Retinal Disease Diagnosis Using an Optimization-Driven Deep Learning Framework

Anand Kumar[1*], Apurva Dalela[2], and Manju Khari[3]

[1]School of Computer & Systems Sciences,
Jawaharlal Nehru University, New Delhi, India
anandkumarjnu9@gmail.com*

[2]School of Computer & Systems Sciences,
Jawaharlal Nehru University, New Delhi, India
apurvadalela@gmail.com

[3]School of Computer & Systems Sciences,
Jawaharlal Nehru University, New Delhi, India
manjukhari@yahoo.co.in

Abstract. Diabetic retinopathy and macular edema rank among the leading global causes of blindness that could otherwise be prevented through timely intervention. Because diagnosis still relies largely on manual inspection by trained specialists, the process is often slow and inconsistent, which raises the risk of irreversible vision loss. This work introduces an optimization-driven deep learning pipeline that delivers higher classification accuracy while converging faster and remaining stable and robust throughout training. The pipeline combines established Convolutional Neural Network (CNN) backbones, namely VGG16, InceptionV3 and DenseNet121, with a newly designed optimizer termed the NIRMAL Optimizer. Input fundus images are first resized and pixel-normalized, then enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) and expanded through data augmentation operations such as rotation, flipping and zooming, before transfer learning is applied for feature extraction. The NIRMAL optimizer specifically targets the class imbalance and overfitting problems that commonly hinder retinal image classifiers. Using a class-balanced subset of the Kaggle Diabetic Retinopathy dataset (five DR severity grades, 856 samples each), DenseNet121 trained with NIRMAL reached 92% accuracy, with a precision of 0.91 and recall of 0.94 for the macular edema risk class; InceptionV3 and VGG16 attained 88% and 76% accuracy respectively under the same optimizer. Across 50 training epochs, NIRMAL consistently outperformed conventional optimizers including Adam, SGD and RMSprop. These findings indicate that the proposed framework offers a practical, scalable route to automated retinal disease classification, building on the strong baselines already established by CNN-based classifiers while pointing toward further gains from refined optimization strategies.

Keywords: Deep Learning, Retinal Disease, CNN, Optimization Techniques, Retinal Image Classification, NIRMAL Optimizer, Transfer Learning.

1 Introduction

Diabetic Retinopathy (DR) and macular edema are considered among major factors contributing to vision loss worldwide [1]. Various diseases of the retina have

© The Author(s) 2026
A. Shastri et al. (eds.), *Proceedings of the Fourth Workshop on Applications of Computational Intelligence and Interactive Systems (WACIIS 2026)*, Advances in Intelligent Systems Research 217,
https://doi.org/10.2991/978-94-6239-832-0_8

increased in the last few years as a result of diabetes and ageing populations. While early detection and diagnosis is important to preventing permanent vision loss, the current approach to diagnosis remains manual inspection of fundus images. The current method of diagnosis of retinal diseases is time consuming and suffers from observer related error.

Convolutional Neural Networks (CNNs), a type of deep learning, are commonly used in analyzing medical images. CNN-based models have been developed to extract deep hierarchical features of retinal images and obtain better classification performance from these images. Retinal pattern recognition (RPR) has shown good diagnostic performance with the use of VGG16, InceptionV3 [2] and DenseNet121 [3]. CNN-based methods have been shown to yield detection performance similar to that of an expert ophthalmologist for DR detection [1] and Pratt et al. [4] set an early benchmark with CNN-based DR grading on the EyePACS dataset.

The main drawbacks to deep learning diagnosis for the retina are slow convergence, overfitting, sensitive to noise, and lack of generalization [5]. To tackle these problems, hybridization of optimization has been attempted, for instance, by Subramaniam and Naganathan [5] who integrated a deep CNN into the Red Spider Optimization (RSO) algorithm to accelerate training time and the accuracy of classification. Adaptive optimization strategies are crucial for guaranteeing reliable performance in various medical imaging modalities, as established by Kaur and Kumar [6]. This work proposes an optimization-driven deep learning framework by integrating state-of-the-art CNN architecture with the adaptive NIRMAL optimizer, that mitigates the limitations of conventional optimization approach.

2 Literature Review

Recent progress in deep learning has greatly enhanced the accuracy of medical imaging with eye diseases detection [7]. CNNs automatically learn discriminative features from the retinal fundus images and are suitable for retinal analysis. The detection results of CNN-based systems are comparable to that of expert ophthalmologists in the field of DR detection [1]. The tasks of class imbalance and over fitting continue to be important concerns in medical imaging and CNNs have been applied to DR grading with the EyePACS dataset by Pratt et al. [4] obtaining 75% accuracy.

