UNIVERSIDAD NACIONAL DE SAN MARTIN

RCSI
https://revistas.unsm.edu.pe/index.php/rcsi
e-ISSN: 2709-992X

Original Article
Artículo Original
Jul-Dec, 2026

Early glaucoma detection using hybrid models based on deep learning in fundus images

Detección temprana de glaucoma mediante modelos híbridos basados en aprendizaje profundo en imágenes de fondo de ojo

Ramírez-Cenepo, Leandro Enrique[1]* **Rojas-Córdova, Kelvin Lleins[1]**
Ruiz-Cueva, John Antony[1] **Leveau-Paredes, Monica Gabriela[1]**
Sánchez-Flores, Jhosep[1] **Llontop-Reategui, Augusto Ricardo[1]**

[1]Grupo de Investigación en Inteligencia Artificial, Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional de San Martín, Tarapoto, Perú
[2]Grupo de Investigación en Salud, Educación, Responsabilidad Social, Facultad de Medicina Humana, Universidad Nacional de San Martín, Tarapoto, Perú

Received: 12 Jan. 2026 | **Accepted:** 22 Jun. 2026 | **Published:** 20 Jul. 2026
Corresponding Autor*: lramirezc@alumno.unsm.edu.pe
How to cite this article: Ramírez-Cenepo, L. E., Ruiz-Cueva, J.A, Sánchez-Flores, J., Rojas-Córdova, K.J., Leveau-Paredes, M.G., Llontop-Reategui, A.R. (2026). Early glaucoma detection using hybrid models based on deep learning in fundus images. *Revista Científica de Sistemas e Informática, 6*(2), e1642. https://doi.org/10.51252/rcsi.v6i2.1642

ABSTRACT

Glaucoma is one of the leading causes of irreversible blindness, making early detection essential to prevent vision loss. This study aimed to develop and evaluate hybrid deep learning models for multiclass glaucoma classification using fundus images. A total of 6,014 retinal images from public databases were used, after undergoing cleaning, balancing, and stratified splitting. The VGG16, ResNet50, and EfficientNet-B0 architectures were employed as feature extractors combined with Cascaded Deep Forest. Evaluation included stratified 10-fold cross-validation and an independent test set. VGG16 + CDF achieved the best performance, with an average accuracy of 0.9320, greater stability, and better generalization. Statistical tests confirmed significant differences among the models, demonstrating that the proposed hybrid approach is a robust alternative for automated glaucoma classification.

Palabras clave: retinal image analysis; multiclass classification; automated diagnosis; transfer learning; computer vision

RESUMEN

El glaucoma es una de las principales causas de ceguera irreversible, por lo que su detección temprana es esencial para prevenir la pérdida visual. La investigación tuvo como objetivo desarrollar y evaluar modelos híbridos de aprendizaje profundo para la clasificación multiclase del glaucoma mediante imágenes de fondo de ojo. Se utilizaron 6014 imágenes retinianas de bases públicas, sometidas a depuración, balanceo y partición estratificada. Las arquitecturas VGG16, ResNet50 y EfficientNet-B0 se emplearon como extractores de características combinados con Cascaded Deep Forest. La evaluación incluyó validación cruzada estratificada de 10 pliegues y una prueba independiente. VGG16 + CDF obtuvo el mejor desempeño, con una accuracy promedio de 0.9320, mayor estabilidad y mejor generalización. Las pruebas estadísticas confirmaron diferencias significativas entre los modelos, demostrando que el enfoque híbrido constituye una alternativa robusta para la clasificación automatizada del glaucoma.

Keywords: análisis de imágenes retinianas; clasificación multiclase; diagnóstico automatizado; transferencia de aprendizaje; visión por computadora

1. INTRODUCTION

Glaucoma is one of the leading causes of irreversible blindness worldwide and represents a major public health concern due to its chronic and silent progression. This optic neuropathy is characterized by the progressive deterioration of the optic nerve, generally associated with increased intraocular pressure, which leads to irreversible loss of peripheral vision. The global burden of the disease is expected to continue increasing, driven by population aging and limitations in early detection (Schuster et al., 2020; Tham et al., 2014). The asymptomatic nature of glaucoma during its early stages results in a high proportion of cases being diagnosed at advanced stages, when visual damage is already irreversible.

Timely glaucoma detection is essential to prevent permanent vision loss; however, conventional diagnostic methods have important limitations. Techniques such as tonometry, optical coherence tomography, and automated perimetry require specialized equipment and highly trained personnel, restricting their accessibility in resource-limited settings (Ha & Park, 2019; Realini et al., 2021; Sidhu & Mansoori, 2024). Furthermore, the interpretation of ophthalmic images largely depends on the specialist's experience, which introduces diagnostic variability and increases the risk of clinical errors (Chaurasia et al., 2025).

In recent years, artificial intelligence, particularly deep learning, has emerged as a promising tool for developing automated systems to support medical diagnosis. Several studies have shown that convolutional neural networks can identify complex patterns in fundus images with high levels of accuracy, achieving performance comparable to that of human specialists in specific tasks (Chiang et al., 2024; Hwang et al., 2025; Ling et al., 2025). Moreover, transfer learning techniques have facilitated the implementation of robust models even in scenarios with limited availability of labeled data (Govindharaj et al., 2025). Complementarily, recent machine learning applications have demonstrated the potential of these techniques to identify complex patterns and construct predictive models in different scientific contexts (Arévalo-Hernández et al., 2023).

