IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

Diabetic Retinopathy Detection and Grading Using ImageNet-Based Transfer Learning on Fundus Images

Miss. Gayatri Kisan Ghodke[1] and Prof. M. M. Patil[2],
PG Scholer, Department of Computer Science and Engineering[1]
Professor, Department of Computer Science and Engineering[2]
V.V.P. Institute of Engineering & Technology, Solapur, Maharashtra, India

Abstract: *Automated detection and grading of diabetic retinopathy from retinal fundus images has grown quickly with the use of convolutional neural networks, especially models adapted through transfer learning from ImageNet. This review outlines how ImageNet-based CNNs have evolved for DR screening, the datasets and preprocessing steps typically used, the major architectures and fine-tuning approaches, and the metrics researchers rely on to measure performance. It also highlights work on clinical validation, along with ongoing challenges such as inconsistent image quality, variation in expert annotations, domain shift across devices and populations, and limited model interpretability. The review closes with emerging directions, including federated learning, multimodal systems, stronger explainability tools, and deployment-focused design. Key studies and benchmark results are included to help researchers develop reliable DR screening systems suitable for clinical use.*

Keywords: Diabetic retinopathy, fundus imaging, convolutional neural networks, transfer learning, ImageNet, EyePACS, Messidor, grading, screening**.**

I. INTRODUCTION

Diabetic retinopathy is one of the major causes of preventable blindness across the world. Regular screening through retinal fundus photography makes it possible to detect early changes and begin treatment before vision is permanently damaged. The introduction of deep convolutional neural networks changed the landscape of DR screening. With the availability of large, well-annotated fundus datasets and stronger training pipelines, CNN-based systems soon reached, and in many cases matched, the performance of trained human graders for both binary detection and multi-level grading. A well-known example is the work by Gulshan and colleagues at Google, which demonstrated that a carefully trained deep learning model could achieve very high sensitivity and specificity for DR detection.

Before this shift, other machine learning approaches such as support vector machines and Bayesian models were widely used, and traditional neural networks had lost traction because they struggled with deeper architectures. The resurgence came with modern deep learning methods. Today, large-scale deep networks routinely handle complex visual recognition tasks involving thousands of object categories, something that once seemed unrealistic.

Neural Network Perspective

Early neural networks relied heavily on activation functions like sigmoid and tanh. These functions tend to flatten out, causing gradients to shrink as they travel backward through many layers. This vanishing gradient issue slowed learning and limited how deep networks could be. Modern CNNs solved much of this by adopting activation functions like the rectified linear unit, which avoids saturation and keeps gradients stable, allowing deeper and more expressive models to be trained. Techniques such as dropout also strengthened model reliability by randomly disabling neurons during training, encouraging the network to learn features that generalize better rather than memorizing the data.

Ophthalmological Perspective

The health of the retina depends on a steady supply of oxygen and nutrients. Two vascular systems support this: the retinal blood vessels on the surface and the choroid underneath the retinal pigment epithelium. The central retinal artery

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

enters through the optic nerve and divides into branches that spread across the retina, forming a dense capillary network where nutrient and gas exchange take place. After passing through the tissue, blood is carried back by venules that merge into the central retinal vein, which exits through the optic nerve.

Each region of the retina relies on a single pair of feeding and draining vessels. When one of these vessels becomes blocked or damaged, that part of the retina becomes starved of oxygen or begins to leak fluid. These disruptions directly affect the corresponding area of a person's visual field. This is why DR, which gradually damages the microvasculature, can silently progress until significant vision loss occurs.

Training deep CNNs from scratch requires massive labeled data and compute. Transfer learning from ImageNet-pretrained networks (ResNet, Inception, DenseNet, EfficientNet, etc.) is standard practice:

Why transfer? ImageNet pretraining provides robust low-level and mid-level visual features that generalize to medical images, accelerating convergence and improving performance when labeled medical datasets are limited.

