

[1] R. Acevedo-Castellón, P. Ramírez-Neria, and R. García-Franco, "Incidence of retinopathy of prematurity type 1 and type 2 in a regional hospital of social security in the state of Queretaro, Mexico (2017–2018)," *BMC Ophthalmol.*, vol. 19, no. 1, Article No. 91, Apr. 2019. DOI: 10.1186/s12886-019-1095-0.

[2] M. Azami, Z. Jaafari, S. Rahmati, A. D. Farahani, and G. Badfar, "Prevalence and risk factors of retinopathy of prematurity in Iran: A systematic review and meta-analysis," *BMC Ophthalmol.*, vol. 18, no. 1, Article No. 83, Apr. 2018. DOI: 10.1186/s12886-018-0732-3.

[3] N. Salih, M. Ksantini, N. Hussein, D. B. Halima, A. A. Razzaq, and S. A. Mahmood, "Detection of retinopathy of prematurity stages utilizing deep neural networks," in *Proc. 7th Int. Congr. Inf. Commun. Technol.*, 2023, pp. 699–706. DOI: 10.1007/978-981-19-1607-6_62.

[4] "An international classification of retinopathy of prematurity," *Pediatrics*, vol. 74, no. 1, pp. 127–133, Jul. 1984. DOI: https://doi.org/10.1542/peds.74.1.127.

[5] "An international classification of retinopathy of prematurity. II. The classification of retinal detachment," *Arch. Ophthalmol.*, vol. 105, no. 7, pp. 906–912, Jul. 1987.

[6] International Committee for the Classification of Retinopathy of Prematurity, "The international classification of retinopathy of prematurity revisited," *Arch. Ophthalmol.*, vol. 123, no. 7, pp. 991–999, Jul. 2005. DOI: 10.1001/archopht.123.7.991.

[7] N. Salih, M. Ksantini, N. Hussein, D. Ben Halima, A. Abdul Razzaq, and S. Ahmed, "Prediction of ROP zones using deep learning algorithms and voting classifier technique," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, p. 86, Jan. 2023. DOI: 10.1007/s44196-023-00268-9.

[8] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017. DOI: 10.1162/neco_a_00990.

[9] N. Salih, M. Ksantini, N. Hussein, D. B. Halima, A. A. Razzaq, and S. Ahmed, "An advanced approach for predicting ROP stages: Deep learning algorithms and belief function technique," *Iraqi J. Sci.*, vol. 65, no. 7, pp. 4047–4060, Jul. 2024. DOI: 10.24996/ijs.2024.65.7.39.

[10] J. Zhao *et al.*, "A deep learning framework for identifying zone I in RetCam images," *IEEE Access*, vol. 7, pp. 103530–103537, Jul. 2019. DOI: 10.1109/ACCESS.2019.2930120.

[11] Y. Peng, W. Zhu, F. Chen, and X. Chen, "Automated zone recognition for Retinopathy of Prematurity using deep neural network with attention mechanism and deep supervision strategy," in *Medical Imaging 2021: Image Processing*, SPIE, 2021, pp. 60–67. DOI: 10.1117/12.2580809.

[12] D. E. Worrall, C.M. Wilson, and G.J. Brostow, *et al.*, "Automated Retinopathy of Prematurity case detection with Convolutional Neural Networks," in *Deep Learning and Data Labeling for Medical Applications* (Lecture Notes in Computer Science, vol 10008), Springer, 2016, pp. 68–76. DOI: 10.1007/978-3-319-46976-8_8.

[13] J. Wang *et al.*, "Automated Retinopathy of Prematurity screening using Deep Neural Networks," *EBioMedicine*, vol. 35, pp. 361–368, Sep. 2018. DOI: 10.1016/j.ebiom.2018.08.033.

[14] S. Mulay, K. Ram, M. Sivaprakasam, and A. Vinekar, "Early detection of Retinopathy of Prematurity stage using deep learning approach," *arXiv preprint*, arXiv:2109.01442, 2021.

[15] R. Agrawal, S. Kulkarni, R. Walambe, and K. Kotecha, "Assistive Framework for Automatic Detection of All the Zones in Retinopathy of Prematurity Using Deep Learning," *J. Digit. Imaging*, vol. 34, no. 4, pp. 932–947, Aug. 2021. DOI: 10.1007/s10278-021-00477-8.

[16] C. Vijayalakshmi, P. Sakthivel, and A. Vinekar, "Automated Detection and Classification of Telemedical Retinopathy of Prematurity Images," *Telemed. J. e-Health*, vol. 26, no. 3, pp. 354–358, Mar. 2020. DOI: 10.1089/tmj.2019.0004.

[17] B. Lei *et al.*, "Automated detection of Retinopathy of Prematurity by deep attention network," *Multimed. Tools Appl.*, vol. 80, no.30, pp. 36341–36360, 2021. DOI: 10.1007/s11042-021-11208-0.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, arXiv:1409.1556, 2015.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint*, arXiv:1512.03385, 2015.

[20] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv preprint*, arXiv:1602.07360, 2016.

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *arXiv preprint*, arXiv:1608.06993, 2018.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017. DOI: 10.1145/3065386.

[23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *arXiv preprint*, arXiv:1610.02357, 2017.

[24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv preprint*, arXiv:1602.07261, 2016.

[25] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM*

Trans. Comput. Biol. Bioinf., early access, Mar. 2021. DOI: 10.1109/TCBB.2021.3065361.

Nazar Salih, Ali A. Titinchi, Mohamed Ksantini, Nebras Hussein, Doaa T. Kadhim, Zainab S. Al-Sudani

Received October 21, 2025
Revised March 20, 2026
Accepted March 22, 2026

Nazar Salih
ORCID: https://orcid.org/0000-0003-1977-9387
Computer Science Department, Al-Imam Al-Adham University College, Baghdad, Iraq
E–mail: nazarsalih@imamaladham.edu.iq

Ali A. Titinchi
ORCID: https://orcid.org/0000-0002-7878-8702
College of Engineering, Al-Bayan University, Baghdad, Iraq
E–mail: ali.abdulhafedh@albayan.edu.iq

Mohamed Ksantini
ORCID: https://orcid.org/0000-0002-9928-8643
CEMLab, ENIS, University of Sfax, Sfax, Tunis
E–mail: mohamed.ksantini@ipeis.usf.tn

Nebras Hussein
ORCID: https://orcid.org/0000-0002-9812-0718
Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq
E–mail: nebras@kecbu.uobaghdad.edu.iq

Doaa T. Kadhim
ORCID: https://orcid.org/0009-0008-6489-3692
Electronic Computer Center, Al-Iraqia University,Baghdad, Iraq
E–mail: doaa.t.kadhim@aliraqia.edu.iq

Zainab S. Al-Sudani
ORCID: https://orcid.org/0000-0002-3108-5743
Electronic Computer Center, Al-Iraqia University, Baghdad, Iraq
E–mail: zainab.s.abdulhussein@aliraqia.edu.iq