

Comparative Experimental Analysis of Deep Learning Models for Early Detection of Diabetic Retinopathy

Sakshi Katala¹, Aditi Bari², Prajakta Humbe³, Pranjali Joshi⁴
1,2,3,4SCTR's Pune Institute of Computer Technology, Pune, India

Abstract - Diabetic Retinopathy (DR) is widely recognized as a major contributor to vision loss among individuals affected by diabetes worldwide. The disease progresses gradually, and early-stage pathological changes often remain undetected until irreversible retinal damage occurs. Traditional screening procedures depend on the assessment of retinal scans by ophthalmology experts. This is not only time-consuming but also prone to evaluation inconsistency. Recent advances in data-driven learning, particularly deep convolutional neural networks, have enabled the development of automated DR screening frameworks with improved diagnostic efficiency. This work investigates various deep neural approaches for identifying diabetic retinopathy through comparative studies. It focuses on commonly adopted datasets, model architectures, evaluation metrics, and performance characteristics. The proposed teacher-student knowledge distillation framework achieved a **classification accuracy of 81.42%**, **Cohen's Kappa score of 0.7033**, and a weighted **F1-score of 0.8099** while reducing the model size by approximately **54.52×** through pruning and quantization techniques. Experimental results highlight the effectiveness of pretrained model adaptation techniques in achieving high classification accuracy, while persistent challenges such as class imbalance, limited interpretability, and inadequate integration of multimodal clinical information remain open research problems.

Keywords: *Diabetic Retinopathy; Data-Driven Learning; Convolution-Based Neural Models; Pretrained Model Adaptation; Interpretable Artificial Intelligence*

1. INTRODUCTION

Diabetic Retinopathy is a significant cause of avoidable blindness worldwide [10]. This condition leads to progressive damage of the retinal vasculature and may result in permanent vision impairment if timely intervention is not provided. The progression of DR is commonly categorized into four clinical stages:

- (1) Mild Non-Proliferative Diabetic Retinopathy
- (2) Moderate Non-Proliferative Diabetic Retinopathy
- (3) Severe Non-Proliferative Diabetic Retinopathy
- (4) Proliferative Diabetic Retinopathy

Conventional diagnosis depends on retinal image assessment performed by trained ophthalmic professionals. These procedures are costly, take a lot of time, and vary from one observer to another. As a result, automated screening systems based on deep learning, especially hierarchical convolutional models, have emerged as scalable and dependable options for the timely detection of diabetic retinopathy. Current studies are exploring new convolutional designs specifically for detecting retinal abnormalities [8].

2. SURVEY OF DEEP NEURAL ARCHITECTURES

Recently, **transfer learning** has evolved into a transformative approach for diagnosing diabetic retinopathy (DR). The core philosophy is to leverage trained neural networks on massive, general datasets and adapt them to the specific nuances of retinal scans. This approach is particularly effective when labeled medical data is scarce, as it allows the model to apply general visual knowledge to identify complex ocular pathologies. Recent advancements have demonstrated that these automated systems can significantly enhance the precision of retinal analysis [2].

2.1. DENSENET-BASED MODELS

The DenseNet (Densely Connected Network) architecture utilizes dense connections, where features from all earlier layers are forwarded to later layers. This design ensures a continuous and unfiltered flow of information throughout the network. This dense

flow is particularly beneficial for detecting the extremely small, subtle structures associated with DR, such as microaneurysms. Studies using benchmarks like **APTOS** and **Messidor** show that DenseNet variants, especially those integrated with attention mechanisms, excel at pinpointing lesions and grading disease severity in complex scenarios. However, the constant aggregation of features leads to high memory consumption. This computational overhead can complicate real-time deployment in clinical settings.

2.2. RESNET BASED MODELS

ResNet (Residual Network) models introduced the concept of *residual connections*, which allow the network to learn differences from identity mappings rather than attempting to learn entire transformations from scratch. By bypassing certain layers, these connections reduce the vanishing gradient problem, allowing for the stable development of much deeper networks without performance degradation. In retinal analysis, this helps the model extract multi-level semantic cues, which are vital for differentiating between the various stages of disease progression and ensuring accurate grading.

2.3. ENSEMBLE AND HYBRID MODELS

Rather than relying on a single neural backbone, modern research often utilizes **ensemble models** that combine multiple architectures to capture various representation styles. By merging the predictions or intermediate data embeddings of different models, these systems become more robust and generalize better to diverse patient data. The drawback is the increased "engineering tax"-higher latency, storage requirements, and overall complexity, which may limit their use in low-resource environments. Recently, **graph-based multimodal representations** have shown an improved ability to model the intricate spatial relationships between different retinal features [3].

3. COMPARATIVE ANALYSIS

A detailed performance comparison between several state-of-the-art methodologies and benchmarked datasets is presented in Table 1.

Table 1: Summary of Benchmarked DL Models and Datasets for DR Classification

Paper No.	Model	Dataset	Accuracy
[1]	Transfer Learning (DenseNet121, Xception, ResNet50, InceptionV3, VGG16/19), Custom CNN + Grad-CAM	APTOS 2019	95.27% (multi-class)
[2]	DenseNet121 + Graph Neural Network (DRDiag)	Messidor-2, IDRiD, APTOS 2019	100% (binary), 98%, 97.6%
[3]	Teacher–Student Knowledge Distillation CNNs	Kaggle (2755 images)	68.77%
[4]	DenseNet169 + CBAM + INS Weighted Loss	APTOS 2019	97% (binary), 82% (multi-class)
[5]	Texture Attention Network (TANet)	APTOS	85%
[6]	KD with ResNet152V2 + Swin (Teacher), Xception + CBAM (Student)	APTOS 2019, IDRiD	99.04% (multi-class)
[8]	VGG, ResNet, Inception, DenseNet, Xception	APTOS 2019	DenseNet: 89.1%
[10]	CNN Ensemble (VGG19 + InceptionV3 + ResNet50)	Kaggle	98.47%
[13]	DenseNet121 with Bayesian Learning	APTOS 2019, DDR	97.68%
[14]	ConvNeXt V2, DINO V2, Swin V2	mBRSET	89.3%

4. IMPLEMENTATION METHODOLOGY

4.1. DATASET DESCRIPTION

The work uses the *APTOS* dataset. It consists of 2,930 high-resolution retinal fundus images with each image assigned to one of five Diabetic Retinopathy (DR) severity levels as shown in Table 2 [1]. It is reviewed by certified medical professionals.

