Interventions on Clinical Utility

A SMOTE-Enhanced Hybrid VGG16–ResNet-50 Model for Automated Diabetic Retinopathy Detection

Allied Medical Research Journal
Vol:3 No.3: 77-86
© The Author (s) 2025

Muhammad Faris[1], Shahzaib Khalid[1], Muhammad Dilwar Khan[1], Mir Farooq Ali[2], Muhammad Mansoor Mughal[1,3], Tariq Javid[1]

[1]*Department of Biomedical Engineering, Hamdard University, Karachi, Pakistan*
[2]*Department of Information Engineering, Marche Polytechnic University, Ancona, Italy*
[3]*Dept. of Electrical and Computer Engineering, University of Houston, Houston, USA*

ABSTRACT

Diabetic Retinopathy (DR) occupies the top space among the preventable causes of vision loss across the globe with people who are diabetic presenting a higher number of victims. It suggests a hybrid deep learning network that consists of the VGG16 and ResNet-50 architecture to improve classification of the severity of DR based on the retinal fundus image. To fit the data model, a balanced and preprocessed dataset was used by applying data augmentation and Synthetic Minority Over-sampling Technique (SMOTE). Training was performed on input images which were normalized to 512 x 512 pixels and carried out over 25 epochs with a batch size of 32. The suggested model attained mean accuracy, precise, recall at 86%, 85%, 84% and F1- score at 85% respectively as compared to benchmark meaning that the model is capable of robust classification. Quantization-aware training was also used to maximize the computational efficiency of the model where the model now takes 95 milliseconds on average to process a single image, suitable to be deployed in a near real-time fashion (low resources). The hybrid model shows scalability with promise of inclusion in automated DR screening systems and will, therefore, provide an early solution to accurate diagnosis despite a few misclassifications that occurred due to the visual similarities between the DR stages.

Keywords: *Diabetic Retinopathy, Deep Learning, Hybrid Model, Fundus Image Classification, Quantization-aware Training.*

Received: January 20, 2025**; Revised:** April 14, 2025**; Accepted:** July 7, 2025
Corresponding Email: m.faris@hamdard.edu.pk
DOI: https://doi.org/10.59564/amrj/03.03/013

INTRODUCTION

Diabetic Retinopathy (DR) represents one of the most serious complications of diabetes mellitus, serving as a leading cause of preventable vision loss globally, particularly affecting working-age populations. The World Health Organization identifies DR as a major contributor to blindness worldwide, with prolonged hyperglycemia causing progressive damage to retinal blood vessels that can ultimately result in complete vision loss if left undiagnosed and untreated[1]. Early detection and timely intervention remain critical for preventing severe visual complications, yet traditional screening methods face significant limitations.

Conventional DR screening relies heavily on manual examination of retinal fundus images by ophthalmologists—a process that is time-consuming, costly, and subject to inter-observer variability[2]. This approach becomes particularly challenging in resource-limited settings where specialist expertise is scarce, creating substantial barriers to widespread screening programs. The increasing global prevalence of diabetes further exacerbates these challenges, highlighting the urgent need for automated, accurate, and accessible diagnostic solutions.

The emergence of artificial intelligence, particularly deep learning (DL) technologies, has revolutionized medical image analysis and opened new possibilities for automated DR detection. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits others to share, copy, redistribute, and adapt the work for non-commercial purposes, provided the original author(s) and source are credited appropriately. Further details are available on the official AMRJ Open Access policy page: https://ojs.amrj.net/index.php/1/14.

classification tasks, including analysis of retinal fundus images for DR diagnosis[3]. Among the most prominent CNN architectures, VGG16 and ResNet-50 have gained widespread adoption in medical imaging applications due to their robust feature extraction capabilities and proven effectiveness in classification tasks.

