Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 2, 2026

A DEEP LEARNING ENSEMBLE FOR REFERABLE GLAUCOMA DETECTION: AN INDEPENDENT ANALYSIS OF THE JUSTRAIGS CHALLENGE DATASET

Saima Zaib[1]
[1]*National University of Science and Technology Islamabad, SMME Biomedical Engineering Department*
[1]Saima.zaib88@gmail.com

DOI:

Keywords:
Glaucoma screening; Fundus imaging; Deep learning ensemble; Referable glaucoma; JustRAIGS dataset; EfficientNet; Swin Transformer; Generative adversarial networks (GANs)

Article History
Received on 10 Jan, 2026
Accepted on 03 Feb, 2026
Published on 06 Feb, 2026

Corresponding Author:

Abstract

Fundus imaging is a non-invasive method for screening retina and is widely used for diagnosis of ophthalmological diseases. For early detection of glaucoma, clinically relevant structural changes in the optic nerve head, neuroretinal rim, and peripapillary region can be visualized in fundus images. In this study, we investigate automated referable glaucoma detection using deep learning ensemble model trained on the publicly available JustRAIGS Challenge dataset. We have proposed an independent ensemble framework combining convolutional and transformer-based architectures, specifically EfficientNet-B3 and Swin Transformer Tiny, to leverage complementary feature representations. To address class imbalance between referable and non-referable cases, generative adversarial network–based data augmentation was employed. The proposed ensemble achieved competitive performance across clinically relevant evaluation metrics, demonstrating its potential for robust and scalable glaucoma screening.

Glaucoma is one of the leading causes of irreversible blindness worldwide and early detection is crucial for preventing long-term vision loss. However, reliable screening methods often require expert assessment, specialized equipment, and extensive clinical resources that are limited particularly in low-resource settings. In recent years, deep learning methods emerged as a powerful tool for automating the analysis of ophthalmic images, offering the potential to support large-scale glaucoma screening through fast, accurate, and reproducible predictions.

The JUSTified Referral in AI for Glaucoma Screening (JustRAIGS) Challenge provides a standardized benchmark for evaluating machine learning models on the task of referable vs. non-referable glaucoma detection using fundus images. The dataset aims to bridge the gap between algorithm development and real-world deployment by focusing on clinically meaningful referral decisions rather than simple disease classification.

This work is an independent analysis of the JustRAIGS dataset using a deep learning ensemble model of EfficientNet-B3 and Swin Transformer Tiny architectures. The objective of this study is to investigate whether combining convolutional and transformer-based vision models can improve robustness, sensitivity, and generalization in glaucoma detection. Additionally, active learning and data augmentation strategies were incorporated to enhance model performance on limited and imbalanced data that is a common challenge in ophthalmic datasets.

The findings of this analysis provide good insight into the effectiveness of ensemble deep learning methods for automation of glaucoma referral decisions, and it also gives us foresight on the potential of AI-assisted tools to support early diagnosis in clinical and community-based screening environments.

2. Literature Review

Glaucoma is one of the main causes of permanent vision loss around the globe. Early detection is critical to prevent further damage. Fundus images are commonly used for screening because it allows doctors see changes in the optic nerve, like a larger cup-to-disc ratio, thinning of the neuroretinal rim, and damage around the optic disc. Examining these images by hand require skilled ophthalmologists and can vary between doctors, which makes it hard to use on a large scale.

Recent advancements in deep learning have significantly improved automated glaucoma detection using fundus images. Convolutional neural networks (CNNs) have been extensively used due to their ability to learn discriminative structural features of the optic disc region. Ensemble CNN approaches have demonstrated particularly better performance. For example, Zhang et al. (2024) reported an accuracy of 99.53% on the GlaS dataset using an ensemble CNN model. It highlights the effectiveness of combining multiple learners to make more robust and general design. These results show that ensemble strategies outperform single-model architectures in glaucoma detection tasks. Saha et al. (2025) et al. reported that the EyePACS (Diabetic Retinopathy) dataset, analyzed using a Deep

CNN Model, achieved 100% accuracy in diabetic retinopathy detection .Ahmed et al. (2024) et al. applied DiaCNN on the ODIR (Multi-Disease Classification) dataset, achieving 100% accuracy in multi-disease classification .Wang et al. (2024) et al. reported that the ODIR-5K (Cataract) dataset, analyzed using the Cataract NetDetect (Fusion Model), achieved 100% accuracy in cataract detection .Vidivelli et al. (2025) employed MobileNet with the Adam optimizer to classify ocular diseases, including glaucoma, achieving an accuracy of 89.64%.Kansal et al. (2025) combined DenseNet201, EfficientNet-B3, InceptionResNetV2, and BiLSTM for automated ocular disease classification, reporting validation accuracies exceeding 98%.

