Detection of Diabetic Retinopathy Using a Multi-Decision Inception-ResNet Blended Hybrid Model

Mr.Muthukumar R [1], Mr.Boobalan M [2], Jerome Mc Jedidiah C [3], Kumaran E[4], Selvakumar S[5]
[1]Assist prof, Dept of Artificial Intelligence And Data Science
[2, 3, 4, 5] Dept of Artificial Intelligence And Data Science
[1, 2, 3, 4, 5] Muthayammal Engineering College

Abstract- *Diabetic retinopathy (DR) represents a critical complication of diabetes mellitus, leading to progressive vision impairment and potential blindness if left undetected. This research presents a novel multi-decision Inception-ResNet blended hybrid model for automated DR detection and classification. The proposed architecture integrates 172 weighted layers, strategically divided into dual-image processing pathways: 86 layers dedicated to color fundus image analysis and 86 layers for grayscale image processing. By employing a multi-layered transfer learning approach with adaptive moment estimation (Adam) and stochastic gradient descent (SGD) optimization techniques, the model achieves comprehensive feature extraction across both sequential and non-sequential image data. The architecture incorporates eight convolutional layers at each processing stage, enabling the extraction of both global and specialized features through chi-square testing mechanisms. Evaluated on the EyePACS and APTOS datasets, the model demonstrates superior performance with a detection accuracy of 98.1%, outperforming existing state-of-the-art approaches. The multi-decision framework effectively classifies DR into five severity stages: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR, providing a robust solution for early-stage diabetic retinopathy detection in clinical settings.*

Keywords: Deep learning, diabetic retinopathy, Inception-ResNet, dual-image processing, transfer learning, fundus images, convolutional neural networks, medical image classification, Adam optimization

I. INTRODUCTION

Diabetes mellitus affects over 537 million adults globally, with diabetic retinopathy (DR) emerging as one of the most severe microvascular complications. DR affects the retinal blood vessels, causing progressive damage that can lead to vision loss or complete blindness. The pathophysiology of DR involves hyperglycemia-induced oxidative stress, inflammation, and vascular endothelial dysfunction, resulting in characteristic retinal lesions including microaneurysms, hemorrhages,

Severe Non-Proliferative DR (Stage 3): Extensive intraretinal hemorrhages, venous beading, and intraretinal microvascular abnormalities

Proliferative DR (Stage 4): Neovascularization, vitreous hemorrhage, and potential retinal detachment Early intervention at stages 1-2 can significantly reduce the risk of progression to vision-threatening stages, making automated detection systems clinically valuable.

Research Gap and Contribution

Despite significant advances in deep learning- based DR detection, several challenges persist. Single-image processing approaches may miss subtle pathological changes visible through complementary analysis. Deep architectures often require extensive computational resources, limiting real-world deployment. Models trained on specific datasets may exhibit reduced performance on diverse populations. Accurately distinguishing between five DR stages remains challenging due to overlapping features.

This research addresses these gaps by introducing a dual-image multilayer processing approach that simultaneously analyzes color fundus and grayscale images through parallel 86- layer processing streams. The hybrid architecture integrates Inception and ResNet models to leverage multi-scale feature extraction and residual learning. A multi-decision framework employs three-phase optimization strategy combining Adam and SGD optimizers across 360 epochs. Comprehensive evaluation on large-scale public datasets demonstrates stage-wise performance exceeding 95% F1-scores.

Research Objectives

This study aims to develop a multi-decision Inception-ResNet blended hybrid model capable of processing both color and grayscale fundus images simultaneously. The implementation features a 172-layer architecture with balanced distribution for dual-image analysis. Transfer learning with multiple optimization strategies enhances

training efficiency. The model achieves accurate five-stage DR classification with superior performance metrics and validates on diverse datasets to ensure generalizability and clinical applicability.

