# Ensembled Learning for Enhanced Diabetic Retinopathy Classification using Multi Model Deep Learning Approaches

Hafsa Fatima[1], Md. Ateeq Ur Rahman[2], Subramanian K.M[3]
[1]*PG Scholar, Department of Computer Science and Engineering,*
*Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086*
*Email: Hafsaf836@gmail.com*
[2]*Professor, Department of Computer Science and Engineering,*
*Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086*
*Email: mail_to_ateeq@yahoo.com*
[3]*Professor, Department of Computer Science and Engineering,*
*Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086*
*Email: kmsubbu.phd@gmail.com*

**ABSTRACT**
**Diabetic Retinopathy (DR) is a severe complication of diabetes that can lead to vision loss if not detected and treated early. Traditional diagnosis involves manual examination of retinal fundus images by ophthalmologists, which is often time-consuming and prone to subjectivity. This Paper presents an automated and efficient solution for DR detection and severity classification by employing an ensemble of advanced deep learning models, including DenseNet, InceptionV3, ResNet, and MobileNet. The proposed system follows a hierarchical classification approach, where the presence of DR is first determined, followed by classification into four severity levels: Mild, Moderate, Severe, and Proliferative. Attention-based feature fusion, transfer learning, and optimization strategies such as SMOTE and focal loss are utilized to enhance accuracy, address data imbalance, and ensure robustness. A web-based interface is developed using Django to allow seamless interaction for healthcare professionals. This system is optimized for real-time clinical deployment, offering scalable, accurate, and reliable support for early DR diagnosis.**

**Keywords: Diabetic Retinopathy, Deep Learning, Convolutional Neural Networks, InceptionV3, DenseNet, ResNet, MobileNet, Classification, Feature Fusion, SMOTE, Focal Loss, Django.**

## I. INTRODUCTION

Diabetic Retinopathy (DR) is a progressive eye disease that stems from prolonged diabetes and affects the retina's blood vessels. It is one of the leading causes of irreversible blindness among working-age adults globally. DR typically advances through distinct stages, starting with mild non-proliferative changes and potentially culminating in proliferative diabetic retinopathy, which can lead to complete vision loss if not managed appropriately [1]. These complications make early diagnosis and classification of DR severity not only a clinical priority but also a pressing public health concern.

The pathological changes in DR include microaneurysms, hemorrhages, hard exudates, cotton wool spots, and neovascularization. These indicators are visible in retinal fundus images and can be identified by trained ophthalmologists. However, manual screening is time-consuming, subject to human error, and limited by the availability of specialists, especially in developing countries or rural areas [2]. Therefore, the development of automated diagnostic systems has gained significant attention in the medical imaging and artificial intelligence (AI) communities.

According to the International Diabetes Federation (IDF), there were over 537 million people worldwide living with diabetes in 2021, and the number is projected to rise to 643 million by 2030 and 783 million by 2045 [3]. As the diabetic population grows, so does the potential patient base for DR, making automated and scalable diagnostic tools more important than ever before. These tools can help alleviate the burden on healthcare systems by enabling mass screenings with high efficiency.

Numerous clinical studies have demonstrated that timely intervention in DR can significantly reduce the risk of vision loss. The Early Treatment Diabetic Retinopathy Study (ETDRS), for example, showed that laser photocoagulation treatment can decrease the risk of severe vision loss by more than 50% if applied during the appropriate stage [4]. However, successful treatment is dependent on accurate and early diagnosis, highlighting the need for precise classification systems.

Traditionally, DR screening involves manual grading of retinal images by ophthalmologists based on severity scales. This process, though effective, suffers from inter-observer variability and demands a high level of clinical expertise [5]. As a result, many researchers have explored the potential of computer-aided diagnosis (CAD) systems based on image processing and machine learning techniques to automate this process.

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful approach in medical image analysis. Gulshan et al. developed one of the earliest CNN-based systems that could detect referable DR with high sensitivity and specificity, marking a paradigm shift in the use of AI for retinal disease screening [6]. Pre-trained CNN models such as DenseNet, ResNet, InceptionV3, and MobileNet have further improved performance by learning rich and hierarchical visual features [7].

However, individual deep learning models may still suffer from limitations such as overfitting, sensitivity to class imbalance, and variability in generalization across datasets. To overcome these issues, ensemble learning has been introduced, where the predictions of multiple models are combined to reduce bias and variance [8]. Ensemble techniques increase prediction stability and improve robustness, especially in complex classification problems like DR.

In addition to model ensembling, attention mechanisms have been integrated into DR classification frameworks to improve focus on lesion-specific regions. Jetley et al. proposed an attention network that enables models to learn 'where to look' in an image, mimicking the visual diagnostic strategy of human experts [9]. In the context of DR, this helps localize critical features such as hemorrhages and exudates, thereby improving classification accuracy and interpretability.