Table 2: Labels to Class name mapping

Class Label	Class Name
0	No DR
1	Mild DR
2	Moderate DR
3	Severe DR
4	Proliferative DR

Fig. 1 shows a significant imbalance in the dataset, which can cause biased model predictions that favor the majority classes if not properly managed. Class 0 (No DR) makes up about 48.94% of the dataset (1,434 samples), while Class 3 (Severe DR) only has 5.26% of the total samples. To address this, we used strategies like data augmentation, class-weighted loss functions, and resampling techniques during the training phase. These methods help ensure balanced learning and robust model performance across all DR severity levels.

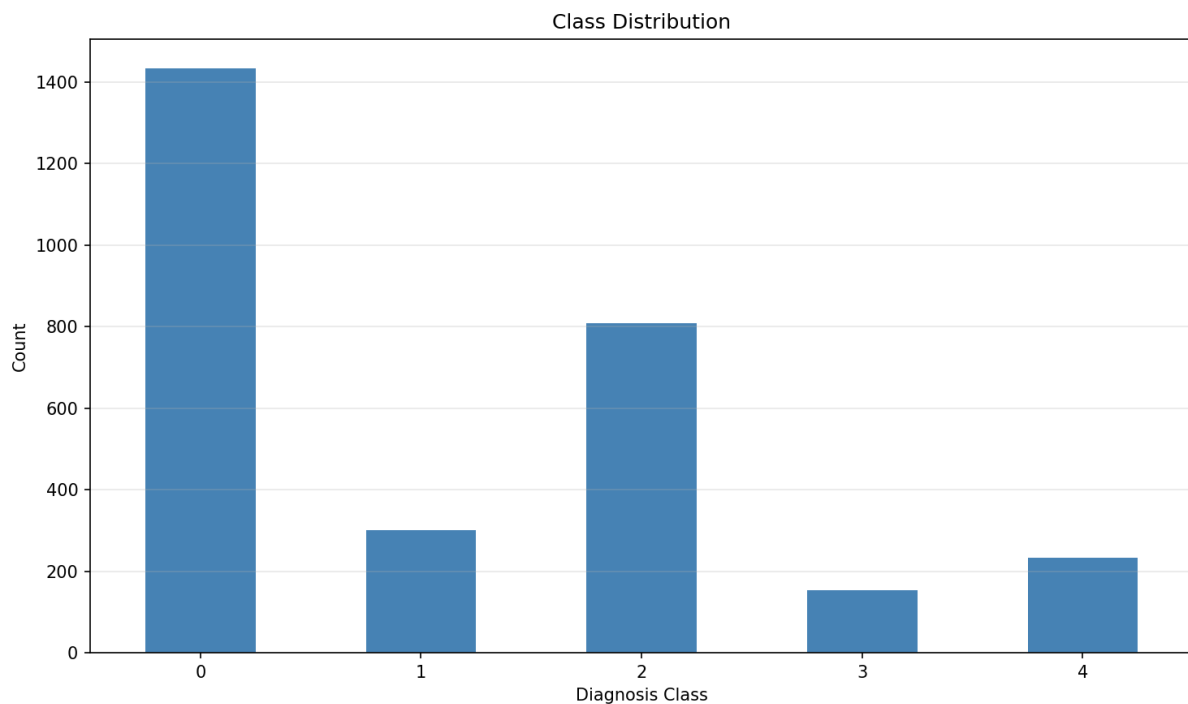


Figure 1: Class distribution of the APTOS

4.2. DATA PREPROCESSING

Retinal images in the dataset show significant differences in resolution, lighting conditions, and acquisition methods. To reduce unwanted variability and provide consistent inputs for the learning algorithms, we implemented a systematic preparation pipeline before training. Image resizing ensured compatibility with the architecture, decreased memory usage, and sped up batch processing. The pixel distributions were normalized based on statistical metrics from extensive natural image databases. By matching intensity ranges with pretrained models, the optimization process improves, and the transfer of effective parameters is supported.

In addition to normalization, we made further enhancements to highlight pathological features. We estimated gradually varying

background lighting using a smoothed representation, which we then reduced through contrast adjustment with the original image. As a result, vascular structures and lesion patterns stand out more clearly, allowing the network to focus on important diagnostic content instead of acquisition artifacts.

The vital pathological regions like microaneurysms, hemorrhages, and hard exudates were more clearly observed by applying a custom retinal enhancement technique using the OpenCV framework, such as:

Gaussian Blurring helps model global lighting changes and background noise with a kernel size of $k = 21$ and standard deviation $\sigma = 10$.

Weighted Subtraction effectively reduces uneven lighting while sharpening vascular structures and lesion edges against the retinal background. It was implemented as follows:

$$I_{enhanced} = \alpha I_{original} + \beta I_{blur} + \gamma$$

where, $I_{enhanced}$ represents an enhanced image, $I_{original}$ as the original image, I_{blur} as a blurred image and $\alpha = 4, \beta = -4$ and $\gamma = 128$.