Nevertheless, despite these advances, most proposed approaches rely on unimodal data, mainly fundus images, which limits the comprehensive representation of the structural and functional complexity of glaucoma (Shi et al., 2025). In addition, challenges persist regarding heterogeneous image quality, the scarcity of well-annotated clinical datasets, and biases in training data, all of which affect model generalizability (Ilesanmi et al., 2023; Świerczyński et al., 2025). These limitations reduce the reliability of the proposed systems in real-world clinical settings.

The problem is particularly severe in low- and middle-income countries, where the availability of ophthalmology specialists is limited and access to diagnostic services remains insufficient (Meethal et al., 2024). In Peru, glaucoma affects approximately 2% of the population over 40 years of age, with a high proportion of cases not being diagnosed in a timely manner, particularly in rural and peri-urban areas (Ministerio de Salud, 2021). This situation highlights the need for accessible technological solutions capable of expanding diagnostic coverage and strengthening strategies for the prevention of avoidable blindness.

Despite the growing number of studies in this field, a relevant scientific gap remains: the lack of robust deep learning models that have been comparatively evaluated and validated under real-world conditions and that achieve clinically useful levels of diagnostic accuracy in local populations. Furthermore, limited evidence is available regarding the relative performance of different convolutional neural network architectures for the early detection of glaucoma using fundus images (Sharma et al., 2025; Shoukat et al., 2023).

Within this context, the objective of the present study was to develop and evaluate hybrid deep learning-based models for the early detection of glaucoma from fundus images, with the aim of identifying the configuration with the best generalization capability for multiclass disease classification.

2. MATERIALS AND METHODS

This research was conducted as an applied study with an explanatory scope, using a non-experimental, cross-sectional design. Its purpose was to develop and evaluate hybrid deep learning-based models for the early detection of glaucoma from fundus images. The methodological approach integrated pretrained convolutional neural networks as feature extractors and a *Cascaded Deep Forest* classifier to assess their ability to discriminate among different stages of the disease (Aljohani & Aburasain, 2024).

The experimental framework was implemented in *Python*. *PyTorch* and *timm* were used for feature extraction through pretrained convolutional neural networks; *scikit-learn* was employed to calculate evaluation metrics and perform validation procedures; and *Optuna* was used for hyperparameter optimization. In addition, *NumPy*, *Pandas*, *Matplotlib*, and *Seaborn* were used for data processing, analysis, and visualization. The experiments were conducted in Google Colab with GPU support to accelerate computational processing.

The methodological process included data preparation and preprocessing, the construction and evaluation of hybrid deep learning-based models, and validation and statistical analysis procedures aimed at assessing their performance in multiclass glaucoma classification (**Figure 1**).

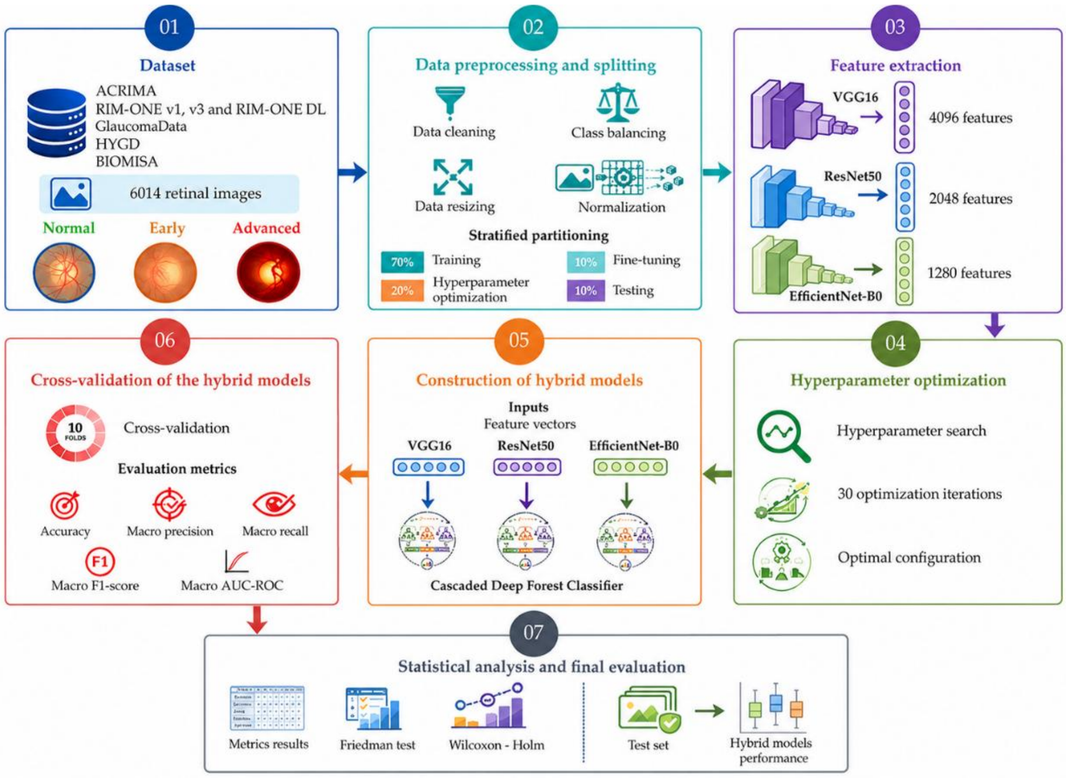

Figure 1. Methodological workflow of the study

Phase 1: Dataset

The study used fundus images obtained by integrating multiple publicly available glaucoma-specific datasets, including ACRIMA (Diaz-Pinto et al., 2019), RIM-ONE v1, v3 and RIM-ONE DL (Fumero Batista et al., 2020), GlaucomaData available on Roboflow Universe (Roboflow, 2026), HYGD (Abramovich et al., 2026) and BIOMISA (Hassan et al., 2018). These datasets contain retinal images classified into different stages of glaucoma and were developed for automated classification tasks using artificial intelligence techniques.