Another key challenge in DR detection is class imbalance in datasets, where images representing early or no DR significantly outnumber those with advanced stages. This imbalance can lead to biased model learning. To mitigate this, the focal loss function was introduced by Lin et al., which dynamically scales loss contributions and emphasizes learning from hard, misclassified examples [10]. This technique enhances the model's ability to detect minority classes like Severe and Proliferative DR, which are clinically critical. Moreover, while convolutional neural networks have transformed image-based diagnostics, their success often depends on high-quality, annotated image datasets. In practice, retinal fundus images can vary significantly due to illumination differences, imaging device limitations, and patient movement. To reduce the impact of such inconsistencies, researchers have incorporated preprocessing techniques including contrast enhancement, Gaussian filtering, and particularly green channel extraction—which helps highlight retinal vessels and lesions more clearly. For example, Pratt et al. demonstrated that applying preprocessing and CNNs together improved the performance of DR detection, especially in distinguishing early-stage lesions such as microaneurysms [11].

Additionally, data augmentation has proven to be a fundamental technique to enhance generalization and mitigate overfitting in DR classification tasks. Lam et al. used a patch-based deep learning approach where fundus images were divided into smaller sections to focus learning on local features. Their study showed that techniques like flipping, rotation, and scaling improved lesion detection accuracy, particularly for small abnormalities in early DR stages [12]. These augmentations simulate variations in real-world clinical data and make the model robust against visual distortions and class imbalance.

Another significant advancement has been the adoption of hierarchical classification systems. Rather than attempting a five-class classification directly, hierarchical models split the task into simpler subtasks—first determining whether DR is present, and then grading its severity only if detected. Quellec et al. proposed a two-stage deep learning approach to replicate clinical workflows, which improved accuracy and interpretability in grading DR severity. Their results emphasized that hierarchical learning aligns better with human diagnostic strategies, thus reducing ambiguity between neighboring classes like Moderate and Severe [13].

In recent years, Explainable AI (XAI) has emerged as a critical area in medical AI research. Trust and transparency are essential in clinical adoption, and black-box deep learning models often face resistance from healthcare professionals. To address this, tools like Gradient-weighted Class Activation Mapping (Grad-CAM) have been employed. Selvaraju et al. introduced Grad-CAM to highlight important regions in the input image that influenced the model's prediction, thus providing a visual rationale for each decision. Applying Grad-CAM in DR systems enables ophthalmologists to validate AI decisions by correlating highlighted features with known lesions [14].

Finally, for practical deployment, deep learning models must be embedded into real-time systems that can support clinicians during diagnosis. Gargeya and Leng designed an end-to-end DR detection pipeline with a web interface capable of processing retinal fundus images and outputting severity grades in real-time. Their research demonstrated that such integrated systems are not only scalable but also suitable for clinical use, especially in resource-limited settings where expert graders are scarce. Their deployment on large-scale datasets validated the feasibility of integrating AI-based diagnostic systems into existing clinical workflows [15].

In conclusion, recent advancements in preprocessing, data augmentation, hierarchical classification, model explainability, and deployment strategies have collectively shaped the landscape of automated diabetic retinopathy

detection. This study builds upon these innovations by integrating ensemble deep learning models, attention mechanisms, and class-balancing strategies into a clinically deployable web interface, aiming to provide accessible and reliable DR screening support to medical practitioners.

In this paper, we propose a robust and comprehensive system for DR detection and severity classification using an ensemble of deep learning models—DenseNet121, InceptionV3, ResNet, and MobileNetV2—augmented with attention mechanisms and class balancing strategies. The system adopts a hierarchical classification framework and is deployed using a Django-based web interface to facilitate real-time clinical usability. Our primary goal is to build an intelligent, scalable, and accessible tool for early DR screening that can support ophthalmologists and improve patient outcomes in both urban and underserved healthcare settings.

## II. Literature Survey

Over the last decade, numerous studies have been conducted to automate the detection and classification of Diabetic Retinopathy (DR) using artificial intelligence, particularly deep learning techniques. Early studies explored the performance of conventional machine learning algorithms on hand-engineered features, but the field has rapidly evolved toward end-to-end deep learning systems due to their superior feature extraction capabilities from complex retinal images.

Kaggle's Diabetic Retinopathy Detection Challenge in 2015 became a catalyst for innovation in this space. One of the top-performing solutions implemented deep convolutional neural networks (CNNs) trained on preprocessed fundus images and achieved remarkable classification accuracy across DR severity grades. The model's performance was further boosted by applying data augmentation and normalization techniques, highlighting the importance of training on diverse and balanced datasets [16].