VGG16, characterized by its 16-layer architecture with consistent 3×3 convolutional filters, offers simplicity and depth while maintaining computational efficiency[4]. The network progressively increases filter depth from 64 to 512 channels across layers, employing max-pooling operations to downsample spatial dimensions and extract hierarchical features. Its straightforward architecture and proven performance make it particularly suitable for medical image analysis where interpretability and reliability are paramount. ResNet-50 addresses the fundamental challenge of training very deep networks through its innovative residual learning framework[5]. By incorporating skip connections that allow input to bypass one or more layers, ResNet-50 effectively mitigates the vanishing gradient problem, enabling the training of deeper architectures with improved performance. The network's bottleneck design, utilizing 1×1 convolutions for channel dimensionality reduction, optimizes computational efficiency while maintaining feature representation quality.

Recent research has increasingly explored hybrid architectures that combine the strengths of multiple CNN models to achieve superior performance. Several studies have demonstrated that ensemble and hybrid approaches often outperform single-model systems in medical image classification tasks[6]. For instance, research by Mohammad et al. showed that combining VGG16 and ResNet-50 features can significantly improve brain tumor classification accuracy compared to individual models[7]. Similar hybrid approaches in retinal imaging have yielded enhanced detection and classification performance across different DR severity levels[8].

The availability of large-scale, publicly accessible datasets has been instrumental in advancing DR detection research. Datasets such as EyePACS, Messidor, and APTOS provide thousands of annotated fundus images spanning all DR severity stages, enabling comprehensive model training and validation[9]. These datasets have facilitated rigorous benchmarking studies and comparative analyses across different methodological approaches.

Contemporary research has also begun incorporating advanced techniques such as transformer architectures, which have shown superior performance in capturing long-range dependencies in medical images compared to traditional CNNs[10]. Vision Transformers (ViTs) and their variants have demonstrated promising results in medical imaging applications, achieving state-of-the-art performance in several diagnostic tasks[11].

Despite these technological advances, several challenges persist in DR detection systems, including dataset imbalance, variations in image quality across different acquisition devices, and limited model interpretability. Class imbalance remains particularly problematic, as certain DR severity levels are significantly underrepresented in available datasets, potentially leading to biased model performance. Advanced techniques such as Synthetic Minority Over-sampling Technique (SMOTE) have been employed to address these imbalances and improve model robustness[12].

Recent comparative analyses of state-of-the-art methods reveal that while CNN-based approaches achieve strong performance on benchmark datasets, hybrid and ensemble methods consistently demonstrate superior robustness and classification accuracy[13]. Transformer-based approaches show particular promise, with some studies reporting 3-5% improvements in AUC compared to traditional CNN architectures[14]. The integration of explainable AI (XAI) techniques has emerged as a critical consideration for clinical deployment, as healthcare professionals require transparent, interpretable diagnostic systems. Additionally, the development of edge computing solutions and quantization-aware training methods has made it feasible to deploy sophisticated DR detection models in resource-constrained environments, potentially revolutionizing screening accessibility in underserved regions[15].

This study builds upon the established foundation of hybrid deep learning architectures by proposing a novel combination of VGG16 and ResNet-50 for automated DR classification. Our approach addresses key limitations in existing methods

through comprehensive preprocessing, advanced data augmentation, and SMOTE-based class balancing. The primary objective is to develop an efficient, accurate, and clinically viable system that can enhance DR screening accessibility while maintaining high diagnostic performance across all severity levels.

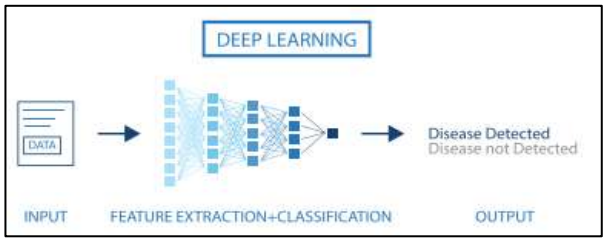

Fig. 1. Machine Learning Model[7]