The integrated dataset comprised 6,014 images distributed across the following categories: normal (3,223 images), early glaucoma (1,120 images), and advanced glaucoma (1,671 images). **Figure 1** presents representative examples of the diagnostic categories used in the study.

Figure 1. Representative samples from the dataset

Phase 2: Data Preprocessing and Splitting

The images underwent a preprocessing procedure aimed at ensuring dataset consistency and compatibility with the deep learning architectures used. First, a data-cleaning stage was performed, including label verification, removal of duplicate records, and inspection of inconsistent or low-quality images. As a result, the dataset was reduced from 6,014 to 5,478 valid images for experimental development (Saha et al., 2023).

Subsequently, to reduce the bias associated with class imbalance among the diagnostic categories, a balancing procedure based on random selection was applied, resulting in a balanced dataset comprising 820 images per class and a total of 2,460 images (Owusu-Adjei et al., 2023).

The selected images were resized to 224 × 224 pixels, normalized using ImageNet parameters, and converted into tensors for computational processing. This procedure standardized the input data and ensured compatibility with the pretrained convolutional architectures employed in the study.

Finally, the balanced dataset was stratified into four independent subsets: training (70%), fine-tuning (10%), hyperparameter optimization (10%), and testing (10%). This strategy enabled model fitting and configuration optimization while preserving a completely independent test set for the final assessment of generalization capability.

Phase 3: Feature Extraction

VGG16, ResNet50, and EfficientNet-B0 were used as deep feature extractors through transfer learning. Initially, the ImageNet-pretrained models were adapted using the fine-tuning subset, allowing their representations to be partially adjusted to the retinal image domain considered in this study. The original classification layers were then removed, retaining only the representations generated by the internal convolutional layers (He et al., 2016; Simonyan & Zisserman, 2014; Tan & Le, 2019).

This approach made it possible to leverage visual patterns previously learned from large datasets while specializing the deep representations for multiclass glaucoma classification. It also reduced the computational complexity associated with fully training deep neural networks.

The deep representations generated by each architecture were transformed into feature vectors and used as inputs to the hybrid classifier. Vector dimensionality varied according to the architecture: VGG16 generated 4,096-dimensional feature vectors, ResNet50 generated 2,048-dimensional vectors, and EfficientNet-B0 generated 1,280-dimensional vectors.

Phase 4: Hyperparameter Optimization

The configuration of the hybrid models was optimized using the Bayesian search algorithm implemented in Optuna. During this stage, multiple combinations of hyperparameters associated with the Cascaded Deep Forest classifier were explored, including the number of cascade layers, number of estimators, maximum tree depth, minimum number of samples per leaf, and feature-selection strategies (Akiba et al., 2019).

Optimization was performed using the subset allocated for hyperparameter tuning to identify the best-performing configurations for each convolutional architecture. Accuracy was used as the primary optimization criterion, prioritizing overall classification performance in the multiclass glaucoma problem.

The search process consisted of 30 optimization trials, allowing different configurations within the hyperparameter space to be explored and the best-performing configuration for each hybrid architecture to be selected. The test set remained completely independent throughout this stage, preventing information leakage and reducing bias in the final assessment of model generalization.

Phase 5: Construction of Hybrid Models

Once the optimal configurations had been identified during the optimization stage, the hybrid models were constructed by integrating the convolutional architectures previously described for deep feature extraction with a Cascaded Deep Forest (CDF) classifier. The vector representations generated by the feature extractors were used as inputs to the hybrid models.

The Cascaded Deep Forest classifier consists of multiple sequential layers of Random Forest and Extra Trees models organized in a cascade structure. At each level, the class probabilities generated by the classifiers were concatenated with the original feature vectors, enabling the construction of progressively more discriminative representations (Breiman, 2001; Geurts et al., 2006; Zhou & Feng, 2019).

This approach combined the automatic feature-extraction capability of deep learning with the generalization ability of decision-tree-based methods, promoting more robust learning for multiclass glaucoma classification.

Phase 6: Cross-Validation of the Hybrid Models

Once the optimal configurations had been established during the optimization stage, the hybrid models were evaluated through stratified 10-fold cross-validation applied to the training set. This procedure preserved the proportion of diagnostic categories within each partition and provided a more robust and consistent assessment of model performance.

In each fold, metrics for assessing multiclass classification performance were calculated, including accuracy, macro precision, macro recall, macro F1-score, and macro AUC-ROC. The results obtained across the different iterations were then averaged to estimate the overall performance of each hybrid model during validation (Hicks et al., 2022).

Phase 7: Statistical Analysis and Final Evaluation

Based on the metrics obtained from stratified 10-fold cross-validation, an inferential statistical analysis was performed using accuracy values as the primary criterion for comparing the VGG16 + CDF, ResNet50 + CDF, and EfficientNet-B0 + CDF hybrid models. The nonparametric Friedman test was used for the overall comparison among configurations because the same partitions were used to evaluate all models. Pairwise comparisons were subsequently performed using the Wilcoxon signed-rank test with Holm correction to identify specific differences between architectures (Demšar, 2006).