II. LITERATURE REVIEW

Traditional Machine Learning Approaches

Early DR detection systems relied on conventional machine learning techniques combined with handcrafted feature extraction. Traditional approaches employed various preprocessing techniques including contrast enhancement, noise reduction, and vessel segmentation. Features were manually extracted based on domain expertise, including statistical texture features (Haralick features, Local Binary Patterns), morphological characteristics (vessel diameter, tortuosity), and color-based features (hemorrhage detection, exudate identification).

Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors (k-NN) were commonly employed for DR classification. While these methods provided interpretable results, their performance was limited by the quality of handcrafted features and inability to capture complex hierarchical patterns.

Deep Learning Revolution in Medical Imaging

The emergence of Convolutional Neural Networks (CNNs) revolutionized medical image analysis. AlexNet, VGGNet, and GoogLeNet demonstrated unprecedented performance in image classification tasks. These architectures introduced concepts of hierarchical feature learning, where early layers detect edges and textures while deeper layers capture semantic information.

Researchers began applying CNNs to fundus image analysis, achieving significant improvements over traditional methods. Key advantages included automatic feature learning without manual engineering, end-to-end training from raw images to diagnostic output, and ability to capture subtle pathological patterns invisible to conventional approaches.

Advanced Architectures for DR Detection

The introduction of residual connections by He et al. addressed the vanishing gradient problem in very deep networks. ResNet's skip connections enable training of networks with hundreds of layers, facilitating better feature propagation and gradient flow. Several studies adapted ResNet architectures for DR detection, achieving improved accuracy through deeper feature representations.

Inception modules, introduced by Google, employ parallel convolutional operations with different kernel sizes, enabling multi-scale feature extraction within a single layer. This architecture is particularly beneficial for medical imaging, where pathological features may appear at various scales. Recent research has explored hybrid architectures combining multiple pre- trained models, including integration of DenseNet, Inception, and ResNet for multi- perspective analysis.

Transfer Learning in Medical Imaging

Transfer learning has become the standard approach for medical image analysis. Models pre-trained on ImageNet provide robust initial feature extractors, which are then fine-tuned on medical datasets. This approach addresses the limited availability of labeled medical data and reduces training time.

Researchers have explored domain adaptation techniques to bridge the gap between natural images (ImageNet) and medical images (fundus photographs), including progressive fine-tuning strategies, learning rate scheduling, and selective layer freezing and unfreezing.

Research Gaps Identified

Analysis of existing literature reveals several critical gaps. Few studies systematically exploit complementary information from color and grayscale representations. Most approaches employ single-path decision-making, potentially missing nuanced diagnostic patterns. Limited exploration exists of adaptive optimization combinations within the same architecture. Many studies focus on binary or three-stage classification, with fewer addressing comprehensive five-stage grading. Models often perform well on training datasets but show degraded performance on external validation sets.

III. METHODOLOGY

Overall Framework Architecture

The proposed multi-decision Inception-ResNet blended hybrid model consists of four major components: a dual-image input processing module that handles both color fundus images (FCIs) and black-white images (BWIs), a multi- layered feature extraction network with 172 weighted layers divided equally for parallel processing, a multi-decision classification module that integrates multiple decision-making

pathways, and an ensemble prediction output that combines predictions from both image streams.

Data Preprocessing Pipeline

Input images are sourced from two major datasets: the EyePACS Dataset containing 88,702 fundus photographs with image resolution varying from 2000×2000 to 4000×4000 pixels, and the APTOS Dataset with 3,662 high-quality retinal images with expert annotations. Selection criteria include minimum image resolution of 312×312 pixels, quality assessment excluding poorly illuminated or blurred images, and balanced distribution ensuring equal representation across all five DR stages.