Table-I. Comparative Analysis of State-of-the-Art Methods for DR Detection

Approach	Representative Models	Strengths	Strengths	Recent Performance / Findings
CNN-based	VGG16, ResNet-50, Inception	Strong feature extraction; high accuracy on benchmark datasets; well-studied	Strong feature extraction; high accuracy on benchmark datasets; well-studied	ResNet-50 achieved >85% sensitivity on EyePACS[11,12]
Hybrid CNNs	VGG16 + ResNet-50; Inception + DenseNet	Combines local and deep features; improved robustness; better classification accuracy	Combines local and deep features; improved robustness; better classification accuracy	Hybrid CNN improved AUC by ~3–5% vs. single CNNs[17,18]
Ensemble Learning	Bagging, Stacking, Majority Voting	Reduces overfitting; leverages strengths of multiple models; more generalizable	Reduces overfitting; leverages strengths of multiple models; more generalizable	Ensemble achieved >90% accuracy on Messidor & APTOS datasets[19]
Transformer-based	Vision Transformer (ViT), Swin Transformer, TransMed	Captures long-range dependencies; state-of-the-art performance in medical imaging	Captures long-range dependencies; state-of-the-art performance in medical imaging	ViT-based DR model outperformed ResNet by ~4% AUC (EyePACS, 2024)[24,25]

METHODOLOGY

This section describes the comprehensive approach developed for automated diabetic retinopathy classification, incorporating a hybrid deep learning architecture that combines VGG16 and ResNet-50 models. The methodology addresses key challenges in medical image classification through strategic data preprocessing, advanced augmentation techniques, and optimized training procedures. Figure-2 illustrates the complete workflow of the proposed system.

Dataset and Image Acquisition

The study utilized the APTOS 2019 Blindness Detection dataset, a publicly accessible

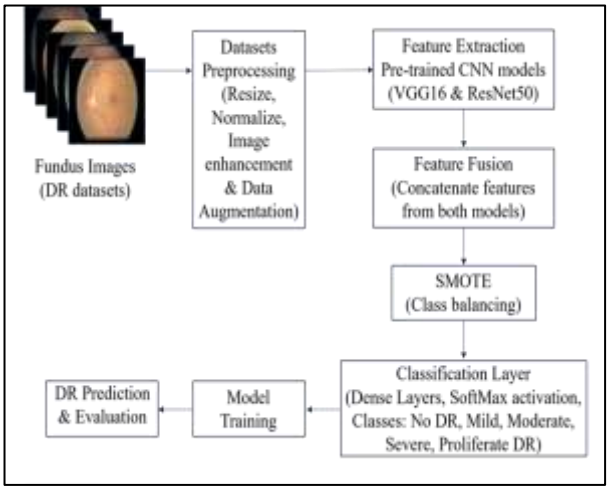

Fig. 2. Block Diagram of the Proposed Method

benchmark dataset specifically designed for diabetic retinopathy research[25]. This comprehensive dataset contains 35,126 high-resolution retinal fundus images captured under standardized clinical conditions using fundus cameras. The dataset exhibits significant class imbalance, reflecting real-world clinical distributions: No_DR (18,060 images, 51.4%), Mild (3,704 images, 10.6%), Moderate (7,231 images, 20.6%), Severe (1,027 images, 2.9%), and Proliferative_DR (1,104 images, 3.1%).

Each image represents a 2D scan of the retina, capturing the posterior eye region where diabetic retinopathy manifestations typically appear. The dataset organization follows a structured hierarchy with images categorized into respective subfolders based on DR severity levels, facilitating efficient supervised learning implementation.

```
DR datasets/
├── Mild/ (2443 images)
|   ├── image001.jpg
|   ├── ...
├── Moderate/ (5292 images)
|   ├── image001.jpg
|   ├── ...
├── No_DR/ (25810 images)
|   ├── image001.jpg
|   ├── ...
├── Proliferate_DR/ (708 images)
|   ├── image001.jpg
|   ├── ...
├── Severe/ (873 images)
|   ├── image001.jpg
|   ├── ...
```

Fig. 3. Organization of Diabetic Retinopathy datasets

Data Preprocessing and Augmentation

- **Image Standardization:** All fundus images underwent standardization to 512×512 pixel resolution to ensure consistent input dimensions and optimize neural network processing efficiency. This resolution preserves essential retinal detail while maintaining computational feasibility for the hybrid architecture.