Additionally, each model was evaluated using the independent test set, which had not been used during training, optimization, or cross-validation. This evaluation enabled the final generalization capability of the models to be estimated under a completely independent scenario.

3. RESULTS AND DISCUSSION

3.1. Final dataset distribution

The cleaning and balancing process yielded a more consistent and balanced dataset for experimental development. **Table 1** presents the image distribution across the various stages of the applied preprocessing.

Table 1. Distribution of the dataset during the cleaning and balancing process

Stage	Normal	Early glaucoma	Advanced glaucoma	Total
Integrated dataset	3223	1120	1671	6014
Cleaned dataset	2979	828	1671	5478
Balanced dataset	820	820	820	2460

3.2. Model Performance

Stratified cross-validation revealed differences in the performance of the hybrid models evaluated for the multiclass classification of glaucoma. **Table 2** presents the average results obtained for the metrics of accuracy, macro-precision, macro-recall, macro-F1-score, and macro-AUC-ROC. In terms of accuracy, VGG16 + CDF achieved the highest average value (0.9320), followed by ResNet50 + CDF (0.8966) and EfficientNetB0 + CDF (0.8310). A similar trend was observed across the other evaluation metrics, with VGG16 + CDF consistently yielding the highest values during cross-validation. These differences highlight variations in the architectures' ability to represent discriminative patterns associated with glaucoma diagnostic categories.

Table 2. Average results from 10-fold stratified cross-validation

Model	Accuracy	Precision macro	Recall macro	F1-score macro	AUC-ROC macro
VGG16 + CDF	0.932041	0.932804	0.932002	0.932036	0.991054
ResNet50 + CDF	0.896646	0.903249	0.896572	0.896878	0.980329
EfficientNetB0 + CDF	0.831012	0.835778	0.831075	0.831260	0.948292

The variation in accuracy across the different folds revealed differences in the stability of the evaluated hybrid models (**Figure 2**). Overall, VGG16 + CDF demonstrated the most consistent performance across iterations, maintaining the highest accuracy values in the majority of the analyzed folds. Meanwhile, ResNet50 + CDF showed intermediate variability, whereas EfficientNetB0 + CDF exhibited the greatest fluctuations and the lowest relative values during the experimental process.

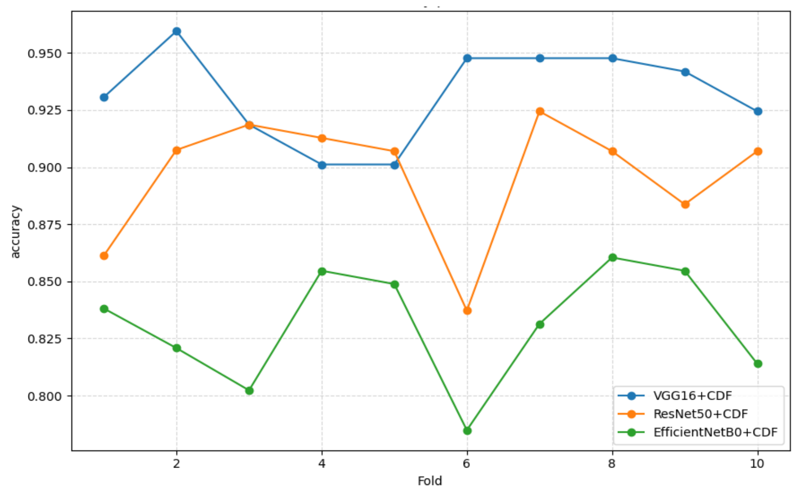

Figure 2. Accuracy behavior during stratified cross-validation

The confusion matrices revealed distinct classification patterns among the evaluated hybrid architectures (**Figure 3**). VGG16 + CDF concentrated the highest number of correct classifications along the main diagonal, performing particularly well in the advanced glaucoma and early glaucoma categories, with 541 and 519 correct classifications, respectively. ResNet50 + CDF maintained competitive performance, although it exhibited a higher number of misclassifications between the early glaucoma and normal categories. Meanwhile, EfficientNetB0 + CDF showed the greatest dispersion of errors, particularly in identifying early glaucoma cases, where more frequent confusion with the other diagnostic categories was observed.