For each input color fundus image FCI, a corresponding grayscale image BWI is generated using the rgb2gray transformation. This weighted conversion preserves luminance information while removing chromatic data, providing complementary feature representations. The conversion formula is:

$$BWI(x,y) = 0.299 \times R(x,y) + 0.587 \times G(x,y) + 0.114 \times B(x,y)$$

To enhance model robustness and prevent overfitting, comprehensive data augmentation is applied including zooming with random zoom factors between 0.8 and 1.2, random cropping to focus on central retinal regions, random rotation angles from -15° to +15°, horizontal flipping with 50% probability, vertical flipping with 30% probability, horizontal translation up to 10% of image width, and vertical translation up to 10% of image height. These augmentations simulate natural variations in fundus image acquisition, improving generalization.

Multi-Layered Network Architecture

The FCI processing stream consists of 86 layers beginning with initial convolutional blocks featuring 64 filters with 7×7 kernels, stride of 2, and ReLU activation, followed by max pooling with 3×3 kernel and stride of 2. Multiple parallel Inception modules employ convolutional paths with different kernel sizes: a 1×1 convolution path, a 1×1 convolution followed by 3×3 convolution path, a 1×1 convolution followed by 5×5 convolution path, and a 3×3 max pooling followed by 1×1 convolution path. Concatenation of all paths provides multi-scale feature representation.

Deep residual blocks with identity shortcuts implement the function $y = F(x, \{Wi\}) + x$, where F represents the residual function and x is the input. The BWI stream mirrors the FCI architecture with adjusted parameters, focusing on intensity gradients and texture patterns while employing similar Inception and ResNet modules to extract complementary features to color information.

TABLE I
PERFORMANCE COMPARISON WITH STATE- OF-THE-ART MODELS

Model	Accuracy	F1-Score	Parameters
VGG-16	87.3%	85.2%	138M
ResNet-50	91.5%	90.1%	25.6M
Inception-v3	93.2%	92.4%	23.8M
DenseNet-121	94.1%	93.2%	8.0M
EfficientNet-B4	95.7%	94.9%	19.3M
Proposed Model	**98.1%**	**97.6%**	**45.2M**

Multi-Decision Classification Framework

The multi-decision framework employs a three-phase optimization strategy. The primary classification path (epochs 1-70) uses Adam optimization for initial training with learning rate of 0.001, momentum values of 0.7, beta parameters β_1=0.9 and β_2=0.999, and epsilon of 10^{-8}.

The secondary classification path (epochs 71- 210) transitions to SGD optimization with momentum, featuring learning rate scheduling that decreases from 0.0001 to 0.0003, momentum that increases from 0.4 to 0.6, and weight decay of 10^{-4}. The regression-based fine-tuning phase (epochs 211-360) uses Adam optimization for final refinement with learning rate of 0.0005, focusing on boundary cases and difficult classifications with L2 regularization of 10^{-3}.

Multi-Stage DR Classification

DR stages are encoded using a dual-labeling mechanism with primary labels Q1: {000, 001, 010, 011, 100} representing No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR respectively. Secondary labels Q2: {00, 01, 10, 11} indicate confidence levels from low (00) to high (11). The combined representation takes the form DR_Stage = Primary_Label.Secondary_Label, for example 001.11 indicates Mild DR with high confidence.

Loss Functions and Optimization

The classification loss employs categorical cross-entropy for multi-class classification. The regression loss uses mean squared error for confidence estimation. The combined total loss function weights classification at $\alpha=0.7$ and regression at $\beta=0.3$. Statistical validation is performed using chi-square testing to compare observed versus expected frequencies across DR stages.

IV. EXPERIMENTAL SETUP

Dataset Description and Splitting

The EyePACS dataset contains 88,702 fundus photographs with varying resolution from 2000×2000 to 4000×4000 pixels. The DR stage distribution includes 25,810 No DR images, 2,443 Mild DR images, 5,292 Moderate DR images, 873 Severe DR images, and 708 Proliferative DR images. The APTOS dataset provides 3,662 high-quality fundus photographs with expert annotations, balanced representation across severity stages, and standardized image resolution of 3000×3000 pixels.

Data splitting follows a stratified sampling approach with 70% allocated to the training set, 15% to the validation set, and 15% to the testing set, maintaining class distribution across all splits.