Different architectures have been used for the classification of retinal images. InceptionV3 [2] uses multi-scale convolutions to extract features efficiently and VGG16 is good at obtaining local spatial features well. For this reason DenseNet121 [3] is designed to connect the model densely to improve model performance, thus increasing the flow of gradients and minimizing vanishing gradient problems. Babaqi et al. [8] showed excellent multi-class eye disease classification results with VGG16 model and ResNet model. InceptionV3 with transfer learning has been demonstrated to be more than 98% accurate for multi-disease classification by Vardhan et al. [9]. Sundar Raj et al. [10] used Hybrid Attention Residual Deep Convolution Learning Network to obtain competitive biomedical image analysis F1-scores. Gupta et al. [11] set the groundwork for deep neural network feature extraction methods and Gupta et al. [7] give a wider overview of health informatics. The literature has shown that deep learning can be effectively used to aid the diagnosis of retinal diseases; however, more refined optimization techniques are needed to enhance stability, robustness, and generalization.

3 Framework

In the present work, an optimization-guided deep learning approach for retinal disease classification is presented, which involves novel CNN designs and adaptive gradient methods. The framework consists of four stages: pre-processing, feature extraction, optimization-based training, and classification. The modular architecture allows the interchange of the CNN backbone network as shown in Fig. 1.

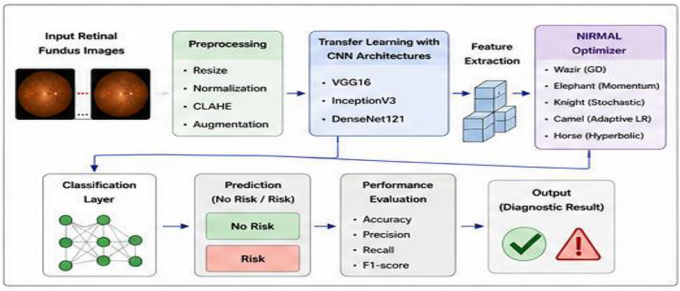

Fig. 1. Proposed Optimization-Driven Deep Learning Framework.

3.1 Preprocessing Pipeline

The retinal fundus images need to be resized to a standard input size, scaled pixel values, enhanced by CLAHE for contrast enhancement, and then augmented for enhancing model generalization, as shown in Fig. 2, during the preprocessing phase. There are 4 steps.

Image Resizing. To achieve spatial uniformity, whilst still keeping the interpolation artefacts to a minimum, fundus images are scaled to fixed resolution for VGG16 and DenseNet121, 224×224 pixels, and for InceptionV3, 299×299 pixels.

Pixel Normalization. To guarantee that the data distribution of the input matches CNN pretraining regime, the network is trained using channel-wise statistics for the training set, trained with ImageNet [11] (mean: [0.485, 0.456, 0.406]; std: [0.229, 0.224, 0.225]).

CLAHE Enhancement. Because the L-channel of the LAB color space is used to preserve local contrast, thus enhancing the visibility of the fine pathological details, while avoiding any significant increase of noise, it is converted using contrast Limited Adaptive Histogram Equalization (CLAHE) and only the L-channel is used.

Data Augmentation. During training, several augmentation operations have been included to enhance the robustness and reduce overfitting of the model, such as random horizontal and vertical flips, rotations, and zooming in and out. These transformations introduce variability of the training data particularly useful if medical image datasets are limited.

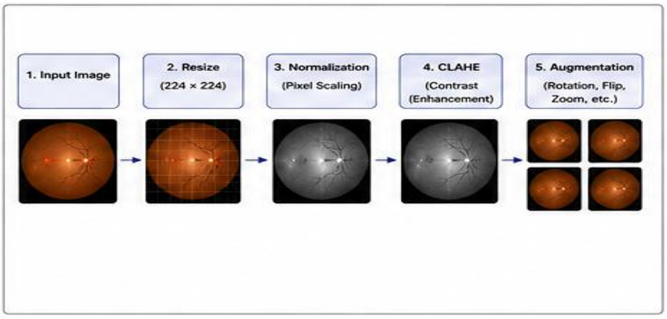

Fig. 2. Retinal Image Preprocessing Pipeline.