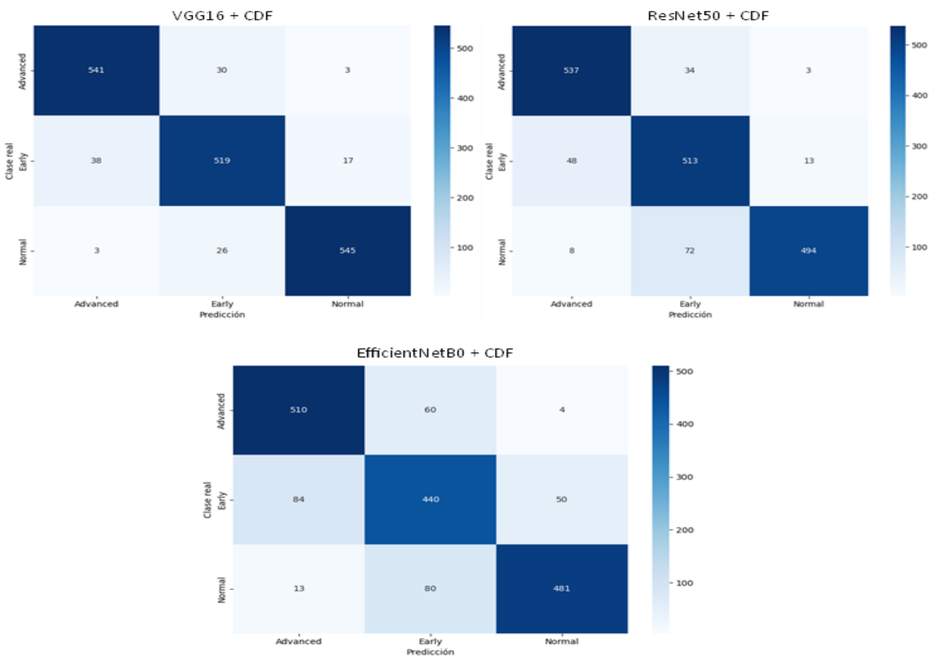

Figure 3. Confusion matrices by model

3.3. Statistical comparison between models

To determine whether the differences observed among the evaluated hybrid models were statistically significant, an inferential analysis was conducted using the accuracy values obtained during 10-fold stratified cross-validation. **Table 3** presents the descriptive statistics for each evaluated architecture. Overall, VGG16 + CDF recorded the highest mean accuracy and the highest confidence interval among the models analyzed, while ResNet50 + CDF showed intermediate performance. Meanwhile, EfficientNetB0 + CDF exhibited the lowest average values during the cross-validation process.

Table 3. Descriptive statistics for accuracy

Model	Mean	Std. deviation	Lower 95% CI	Upper 95% CI
VGG16 + CDF	0.9320	0.0204	0.9175	0.9466
ResNet50 + CDF	0.8966	0.0277	0.8768	0.9165
EfficientNetB0 + CDF	0.8310	0.0252	0.8130	0.8490

Prior to the comparative analysis, the distribution of accuracy values was assessed using the Shapiro-Wilk test. The results showed that at least one of the evaluated architectures did not meet the normality assumption ($p < 0.05$); therefore, non-parametric tests were selected for the statistical comparison between models.

The Friedman test revealed statistically significant differences among the evaluated hybrid models (**Table 4**). The results yielded a value of $\chi^2 = 16.67$ and a significance level of $p = 0.00024$, indicating significant differences in the performance of the architectures during stratified cross-validation. Additionally, Kendall's coefficient of concordance ($W = 0.83$) demonstrated a high level of consistency in the ranking of the models across the different folds evaluated.

Table 4. Friedman test result

Metric	Friedman,s χ^2	p-value	Kendall's W
accuracy	16.666667	0.00024	0.833333

Post-hoc comparisons using the Wilcoxon test with Holm correction revealed statistically significant differences among the evaluated hybrid architectures (**Table 5**). Specifically, VGG16 + CDF showed significant differences compared to ResNet50 + CDF and EfficientNetB0 + CDF, also exhibiting the largest mean difference relative to the latter architecture. Likewise, ResNet50 + CDF showed significant differences compared to EfficientNetB0 + CDF. These results confirm the differences previously observed during cross-validation and support the relative superiority of VGG16 + CDF in terms of overall performance.

Table 5. Wilcoxon post hoc comparisons with Holm correction

Comparison	Mean difference	Raw p-value	Adjusted p-value (Holm)
VGG16 + CDF vs ResNet50 + CDF	0.035395	0.019531	0.019531
VGG16 + CDF vs EfficientNetB0 + CDF	0.101028	0.001953	0.005859
ResNet50 + CDF vs EfficientNetB0 + CDF	0.065634	0.001953	0.005859

The distribution of accuracy values obtained during cross-validation revealed consistent differences among the hybrid architectures evaluated (**Figure 4**). VGG16 + CDF exhibited the highest median and lower relative dispersion compared to the other models, whereas EfficientNetB0 + CDF showed the lowest values and greater variability across folds. These results reinforce the differences identified through inferential statistical tests.