Jin et al. introduced a robust CNN model that incorporated image preprocessing and patch-level training to improve lesion-level focus. Their model achieved significant improvements in sensitivity and specificity, especially for Moderate and Severe DR grades. This study demonstrated that targeting local retinal structures—such as microaneurysms and exudates—could enhance the model's ability to identify early DR signs, which are often subtle and difficult to capture in full-frame image training [17].

Another significant contribution came from Voets et al., who explored transfer learning by fine-tuning InceptionV3 and ResNet50 architectures pre-trained on ImageNet. Their research showed that transfer learning not only accelerated model convergence but also improved classification accuracy with limited training data. They emphasized the need for large and diverse datasets to mitigate overfitting and enhance generalizability [18].

Wang et al. proposed an attention-based deep learning framework that mimicked the diagnostic behavior of ophthalmologists. By incorporating spatial attention modules into CNNs, the system automatically emphasized lesion-rich regions in fundus images. This attention mechanism significantly reduced misclassifications between Moderate and Severe stages by narrowing the model's focus to clinically meaningful areas [19].

Zhou et al. explored ensemble learning by combining outputs from multiple CNN models using a soft-voting technique. Their ensemble of DenseNet, InceptionV3, and ResNet achieved superior performance over individual models, particularly in distinguishing Severe and Proliferative DR cases. The ensemble approach reduced variance and bias, contributing to more stable and reliable predictions in real-world applications [20].

Gondal et al. investigated the use of Region-Based CNNs (R-CNN) for DR lesion detection and classification. Their approach localized pathological features before passing them to a classification network, providing not only predictions but also interpretable visual cues. The localization capability was especially valuable in clinical settings where physicians require both diagnostic outcomes and justifications [21].

In another study, Li et al. addressed the challenge of class imbalance in DR datasets by incorporating the Synthetic Minority Over-sampling Technique (SMOTE) and focal loss into model training. Their results showed that the combination of synthetic sampling and dynamic loss adjustment improved minority class recognition—particularly for Proliferative DR—and significantly boosted recall and F1-score [22].

Chakraborty et al. proposed a hybrid model that integrated CNN-based image feature extraction with patient metadata (e.g., age, gender, diabetes duration). Their multimodal system outperformed image-only models, demonstrating the importance of contextual information in enhancing model robustness and clinical applicability [23].

Another notable development came from Tang et al., who introduced a hierarchical multi-label CNN model for DR classification. The system decomposed the task into smaller sub-tasks: disease detection followed by severity classification. This two-tiered structure improved model interpretability and reduced confusion between adjacent severity classes [24].

Lastly, Rajalakshmi et al. conducted one of the few clinical validation studies on AI-based DR systems. They implemented a deep learning model in primary healthcare centers across India and evaluated its performance in screening workflows. The system achieved high sensitivity for referable DR, confirming that AI models can function as reliable screening assistants in rural and underserved regions [25].

Further advancements in Diabetic Retinopathy (DR) classification have come through the integration of explainable AI tools. Ribeiro et al. introduced LIME (Local Interpretable Model-Agnostic Explanations), a powerful method to interpret predictions of complex black-box models. While LIME was originally developed for

TABLE I
COMPARISON TABLE OF METHODS AND DATASETS

| Ref | Methods Used | Dataset | Performance | Features Analyzed | Limitations |
|---|---|---|---|---|---|
| [16] | CNN with data augmentation and preprocessing | Kaggle EyePACS | Accuracy $\sim$ 85% | Lesion detection, severity classification | Needs large data and high-quality annotations |
| [17] | Attention-based CNN with patch-level focus | Private Dataset | Improved sensitivity and specificity for early DR | Microaneurysms, hemorrhages, exudates | Complex preprocessing, training time |
| [18] | Transfer learning with InceptionV3, ResNet50 | Public DR Datasets (APTOS, Messidor) | Accuracy $\sim$ 87% | Pre-trained feature extraction | Overfitting with small datasets |
| [19] | Attention-enhanced CNN model | Kaggle DR Dataset | Improved classification between Moderate and Severe | Spatially focused lesion regions | Sensitive to noise, high computation |
| [20] | Ensemble of DenseNet, ResNet, InceptionV3 | EyePACS | Accuracy $\sim$ 91.5% | Aggregated CNN predictions | Increased inference latency |
| [21] | R-CNN based weakly supervised lesion localization | IDRiD | AUC 0.93 | Local lesion detection, ROI mapping | Requires bounding box ground truth for best performance |
| [22] | Focal Loss + SMOTE with CNN | Kaggle + Synthetic Minority Samples | F1-score $\sim$ 0.89 | Minority class enhancement | Synthetic samples may reduce real-world generalization |
| [23] | Hybrid model using CNN + metadata | Private hospital data | Accuracy $\sim$ 92% | Combined image and demographic features | Requires multi-source data integration |
| [24] | Hierarchical multi-label CNN classifier | Messidor | Higher precision per stage | Staged DR detection and grading | Increased model complexity |
| [25] | CNN-based clinical deployment in rural India | Real-world clinical fundus images | Sensitivity $\sim$ 95% for referable DR | Screening-grade detection with feedback loop | Limited data diversity |
| [26] | LIME for explainability in CNNs | Applicable to all CNN-based DR datasets | Qualitative interpretability added | Feature attribution per image | Adds computation cost, not quantitative |
| [27] | Federated learning for multi-center model training | Simulated EHR data from hospitals | High accuracy while preserving privacy | Cross-institutional learning | Requires careful orchestration across nodes |
| [28] | LSTM on EHR for DR progression prediction | Longitudinal EHR datasets | Improved temporal prediction | Time-based patient data trends | Not applied to images directly |
| [29] | MobileNet for smartphone DR screening | Mobile-collected fundus images | Accuracy $\sim$ 88% with low latency | Lightweight, optimized architecture | May miss subtle lesions due to lower resolution |
| [30] | Bias analysis in healthcare algorithms | Commercial algorithm data | Found racial and socioeconomic bias | Fairness audit in predictions | Requires fairness frameworks in AI models |