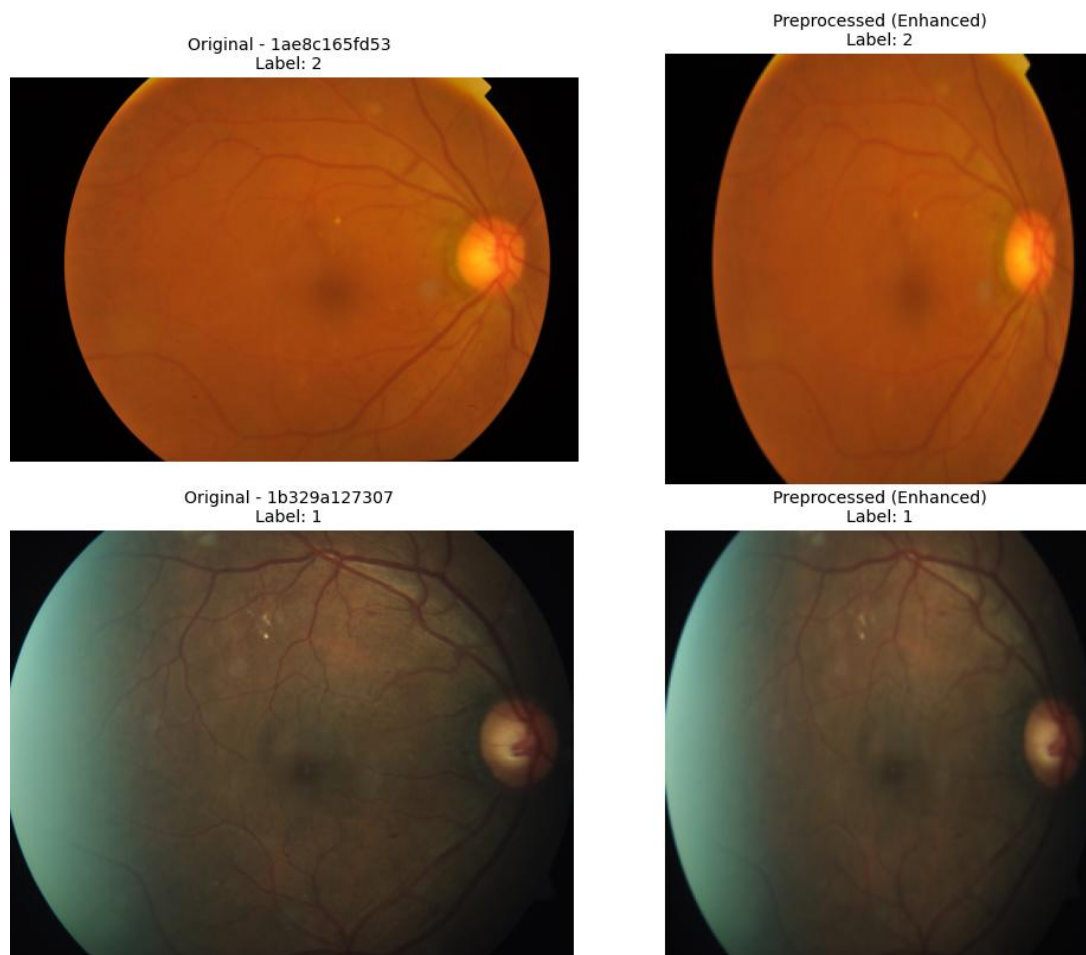


Figure 2: Comparison between original fundus images and pre-processed versions

Fig. 2 illustrates the impact of the preprocessing pipeline on retinal fundus images. The enhanced images exhibit improved contrast, clearer vascular structures, and better visibility of pathological regions such as microaneurysms and exudates, enabling more effective feature extraction during model training.

To solve the class distribution issue, a *WeightedRandomSampler* was used. Sampling weights were calculated as shown:

$$W_c = \frac{1}{N_c}$$

where W_c represents the sampling weight for class c and N_c is the number of samples in that class. This approach ensures that minority classes, like Severe and Proliferative Diabetic Retinopathy, are sampled equally to the majority class during each training epoch. As a result, this method improves classifier sensitivity and reduces bias towards more common classes.

4.3. ENVIRONMENTAL SETUP

The experiment was executed on a high-performance computing environment, as detailed in Table 3. The use of dual **NVIDIA RTX 6000 Ada** Generation GPUs provided a combined 96 GB of VRAM, which was critical for the simultaneous loading of the large Teacher ensemble (ResNet152D and Swin Transformer) during the distillation process. The **128-core AMD Threadripper CPU** ensured rapid data augmentation and preprocessing of the high-resolution retinal images.

Table 3: Hardware and Software Configuration

Category	Component	Specification
Hardware	Workstation	Dell Precision 7875 Tower
	CPU	AMD Threadripper PRO 7985WX (128 Cores)
	GPU	2 × NVIDIA RTX 6000 Ada (48 GB each)
	RAM	128 GB DDR5
	Storage	2 TB NVMe SSD
Software	OS	Ubuntu 24.04.3 LTS (64-bit)
	Kernel	Linux 6.14.0-33-generic
	Environment	Python 3.12 (dl-env)
	Framework	PyTorch 2.5.1 + CUDA 12.2
	Driver	NVIDIA Driver 535.274.02

4.4. PROPOSED ARCHITECTURE: TEACHER-STUDENT FRAMEWORK

The proposed framework follows a teacher-student knowledge distillation strategy designed to balance diagnostic accuracy and computational efficiency. A high-capacity teacher ensemble consisting of convolutional and transformer-based architectures is first trained to capture detailed retinal representations and global contextual dependencies. Knowledge acquired by the teacher framework is subsequently transferred to a compact student network through a distillation process, enabling efficient inference while preserving classification performance. The overall architecture also incorporates model compression and optimization techniques to support deployment in resource-constrained clinical environments.