Hardware and Software Configuration

The experimental setup utilizes an NVIDIA Tesla V100 GPU with 32GB memory, Intel Xeon E5- 2690 v4 CPU with 28 cores, 256GB DDR4 RAM, and 2TB SSD storage for fast data loading. The software environment comprises TensorFlow 2.8 or PyTorch 1.10 frameworks, Python 3.8 programming language, Keras and TensorBoard for deep learning, OpenCV 4.5 and PIL for image processing, Albumentations for data augmentation, and SciPy with scikit-learn for statistical analysis.

Training Configuration

Training hyperparameters include a batch size of 32 for color images plus 32 for grayscale images, totaling 360 epochs with initial learning rate of 0.001. The learning rate scheduler employs ReduceLROnPlateau with patience of 4 epochs, and early stopping monitors validation loss with patience of 10 epochs.

Performance Evaluation Metrics

Classification performance is evaluated using accuracy, calculated as the ratio of correct predictions to total predictions. Precision measures the proportion of true positive predictions among all positive predictions. Recall (sensitivity) quantifies the proportion of actual positives correctly identified. F1-Score provides the harmonic mean of precision and recall. Specificity measures the proportion of actual negatives correctly identified.

Multi-class evaluation employs Kappa Score to measure inter-rater agreement, AUC-ROC for area under the receiver operating characteristic curve, and confusion matrix for detailed misclassification analysis. Clinical metrics include referral accuracy for correct identification of referable DR (Moderate, Severe, Proliferative stages), early detection rate measuring sensitivity for Mild DR cases, and false negative rate which is critical for patient safety.

V. RESULTS AND DISCUSSION

Overall Performance

The proposed multi-decision Inception-ResNet blended hybrid model achieved outstanding results with overall accuracy of 98.1%, average precision of 97.8%, average recall of 97.5%, average F1-Score of 97.6%, and Kappa Score of 0.976. These results demonstrate the model's exceptional capability in detecting and classifying diabetic retinopathy across all severity stages.

Stage-Wise Performance Analysis

Detailed analysis of individual DR stages reveals consistent high performance. No DR classification achieved 99.2% precision, 99.5% recall, and 99.4% F1-Score across 3,872 test images. Mild DR detection showed 95.8%

Score. The complete multi-path proposed approach achieved 98.1% accuracy and 97.6% F1-Score, confirming the value of the multi- decision strategy.

TABLE III ABLATION STUDY RESULTS

Configuration	Accuracy	Improvement
Colorimages only	95.2%	Baseline
Grayscaleonly	92.8%	-2.4%
Dual-image (proposed)	98.1%	+2.9%

precision, 94.3% recall, and 95.0% F1-Score on 367 images. Moderate DR classification reached 97.1% precision, 96.8% recall, and 96.9% F1- Score with 794 samples. Severe DR achieved 96.5% precision, 95.2% recall, and 95.8% F1-

Score on 131 images. Proliferative DR detection demonstrated 98.3% precision, 97.9% recall, and 98.1% F1-Score across 106 test cases.

TABLE II
STAGE-WISE PERFORMANCE ME777TRICS

DRStage	Precision	Recall	F1-Score	Support
NoDR	99.2%	99.5%	99.4%	3,872
MildDR	95.8%	94.3%	95.0%	367
Moderate DR	97.1%	96.8%	96.9%	794
SevereDR	96.5%	95.2%	95.8%	131
Proliferative DR	98.3%	97.9%	98.1%	106

Comparison with State-of-the-Art

Comparative analysis with existing state-of-the- art models demonstrates the superiority of the proposed approach. VGG-16 achieved 87.3% accuracy and 85.2% F1-Score with 138M parameters. ResNet-50 reached 91.5% accuracy and 90.1% F1-Score using 25.6M parameters. Inception-v3 obtained 93.2% accuracy and 92.4% F1-Score with 23.8M parameters. DenseNet-121 achieved 94.1% accuracy and 93.2% F1-Score using only 8.0M parameters. EfficientNet-B4 reached 95.7% accuracy and 94.9% F1-Score with 19.3M parameters. The proposed model significantly outperforms all baseline methods with 98.1% accuracy and 97.6% F1-Score using 45.2M parameters.