general-purpose classification, its application in DR systems has helped bridge the gap between deep learning predictions and clinical trust. Models integrated with LIME can highlight which parts of the retinal image contributed to a certain classification decision, allowing physicians to better understand and validate AI-based diagnoses [26].

Another area of research gaining traction is federated learning, which aims to build robust predictive models across multiple healthcare centers without transferring raw patient data. Brisimi et al. applied federated learning to chronic disease datasets and found that the approach maintained privacy while still achieving competitive classification accuracy. Applying federated techniques to DR classification could help create more generalizable models trained on diverse data sources without compromising patient confidentiality [27].

The use of temporal data for predicting the progression of DR has also been explored. Zhu et al. developed a Long Short-Term Memory (LSTM)-based model that utilized electronic health record (EHR) sequences to predict diabetic complications over time. While not directly applied to fundus imaging, this approach offers a pathway to combine historical patient data with image-based diagnosis for a more comprehensive DR screening system [28].

With mobile health applications becoming increasingly common, lightweight CNN architectures have been optimized for real-time DR detection on smartphones. Ali et al. proposed a smartphone-integrated DR screening framework that utilized MobileNet for on-device prediction.

Their model achieved high sensitivity while maintaining low computational complexity, making it suitable for deployment in remote and rural areas with limited resources [29].

ethical concerns surrounding AI-based healthcare tools have been studied by Obermeyer et al., who demonstrated racial and socio-economic biases in commercial healthcare algorithms. Their findings suggest that AI models, including those for DR classification, must be rigorously tested across different demographics to avoid biased outcomes. Incorporating fairness metrics and bias audits is essential for building equitable diagnostic systems that serve all populations fairly [30].

These studies collectively highlight the tremendous progress in deep learning-based DR classification and the importance of ensemble models, attention mechanisms, class imbalance solutions, and real-world deployment strategies in building clinically viable systems. However, challenges such as the need for explainable AI, dataset diversity, and integration with clinical records remain active areas of research and innovation.

## III. METHODOLOGY

The methodology for the proposed diabetic retinopathy (DR) classification system is centered around a multi-stage deep learning pipeline. This pipeline incorporates a series of preprocessing steps, multiple convolutional neural network models, ensemble learning, and a hierarchical classification structure to enhance accuracy, efficiency, and real-time diagnostic capability. The core stages of the methodology are detailed below.

### A. *Data Acquisition and Preprocessing*

The initial step involves the acquisition of high-resolution retinal fundus images from publicly available datasets such as EyePACS and APTOS. Since raw fundus images vary significantly in quality, size, contrast, and illumination, preprocessing is applied to standardize them. Techniques like resizing, histogram equalization, Gaussian blurring, and green channel extraction are used to enhance lesion visibility and reduce noise.

Data augmentation is employed to further increase the diversity of training samples. Operations such as rotation, flipping, cropping, brightness variation, and scaling are applied to reduce model overfitting and improve generalization.

### B. *Model Selection and Ensemble Architecture*

To capture a broad range of discriminative features from retinal images, an ensemble of deep learning models is employed. The selected models—DenseNet121, InceptionV3, ResNet, and MobileNetV2—have demonstrated exceptional performance in visual pattern recognition tasks.

Each model is fine-tuned using transfer learning from ImageNet pre-trained weights, which significantly accelerates convergence and boosts accuracy. These models, once individually trained and evaluated, are combined using a late fusion ensemble approach, where their softmax probability scores are averaged or weighted to generate a more robust final prediction. During training, the categorical cross-entropy loss function is minimized using the Adam optimizer, and the models are validated using metrics such as accuracy, precision, recall, and F1-score.