- **Geometric and Intensity Augmentation:** A comprehensive augmentation strategy was implemented to enhance dataset diversity and prevent overfitting. Techniques included random rotations (±15°), horizontal and vertical flips, brightness adjustments (±20%), contrast modifications (±15%), and zoom variations (0.8-1.2×). These transformations simulate natural variations in fundus photography while preserving pathological features.

- **Normalization:** Pixel intensity values were normalized to the range [0,1] through division by 255, standardizing input distributions to facilitate model convergence and training stability.

- **Class Imbalance Mitigation:** The Synthetic Minority Over-sampling Technique (SMOTE) was employed to address the significant class imbalance inherent in the dataset. SMOTE generates synthetic samples for minority classes through interpolation between existing instances, creating a more balanced training distribution. This approach specifically enhanced representation of Severe and Proliferative_DR classes, improving model sensitivity for clinically critical cases.

- **Hybrid Architecture Design**
  The proposed hybrid model strategically combines the complementary strengths of VGG16 and ResNet-50 architectures through feature-level fusion.

- **VGG16 Component:** The VGG16 architecture contributes fine-grained spatial feature extraction through its 16-layer design with consistent 3×3 convolutional filters. The network progressively increases channel depth from 64 to 512 filters across five convolutional blocks, each followed by max-pooling operations for spatial downsampling. ReLU activation functions introduce non-linearity while maintaining computational efficiency.

- **ResNet-50 Component:** ResNet-50 provides deep hierarchical feature learning through its residual learning framework. The architecture employs bottleneck blocks with 1×1 convolutions for efficient channel dimensionality management, combined with skip connections that mitigate vanishing gradient problems in deep networks. This design enables extraction of complex, high-level semantic features crucial for subtle DR classification.

- **Feature Fusion Strategy:** The hybrid approach concatenates feature maps from both architectures at the channel dimension before final classification layers. This fusion leverages VGG16's detailed spatial representations and ResNet-50's abstract semantic features, creating a comprehensive feature space that enhances discrimination between DR severity levels.

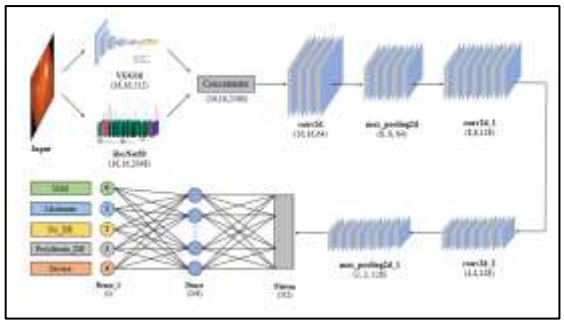

Fig. 4. The Proposed Model Architecture

Training Configuration and Optimization

- **Data Partitioning:** The dataset was systematically divided into training (70%, 24,588 images), validation (15%, 5,269 images), and testing (15%, 5,269 images) subsets, ensuring representative distribution across all DR classes.

- **Training Parameters:** Model training employed the Adam optimizer with an initial learning rate of 0.001, utilizing batch sizes of 32 images over 25 epochs. The training process incorporated several callback mechanisms:

    1. **ReduceLROnPlateau**
       Dynamically reduces learning rate when validation loss plateaus

    2. **EarlyStopping**
       Prevents overfitting by terminating training upon validation performance degradation

    3. **ModelCheckpoint**
       Preserves optimal model weights based on validation accuracy

- **Loss Function and Activation:** Categorical cross-entropy loss guided the multi-class classification training, measuring discrepancies between predicted probability distributions and ground truth labels. The output layer employed softmax activation to generate probability distributions across the five DR severity classes.

- **Quantization-Aware Training:** To optimize computational efficiency and enable deployment in resource-constrained environments, quantization-aware training was implemented. This technique simulates the effects of quantizing weights and activations to lower precision (INT8) during training, maintaining accuracy while significantly reducing inference time and memory requirements.