Ablation Studies

Ablation studies quantify the contribution of key architectural components. Processing color images only achieved 95.2% accuracy as the baseline. Using grayscale images alone resulted in 92.8% accuracy, showing a 2.4% decrease. The proposed dual-image approach achieved 98.1% accuracy, demonstrating a 2.9% improvement over single-image processing.

Analysis of the multi-decision framework reveals that a single path using Adam optimization only achieved 96.3% accuracy and 95.7% F1-Score. A dual path combining Adam and SGD reached 97.5% accuracy and 96.9% F1-

Computational Efficiency

Computational analysis reveals practical deployment feasibility. Training time totaled 36 hours on NVIDIA V100 GPU. Inference time per image pair is 0.18 seconds, enabling real-time screening applications. The compressed model size is 176 MB, suitable for deployment on standard hardware. The model requires 24.3 billion floating-point operations (FLOPs) per inference.

Clinical Validation

Clinical validation metrics demonstrate strong applicability for real-world screening. Referral accuracy shows 97.8% sensitivity for referable DR (stages 2-4) and 98.9% specificity for non- referable DR, with positive predictive value of 96.7%. Early detection performance includes 94.3% detection rate for Mild DR and only 5.7% false negative rate for early stages, critical for preventing disease progression.

Failure Case Analysis

Analysis of misclassified cases identifies common failure modes. Poor image quality including blurred or over/under-exposed images accounts for approximately 40% of failures. Rare lesion patterns with atypical presentations not well-represented in training data contribute 25% of errors. Boundary cases with features spanning multiple stages represent 20% of failures. Artifact interference from dust, reflections, or imaging artifacts causes 15% of misclassifications. These findings suggest targeted improvements for future model iterations.

Generalization Performance

Cross-dataset validation assesses model generalization. Training on EyePACS and testing on APTOS achieved 96.4% accuracy. Training on APTOS and testing on EyePACS reached 95.8% accuracy. Combined training on both datasets yielded 98.1% accuracy. The minimal performance degradation (less than 2.5%) in cross-dataset evaluation demonstrates strong generalization capability across different acquisition protocols and patient populations.

VI. DISCUSSION

Key Findings and Contributions

This research makes several significant contributions to automated diabetic retinopathy detection. The dual-image processing innovation systematically integrates color and grayscale fundus images, providing complementary feature representations. Color images capture vascular abnormalities

and hemorrhage characteristics, while grayscale images emphasize texture patterns and microaneurysms. This dual approach achieves 2.9% accuracy improvement over single-image methods, demonstrating the value of multi-modal feature extraction.

The multi-decision framework advantage manifests through the three-phase optimization strategy. Phase 1 enables rapid initial convergence with adaptive learning rates using Adam optimizer. Phase 2 provides fine-grained parameter tuning with momentum-based updates through SGD. Phase 3 performs final refinement for boundary cases returning to Adam optimization. This approach outperforms single-optimizer strategies by 1.8%, validating the multi-decision concept.

Architectural efficiency emerges from the 172- layer symmetric architecture with 86 layers for each image stream. This design provides balanced computational load across both processing pathways, enables efficient parallel processing capabilities, and reduces training time compared to sequential processing approaches.

Clinical Implications

The model's clinical applicability extends across multiple healthcare scenarios. For screening applications, the high sensitivity (97.8%) for referable DR makes it suitable for population- based screening programs, telemedicine platforms in remote areas, and primary care settings with limited ophthalmology access. The 94.3% detection rate for Mild DR enables early intervention support through timely referral for preventive treatment, reduced progression to vision-threatening stages, and cost-effective healthcare resource allocation.