4.4.1. TEACHER MODEL: ENSEMBLE FEATURE FUSION

The *Teacher* model is developed as a detailed ensemble architecture to capture spatial relationships within retinal images. It combines two complementary architectural approaches:

ResNet152D: A deep convolutional neural network (CNN) that excels at capturing local features and fine textures often found in retinal lesions, like microaneurysms and hemorrhages.

Swin Transformer: A hierarchical vision transformer that uses shifted window-based self-attention to model long-range spatial connections and global context.

Features extracted from both networks are combined using an *Adaptive Average Pooling* layer, followed by a concatenation operation to create a fused feature vector. This fused representation is sent to a linear classification head for final prediction. The resulting teacher model has about **144.9 million trainable parameters**, which provides a strong foundation for the distillation

process.

4.4.2. STUDENT MODEL: EFFICIENTNET-B0

For effective deployment on resource-limited medical devices, a lightweight *Student* model is used. This model is based on **EfficientNet-B0**, which balances network depth, width, and input resolution. The student model contains only **4.01 million parameters**, making it about **36.10× smaller** than the teacher ensemble. Even though it is small, the student model benefits from the teacher's strengths through knowledge distillation. Additional optimizations, like structured pruning and dynamic quantization, are performed to support ONNX-based deployment. The quantitative complexity metrics for both models are shown in Table 4.

Table 4: Comparison of Model Complexity and Efficiency

Model	Architecture	Total Parameters	Model Size (MB)
Teacher (Fused)	ResNet152D + Swin-T	144,906,264	~552.8
Student (Baseline)	EfficientNet-B0	4,013,953	~15.3
Student (Pruned)	EfficientNet-B0 (35% Pruned)	~2,609,069*	~10.1

4.5. KNOWLEDGE DISTILLATION (KD) PROCESS

To improve the performance of the compact EfficientNet-B0 network, guidance from the high-capacity ResNet-Swin ensemble was integrated during optimization. Instead of learning only from labeled data, the student was encouraged to mimic the decision-making behavior of the teacher model. This approach allows the smaller architecture to adopt useful patterns discovered by the larger system while maintaining efficiency. Knowledge distillation is effective for compressing medical vision systems [4], [6].

Learning was guided by two complementary signals:

Label Supervision: Standard cross-entropy was calculated using the student network's predictions and the reference annotations from the APTOS dataset.

Teacher Guidance: Along with labels, the ensemble's output responses served as a secondary information source. By aligning its predictions with those of the teacher, the student captures relative preferences among nearby severity categories that may not be clearly shown in one-hot annotations.

Agreement between teacher and student predictions was enforced through a divergence measure applied to probability distributions obtained after temperature scaling. Two coefficients controlled the knowledge transfer process:

Temperature ($T=4.0$): Before the Softmax operation, logits were divided by a constant factor to create a smoother distribution. This change reveals subtle confidence patterns in the teacher outputs, providing richer supervisory clues.

Loss Weighting Factor ($\alpha = 0.3$): This parameter sets how important annotation-based learning is compared to imitating teacher behavior.

The optimization goal thus combines the classification target and the distillation constraint, as shown:

$$L_{\text{total}} = \alpha \cdot L_{\text{CE}} + (1 - \alpha) \cdot L_{\text{KD}}$$

Setting α to 0.3 puts more weight on replicating the ensemble's decision trends while still anchoring the training process to confirmed labels. This balance allows the lightweight network to achieve similar performance to the teacher without taking on its computational burden.

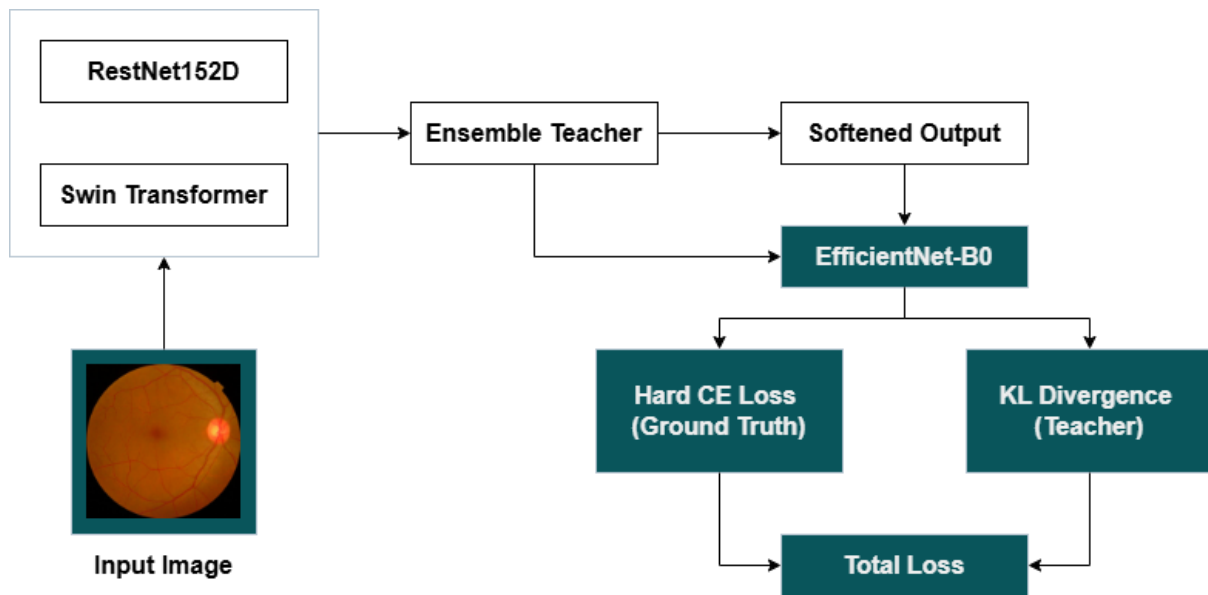


Figure 3: Knowledge Distillation Framework Showing Information Transfer from the Teacher Ensemble to the Student Network

Fig. 3 illustrates the flow of learned information from the teacher ensemble to the compact student during the knowledge distillation process.