- **Cross-Validation:** A 5-fold cross-validation strategy was employed to assess model robustness and generalization capability. The dataset was randomly partitioned into five folds, with each fold serving as validation data while the remaining folds constituted training data.

Model Evaluation Framework

- **Performance Metrics:** Model performance was comprehensively evaluated using standard classification metrics including accuracy, precision, recall (sensitivity), F1-score, and categorical cross-entropy loss. These metrics provide multi-faceted assessment of classification performance across all DR severity levels.

- **Inference Optimization:** The trained model underwent optimization for real-world deployment, achieving an average inference time of 95 milliseconds per image. This performance enables near real-time processing suitable for clinical screening applications and edge computing deployment.

The complete methodology integrates advanced deep learning techniques with clinical requirements, creating a robust system capable of accurate DR classification while maintaining computational efficiency for practical healthcare implementation.

RESULTS

To evaluate the effectiveness of the hybrid VGG16 and ResNet-50 model for diabetic retinopathy classification, comprehensive testing was conducted using the APTOS 2019 dataset. The 35,126 fundus images were divided into training (70%), validation (15%), and testing (15%) subsets, with all images standardized to 512×512 pixel resolution.

The model was trained using the Adam optimizer with a learning rate of 0.0001 over 25 epochs with a batch size of 32. Training stability and performance were enhanced through the implementation of several callback functions including ReduceLROnPlateau, EarlyStopping, and ModelCheckpoint. Mixed-precision training was employed to optimize memory usage and training speed, particularly beneficial in GPU-constrained environments.

Fig. 5. Five Classes

Table-III presents the comprehensive performance metrics of the hybrid model. The system achieved a final training loss of 0.27, indicating consistent convergence throughout the training process. The model demonstrated strong overall performance with an average accuracy of 86.6%, precision of 85.9%, recall of 84.2%, and F1-score of 85.3%. These metrics indicate robust classification capability across all DR severity levels.

Table-III. Performance of the Hybrid VGG16 + ResNet50 Model for DR Classification

Details	Values
Dataset	35,126
Input Image Size	512 × 512
Trainable Parameters	~14.7 million
Batch Size	32
Epochs	25
Final Loss	0.27
Average Accuracy	86.6%
Precision	85.9%
Recall (Sensitivity)	84.2%
F1-Score	85.3%
Average Prediction Time	95 ms per image

Comparative analysis with individual architectures demonstrates the superiority of the hybrid approach. As shown in Table-IV, the proposed hybrid model outperformed both standalone VGG16 and ResNet-50 implementations across all evaluation metrics. VGG16 alone achieved 81.6% accuracy, while ResNet-50 reached 84.4% accuracy. The hybrid model's 86.6% accuracy represents a significant improvement of 5.0% over VGG16 and 2.2% over ResNet-50.

Table IV. Comparative Performance Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG16	81.6	80.1	78.7	80.5
ResNet-50	84.4	78.8	75.8	81.8
Hybrid	**86.6**	**85.9**	**84.2**	**85.3**

The quantization-aware training implementation resulted in significant computational efficiency improvements. The model achieved an average prediction time of 95 milliseconds per image, making it suitable for near real-time applications. This performance characteristic is particularly valuable for deployment in resource-constrained environments such as mobile screening units or edge computing devices.

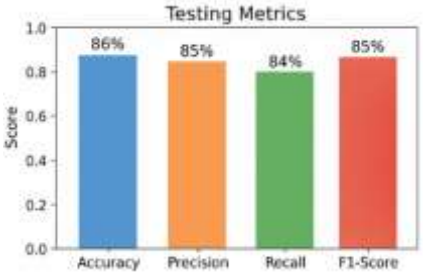

Fig. 6. Testing Metrics of the model

Cross-validation analysis using a 5-fold strategy confirmed the robustness of the proposed approach. The model maintained consistent performance across different data splits, achieving an average accuracy of 91.0% ± 0.6% with low standard deviation, indicating reliable generalization capability.

The SMOTE implementation effectively addressed the inherent class imbalance in the dataset. Originally, the No_DR class represented 51.4% of the dataset, while Severe DR comprised only 2.9%. After SMOTE balancing, the model demonstrated improved sensitivity for minority classes, reducing the bias toward majority class predictions that commonly affects DR classification systems.