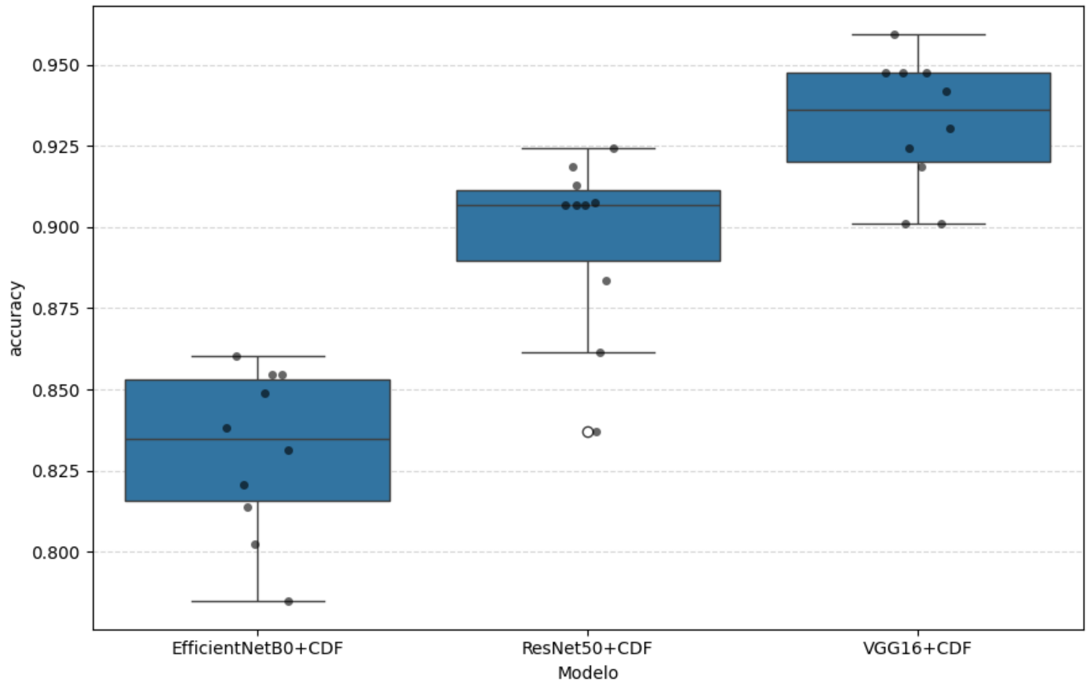

Figure 4. Accuracy distribution by model

3.4. Generalization on the test set

Evaluation on the independent test set revealed differences in the generalization capability of the hybrid models assessed (**Figure 5**). Overall, VGG16 + CDF maintained the highest accuracy values during both cross-validation and testing, while also exhibiting lower relative dispersion across folds. ResNet50 + CDF showed intermediate performance, whereas EfficientNetB0 + CDF recorded the lowest overall performance and greater relative variability. Additionally, a decrease in accuracy was observed on the test set compared to the average values obtained during cross-validation—an expected outcome when evaluating images not used during model training and optimization. These results demonstrate that VGG16 + CDF exhibited the most consistent generalization capability for multiclass glaucoma classification.

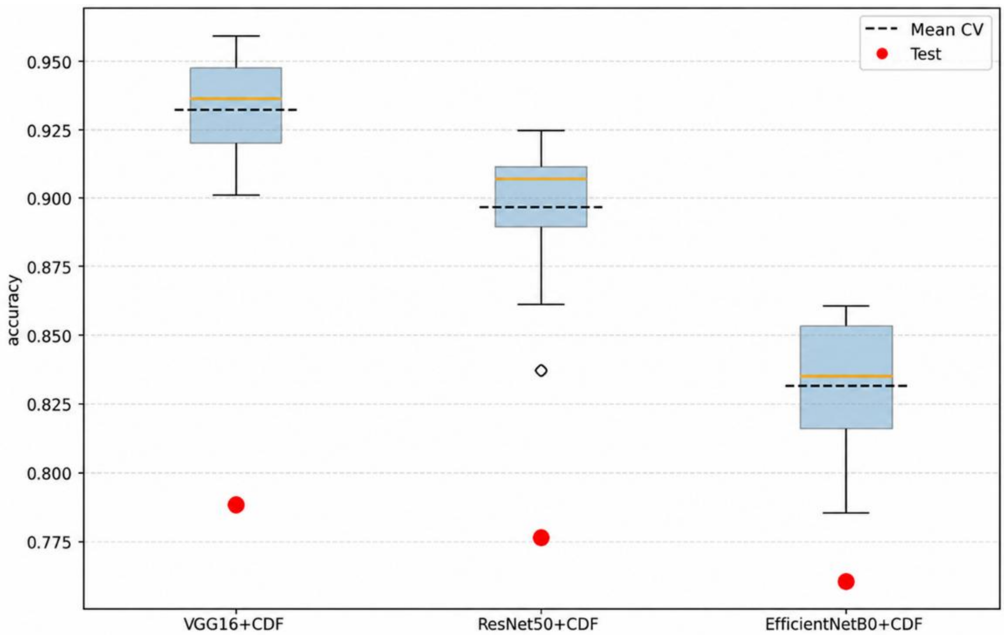

Figure 5. Model performance on the test set

3.5. Discussion

The evaluated hybrid models demonstrated competitive performance in the multi-class classification of glaucoma using fundus images, showing that combining deep feature extraction with classification via Cascaded Deep Forest constitutes an effective strategy for the automated detection of the disease. Among the architectures analyzed, VGG16 + CDF exhibited the best performance in both cross-validation and the independent test set, while also demonstrating greater stability across folds and more consistent generalization capability compared to ResNet50 + CDF and EfficientNetB0 + CDF.

The results obtained align with those reported by Sharma et al. (2025), who developed a hybrid system for automated glaucoma screening using fundus images, achieving high levels of sensitivity and specificity. Similarly, the present research demonstrated that integrating convolutional architectures with Cascaded Deep Forest maintained robust performance during cross-validation and evaluation on unseen data. Chaurasia et al. (2025) noted that model performance tends to decline when evaluated on external or independent datasets a pattern also observed in this study when comparing cross-validation results with the test set.

Compared to approaches based exclusively on deep convolutional networks, the results suggest significant advantages of the proposed hybrid model. Shoukat et al. (2023) reported high accuracy levels using ResNet50 for glaucoma detection; however, in this study, the ResNet50 + CDF combination showed inferior performance compared to VGG16 + CDF. This behavior demonstrates that the performance of a convolutional architecture can vary depending on the classification strategy employed and the characteristics of the dataset used. Furthermore, the confusion matrices revealed greater difficulty in distinguishing between the early-stage and advanced-stage glaucoma categories a finding consistent with the clinical complexity of the disease and the presence of visual patterns that partially overlap across stages.