Workflow integration benefits from the 0.18- second inference time, supporting real-time screening during patient consultations, high- throughput analysis in screening centers, and seamless integration with existing electronic health record systems. These capabilities position the model as a practical clinical decision support tool.

Limitations and Challenges

Several limitations warrant consideration. Data dependency remains a concern as performance relies on high-quality labeled data. Imbalanced class distribution, especially for Severe and Proliferative stages, may affect rare case detection. Limited diversity in demographic representation could impact generalization to underrepresented populations.

Computational requirements present deployment challenges. The 172-layer architecture requires substantial GPU memory, limiting deployment on resource-constrained devices. Training duration of 36 hours may restrict iterative development cycles. Model optimization techniques such as pruning, quantization, and knowledge distillation could address these limitations.

Interpretability concerns arise from the deep learning black-box nature. Limited explainability for clinical decision-making may hinder clinical adoption. Regulatory approval processes often require interpretable predictions and transparent decision pathways. Integration of attention mechanisms and gradient-based visualization techniques could improve model interpretability.

Generalization to unseen populations requires validation across different ethnic groups, imaging equipment variations, and acquisition protocol diversity. Adaptation to new clinical settings may necessitate retraining or fine-tuning procedures.

Comparison with Existing Approaches

The proposed model demonstrates several advantages over traditional methods. Unlike handcrafted feature-based approaches, it automatically learns optimal feature representations, captures subtle patterns invisible to manual detection, and adapts to diverse image characteristics without domain expertise.

Compared to single-model architectures using only ResNet or Inception, the hybrid design benefits from multi-scale feature extraction through Inception modules, deep residual learning from ResNet architecture, and complementary strengths through integrated design. While ensemble methods combine multiple models, the proposed hybrid integrates advantages within a single architecture, achieving reduced inference time compared to sequential ensemble predictions and more efficient memory utilization.

Future Research Directions

Several promising research directions emerge from this work. Explainable AI integration should implement attention mechanisms to highlight diagnostic regions, generate heatmaps showing lesion locations, and develop natural language explanations for predictions to enhance clinical interpretability.

Multi-modal fusion could integrate optical coherence tomography (OCT) images, combine fundus photography with

fluorescein angiography, and incorporate patient clinical data including HbA1c levels and diabetes duration for comprehensive risk assessment.

Longitudinal analysis capabilities would enable tracking disease progression over time, predicting future DR development risk, and optimizing screening interval recommendations based on individual patient trajectories.

Model compression and optimization through pruning techniques could reduce model size, quantization would enable edge device deployment, and lightweight versions would support mobile application development for point-of-care screening.

Clinical trial validation should conduct prospective studies in real-world settings, compare performance with ophthalmologist diagnoses, and assess impact on patient outcomes and healthcare costs to establish clinical utility and cost-effectiveness.

VII. CONCLUSION

This research presents a novel multi-decision Inception-ResNet blended hybrid model for automated diabetic retinopathy detection, achieving state-of-the-art performance of 98.1% accuracy. The key innovations include dual- image processing architecture with systematic integration of color fundus and grayscale images through parallel 86-layer processing streams, capturing complementary diagnostic features. The multi-decision framework employs three- phase optimization strategy combining Adam and SGD optimizers across 360 epochs, enabling effective exploration and fine-grained parameter tuning.

Comprehensive five-stage classification accurately discriminates among No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR with stage-wise F1-scores exceeding 95%. Clinical viability is demonstrated through high sensitivity (97.8%) for referable DR and fast inference time (0.18 seconds), supporting real- world screening applications.

The model outperforms existing approaches including VGG-16, ResNet-50, Inception-v3, DenseNet-121, and EfficientNet-B4, demonstrating the effectiveness of the proposed hybrid architecture and multi-decision framework. Ablation studies confirm the significant contributions of both dual-image processing (+2.9% accuracy) and multi-path optimization (+0.6% accuracy).