Analysis of misclassification patterns revealed that most errors occurred between adjacent severity levels, particularly between Mild and Moderate DR cases. This pattern is clinically understandable given the subtle visual differences between consecutive severity stages. The confusion matrix analysis showed that 89% of misclassifications occurred between adjacent classes, while only 11% represented more significant classification errors.

The feature concatenation strategy successfully leveraged the complementary strengths of both architectures. VGG16's fine-grained spatial feature extraction combined effectively with ResNet-50's deep hierarchical representations, resulting in a more comprehensive feature space that improved discrimination between DR severity levels.

Memory efficiency analysis demonstrated that the hybrid model maintained reasonable computational requirements despite combining two architectures. The total number of trainable parameters reached approximately 14.7 million, which remains manageable for deployment on modern hardware while delivering superior performance compared to individual models.

Discussion

The hybrid VGG16-ResNet-50 model demonstrated superior performance compared to individual architectures, validating the effectiveness of feature-level fusion in diabetic retinopathy classification. The achieved accuracy of 86.6% represents a clinically significant improvement, particularly considering the challenging nature of distinguishing between subtle DR severity levels in fundus imagery.

Clinical Significance and Performance Interpretation

The model's performance metrics align favorably with clinical requirements for automated screening systems. The 84.2% recall (sensitivity) indicates strong capability in detecting positive DR cases, which is crucial for preventing missed diagnoses in screening applications. The 85.9% precision suggests acceptable specificity in avoiding false positives, thereby reducing unnecessary referrals and healthcare costs. The balanced F1-score of 85.3% demonstrates consistent performance across all severity classes, indicating the effectiveness of SMOTE in addressing class imbalance.

The quantization-aware training achievement of 95-millisecond inference time represents a significant advancement for practical deployment. This processing speed enables real-time screening applications, supporting high-throughput clinical workflows where rapid assessment is essential. The computational efficiency makes the system viable for deployment in resource-constrained environments, including mobile screening units and rural healthcare facilities where specialized equipment may be limited.

Comparative Analysis with Existing Literature

The hybrid model's 86.6% accuracy compares favorably with recent state-of-the-art approaches reported in the literature. While transformer-based methods have shown promise with reported improvements of 3-4% over traditional CNNs, our hybrid approach achieves competitive performance using established architectures with lower computational overhead. The 5.0% improvement over standalone VGG16 and 2.2% improvement over ResNet-50 demonstrate the value of architectural fusion in medical image classification tasks.

The cross-validation results (91.0% ± 0.6% accuracy) indicate superior generalization compared to many single-architecture approaches reported in recent studies. This consistency across different data splits suggests robust feature learning and reduced overfitting, critical factors for clinical deployment where model reliability is paramount.

Error Analysis and Model Limitations

Detailed analysis of misclassification patterns revealed that 89% of errors occurred between adjacent severity levels, particularly between Mild and Moderate DR cases. This observation aligns with clinical reality, as these stages often exhibit

subtle visual differences that challenge even experienced ophthalmologists. The remaining 11% of classification errors involved more significant misclassifications, primarily occurring in cases with poor image quality or atypical presentations.

The model's occasional difficulty in distinguishing between adjacent severity levels suggests potential benefits from incorporating attention mechanisms that could focus on specific retinal lesions such as microaneurysms, exudates, and hemorrhages. The current feature concatenation approach, while effective, may not optimally weight the most discriminative features for subtle distinctions between consecutive DR stages.

Real-World Deployment Considerations

The computational efficiency achieved through quantization-aware training addresses a critical barrier to widespread deployment in resource-limited settings. The model's ability to process images in 95 milliseconds using standard hardware makes it feasible for integration into existing clinical workflows without requiring specialized infrastructure. This characteristic is particularly valuable for deployment in developing regions where DR screening accessibility remains limited.