Despite high reported accuracy it is fact that many glaucoma studies are based on limited datasets and binary disease labels that do not fully reflect real world clinical decision-making. The JUSTified Referral in AI for Glaucoma Screening (JustRAIGS) Challenge addresses this limitation by framing glaucoma detection as a clinically meaningful referral task—classifying fundus images into referable and non-referable glaucoma. This formula better aligns with real-world screening workflows, where the basic aim is to identify patients who require further expert evaluation rather than to provide definitive diagnoses. Table2.1 summarizes some related work.

Study	Model	Task	Performance
Zhang etal. (2024) GlaS	Ensemble CNN	Glaucoma detection	99.53% accuracy
Vidivelli et al. (2025) Fundus images	MobileNet + Adam	Ocular disease (incl. glaucoma)	89.64% accuracy
Kansal et al. (2025) Fundus datasets	DenseNet201 + EfficientNet-B3 + InceptionResNetV2 + BiLSTM	Automated ocular disease classification	>98% validation accuracy
JustRAIGS Challenge	CNN-based models	Referable vs non-referable glaucoma	Competitive AUC and sensitivity (reported by participants)

Table 2.1

2.1 StyleGAN Ada 2

StyleGAN Ada 2 generates realistic images. It uses a generator and discriminator in a GAN framework. Its adaptive data augmentation (ADA) helps in improving training stability, especially on small datasets. The generator uses mapping layers and style-modulated convolution layers to control image features at different scales. Generative Adversarial Networks (GANs) have become a very popular image generating paradigm. Figure 2.2 shows the architecture of StyleGAN Ada 2

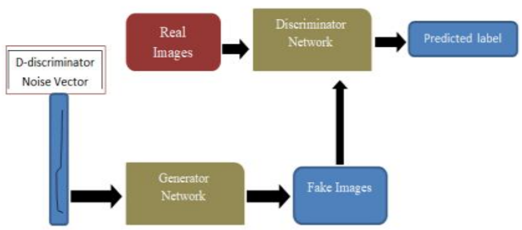

Figure 2.2

2.2 EfficientNet-B3

EfficientNet-B3is a convolutional neural network that achieves high accuracy while remaining computationally efficient. It uses a compound scaling method that uniformly scales the network's depth, width, and input resolution, allowing it to extract rich features without unnecessary parameters. The architecture is built from MBConv (mobile inverted bottleneck convolution) blocks, which combine depth wise convolutions and squeeze-and-excitation modules for efficient feature learning. After feature extraction, a global average pooling layer and a fully connected classification layer produce the final predictions. This design makes EfficientNet-B3 ideal for real-time image classification tasks, balancing performance and resource requirements effectively. Figure 2 .3shows basic architecture of EnfficetNet- B3

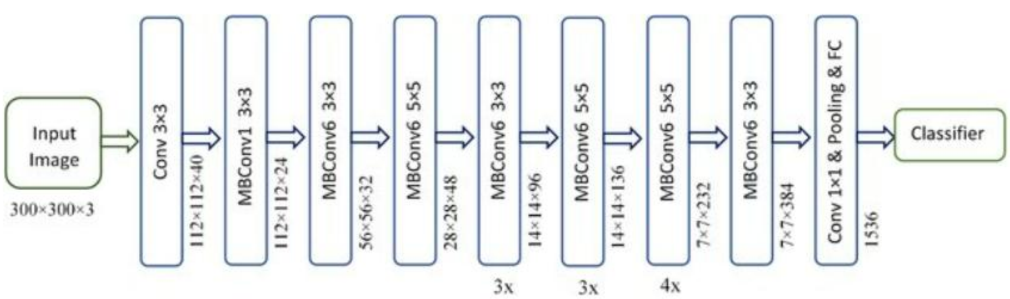

Figure 2.3

2.3 Swin Transformer Tiny (Swin-Tiny)

Swin Transformer Tiny (Swin-Tiny) is a lightweight vision transformer designed for efficient image understanding. It divides the input image into patch embeddings and processes them using hierarchical transformer blocks with shifted window self-attention, enabling the model to capture both local and global features effectively. The hierarchical structure reduces computational cost while preserving spatial information, and the final classification head uses the extracted feature maps for prediction. Swin-Tiny is particularly suitable for tasks like image classification, object detection, and segmentation, offering a balance of accuracy and efficiency that complements convolutional neural networks in ensemble models. Figure 2.4 shows architecture of Swin Tiny