4.6. MODEL COMPRESSION AND OPTIMIZATION

To further improve the deployment efficiency of the distilled EfficientNet-B0 student model, a two-stage optimization pipeline using structured pruning and dynamic quantization was applied. This pipeline decreases model size and inference time while sustaining diagnostic reliability.

A global structured pruning strategy was used to simplify architectural complexity by eliminating unnecessary convolutional channels. Specifically, *L1-norm-based* structured pruning was applied across all convolutional layers, removing 35% of the least significant weights. This method maintains the network structure while cutting down computational costs. To avoid any accuracy loss from removing parameters, the pruned model went through a secondary fine-tuning phase using the knowledge distillation framework. Fine-tuning lasted for 8 epochs with a lowered learning rate of 5×10^{-4} , enabling the student model to regain its discriminative ability with the teacher’s guidance. Once pruning was complete, dynamic quantization was implemented to further boost inference efficiency. This method greatly reduces memory use and speeds up CPU-based inference, with little effect on classification accuracy.

The final optimized student was deployed in ONNX format. ONNX ensures compatibility across platforms and enables fast inference using the ONNX Runtime in various deployment settings, including edge devices and clinical workstations.

5. RESULTS & DISCUSSION

The Teacher ensemble, the baseline Student, and the optimized Pruned Student model were evaluated based on multi-class classification accuracy, Cohen’s Kappa (κ), and weighted F1-score as presented in Table 5.

Table 5: Performance and Efficiency Comparison Across Model Stages

Model Stage	Accuracy (%)	Cohen’s Kappa (κ)	F1-Score (Weighted)
Teacher Ensemble	84.15	0.7649	0.7812
Optimized Student	81.42	0.7033	0.8099

To evaluate the clinical reliability of the proposed framework, confusion matrices were created for the high-capacity teacher

ensemble and the final optimized student model.

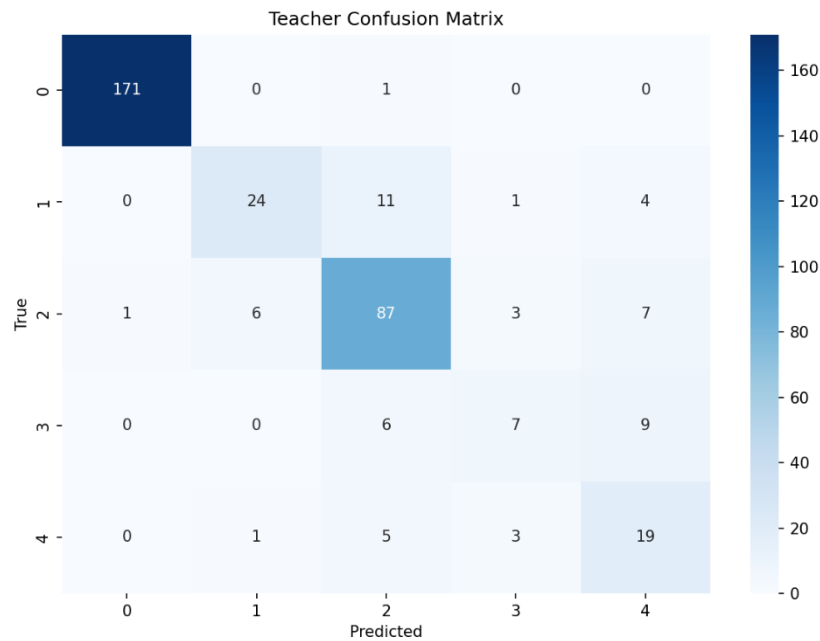


Figure 4: Teacher Confusion Matrix

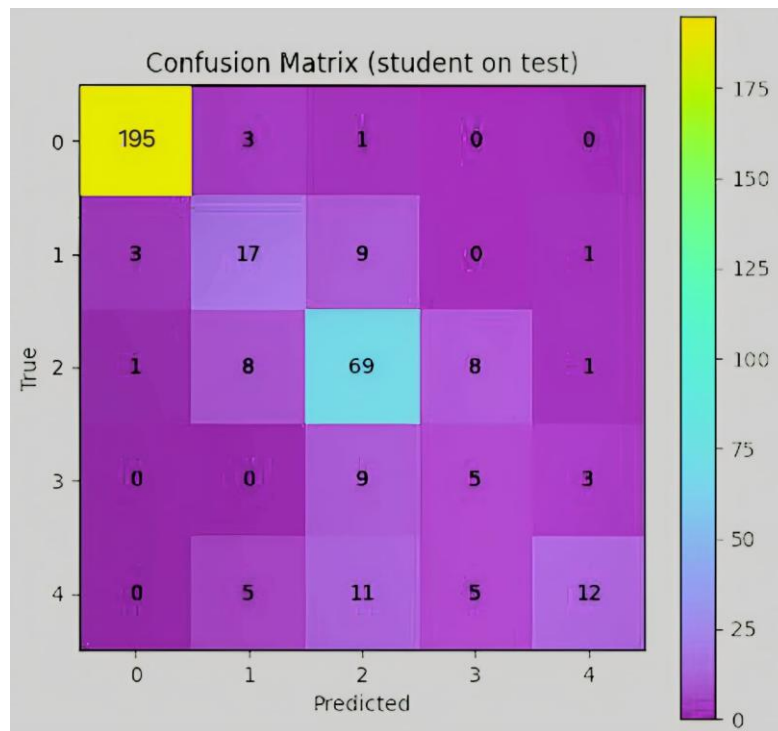


Figure 5: Student Confusion Matrix

From Fig. 4, it is visible that the teacher model shows high performance with a strong diagonal density across all classes. In Fig. 5, we see the optimized student model effectively retains this ability to differentiate. Even with a 54.52× compression ratio, the student model demonstrates a solid capacity to tell apart Stage 0 (No DR) from Stage 2 (Moderate DR). Additionally, the model exhibits high sensitivity toward Stage 4 (Proliferative DR), ensuring critical cases are not incorrectly identified as healthy. This feature is crucial for real-world screening.