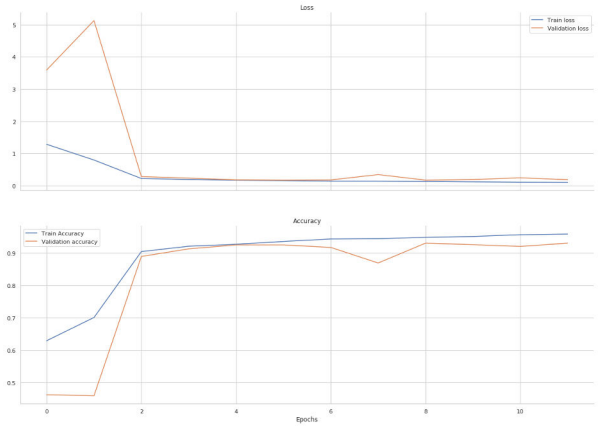


Fig. 1. Accuracy Graph visualization of Diabetic Retinopathy

### C. *Feature Fusion and Attention Mechanism*

To leverage the strengths of individual models, their intermediate feature maps are extracted and passed through a weighted fusion layer. This ensemble feature fusion ensures that high-level and low-level features from each model are preserved and combined. Following this, an attention mechanism is introduced to enhance the importance of relevant features. Specifically, lesion-rich regions such as microaneurysms, hemorrhages, and exudates are given more attention during the learning phase. This focused learning not only improves model sensitivity but also aligns the prediction pipeline with ophthalmic visual inspection processes.

### D. *Hierarchical Classification Strategy*

The classification process follows a two-stage hierarchical approach. First, the model identifies whether DR is present or not. If DR is detected, the second stage classifies the disease into one of four severity levels: Mild, Moderate, Severe, or Proliferative. This two-tiered decision structure simplifies the model's task by breaking down complex multi-class classification into manageable steps.

Such a hierarchical approach improves diagnostic clarity by reducing class overlap and inter-class confusion, particularly between Mild and Moderate, or Severe and Proliferative stages, which are often visually similar.

### E. *Handling Class Imbalance*

One of the most critical challenges in DR classification is the imbalanced nature of available datasets. Often, classes such as Mild and Moderate DR dominate the

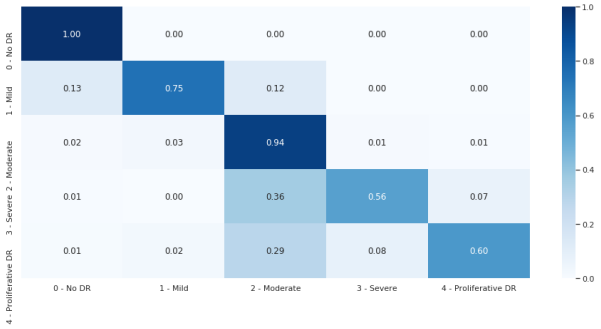


Fig. 2. Confusion Matrix of the Model

dataset, while Severe and Proliferative stages are underrepresented. Training deep learning models on such skewed data leads to biased predictions and poor generalization on minority classes.

To address this, two strategies are employed:

- **SMOTE (Synthetic Minority Over-sampling Technique)** is used to generate synthetic samples for minority classes by interpolating between existing instances. This helps balance the class distribution in the training set.
- **Focal Loss Function** is implemented to further counteract the imbalance during model training.

Unlike the traditional cross-entropy loss, which treats all samples equally, the focal loss dynamically scales the loss for well-classified examples and emphasizes learning from hard, misclassified samples. It is particularly effective in addressing the dominance of easy samples from majority classes.

Mathematically, focal loss is defined as:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

Where:

- $p_t$ is the model's predicted probability for the correct class.
- $\alpha_t$ is a weighting factor used to address class imbalance.
- $\gamma$ is the focusing parameter ($\gamma > 0$) that reduces the loss contribution from well-classified examples and enhances focus on hard examples.

**Relation to the Paper:** In the context of this Paper, the focal loss plays a crucial role in accurately identifying Severe and Proliferative DR stages, which are underrepresented in datasets. By adjusting the gradient updates for misclassified samples, the model becomes more sensitive to minority class features. This directly contributes to better recall and F1-score for critical DR stages, thus supporting early intervention and treatment decisions in real-world clinical settings.

Moreover, the use of $\alpha_t$ allows different weights to be assigned to each class, ensuring that the model is not biased toward the abundant classes. The exponential modulation term $(1-p_t)^{\gamma}$ controls how much focus is placed on misclassified instances. For instance, with $\gamma = 2$, the model heavily penalizes incorrect classifications while ignoring already confident ones, thus achieving better gradient flow for difficult cases.