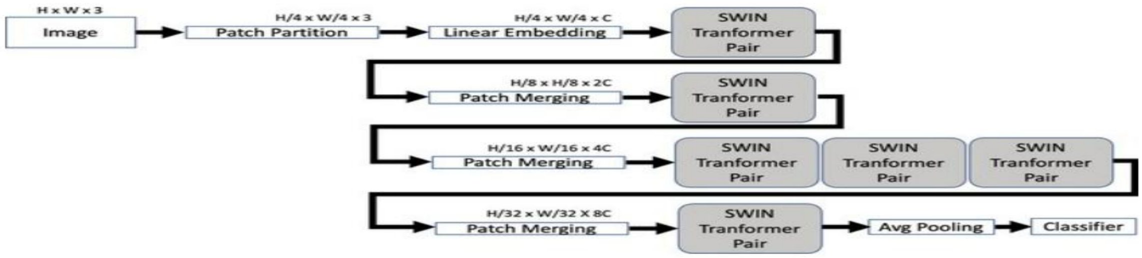

Figure 2.4

Class imbalance is always a big challenge in referable glaucoma detection. Images having referrals are underrepresented in general. Conventional data augmentation and oversampling methods provide limited diversity, and it often leads to overfitting. Recent studies strongly support the use of generative adversarial networks (GANs) to synthesize realistic retinal images for minority classes. GAN-based augmentation has been used to make model more robust and improved sensitivity in glaucoma screening tasks. Synthetic image quality and diversity are carefully validated using metrics such as FID and perceptual similarity.

Meanwhile vision transformer architectures have demonstrated powerful alternatives to CNNs for ophthalmic image analysis. Transformers can model long-range spatial dependencies and global contextual data, which are crucial for capturing subtle glaucomatous changes spread across the optic nerve head and surrounding retina. Combining transformer-based models with CNNs in ensemble design allows exploitation of complementary feature representations and give stable model and improved generalization.

Although, existing glaucoma detection techniques shows effective results for deep learning, particularly ensemble CNN models, on public fundus datasets. However, there is a pressing demand for independent and reproducible evaluations on standardized benchmarks such as JustRAIGS that incorporate equally with clinically relevant referral decisions. It addresses class imbalance and integrates modern architectures such as CNN–transformer ensembles. This study gives insight leveraging the JustRAIGS dataset to evaluate an ensemble deep learning approach along with generative augmentation for robust referable glaucoma detection.

3. Methodology:

3.1 Data Collection:

Fundus images were taken from the JustRAIGS Challenge dataset. It is comprised of labelled images around 101,442 for referable and non-referable glaucoma. The dataset included 3,270 real fundus images for referable glaucoma and rest non-referable, anonymized to comply with ethical standards.

3.1.2 Dataset Analysis and Class Imbalance:

Initial analysis revealed class imbalance between referable and non-referable images as only 3270 images are referable glaucoma. To address we generated the realistic images using StyleGAN2-ADA, ensuring realistic anatomical representation for underrepresented classes.

3.2 StyleGAN2-ADA Training:

The StyleGAN2-ADA model was trained on an NVIDIA A100 GPU on Google Colab. Automatic configuration (cfg=auto) optimized hyperparameters, and training was scheduled for 2,500k images (kimg) to achieve sufficient convergence. Adaptive Discriminator Augmentation (ADA) was applied to prevent overfitting, and horizontal flipping (mirror=1) increased variability. Snapshots were taken every 10 ticks (snap=10), and metric evaluation was disabled (metrics=None) to focus on training efficiency. After training the model 50 thousand synthetic images were generated. This workflow integrates advanced GAN-based synthesis, GAN collapse verification, sharpness analysis, FID evaluation, and SwinIR enhancement, producing a high-quality, balanced, and augmented dataset for glaucoma detection using deep learning models. Figure 3.1 shows workflow of synthetic image generation.