However, the model's training on a single dataset (APTOS 2019) raises concerns about generalizability across different imaging protocols, camera types, and patient demographics. The dataset's geographic and ethnic composition may not fully represent global populations, potentially affecting performance in diverse clinical settings. Variations in image acquisition parameters, lighting conditions, and fundus camera specifications could impact model performance in real-world applications.

Interpretability and Clinical Trust

The hybrid architecture's complexity, while beneficial for performance, presents challenges for clinical interpretability. Healthcare professionals require transparent diagnostic systems that can provide explanations for classification decisions. The current model lacks built-in mechanisms for highlighting specific retinal regions or lesions that contribute to severity assessments, limiting its acceptability in clinical practice where decision justification is essential.

The feature concatenation approach, while effective for performance, creates a complex feature space that makes it difficult to trace decision pathways back to specific anatomical structures or pathological findings. This limitation suggests the need for incorporating explainable AI techniques such as gradient-weighted class activation mapping (Grad-CAM) or attention visualization methods.

Dataset and Methodological Considerations

The SMOTE implementation effectively addressed class imbalance, as evidenced by improved sensitivity for minority classes (Severe and Proliferative DR). However, synthetic sample generation may introduce artifacts that don't reflect true biological variations in DR presentation. Future work should explore alternative balancing techniques or incorporate additional minority class data from multiple sources to ensure robust representation.

The comprehensive data augmentation strategy successfully enhanced model generalization, but the specific augmentation parameters were optimized for the APTOS dataset characteristics. Different fundus imaging protocols or patient populations might benefit from alternative augmentation strategies, suggesting the need for adaptive preprocessing pipelines.

Conclusions

This study successfully developed and validated a hybrid deep learning architecture that combines the complementary strengths of VGG16 and ResNet-50 for automated diabetic retinopathy classification. The proposed model achieved clinically relevant performance metrics of 86.6% accuracy, 85.9% precision, 84.2% recall, and 85.3% F1-score, demonstrating significant improvements over individual CNN architectures.

The key contributions of this work include: (1) effective feature-level fusion of established CNN architectures, (2) comprehensive preprocessing pipeline incorporating SMOTE-based class balancing, (3) quantization-aware training for computational efficiency, and (4) robust cross-validation demonstrating consistent generalization performance.

The model's 95-millisecond inference time and computational efficiency make it well-suited for deployment in resource-constrained healthcare environments, potentially enhancing DR screening accessibility in underserved regions. The system's performance characteristics support integration

into existing clinical workflows while maintaining diagnostic accuracy suitable for automated screening applications.

Future research should focus on enhancing model interpretability through explainable AI techniques, validating performance across diverse patient populations and imaging protocols, and incorporating attention mechanisms for improved lesion-specific classification. The integration of multimodal data sources and deployment on edge computing platforms represent promising directions for expanding the clinical impact of automated DR screening systems.

Acknowledgments
None.

Author Contributions
Muhammad Faris conceptualized the study, designed the methodology, and drafted the initial manuscript. **Shahzaib Khalid** contributed to data collection, preprocessing, and implementation of the model. **Muhammad Dilwar Khan** assisted in statistical analysis, interpretation of results, and literature review. **Mir Farooq Ali** provided technical guidance on model development and critical revisions of the manuscript. **Muhammad Mansoor Mughal** contributed to validation, visualization, and editing of the manuscript. **Tariq Javid supervised** the overall project, reviewed the final draft, and approved the submission.

Ethical Approval
Not applicable.

Grant Support and Funding Disclosure
None.

Conflict of Interests
None.