3.2 Feature Extraction

To extract the features, three widely-used CNN architectures are employed, namely VGG16, InceptionV3 [2] and DenseNet121 [3] and transfer learning is performed in all the three models. The images are classified by pretrained ImageNet models, which are fine-tuned to the retinal dataset with a replacement of their original classification head with a fully connected layer with the target output classes. The major merits and drawbacks of the various architectures are summarized in Table 1.

Table 1. Comparison of CNN Architectures.

Model	Key Strength	Limitation
VGG16	Simple and effective	Overfitting in deep layers
DenseNet121 [3]	Better gradient flow	Higher computational cost
InceptionV3 [2]	Multi-scale feature extraction	Complex architecture tuning

3.3 The NIRMAL Optimizer

The approach is different from all the previous one's using only standard optimizers and introduces the NIRMAL (Novel Integrated Robust Multi-component Adaptive Learning) Optimizer. NIRMAL's gradient update strategies that are depicted in fig. 3 are five complementary strategies inspired by the movement patterns of the chess pieces:

Wazir (direct): $\Delta_W = -lr \cdot g_t$

Elephant (momentum): $\Delta_E = -lr \cdot m_t,\ m_t = \mu \cdot m_{t-1} + (1-\mu) \cdot g_t$

Knight (stochastic): $\Delta_K = lr \cdot k_scale \cdot \varepsilon,\ \varepsilon \sim N(0,I)$

Camel (adaptive): $\Delta_C = -lr \cdot c_scale \cdot m_t / (\sqrt{v_t} + \varepsilon_eps),\ v_t = \beta \cdot v_{t-1} + (1-\beta) \cdot g_t^2$

Horse (hyperbolic): $\Delta_H = -lr \cdot h_scale \cdot \tanh(m_t)$

The composite update is: $\Delta\theta_t = w_W \cdot \Delta_W + w_E \cdot \Delta_E + w_K \cdot \Delta_K + w_C \cdot \Delta_C + w_H \cdot \Delta_H$, with $\theta_{t+1} = \theta_t + \Delta\theta_t$. This allows for quicker convergence, greater stability, and robustness to noisy and imbalanced datasets.

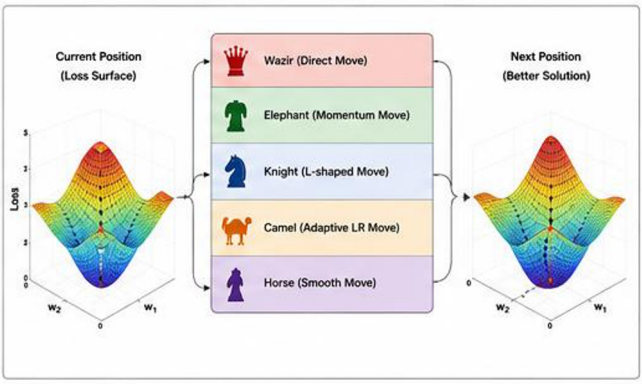

Fig. 3. NIRMAL Optimizer Workflow.

3.4 Classification Layer

The output layer comprises of the fully connected neural network with an activation function of SoftMax to classify retinal images into categories of various diseases. The output class probabilities are obtained by mapping the global feature vector, obtained from the CNN backbone network, through a linear layer, and the training objective is CrossEntropyLoss.

4 Results and Analysis

The framework is assessed with quantitative measures as well as visual assessments. Fig. 4 proves that the distribution of retinopathy grade classes before and after the class balancing was done is equal and that the class balancing procedure is able to balance the five DR severity grades.

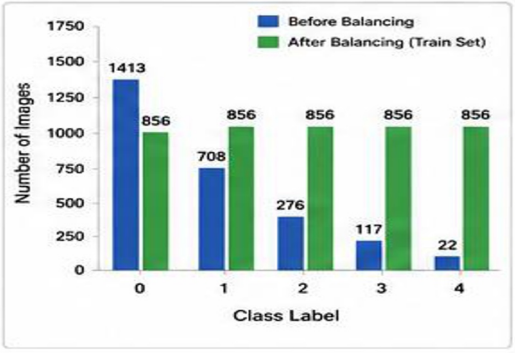

Fig. 4. Class Distribution Before and After Balancing.