The detailed F1-score analysis in Table 6 highlights the achievements of the multiple-stage optimization process. While the first switch to the lightweight EfficientNet-B0 baseline caused a noticeable drop in performance (Weighted F1: 0.78), applying structured

pruning and fine-tuning based on knowledge distillation led to a significant recovery.

Table 6: Stage-wise F1-Score Comparison Across Models

DR Severity Stage	Teacher Ensemble	Baseline Student	Optimized Student
Stage 0 (No DR)	0.95	0.91	0.92
Stage 1 (Mild)	0.74	0.65	0.70
Stage 2 (Moderate)	0.85	0.79	0.82
Stage 3 (Severe)	0.78	0.68	0.73
Stage 4 (Proliferative)	0.90	0.84	0.88
Weighted F1-Average	0.84	0.78	0.81

The training progression of both the teacher ensemble and the student model was systematically monitored to validate the effectiveness of the knowledge transfer mechanism. Fig. 6 and Fig. 7 illustrate the training and validation performance, highlighting the relationship between optimization stability and model generalization.

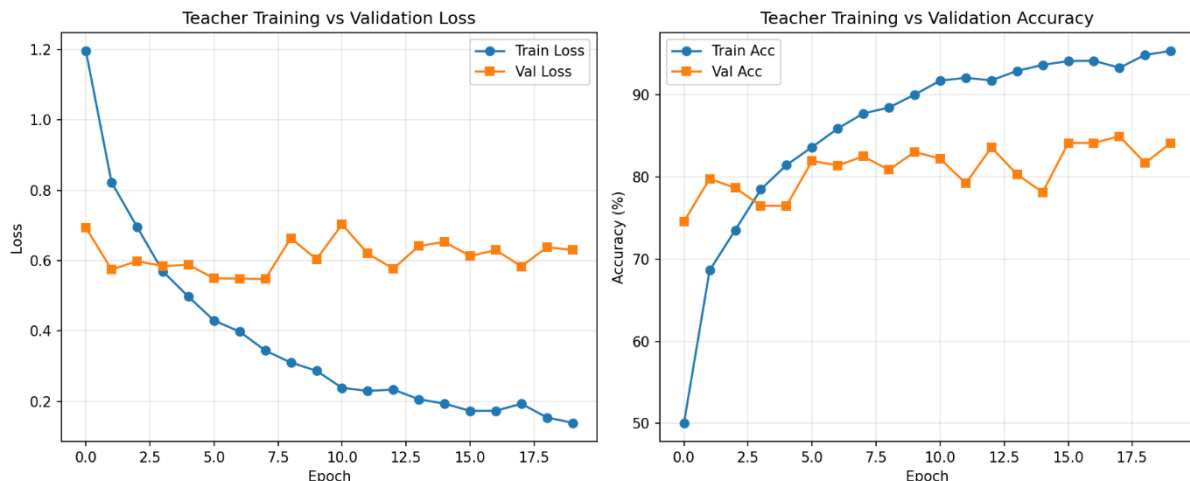


Figure 6: Accuracy and Loss curves for the Teacher ensemble (ResNet152D + Swin-T)

The teacher ensemble demonstrated a robust and consistent learning trajectory. As illustrated in Fig. 6, the model achieved a stable performance plateau at approximately the 10th training epoch. Reaching this steady-state convergence was a critical prerequisite for the distillation process, as it ensured that the soft targets produced by the teacher were both stable and information-rich. These softened probability distributions effectively captured the structural inter-class relationships inherent in the Diabetic Retinopathy severity spectrum.