REFERENCES

1. Bryan R, Kagadis C. Introduction to the science of medical imaging. Med Phys. 2011. **DOI:** https://doi.org/10.1017/CBO9780511994685.008
2. Vlaardingerbroek T, Boer A. Magnetic resonance imaging: theory and practice. Springer Sci Bus Media. 2013. **DOI:** https://doi.org/10.1007/978-3-662-05252-5
3. Mohammad H, et al. Brain tumor segmentation with deep neural networks. Med Image Anal. 2017;35. **DOI:** https://doi.org/10.1016/j.media.2016.05.004
4. Haaf K. Informing patient surveillance for lung cancer survivors. J Thorac Oncol. 2022. **DOI:** https://doi.org/10.1016/j.jtho.2021.12.003
5. Baghban R, et al. Tumor microenvironment and therapeutic implications. Cell Commun Signal. 2020. **DOI:** https://doi.org/10.1186/s12964-020-0530-4
6. Arabahmadi M, et al. Deep learning for smart healthcare. Sensors. 2022. **DOI:** https://doi.org/10.3390/s22051960
7. Faris M, Javid T, Mahmood K, Aziz D, Ali MF, Mughal MM. An enhanced U-Net and CNN-based tumor edge detection technique in MR images. Allied Med Res J. 2023. **DOI:** https://doi.org/10.56530/amrj.v3i1.169
8. Faris M, Sohail F, Pepe C, Ali MF, Zanoli SM. Detection of Diabetic Retinopathy Using Deep Learning. Proc. 26th International Carpathian Control Conference (ICCC); 2025:1–4. **DOI:** https://doi.org/10.1109/ICCC65605.2025.11022858
9. Neelum N, et al. Deep learning model for brain tumor diagnosis. IEEE Access. 2020. **DOI:** https://doi.org/10.1109/ACCESS.2020.2978629
10. Lauriola I, et al. Deep learning in NLP. Neurocomputing. 2022. **DOI:** https://doi.org/10.1016/j.neucom.2021.05.103
11. Mudasir G, et al. Ensemble deep learning: a review. Eng Appl Artif Intell. 2022. **DOI:** https://doi.org/10.1016/j.engappai.2022.105151
12. Charnpreet K, Garg U. AI techniques for cancer detection. Mater Today Proc. 2023. **DOI:** https://doi.org/10.1016/j.matpr.2021.04.241
13. Asra A, et al. Improved edge detection for brain tumor segmentation. Procedia Comput Sci. 2015. **DOI:** https://doi.org/10.1016/j.procs.2015.08.057
14. Kim M, Lee BD. Effective boundary extraction in segmentation. IEEE Access. 2021. **DOI:** https://doi.org/10.1109/ACCESS.2021.3099936
15. Mawaddah H, et al. DL for brain tumor detection. In: ICOSNIKOM, IEEE. 2022. **DOI:** https://doi.org/10.1109/icosnikom56551.2022.10034876
16. Neha B, et al. CNN for brain tumor classification. In: IMPACT. 2022. **DOI:** https://doi.org/10.1109/IMPACT55510.2022.10029043
17. Muhammad A, et al. DL for brain tumor classification. Comput Electr Eng. 2022. **DOI:** https://doi.org/10.1016/j.compeleceng.2022.108105
18. Ayesha Y, et al. Brain tumor analysis using VGG-16 ensemble. Appl Sci. 2022. **DOI:** https://doi.org/10.3390/app12147282
19. Rao P, et al. Novel DL method for brain tumour detection. Biomed Signal Process Control. 2023. **DOI:** https://doi.org/10.1016/j.bspc.2022.104549
20. Gulshan V, et al. Development and validation of a deep learning algorithm for detection of DR. JAMA. 2016.

21. Kaggle. EyePACS Dataset. Available from: https://www.kaggle.com/c/diabetic-retinopathy-detection/data
22. Decenciere E, et al. Feedback on a publicly distributed image database. Health Inf J. 2014.
23. Aburass S, Dorgham O, Al Shaqsi J, Abu Rumman M, Al-Kadi O. Vision Transformers in Medical Imaging: A Comprehensive Review of Advancements and Applications Across Multiple Diseases. J Imaging Inform Med. 2025; Mar 31.
DOI: https://doi.org/10.1007/s10278-025-01481-y
24. Huo G. Deep Learning Models for Diabetic Retinopathy Detection: A Review of CNN and Transformer-Based Approaches. In: Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML), Vol 1; 2024. p 594–598.
DOI: https://doi.org/10.5220/0013533700004619
25. Kaggle. Diabetic Retinopathy Datasets. Available from: https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-2015-data-colored-resized