Although recent studies such as those by Hwang et al. (2025) and Świerczyński et al. (2025) highlight the potential of multimodal and multisensory approaches to enhance automated glaucoma diagnosis, the results of this research demonstrate that competitive performance can be achieved using only fundus images and hybrid models with lower computational complexity. This increases the potential applicability of the proposed approach in resource-limited settings and identifies the incorporation of multimodal strategies as a future direction to further strengthen the system's clinical generalization capability.

CONCLUSIONS

The hybrid approach, based on convolutional neural networks for feature extraction and Cascaded Deep Forest for classification, demonstrated competitive performance in the multi-class classification of glaucoma using fundus images. The results showed that combining deep learning with tree-based methods is effective for automated glaucoma detection. Among the evaluated architectures, VGG16 + CDF achieved the best performance in both cross-validation and the independent test set, exhibiting greater stability and generalization capability than ResNet50 + CDF and EfficientNetB0 + CDF. Furthermore, statistical tests confirmed significant differences between the models, supporting the robustness of the approach.

Confusion matrices revealed greater difficulty in distinguishing early-stage from advanced-stage glaucoma, owing to the complexity of the disease's progression. Nevertheless, the model shows potential as a support tool for automated screening. Incorporating multimodal approaches, interpretability techniques, and more diverse datasets is recommended to enhance the model's generalization and clinical applicability.

FUNDING

The authors received no sponsorship to conduct this study/article.

CONFLICT OF INTEREST

There is no conflict of interest related to the subject matter of the work.

AUTHOR CONTRIBUTIONS

Conceptualization: Ramírez-Cenepo, L. E.; Rojas-Córdova, K. L.; Llontop-Reategui, A. R. Data curation: Ramírez-Cenepo, L. E.; Ruiz-Cueva, J. A.; Sánchez-Flores, J. Formal analysis: Ramírez-Cenepo, L. E.; Rojas-Córdova, K. L. Investigation: Ramírez-Cenepo, L. E.; Ruiz-Cueva, J. A.; Sánchez-Flores, J.; Llontop-Reategui, A. R. Methodology: Ramírez-Cenepo, L. E.; Rojas-Córdova, K. L.; Leveau-Paredes, M. G. Project administration: Ramírez-Cenepo, L. E. Software: Sánchez-Flores, J.; Ramírez-Cenepo, L. E. Validation: Ramírez-Cenepo, L. E.; Sánchez-Flores, J. Visualization: Rojas-Córdova, K. L.; Leveau-Paredes, M. G. Writing – original draft: Ramírez-Cenepo, L. E.; Rojas-Córdova, K. L. Writing – review and editing: Ramírez-Cenepo, L. E.; Ruiz-Cueva, J. A.; Leveau-Paredes, M. G.

REFERENCES

Abramovich, O., Pizem, H., Fhima, J., Berkowitz, E., Gofrit, B., Baskin, M., Meisel, M., Eijgen, J. Van, Blumenthal, E., & Behar, J. (2026). Hillel Yaffe Glaucoma Dataset (HYGD): A Gold-Standard Annotated Fundus Dataset for Glaucoma Detection. *PhysioNet*. https://doi.org/https://doi.org/10.13026/m92s-0z95

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Aljohani, A., & Aburasain, R. Y. (2024). A hybrid framework for glaucoma detection through federated machine learning and deep learning models. *BMC Medical Informatics and Decision Making*, *24*(1), 115. https://doi.org/10.1186/s12911-024-02518-y

Arévalo-Hernández, C. O., Arévalo-Gardini, E., Arévalo-López, L. A., Tuesta-Hidalgo, O., Romero-Vela, D. S., & Ruiz-Camus, C. E. (2023). Predicción de la fertilidad del suelo mediante aprendizaje automático en la provincia de Alto Amazonas, Perú. *Revista Peruana de Investigación Agropecuaria*, *3*(2), e63. https://doi.org/10.56926/repia.v3i2.63

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chaurasia, A. K., Liu, G.-S., Greatbatch, C. J., Gharahkhani, P., Craig, J. E., Mackey, D. A., MacGregor, S., & Hewitt, A. W. (2025). A generalised computer vision model for improved glaucoma screening using fundus images. *Eye*, *39*(1), 109–117. https://doi.org/10.1038/s41433-024-03388-4

Chiang, Y.-Y., Chen, C.-L., & Chen, Y.-H. (2024). Deep Learning Evaluation of Glaucoma Detection Using Fundus Photographs in Highly Myopic Populations. *Biomedicines*, *12*(7), 1394. https://doi.org/10.3390/biomedicines12071394

Demšar, J. (2006). *Statistical Comparisons of Classifiers over Multiple Data Sets*. JMLR.

https://www.jmlr.org/papers/v7/demsar06a.html

Fumero Batista, F. J., Diaz-Aleman, T., Sigut, J., Alayon, S., Arnay, R., & Angel-Pereira, D. (2020). RIM-ONE DL: A Unified Retinal Image Database for Assessing Glaucoma Using Deep Learning. *Image Analysis & Stereology*, *39*(3), 161–167. https://doi.org/10.5566/ias.2346

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Govindharaj, I., Santhakumar, D., Pugazharasi, K., Ravichandran, S., Vijaya Prabhu, R., & Raja, J. (2025). Enhancing glaucoma diagnosis: Generative adversarial networks in synthesized imagery and classification with pretrained MobileNetV2. *MethodsX*, *14*, 103116. https://doi.org/10.1016/j.mex.2024.103116