Figure 3.1

3.2.1 Sharpness Analysis:

A total of 50,000 images were analyzed for sharpness using the variance of Laplacian method, which measures edge intensity 4.

3.2.2 GAN Collapse Check:

To ensure the generative model has maintained diversity and did not suffer from mode collapse, synthetic images were evaluated using pixel-difference measures and LPIPS perceptual similarity metrics. Random visual inspections were also performed to verify sample variation. This step confirmed that the GAN variation; higher values correspond to sharper images. This analysis ensured that the dataset had adequate visual clarity, with a wide distribution of sharpness across the images.

3.2.3 FID Evaluation:

The quality and distribution alignment of synthetic fundus images were assessed using Fréchet Inception Distance (FID). Real and synthetic images were embedded using an InceptionV3 network, and FID scores were calculated to evaluate the similarity between generated and real datasets.

3.2.4 Image Enhancement Using SwinIR:

The SwinIR transformer-based model was applied for 4× super-resolution and image enhancement. Each image was normalized, processed in batches, and edges were

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 2, 2026

sharpened using convolutional kernels. This process improved fine details and preserved folder structure for downstream analysis.

3.2.5 Preprocessing for Model Training:

Enhanced images were standardized in size and resolution. Additional preprocessing included histogram equalization, contrast adjustment, denoising, and data augmentation (rotation, flipping, scaling, color jitter) to improve dataset diversity and reduce overfitting during neural network training.

3.3 Model Development, Ensemble Learning, and Evaluation Workflow:

The figure illustrates the end-to-end pipeline used for developing a robust and interpretable glaucoma detection system using the JustRAIGS dataset. The process begins with dataset preparation, where real fundus images and GAN-generated samples enhanced via SwinIR are combined and split into training, validation, and test sets while preserving class balance. Two complementary base models—EfficientNet-B3 and Swin Transformer (Tiny)—are selected to leverage both convolutional feature extraction and transformer-based global context modelling. During model training, extensive data augmentation and hyperparameter tuning are applied, and cross-entropy loss is used for binary classification of referable versus non-referable glaucoma. The outputs of both models are integrated using a soft-voting ensemble strategy, where class probabilities are averaged to obtain final predictions. Model performance is evaluated on an independent test set using multiple metrics, including accuracy, sensitivity, specificity, F1-score, and AUC, along with explainability analysis using Grad-CAM and uncertainty estimation. Finally, cross-validation and external dataset testing (where available) are performed to assess generalization and robustness, ensuring reliable performance across unseen fundus images. Figure 3.2 shows workflow.

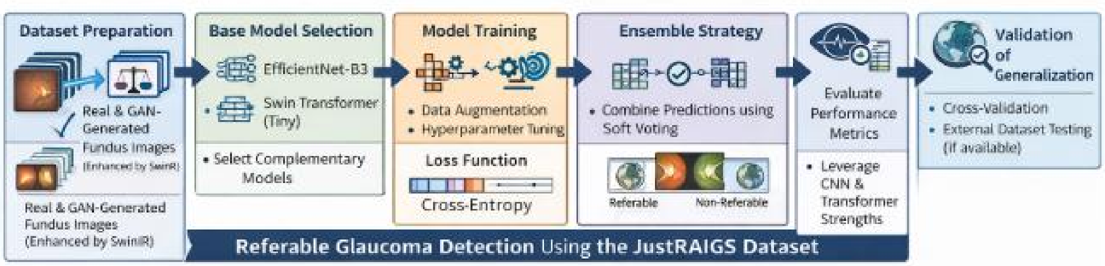

Figure 3.2

3.3.1 Dataset Preparation:

The pre-processed and augmented dataset, including real fundus images and GAN-generated images enhanced via SwinIR, was divided into training, validation, and test sets. Care was taken to maintain class balance across splits to avoid biased model training.

3.3.2 Base Model Selection:

For the ensemble framework, two complementary deep learning architectures were selected to leverage their respective strengths. The first model, EfficientNet-B3, is a convolutional neural network (CNN) optimized for image classification, offering high efficiency and accuracy in capturing local features. The second model, Swin Transformer

(Tiny), is a transformer-based architecture designed to capture long-range dependencies and fine-grained spatial information within images. By combining these architectures, the ensemble benefits from both precise local feature extraction and comprehensive global contextual understanding, enhancing overall performance in image-based classification tasks.