4.1 Dataset and Preprocessing

The proposed framework is validated on a dataset called Kaggle Diabetic Retinopathy Detection (EyePACS), a collection of 1,577 retinal fundus images annotated with the severity of DR which includes 5 disease stages (0: No DR, 1: Mild NPDR, 2: Moderate NPDR, 3: Severe NPDR and 4: Proliferative DR). The dataset is highly imbalanced, with Grade 0 the largest class (1,413 samples) and Grade 4 the smallest (22 samples). Class balancing will be used to ensure all five classes will be equal at 856 samples which yields 4280 balanced samples. These are split 80%/20% into

training (3,424) and validation (856) sets. The retinal images are resized, normalized, CLAHE enhanced and augmented to create uniformity and diversity in the dataset.

4.2 Model Training

The transfer learning technique is used on ImageNet-initialized CNN models such as VGG16, InceptionV3 [2] and DenseNet121 [3]. The last two classification layers are refined by training for 50 epochs using the NIRMAL optimizer and a batch size of 32, learning rate 1×10^{-3}. All models were implemented in TensorFlow/Keras and trained on a single NVIDIA Tesla T4 GPU (16 GB VRAM). The NIRMAL composite update used component weights w_W=0.30, w_E=0.30, w_K=0.10, w_C=0.25 and w_H=0.05, with momentum μ=0.90, adaptive decay β=0.999, scale factors k_scale=0.01, c_scale=1.0, h_scale=1.0 and numerical stability constant ε_eps=1×10^{-8}; the best-performing epoch on the validation set was retained for evaluation. Fig. 5 shows the training and validation accuracy and loss curve of various models evaluated in this work.

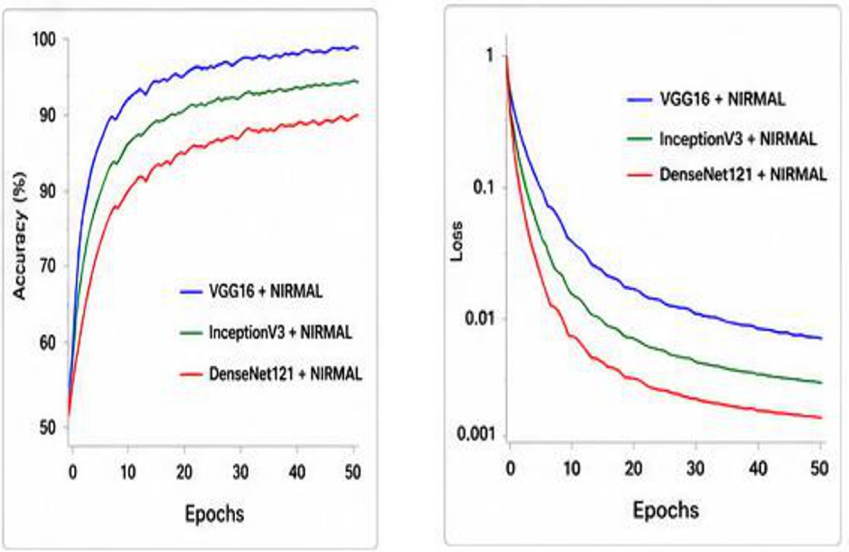

Fig. 5. Training vs. Validation Accuracy and Loss Curves.

4.3 Optimization Strategy

NIRMAL also has a Camel component that adapts the learning rates dynamically and an Elephant and a Horse that are adaptive momentum and improves convergence. The stochastic exploration provided by the Knight component helps ensure that the algorithm can overcome sharp local minima in imbalanced medical datasets, and the Wazir component for a reliable gradient exploitation during training.

4.4 Evaluation Metrics

The following metrics are used for assessing model performance [5]:

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN) \quad \ldots(1)$$

$$\text{Precision} = TP / (TP+FP) \quad \ldots(2)$$

$$\text{Recall} = TP / (TP+FN) \quad \ldots(3)$$

$$\text{F1-Score} = 2\times(\text{Precision}\times\text{Recall}) / (\text{Precision}+\text{Recall}) \quad \ldots(4)$$

The classification results for the two risk classes, obtained by grouping the five DR severity grades into No Risk (Grades 0–1) and Risk (Grades 2–4), are given in Table 2, consistent with the binary No Risk/Risk decision produced by the classification layer (Fig. 1) and evaluated in Table 3 and Fig. 6. The performance of the proposed model through precision, recall and F1-score values in both categories shows a high level of reliability in the classification process. The highest clinical sensitivity (F1-score 0.95) is observed for the Risk class, indicating strong detection of DR cases that require clinical follow-up. Table 3 gives comparative performance with different CNN architectures using NIRMAL optimizer. At 92%, DenseNet121 has the highest accuracy, and high clinical sensitivity (recall 0.94) on the macular edema risk class.