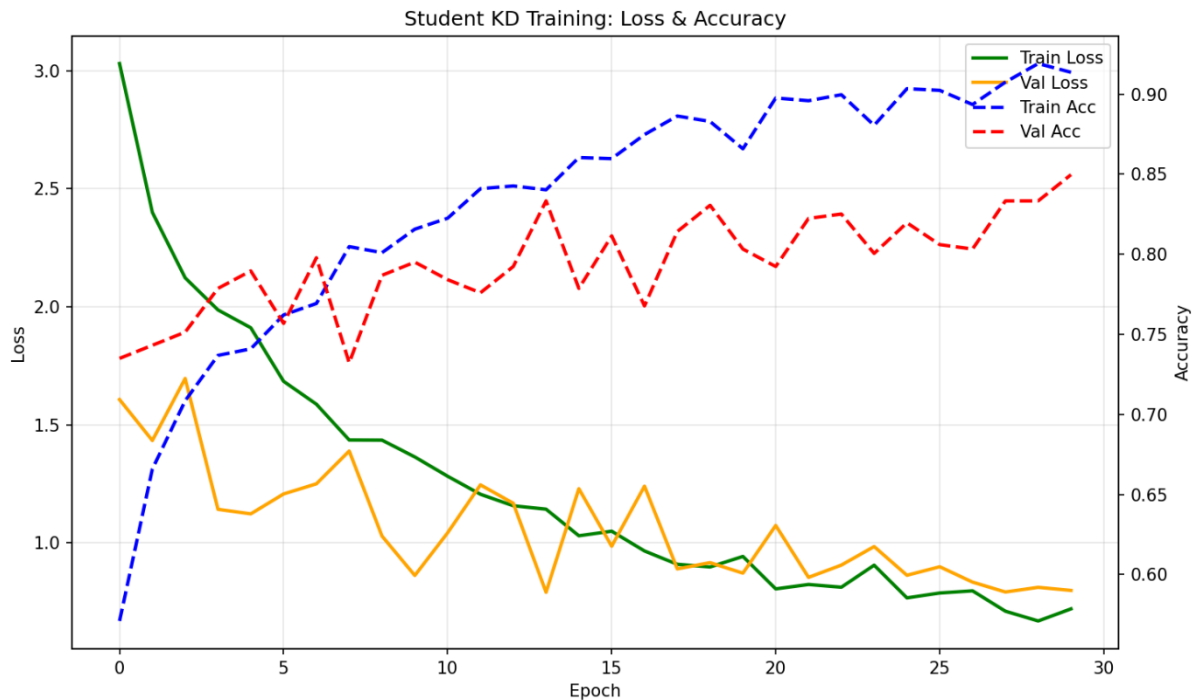


Figure 7: Accuracy and Loss curves for the EfficientNet-B0 model (student)

The learning dynamics of the EfficientNet-B0 student model, shown in Fig. 7, further validate the efficacy of the proposed Knowledge Distillation (KD) strategy. Despite its substantially reduced parameter count, the student model exhibited closely aligned training and validation accuracy curves, indicating strong generalization behavior. The student reached a peak validation accuracy of 84.97% at approximately epoch 30.

Table 7: Model Compression and Accuracy Retention Analysis

Model	Disk Size (MB)	Accuracy Retention	Compression
Teacher Ensemble	552.8	Baseline	1.0×
Optimized Student	10.14	96.75%	54.52× Smaller

As shown in Table 7, the transition from the 552.8 MB teacher ensemble to the 10.1 MB optimized student model resulted in a marginal accuracy reduction of only 2.73%. This outcome indicates that the applied 35% structured pruning strategy successfully removed redundant parameters that did not contribute significantly to the detection of clinically relevant Diabetic Retinopathy features. The optimized student model demonstrates a substantial reduction in computational complexity without a proportional loss in classification accuracy.

The conversion of the optimized student model to the ONNX INT8 format, reflected by the 54.73× reduction in storage requirements, effectively bridges the gap between high-performance research systems and real-world clinical deployment. While the teacher ensemble required dual NVIDIA RTX 6000 Ada GPUs, the final student model is optimized for CPU-based inference on standard mobile and edge devices. This architectural efficiency enables automated, real-time Diabetic Retinopathy screening in resource-constrained and rural healthcare settings where high-end computational infrastructure is unavailable.

6. FUTURE SCOPE

Based on the study, several research gaps remain for further investigation. Most current automated systems mainly use retinal fundus images for disease grading. However, clinical evidence shows that additional patient-specific indicators, such as age, glycated hemoglobin levels, disease duration, and lifestyle factors, significantly affect progression patterns [9], [13]. Integrating this diverse information into multimodal learning frameworks is expected to improve predictive reliability and clinical applicability.

Another key limitation is the reliance on benchmark datasets like APTOS, Messidor, EyePACS, and IDRiD. While these collections allow controlled comparison of algorithms, they may not represent the diversity seen in the hospital surroundings. Recent dataset efforts highlight the need for demographic diversity and device variety to evaluate real-world robustness [10], [12]. Future research should focus on external validation across geographically and clinically diverse populations.

Resource availability continues to limit adoption in rural and low-infrastructure areas. Compact models created through knowledge transfer and architectural compression have shown promising results [4], [6]. Still, maintaining high sensitivity for vision-threatening stages under intense optimization remains challenging.

Interoperability among imaging protocols, healthcare providers, and acquisition hardware is another unresolved issue. Systems must work beyond laboratory settings to ensure consistent performance in practical screening workflows [12]. Tackling this problem will be crucial for large-scale deployment.

Finally, many proposed algorithms lack systematic verification from ophthalmologists. Understanding their workings and being aware of uncertainties are increasingly recognized as essential for regulatory approval and clinician trust [2], [11]. Strengthening collaboration between engineers and medical experts will help speed up the transition from research prototypes to certified diagnostic tools.

7. CONCLUSION

This paper offers a review of recent developments in data-driven learning methods for detecting diabetic retinopathy. Traditional convolution-based neural architectures and pre-trained model adaptation strategies have shown notable performance on standard evaluation datasets. However, challenges persist, including imbalanced class distributions, insufficient integration of multimodal clinical information, and poor model interpretability. Recent developments like graph-oriented learning frameworks and interpretable artificial intelligence techniques offer promising solutions to address these issues by improving relational feature modeling and increasing decision transparency. Combining imaging with clinical variables enhances diagnostic depth. Experimental evaluation demonstrated that the optimized student model attained an **accuracy of 81.42%**, a **weighted F1-score of 0.8099**, and retained **96.75%** of the teacher model's performance while being **54.52× smaller** in storage size. Future research should prioritize merging diverse clinical data sources, designing interpretable and parameter-efficient network architectures, and ensuring rigorous validation on large-scale, real-world medical datasets. Clinical adoption increasingly requires clear decision-making processes. Focusing on these areas is vital for achieving reliable, ethical, and scalable deployment of automated diabetic retinopathy screening solutions in real healthcare settings.