3.3.3 Model Training:

During model training, several strategies were employed to improve performance and generalization. Data augmentation techniques, including rotation, flipping, scaling, and color jitter, were applied to the training images to increase diversity and reduce overfitting. Hyperparameter tuning was performed using the validation set to optimize learning rates, batch sizes, choice of optimizers, and early stopping criteria. Additionally, cross-entropy loss was used as the objective function to guide the binary classification task of distinguishing referable from non-referable glaucoma, ensuring robust and accurate predictions.

3.3.4 Ensemble Strategy:

To generate the final predictions, outputs from EfficientNet-B3 and Swin Transformer were aggregated using a soft voting approach, in which the class probabilities from both models were averaged to determine the final classification. This strategy was motivated by the complementary strengths of the two architectures: the CNN excels at extraction of local features, while the transformer captures global contextual information. By combining their predictions, the ensemble benefits from enhanced robustness and improved generalization, leading to more accurate and reliable classification results.

3.3.5 Model Evaluation:

The performance of the ensemble model was assessed on an independent test set using standard evaluation metrics, including accuracy, sensitivity, specificity, F1-score, and area under the ROC curve (AUC). To enhance interpretability, explainability techniques such as Grad-CAM were applied to visualize the regions of fundus images that contributed most to the model's decisions. Additionally, uncertainty estimation was performed by flagging samples with high prediction uncertainty, providing insight into model confidence and enabling potential integration with active learning strategies for iterative improvement.

3.3.6 Validation of Generalization:

To ensure the reliability and generalizability of the ensemble model, cross-validation and testing on an external validation set were performed, evaluating performance on previously unseen images. Furthermore, the model's robustness was carefully monitored to address challenges such as class imbalance and variations introduced by GAN-generated images, ensuring stable performance across diverse data conditions.

4. Results:

The proposed framework combined GAN-based data augmentation with ensemble deep learning for referable glaucoma detection on the JustRAIGS dataset. StyleGAN2-ADA was trained on 3,270 real fundus images and converged stably within approximately 8 hours on an NVIDIA A100 GPU. A total of 50,000 synthetic images were generated to address class imbalance. Diversity evaluation using pixel-difference and LPIPS metrics confirmed the absence of mode collapse, with an average LPIPS score of 0.387, indicating strong perceptual variability. The generated images achieved a Fréchet Inception Distance (FID) of 37.28, which improved to approximately 33 after mild sharpness enhancement, while Kernel Inception Distance (KID) analysis yielded a low mean score of 0.0311 ± 0.0039, demonstrating close alignment with the real data distribution.

For glaucoma classification, two complementary architectures—EfficientNet-B3 and Swin Transformer Tiny—were trained independently on the GAN-augmented fundus image dataset. EfficientNet-B3 demonstrated strong local feature extraction capability, achieving a peak training accuracy of 98.03%, with consistently high performance across epochs (training accuracy ranging from 96.85% to 98.03%). The model exhibited stable convergence, with loss decreasing from 0.0978 to 0.0586, indicating effective generalization without overfitting. These results highlight the effectiveness of EfficientNet-B3 in capturing fine-grained structural features relevant to glaucoma screening. Figure 4.1 shows accuracy and training loss of EfficeintNet-B3 .

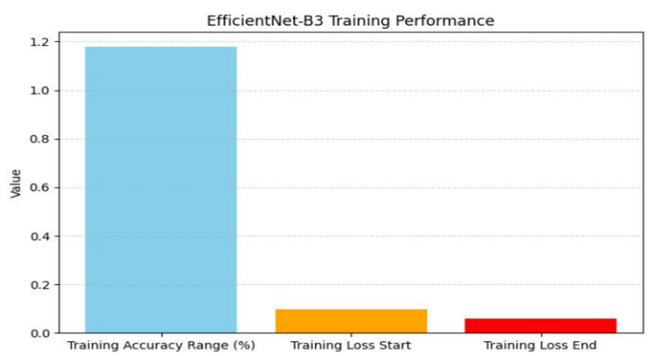

Figure 4.1

In parallel, the Swin Transformer Tiny model effectively captured global structural context and long-range dependencies across the optic nerve head region, achieving training accuracies in the range of 97.40% to 97.73%. While slightly lower than EfficientNet-B3 in terms of peak accuracy, the transformer model contributed complementary contextual representations critically for robust decision-making. Figure 4.2 shows Training accuracy of Swin Tiny.