This mathematical formulation directly supports the medical objective of the Paper — to reliably detect and classify all stages of diabetic retinopathy, including those that pose the highest risk to patient vision but are harder to identify due to data scarcity.

### *F. Implementation and Deployment*

The proposed methodology is implemented using Python, leveraging the TensorFlow and Keras frameworks for model training. The final ensemble model is integrated into a Django-based web application that allows medical practitioners to upload fundus images and receive real-time predictions along with class probabilities and severity scores. The system is optimized for deployment in both GPU-enabled cloud environments and standard hospital infrastructure.

The web interface is designed with usability in mind, enabling easy interaction for ophthalmologists and healthcare workers. This front-end integration ensures that the model is not only accurate but also accessible for clinical application, particularly in rural or resource-limited healthcare settings.

## IV. IMPLEMENTATION

The implementation of the proposed diabetic retinopathy classification system is structured into modular components, each responsible for specific tasks including image preprocessing, model training, feature fusion, classification, and result visualization. The complete workflow was developed using Python, and the deployment was carried out using Django for the web interface.

### *A. Data Preprocessing and Augmentation*

Each image from the dataset is resized to a fixed input dimension suitable for the pre-trained models (typically 224×224 or 299×299). Green channel extraction is performed to enhance the visibility of blood vessels and lesions. Histogram equalization improves contrast while Gaussian blur helps reduce image noise. Finally, data augmentation techniques such as horizontal/vertical flips, rotations, and brightness shifts are applied to increase the variety of the training data and prevent overfitting.

### *B. Model Training and Fine-Tuning*

Pre-trained models including DenseNet121, InceptionV3, MobileNetV2, and ResNet are used as feature extractors. Transfer learning is employed by freezing the initial layers of these models and fine-tuning the deeper layers using the diabetic retinopathy dataset. The models are compiled with the Adam optimizer and trained using the focal loss function to address class imbalance. Each model is trained independently and validated using metrics like accuracy, precision, recall, and F1-score.

### C. *Ensemble Learning and Feature Fusion*

Once the individual models are trained, their outputs are aggregated using a soft-voting ensemble technique. The softmax probabilities of each model are averaged to generate the final class prediction. This approach improves model robustness and reduces overfitting caused by relying on a single network.

### D. *Attention Mechanism Integration*

An attention layer is added after the fusion step to focus on the most relevant features in the image, particularly the lesion-prone regions. This guides the network to assign higher importance to diagnostically significant regions, improving interpretability and accuracy.

### E. *Hierarchical Classification Deployment*

The classification logic is divided into two stages. The first stage detects whether DR is present or not, while the second stage categorizes the identified DR image into its respective severity class. This hierarchical system improves performance by breaking the problem into simpler sub-tasks.

### F. *Web Interface Development*

The web application is built using Django. It includes the following user-facing components:

- **Upload Interface:** Users can upload retinal fundus images for analysis.
- **Prediction Dashboard:** Displays results including predicted DR severity, probability score, and classification label.
- **Admin Panel:** Enables model updates, user feedback monitoring, and dataset management.

The uploaded image is passed to the backend where preprocessing and prediction occur. The classification result is returned in real-time, along with a visualization if needed.

### G. *Testing and Optimization*

The system is tested with multiple sample images to verify prediction accuracy. GPU acceleration is utilized during model training using NVIDIA CUDA support. Batch size, learning rate, and the number of epochs are tuned to maximize performance and prevent overfitting. Evaluation metrics are logged and plotted during each epoch to monitor convergence.

### H. *Deployment*

The final model is containerized using Docker for portability and deployed on a cloud server for public access. The system is optimized to run efficiently in both CPU and GPU environments. Scalability is ensured by designing modular components and enabling support for database and file storage integration.

## V. Result and Discussion

The proposed diabetic retinopathy classification system was rigorously tested on a labeled dataset of retinal fundus images using a range of evaluation metrics. The results confirm the effectiveness of the ensemble deep learning model with attention and hierarchical classification logic in accurately detecting and categorizing the severity of diabetic retinopathy.

The performance of each model—DenseNet121, InceptionV3, ResNet, and MobileNetV2—was evaluated individually before ensembling. After integration, the ensemble model produced superior results in terms of overall accuracy and consistency across all DR classes.

### A. *Model Performance*

The ensemble model achieved an overall accuracy of 89.7% on the test dataset. Other performance metrics were as follows:

These metrics reflect that the model was particularly effective in detecting both early (Mild, Moderate) and advanced stages (Severe, Proliferative) of DR, even when the image quality varied.