While the model shows exceptional performance, several limitations warrant attention including computational requirements for training and inference, limited interpretability for clinical decision support, and the need for validation across diverse populations and imaging protocols. Future work should focus on explainable AI integration, multi-modal data fusion, longitudinal analysis capabilities, model compression for edge deployment, and prospective clinical validation.

The proposed system represents a significant advancement toward automated, accurate, and accessible diabetic retinopathy screening, with potential to reduce preventable vision loss through early detection and timely intervention. As deep learning continues to evolve, integration of such systems into routine clinical practice promises to transform ophthalmologic care delivery, particularly in resource-limited settings where access to specialist expertise remains challenging.

VIII. ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by their respective institutions. We thank the EyePACS and APTOS dataset contributors for making their data publicly available for research purposes. We also acknowledge the valuable feedback from ophthalmologists and clinical experts who provided domain knowledge and validation insights.

REFERENCES

[1] G. Lim, M. L. Cheng, W. Hsu, and M. L. Lee, "Integrated optic disc and cup segmentation with deep learning," in *IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2015, pp. 162-169.

[2] M. D. Abràmoff, Y. Lou, A. Erginay, et al., "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Ophthalmology & Visual Science*, vol. 57, no. 13, pp. 5200-5206, 2016.

[3] V. Gulshan, L. Peng, M. Coram, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402-2410, December 2016.

[4] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962-969, July 2017.

[5] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, "Zoom-in-net: Deep mining lesions for diabetic retinopathy detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 267-275.

[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for

computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016,pp. 770-778.

[8] P. Porwal, S. Pachade, R. Kamble, et al., "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, July 2018.

[9] D. S. W. Ting, C. Y. Cheung, G. Lim, et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211-2223, December 2017.

[10] S. Qummar, F. G. Khan, S. Shah, et al., "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530-150539, 2019.

[11] N. Eftekhari, H. R. Pourreza, M. Masoudi, K. Ghiasi-Shirazi, and E. Saeedi, "Microaneurysm detection in fundus images using a two-step convolutional neural network," *BioMedical Engineering OnLine*, vol. 18, no. 1, pp. 1-16,December 2019.

[12] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, 2020, Art. no. 100377.

[13] S. Majumder and N. Kehtarnavaz, "Multitasking deep learning model for detection of five stages of diabetic retinopathy," *IEEE Access*, vol. 9, pp. 123220-123230, 2021.

[14] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C. H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164-118173, 2020.

[15] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification,"*Computers and Electrical Engineering*, vol. 72,pp. 274-282, November 2018.

[16] M. Mateen, J. Wen, M. Hassan, N. Nasrullah, S. Sun, and S. Hayat, "Automatic detection of diabetic retinopathy: A review on datasets, methods and evaluation metrics," *IEEE Access*, vol. 8, pp. 48784-48811, 2020.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[18] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International Conference on Machine Learning*, 2013, pp. 1139-1147.

[19] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, andL. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009,pp. 248-255.

[20] C. Lim, R. Shin, K. C. Lee, et al., "Computer- aided diagnosis of diabetic retinopathy based on fundus image," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 665-675, August 2019.

[21] S. K. Henge, N. R. Viraati, M. Alhussein, A. S. Kushwaha, K. Aurangzeb, and R. Singh, "Detection of diabetic retinopathy using a multi- decision inception-ResNet-blended hybrid model," *IEEE Access*, vol. 13, pp. 8988-9005, January 2025.

[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015.

[23] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP- BMEI)*, 2017, pp. 1-11.

[24] N. G. Reddy, R. Venkatesh, C. Jayadev, et al., "Diabetic retinopathy and diabetic macular edema in people with early-onset diabetes," *Clinical Diabetes*, vol. 40, no. 2, pp. 222-232, April 2022.

[25] World Health Organization, "Global report on diabetes," WHO Press, Geneva, Switzerland, 2016.