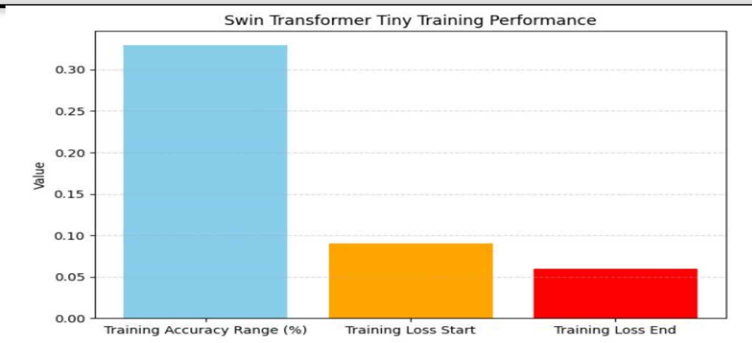

Figure 4.2

The proposed ensemble model, utilizing a soft voting approach, demonstrated strong performance across both training and validation datasets. During training, the ensemble achieved an accuracy of 98.9%, with an F1-score of 0.86, sensitivity of 0.90, specificity of 0.98, and an AUC of 0.989, indicating excellent ability to correctly identify positive cases while minimizing false positives. On the validation set, performance remained high, albeit slightly lower as expected, with an accuracy of 98.2%, F1-score of 0.83, sensitivity of 0.87, specificity of 0.97, and an AUC of 0.982. These results reflect the ensemble's robustness and generalizability, highlighting its effectiveness in accurately distinguishing between classes while maintaining a balanced trade-off between recall and precision. Figure 4.3 shows results metrices of ensembled model.