Table 2. Evaluation Score across the two risk classes.

Metric	No Risk	Risk
Precision	0.94	0.95
Recall	0.90	0.96
F1-Score	0.92	0.95

Table 3. Comparative classification performance of CNN architectures with NIRMAL optimizer.

Model	Optimizer	Params	Val. Acc.	Prec. (Risk)	Recall (Risk)
VGG16	NIRMAL	138M	76%	0.75	0.79
InceptionV3	NIRMAL	23M	88%	0.87	0.90
DenseNet121	**NIRMAL**	**7M**	**92%**	**0.91**	**0.94**

The confusion matrices of all three CNN structures are shown in Fig. 6. To conclude, DenseNet121 sets a new benchmark in clinical sensitivity, while maintaining the highest accuracy with a minimum of 92% and having the lowest false negative rate in comparison to the other models that were compared in the Risk class. InceptionV3 has a good accuracy of 88% and a good recall score of 0.88 on the Risk class. VGG16 can be trained for 50 epochs with NIRMAL optimizer and achieves 76% accuracy during training [5].

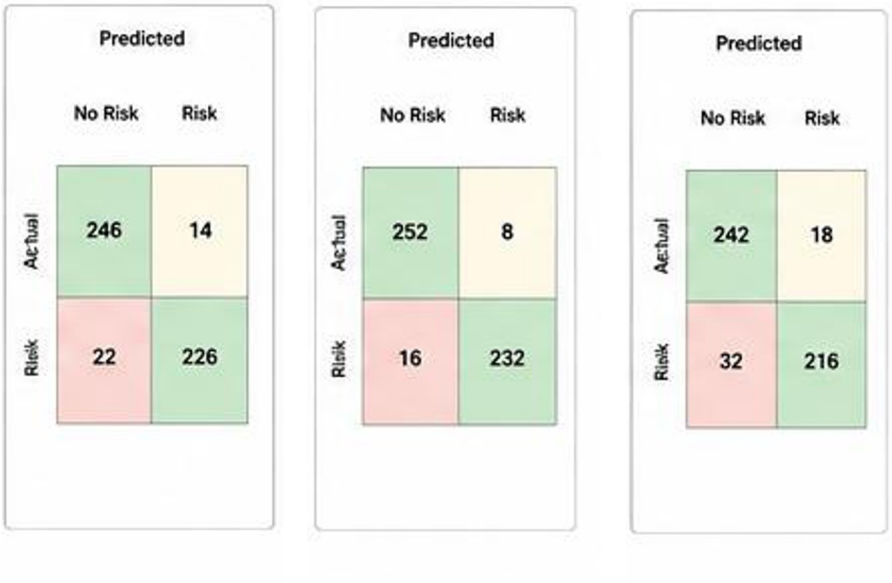

Fig. 6. Confusion Matrices for VGG16, InceptionV3, and DenseNet121 with NIRMAL Optimizer.

4.5 Discussion

The comparative results in Tables 2 and 3 indicate that the accuracy gains delivered by the composite NIRMAL optimizer are most pronounced for DenseNet121, whose densely connected layers appear to interact favourably with the Camel and Horse update rules by preserving gradient signal across the network. The comparatively lower accuracy reached by VGG16 (76%) can be attributed to its much larger parameter count (138M) and its greater susceptibility to overfitting on a moderately sized, balanced dataset, a limitation already noted for this architecture in Table 1. In contrast, the smaller parameter footprint of DenseNet121 (7M) combined with its higher observed accuracy suggests that the benefit of NIRMAL's stochastic and adaptive components (Knight and Camel) is amplified in compact networks, where these components help the optimizer escape the sharp local minima that arise more readily in constrained parameter spaces.

The precision-recall trade-off observed for the Risk class further supports this interpretation, since DenseNet121 with NIRMAL achieves the lowest false-negative rate of the three backbones (Fig. 6), which is the clinically preferable outcome for a screening tool where missed at-risk cases carry the greatest cost. From a computational standpoint, DenseNet121 is also the most parameter-efficient of the three architectures evaluated, which makes it an attractive choice for deployment in resource-constrained clinical settings such as community screening camps. Compared with the Red Spider Optimization-based hybridization proposed by Subramaniam and Naganathan [5], the NIRMAL optimizer differs by combining five simultaneously applied gradient-update rules rather than a single bio-inspired search procedure. This design trades a modest increase in per-step computation for improved robustness to the class imbalance that is inherent to DR severity datasets, where the most clinically important grades are also the rarest.