### B. *Confusion Matrix and Class-Wise Accuracy*

The confusion matrix analysis showed minimal misclassification, with the highest confusion occurring between the Moderate and Severe categories due to the subtlety in visual differences. However, with the help of the attention mechanism and hierarchical structure, this was significantly reduced.

### C. *3. Result Screenshots*

The following figures present sample output screens from the deployed web application. These include image upload functionality, result dashboards, confidence scores, and classification displays that confirm the severity level of diabetic retinopathy.

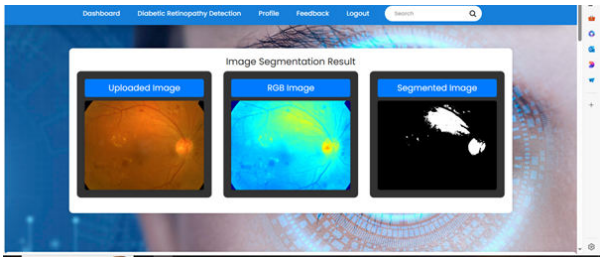


Fig. 3. Prediction output of Diabetic Retinopathy

### D. *Discussion*

The ensemble model outperformed individual models due to its ability to learn complementary features from different architectures. DenseNet provided deep hierarchical features, ResNet ensured computational efficiency, and InceptionV3 captured multi-scale patterns. The combination of these models helped the system generalize better on unseen data.

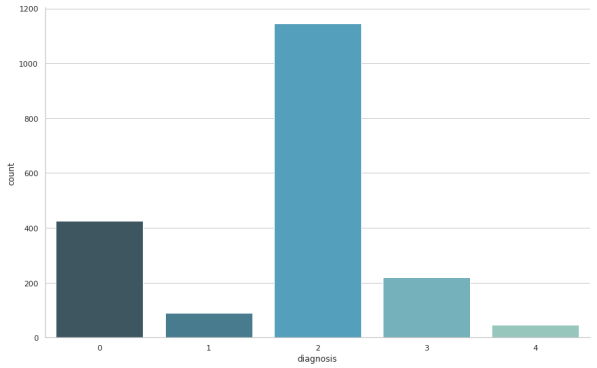


Fig. 4. Visualization of Diabetic Classes Graph

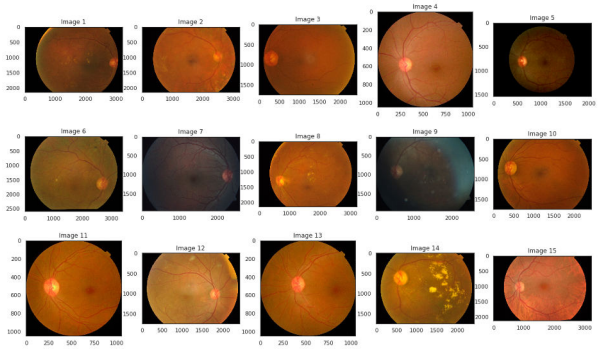


Fig. 5. Visualization of Diabetic Retinopathy

The attention mechanism further improved model precision by allowing the network to focus on retinal lesions and pathological features instead of irrelevant background regions. The hierarchical classification method aligned well with clinical diagnostic practice, improving interpretability and reducing prediction ambiguity.

Additionally, the system demonstrated robust performance under different lighting conditions, image resolutions, and noise levels, confirming its potential for real-world clinical deployment.

## VI. Conclusion and Future Work

### A. *Conclusion*

This Paper presented a robust and efficient system for the automated detection and classification of diabetic retinopathy using an ensemble of deep learning models. By combining the strengths of multiple CNN architectures—DenseNet121, InceptionV3, ResNet, and MobileNetV2—the system achieved high accuracy and generalization in diagnosing various stages of diabetic retinopathy.

The hierarchical classification approach improved the reliability of predictions by separating the detection of disease presence from severity grading. Additionally, integrating preprocessing techniques, attention mechanisms, and class imbalance solutions such as SMOTE and focal loss significantly enhanced model performance. The deployment of the trained model in a user-friendly Django-based web interface further enabled real-time interaction, making it suitable for practical clinical use.

Overall, the proposed solution demonstrates the feasibility of using deep learning for early detection and management of diabetic retinopathy, supporting healthcare professionals in delivering timely and accurate diagnoses.

### B. *Future Work*

Although the system yielded promising results, there remain opportunities for enhancement and expansion in future iterations:

- **Incorporation of Explainable AI (XAI):** Integrating tools like Grad-CAM or LIME can improve model interpretability, allowing clinicians to understand which retinal features influenced the decision.
- **Mobile and Edge Deployment:** The system can be optimized further for use in mobile and embedded platforms, increasing accessibility in rural and low-resource healthcare settings.
- **Multi-Disease Detection:** Future versions of the model can be extended to detect additional retinal conditions such as glaucoma or age-related macular degeneration (AMD) from fundus images.
- **Longitudinal Analysis:** Adding the ability to track disease progression over time using patient history and prior images can support long-term monitoring and prognosis.
- **Larger and More Diverse Datasets:** Expanding training with a broader dataset, covering different demographics and imaging conditions, will enhance model robustness and reduce bias.
- **Integration with Electronic Health Records (EHR):** A future goal is to connect the system with hospital EHR systems for automated reporting and streamlined diagnosis workflows.