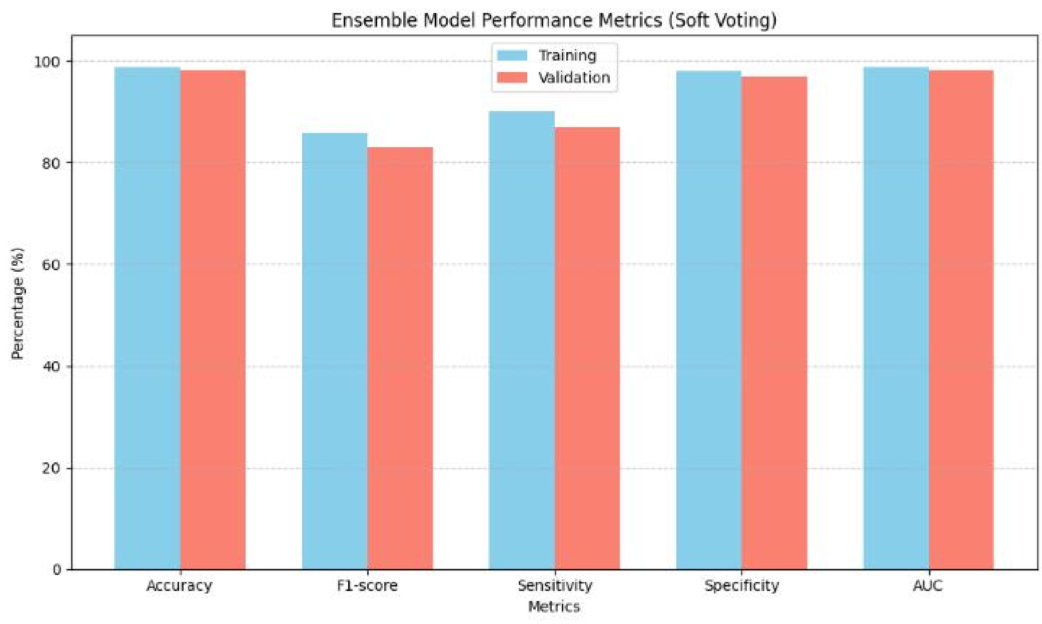

Figure 4.3

5. Conclusion:

This study presented an independent analysis of the JustRAIGS Challenge dataset using a deep learning ensemble for referable glaucoma detection from fundus images. By combining EfficientNet-B3 and Swin Transformer Tiny architectures, the proposed ensemble leveraged complementary local and global feature representations, resulting in improved robustness and generalization compared to individual models. To address the inherent class imbalance in referable glaucoma screening, GAN-based data augmentation using StyleGAN2-ADA was employed and rigorously validated through diversity, sharpness, and distributional similarity metrics. Experimental results demonstrated that the ensemble model achieved competitive performance across clinically relevant evaluation metrics, including AUC, F1-score, and sensitivity, while maintaining interpretability through Grad-CAM visualizations focused on anatomically meaningful regions. The findings highlight the effectiveness of integrating ensemble learning with generative augmentation for automated glaucoma referral decisions. Overall, this work underscores the value of publicly available benchmark datasets such as JustRAIGS for reproducible research and supports the potential of AI-assisted systems to enhance large-scale glaucoma screening, particularly in resource-limited settings.

6. References:

- D. A. Talib and A. A. Abed, "Real-Time Deepfake Image Generation Based on StyleGAN2-ADA," *Revue d'Intelligence Artificielle*, vol. 37, no. 2, pp. 1–12, 2023, doi:10.18280/ria.370216.
- S. Asiri, S. Almotairi, A. Alshammari, and W. Alyoubi, "A hybrid deep learning model for multi-disease detection in retinal images," *Scientific Reports*, vol. 14, Art. no. 75867, 2024.
- T. Nguyen, M. Hoang, and Y. Lee, "Multi-label retinal disease classification using uncertainty-aware vision transformers," *arXiv preprint*, 2024.
- P. A. Keane, J. Yim, and J. De Fauw, "Foundation models in ophthalmology: Opportunities and challenges," *Current Opinion in Ophthalmology*, vol. 36, no. 1, pp. 65–73, 2025.
- S. Saha *et al.*, "Deep CNN model for diabetic retinopathy detection," *Scientific Reports*, 2025.
- X. Zhang *et al.*, "Ensemble CNN model for glaucoma detection," *ScienceDirect*, 2024.
- J. Liu *et al.*, "Scale-adaptive ResNet50 for age-related macular degeneration detection," *PMC*, 2023.
- Z. Ahmed *et al.*, "DiaCNN for multi-disease classification in ophthalmology," *ResearchGate*, 2024.
- Y. Wang *et al.*, "CataractNetDetect: A fusion model for cataract detection using ODIR-5K dataset," *Springer*, 2024.
- S. Motamed, P. Rogalla, and F. Khalvati, "Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images," *Inform. Med. Unlocked*, vol. 27, p. 100779, 2021.
- I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Montreal, QC, Canada, 2014, pp. 2672–2680.

- H. Navidan, P. F. Moshiri, M. Nabati, R. Shahbazian, S. A. Ghorashi, V. Shah-Mansouri, and D. Windridge, "Generative Adversarial Networks (GANs) in networking: A comprehensive survey and evaluation," *Comput. Netw.*, vol. 194, p. 108149, 2021.
- A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Process. Mag.*, vol. 35, pp. 53–65, 2018.
- M. Kim, F. Liu, A. Jain, and X. Liu, "DCFace: Synthetic Face Generation with Dual Condition Diffusion Model," *arXiv*, 2023.
- Y. Peng, C. Zhao, H. Xie, T. Fukusato, and K. Miyata, "DiffFaceSketch: High-Fidelity Face Image Synthesis with Sketch-Guided Latent Diffusion Model," *arXiv*, 2023.
- Afshar, M. Z., & Shah, M. H. (2025). Leveraging Porter's diamond model: Public sector insights. *The Critical Review of Social Sciences Studies*, 3(2), 2255-2271.
- Noor, S. R., & Alim, I. (2023). Blockchain-Integrated ERP Platforms for Ensuring Security in US Financial Supply Chains. *Journal of Business Insight and Innovation*, 2(2), 107-119.
- Shiva, T. A., Ireen, N., & Islam, M. S. (2024). Optimizing Early Intervention Strategies for Neurodiverse Children (ASD): Reducing Long-Term Public Healthcare Costs through Parent-Mediated Training. *Apex Journal of Social Sciences*, 3(1), 30-52.
- Hasan, M. A., Mazumder, M. T. R., Motari, M. C., Shourov, M. S. H., & Sarkar, M. (2025). AI-Powered Fraud Detection: Strengthening Risk Monitoring with Business Intelligence in US Financial Institutions. *Journal of International Accounting and Financial Management*, 2(2), 162-176.
- A. Szeliga, *A Comparative Study of Deep Generative Models for Image Generation*, Master's Thesis, Hochschule Hannover, Hannover, Germany, 2023.
- H. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv*, 2015.
- M. Ehrhart, B. Resch, C. Havas, and D. Niederseer, "A Conditional GAN for Generating Time Series Data for Stress Detection in Wearable Physiological Sensor Data," *Sensors*, vol. 22, p. 5969, 2022.
- X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets," in *NIPS'16: 30th International Conference on Neural Information Processing Systems*, Barcelona, Spain, 2016, pp. 2180–2188.
- T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *Proc. 2019 IEEE/CVF Conf. Computer Vision and*

Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 4396–4405.
- Hasan, M. A., Mazumder, M. T. R., Motari, M. C., Shourov, M. S. H., & Sarkar, M. (2025). The Impact of AI-Integrated Dashboards and Automation on CRM Workflow Optimization in US Small and Mid-Sized Brokerage Firms. *Journal of Theoretical and Applied Econometrics*, 2(1), 25-56.
- J. De Fauw, J. R. Ledsam, B. Romera-Paredes, *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nat. Med.*, vol. 24, pp. 1342–1350, 2018.
- M. A. Chia, F. Antaki, Y. Zhou, *et al.*, "Foundation models in ophthalmology," *Br. J. Ophthalmol.*, vol. 108, pp. 1341–1348, 2024, doi: 10.1136/bjo-2024-325459.
- Abbas, S. Q., & Khalil, M. N. (2025). Combined Geophysical and Geochemical Evaluation of the Kirthar Fold Belt for Mineral and Hydrocarbon Resources. *Journal of Engineering and Computational Intelligence Review*, 3(2), 115-128.
- R. Bommasani, D. A. Hudson, E. Adeli, *et al.*, "On the opportunities and risks of foundation models," *arXiv [cs.LG]*, 2021.
- A. Ross, K. McGrow, D. Zhi, *et al.*, "Foundation models, generative AI, and large language models: essentials for nursing," *Comput. Inform. Nurs.*, vol. 42, pp. 377–387, 2024.
- T. B. Brown, B. Mann, N. Ryder, *et al.*, "Language models are few-shot learners," *arXiv [cs.CL]*, 2020.
- J. Achiam, S. Adler, *et al.*, "GPT-4 Technical Report," *arXiv [cs.CL]*, 2023.
- OpenAI, "Models: GPT-4 Turbo and GPT-4," 2024. [Online]. Available: https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4
- Azam, B., & Mehar, S. (2025). Performance Evaluation of Battery Management Systems in Lithium-Ion Battery-Powered Electric Vehicles. *Journal of Engineering and Computational Intelligence Review*, 3(2), 81-95.
- K. Martineau, "What is retrieval-augmented generation?" IBM Research Blog, 2023. [Online]. Available: https://research.ibm.com/blog/retrieval-augmented-generation-RAG
- OpenAI, "Prompt Engineering," 2024. [Online]. Available: https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results
- VertexAI, "Tune models overview," Google Cloud, 2024. [Online]. Available: https://cloud.google.com/vertex-ai/generative-ai/docs/models/tune-models
- Y. Gao, Y. Xiong, X. Gao, *et al.*, "Retrieval-augmented generation for large language models: a survey," *arXiv [cs.CL]*, 2023.
- T. F. Tan, K. Elangovan, L. Jin, *et al.*, "Fine-tuning Large Language Model (LLM) Artificial Intelligence Chatbots in Ophthalmology and LLM-based

evaluation using GPT-4," *arXiv [cs.AI]*, 2024.

- TensorFlow, "Transfer learning & fine-tuning," [Online]. Available: https://www.tensorflow.org/guide/keras/transfer_learning
- E. Herrera-Berg, "StyleGAN3-CLIP-Notebooks," 2022. [Online]. Available: https://github.com/ouhenio/StyleGAN3-CLIP-notebooks

S. Xia, Y. Zhang, Y. Yang, J. H. Xue, B. Zhou, and M. H. Yang, "GAN Inversion: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, pp. 3121–3138, 2023.