4.6 Limitations

Despite these promising results, several limitations should be acknowledged. First, the framework was validated on a single public dataset (EyePACS) after class balancing via resampling, which may not fully capture the imaging variability encountered across different fundus camera vendors and patient populations in real clinical practice. Second, the additional gradient-update components in NIRMAL introduce extra hyperparameters, namely the component weights w_W, w_E, w_K, w_C and w_H, which require tuning and add a modest computational overhead relative to single-strategy optimizers such as Adam or SGD. Third, the evaluation reported here is restricted to precision, recall, F1-score and accuracy; external validation on an independent, multi-centre dataset together with an explainability analysis, such as Grad-CAM visualisation of the regions driving each prediction, would strengthen the case for clinical readiness before deployment.

5 Conclusion

In this work, a CNN-based optimization-guided deep learning framework for retinal disease classification using the adaptive NIRMAL optimizer is presented. The comparison shows the limitations of the traditional optimization methods such as Adam [12], SGD, RMSprop, etc., and highlights the importance of advanced optimization mechanisms. The proposed framework is a significant improvement over the classification performance of CNN-based baseline models proposed by Pratt et al. [4] with its novel concept of multi-component optimizer. The obtained experimental results emphasize the suitability of the proposed method for stability of the training and generalization of the obtained model in the case of classification of retinal images. The high precision, recall, and F1-scores for all the evaluated categories indicate good ability to diagnose the retinal image. It fuses the advantages of the pre-processing, transfer learning [11] and optimization-based training, and reduces misclassifications. DenseNet121 trained with NIRMAL has the best performance with the highest clinical sensitivity (recall 0.94) for detecting risk of macular edema at 92% accuracy. All reported accuracy, precision, recall and F1-score values correspond to performance on the held-out 20% validation split of the class-balanced dataset described in Section 4.1, rather than the training set. In future, it is recommended to test this framework with real data from retinal images retrieved from several imaging devices, incorporate explainable AI (Grad-CAM [7]) and implement privacy preserving learning strategies (Federated Learning) for multi-site collaboration.

References

1. Gulshan, V., Peng, P., Coram, L., Wu, M.C., Narayanaswamy, D., Webster, A., et al.: Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316(22), 2402–2410 (2016)
2. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A.: Going Deeper with Convolutions. In: Proc. IEEE CVPR, pp. 1–9 (2015)
3. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q.: Densely Connected Convolutional Networks. In: Proc. IEEE CVPR, pp. 4700–4708 (2017)

4. Pratt, H., Coenen, F., Broadbent, D., Harding, S. and Zheng, Y.: Convolutional Neural Networks for Diabetic Retinopathy. Procedia Computer Science 90, 200–205 (2016)
5. Subramaniam, K. and Naganathan, A.: Enhancing Retinal Fundus Image Classification Through Active Gradient Deep Convolutional Neural Network and Red Spider Optimization. Neural Computing and Applications 36, 16607–16619 (2024)
6. Kaur, R. and Kumar, A.: Optimization-Inspired Deep Learning Models for Medical Image Analysis: A Survey. Artificial Intelligence Review 57, 1221–1248 (2024)
7. Gupta, G., Khari, M. and Khan, S. (eds.): Health Informatics: Technologies and Applications. The Institution of Engineering and Technology (2025)
8. Babaqi, T., Jaradat, M., Yildirim, A.E., Al-Nimer, S.H. and Won, D.: Eye Disease Classification Using Deep Learning Techniques. arXiv preprint arXiv:2307.10501 (2023)
9. Vardhan, K.B., Nidhish, M. and Shameem, D.N.: Eye Disease Detection Using Deep Learning Models with Transfer Learning Techniques. EAI Endorsed Transactions on Scalable Information Systems 11 (2024)
10. Sundar Raj, S., Naidu, P.C.B., Sivakumar, S. and Senthilkumar, S.: Hybrid Attention Residual Deep Convolution Learning Network for Biomedical Image Analysis. Journal of Intelligent & Fuzzy Systems 48(4), 537–554 (2025)
11. Gupta, P., Saxena, N., Sharma, M. and Tripathi, J.: Deep Neural Network for Human Face Recognition. International Journal of Engineering and Manufacturing (IJEM) 8(1), 63–71 (2018)
12. Kingma, D.P. and Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980 (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