Ha, A., & Park, K. H. (2019). Optical Coherence Tomography for the Diagnosis and Monitoring of Glaucoma. *Asia-Pacific Journal of Ophthalmology*. https://doi.org/10.22608/APO.201902

Hassan, T., Akram, M. U., Masood, M. F., & Yasin, U. (2018). *BIOMISA Retinal Image Database for Macular and Ocular Syndromes* (pp. 695–705). https://doi.org/10.1007/978-3-319-93000-8_79

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, *12*(1), 5979. https://doi.org/10.1038/s41598-022-09954-8

Hwang, E. E., Chen, D., Han, Y., Jia, L., & Shan, J. (2025). Utilization of Image-Based Deep Learning in Multimodal Glaucoma Detection Neural Network from a Primary Patient Cohort. *Ophthalmology Science*, *5*(3), 100703. https://doi.org/10.1016/j.xops.2025.100703

Ilesanmi, A. E., Ilesanmi, T., & Gbotoso, G. A. (2023). A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks. *Healthcare Analytics*, *4*, 100261. https://doi.org/10.1016/j.health.2023.100261

Ling, X. C., Chen, H. S.-L., Yeh, P.-H., Cheng, Y.-C., Huang, C.-Y., Shen, S.-C., & Lee, Y.-S. (2025). Deep Learning in Glaucoma Detection and Progression Prediction: A Systematic Review and Meta-Analysis. *Biomedicines*, *13*(2), 420. https://doi.org/10.3390/biomedicines13020420

Meethal, N. S. K., Sisodia, V. P. S., George, R., & Khanna, R. C. (2024). Barriers and Potential Solutions to Glaucoma Screening in the Developing World: A Review. *Journal of Glaucoma*, *33*(8S), S33–S38. https://doi.org/10.1097/IJG.0000000000002404

Ministerio de Salud. (2021). *Minsa: Más del 50 % de los pacientes que tiene glaucoma no sabe que lo padece*. Ministerio de Salud. https://www.gob.pe/institucion/minsa/noticias/346283-minsa-mas-del-50-de-los-pacientes-que-tiene-glaucoma-no-sabe-que-lo-padece?utm_source

Owusu-Adjei, M., Ben Hayfron-Acquah, J., Frimpong, T., & Abdul-Salaam, G. (2023). Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. *PLOS Digital Health*, *2*(11), e0000290. https://doi.org/10.1371/journal.pdig.0000290

Realini, T., McMillan, B., Gross, R. L., Devience, E., & Balasubramani, G. K. (2021). Assessing the Reliability of Intraocular Pressure Measurements Using Rebound Tonometry. *Journal of Glaucoma*, *30*(8), 629–633. https://doi.org/10.1097/IJG.0000000000001892

Roboflow. (2026). *GlaucomaData Computer Vision Model*. Universe Roboflow.

https://universe.roboflow.com/classification-8wwxf/glaucomadata/dataset/1

Saha, S., Vignarajan, J., & Frost, S. (2023). A fast and fully automated system for glaucoma detection using color fundus photographs. *Scientific Reports*, *13*(1), 18408. https://doi.org/10.1038/s41598-023-44473-0

Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T., & Pfeiffer, N. (2020). The Diagnosis and Treatment of Glaucoma. *Deutsches Ärzteblatt International*. https://doi.org/10.3238/arztebl.2020.0225

Sharma, P., Takahashi, N., Ninomiya, T., Sato, M., Miya, T., Tsuda, S., & Nakazawa, T. (2025). A hybrid multi model artificial intelligence approach for glaucoma screening using fundus images. *Npj Digital Medicine*, *8*(1), 130. https://doi.org/10.1038/s41746-025-01473-w

Shi, D., Zhang, W., Yang, J., Huang, S., Chen, X., Xu, P., Jin, K., Lin, S., Wei, J., Yusufu, M., Liu, S., Zhang, Q., Ge, Z., Xu, X., & He, M. (2025). A multimodal visual–language foundation model for computational ophthalmology. *Npj Digital Medicine*, *8*(1), 381. https://doi.org/10.1038/s41746-025-01772-2

Shoukat, A., Akbar, S., Hassan, S. A., Iqbal, S., Mehmood, A., & Ilyas, Q. M. (2023). Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach. *Diagnostics*, *13*(10), 1738. https://doi.org/10.3390/diagnostics13101738

Sidhu, Z., & Mansoori, T. (2024). Artificial intelligence in glaucoma detection using color fundus photographs. *Indian Journal of Ophthalmology*, *72*(3), 408–411. https://doi.org/10.4103/IJO.IJO_613_23

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Vision and Pattern Recognition*, *1*. https://doi.org/https://doi.org/10.48550/arXiv.1409.1556

Świerczyński, H., Pukacki, J., Szczęsny, S., Mazurek, C., & Wasilewicz, R. (2025). Application of machine learning techniques in GlaucomAI system for glaucoma diagnosis and collaborative research support. *Scientific Reports*, *15*(1), 7940. https://doi.org/10.1038/s41598-025-89893-2

Tan, M., & Le, Q. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Proceedings of the 36th International Conference on Machine Learning, PMLR. https://proceedings.mlr.press/v97/tan19a.html

Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., & Cheng, C.-Y. (2014). Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040. *Ophthalmology*, *121*(11), 2081–2090. https://doi.org/10.1016/j.ophtha.2014.05.013

Zhou, Z.-H., & Feng, J. (2019). Deep forest. *National Science Review*, *6*(1), 74–86. https://doi.org/10.1093/nsr/nwy108