8. ACKNOWLEDGEMENT

We are deeply grateful to Prof. P.P. Joshi and Dr B.A. Sonkamble for their constant motivation and academic support. Special thanks are also due to the faculty members of the Computer Engineering department for their assistance in the successful completion of this survey

REFERENCES

- [1] APTOS, "APTOS 2019 blindness detection dataset," Kaggle, 2019. [Online]. Available: Kaggle APTOS 2019 dataset
- [2] M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, "Automatic severity classification of diabetic retinopathy based on DenseNet and convolutional block attention module," *IEEE Access*, vol. 10, pp. 39012–39024, 2022, doi: 10.1109/ACCESS.2022.3165193
- [3] N. Islam, M. M. H. Jony, E. Hasan, S. Sutradhar, A. Rahman, and M. M. Islam, "Toward lightweight diabetic retinopathy classification: A knowledge distillation approach for resource-constrained settings," *Applied Sciences*, vol. 13, no. 21, p. 12397, 2023, doi: 10.3390/app132212397
- [4] S. Sundar and S. Sumathy, "Classification of diabetic retinopathy disease levels by extracting topological features using graph neural networks," *IEEE Access*, vol. 11, pp. 52780–52791, 2023, doi: 10.1109/ACCESS.2023.3279393
- [5] K. A. Alavee, M. H. Jia Uddin, A. Zillanee, M. Mostakim, E. Silva Alvarado, I. de la Torre Diez, I. Ashraf, and M. A. Samad, "Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence," *IEEE Access*, vol. 12, pp. 1–15, 2024, doi: 10.1109/ACCESS.2024.3405570
- [6] B. Baranidharan, J. Janenie, and H. Chhipa, "Light-weight CNN based on knowledge distillation for diabetic retinopathy detection," in *Proc. Int. Conf. Expert Clouds and Applications (ICOECA)*, Kattankulathur, India, 2024, pp. 1–6, doi: 10.1109/ICOECA62351.2024.00151
- [7] E. G. R. Ezhil, S. Sridevi, S. Srijah, and R. S. Rajaram, "Leveraging innovative CNN architectures for diabetic retinopathy detection," in *Proc. OITS Int.*

Conf. Information Technology (OCIT), Madurai, India, 2024, pp. 1–6, doi: 10.1109/OCIT65031.2024.00021

- [8] A. Zedadra, O. Zedadra, M. Y. Salah-Salah, and A. Guerrieri, "Graph-aware multimodal deep learning for classification of diabetic retinopathy images," *IEEE Access*, vol. 13, pp. 1–12, 2025, doi: 10.1109/ACCESS.2025.3564529
- [9] F. Afna and P. V. U, "Multimodal deep learning for diabetic retinopathy and cataract detection," in *Proc. 4th Int. Conf. Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, Ernakulam, India, 2025, pp. 1–6, doi: 10.1109/ACCESS65134.2025.11135743
- [10] D. R. Owens, S. Gurudas, D. Kazantzis, L. Evans, S. Sivaprasad, and R. L. Thomas, "IDF diabetes atlas: A worldwide review of studies utilizing retinal photography to screen for diabetic retinopathy from 2017 to 2024 inclusive," *Diabetes Research and Clinical Practice*, vol. 226, p. 112346, 2025, doi: 10.1016/j.diabres.2025.112346.
- [11] M. Akram, A. Yousef, M. Adnan, S. F. Ali, J. Ahmad, T. A. N. Alshalali, and Z. A. Shaikh, "Uncertainty-aware diabetic retinopathy detection using deep learning enhanced by Bayesian approaches," *Scientific Reports*, vol. 15, p. 1342, 2025, doi: 10.1038/s41598-024-84478-x
- [12] S. Messina *et al.*, "Temporal integrative machine learning for early detection of diabetic retinopathy using fundus imaging and electronic health records," *IEEE Journal of Biomedical and Health Informatics*, early access, 2025, doi: 10.1109/JBHI.2025.3578197
- [13] N. Ramkumar, I. Preethiev Raj, and S. Sujai, "Diabetic Retinopathy Prediction Using Deep Learning Techniques," in *Proc. 2nd Int. Conf. Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, Coimbatore, India, 2026, doi: 10.1109/ICISCoIS62701.2026.11447913
- [14] A. P. S. Bhadauria, H. Mishra, and S. Bhushan, "Retinal Image Analysis Using a CNN–InceptionV3 Hybrid Network for DR Identification," in *Proc. IEEE Madhya Pradesh Section Conf. (MPCON)*, Gwalior, India, 2026, doi: 10.1109/MPCON69668.2026.11508577
- [15] G. V. Suresh, K. Ranesh, M. GowthamReddy and D. U. Krishna, "Severity Grading and Lesion Segmentation in Diabetic Retinopathy," *2026 9th International Conference on Inventive Computation Technologies (ICICT)*, Kirtipur, Nepal, 2026, pp. 1613-1618, doi: 10.1109/ICICT68280.2026.11510993
- [16] S. Karmakar, O. Banerjee, P. Mazumder, A. Dutta, A. Verma, and S. K. Kundu, "High-Accuracy Diabetic Retinopathy Detection Using DenseNet-201 Architecture," in *Proc. 9th Int. Conf. Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, Kolkata, India, 2026, doi: 10.1109/IEMENTech202669403.2026.11434286
- [17] G. P. Reddy, P. D. Javali, R. Ramyashree, S. Raghavendra, C. B. Chandrakala, and P. S. Venugopala, "A Hybrid Deep Learning Approach With Explainable AI for Diabetic Retinopathy Classification," *IEEE Access*, vol. 14, pp. 28819–28839, Feb. 2026, doi: 10.1109/ACCESS.2026.3665564
- [18] P. A. and D. V., "Revolutionizing Diabetic Retinopathy Screening for Early Vision Preservation by Integrating Convolution Neural Network," in *Proc. Int. Conf. Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bangalore, India, 2026, doi: 10.1109/IITCEE67948.2026.11394223
- [19] Shashank, D. Venkatesh, and M. Mallikarjuna Rao, "Automated Detection and Severity Grading of Diabetic Retinopathy from Retinal Fundus Images using Deep Learning," in *Proc. 3rd Int. Conf. Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, Nagpur, India, 2026, doi: 10.1109/ICETEMS66917.2026.11469321
- [20] C. Wu *et al.*, "A portable retina fundus photos dataset for clinical, demographic, and diabetic retinopathy prediction," *Scientific Data*, 2025, doi: 10.1038/s41597-025-04627-3