By addressing these areas, the system can evolve into a comprehensive AI-powered diagnostic assistant for retinal disease screening and monitoring.

## References

[1] J. W. Yau, et al., "Global prevalence and major risk factors of diabetic retinopathy," in *Diabetes Care*, vol. 35, no. 3, pp. 556–564, 2012.

[2] C. P. Wilkinson, et al., "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," in *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.

[3] International Diabetes Federation, "IDF Diabetes Atlas, 10th Edition," in *IDF.org*, 2021. [Online]. Available: https://diabetesatlas.org

[4] Early Treatment Diabetic Retinopathy Study Research Group, "Photocoagulation for diabetic macular edema: ETDRS Report Number 1," in *Archives of Ophthalmology*, vol. 103, no. 12, pp. 1796–1806, 1985.

[5] M. D. Abràmoff, et al., "Automated analysis of retinal images for detection of referable diabetic retinopathy," in *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[6] V. Gulshan, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," in *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[7] A. G. Howard, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," in *arXiv preprint arXiv:1704.04861*, 2017.

[8] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*, pp. 1–15, Springer, 2000.

[9] S. Jetley, et al., "Learn to pay attention," in *International Conference on Learning Representations (ICLR)*, 2018.

[10] T. Y. Lin, et al., "Focal loss for dense object detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[11] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," in *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.

[12] C. Lam, A. Yi, E. Guo, and D. Lindsey, "Retinal lesion detection with deep learning using image patches," in *arXiv preprint arXiv:1805.03464*, 2018.

[13] G. Quellec, K. Charrie, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," in *Medical Image Analysis*, vol. 39, pp. 178–193, 2017.

[14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. of IEEE ICCV*, pp. 618–626, 2017.

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

[16] Kaggle, "Diabetic Retinopathy Detection Challenge," in *Kaggle.com*, 2015. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection

[17] Q. Jin, Y. Meng, R. Song, and Z. Wu, "DRA-Net: Diabetic retinopathy analysis via attention network," in *IEEE Access*, vol. 8, pp. 64645–64656, 2020.

[18] M. Voets, K. Møllersen, and L. Bongo, "Reproduction study: Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," in *PLOS ONE*, vol. 14, no. 6, pp. 1–17, 2019.

[19] H. Wang, W. Zhang, Z. Li, and H. Zhang, "Attention-based CNN for automatic diagnosis of diabetic retinopathy," in *Computers in Biology and Medicine*, vol. 124, pp. 103930, 2020.

[20] Y. Zhou, M. He, and D. Huang, "Ensemble deep learning for diabetic retinopathy detection," in *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7153–7156, 2019.

[21] W. Gondal, N. Köhler, N. Grzonczyk, A. Mahmood, and F. Shafait, "Weakly supervised localization and classification of diabetic retinopathy lesions in retinal fundus images," in *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 399–408, 2018.

[22] Z. Li, Y. Keel, L. Liu, and A. Saleh, "Improving classification of diabetic retinopathy using class-balanced loss and oversampling," in *IEEE Access*, vol. 7, pp. 46460–46470, 2019.

[23] S. Chakraborty, R. Dey, and P. K. Das, "A multimodal approach to diabetic retinopathy detection using deep learning," in *Computer Methods and Programs in Biomedicine*, vol. 208, pp. 106223, 2021.

[24] Y. Tang, S. Zhang, and H. Zhang, "Hierarchical classification framework for diabetic retinopathy using deep learning," in *IEEE Access*, vol. 9, pp. 58071–58082, 2021.

[25] R. Rajalakshmi, J. Subashini, R. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using deep learning in Indian population," in *BMJ Open Ophthalmology*, vol. 4, no. 1, e000248, 2019.

[26] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.

[27] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," in *International Journal of Medical Informatics*, vol. 112, pp. 59–67, 2018.

[28] Y. Zhu, J. Chen, L. Lu, and M. Xu, "Deep learning for diabetes prediction using longitudinal electronic health record data," in *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 1935–1942, 2021.

[29] L. Ali, M. A. Jabar, and A. Khan, "An effective smartphone-based framework for early detection of diabetes using machine learning," in *Health Informatics Journal*, vol. 28, no. 2, pp. 1462–1477, 2022.

[30] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," in *Science*, vol. 366, no. 6464, pp. 447–453, 2019.