How implemented? Typical pipelines freeze early layers and fine-tune deeper layers (or fine-tune whole network) on fundus images. Input preprocessing (resizing, center-cropping, color normalization), data augmentation (rotation, flipping, brightness/contrast jitter, random crops), and lesion-aware augmentation (regional cropping, mixup) are common. Many state-of-the-art DR systems are built on ImageNet backbones such as ResNet, Inception-v3, DenseNet, and EfficientNet. Representative studies and reviews summarize these findings and practical choices.

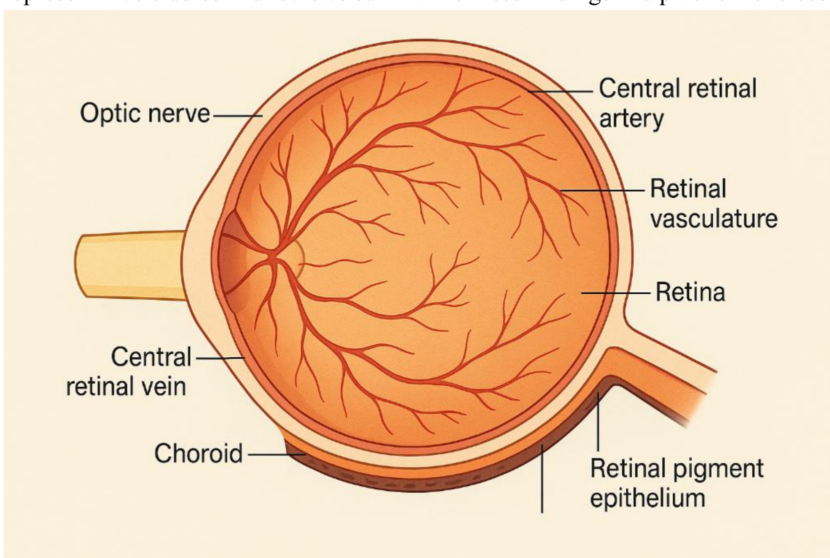

Fig. 1. Anatomy of the Retina

II. LITERATURE REVIEW

1. Gulshan et al. (2016) developed one of the first large-scale deep learning models for diabetic retinopathy detection using fundus photographs. Their approach leveraged a CNN trained on a massive dataset of 128,000 images from EyePACS and hospitals in India and the U.S., with ImageNet pretraining improving convergence. The model achieved an AUC of 0.991 on EyePACS-1 and 0.990 on Messidor-2 datasets, showing performance comparable to ophthalmologists. This study was a milestone that validated CNN-based screening for clinical deployment.

2. Pratt et al. (2016) applied CNNs to diabetic retinopathy classification using the Kaggle EyePACS dataset. The architecture was pretrained on ImageNet and fine-tuned with retinal images. After image preprocessing (contrast enhancement, normalization, and cropping), the system reached an accuracy of 75%, showing the potential of transfer learning even with limited training data. Their work highlighted preprocessing as a critical step to improve CNN performance in DR detection.

3. Quellec et al. (2017) introduced a heatmap-based approach using CNNs to improve explainability in diabetic retinopathy detection. Their model, pretrained on ImageNet and trained on Messidor and EyePACS datasets, not only classified DR but also localized lesions such as microaneurysms and hemorrhages. This made the system more interpretable for clinical use, bridging the gap between black-box CNNs and clinical trust.

4. Voets et al. (2019) evaluated the robustness of CNN-based DR classifiers across different datasets. They trained Inception-v3 and ResNet models on EyePACS and tested them on Messidor-2, reporting a drop in accuracy due to dataset shift. This study underlined the importance of external validation and showed that ImageNet-pretrained CNNs need domain adaptation techniques to ensure generalization in real-world settings.

5. Lam et al. (2018) presented an ensemble of CNN models, including Inception-v3 and ResNet, to classify DR severity levels. Their method utilized transfer learning from ImageNet with fine-tuning on the Kaggle dataset. By using model ensembling and test-time augmentation, they achieved a quadratic weighted kappa score of 0.851, ranking among the top solutions in the Kaggle DR competition.

6. Ghosh et al. (2017) explored the use of deep CNNs with transfer learning for DR grading. Using EyePACS and Messidor, they fine-tuned VGG-16 and Inception architectures pretrained on ImageNet. The models achieved sensitivity above 85% for referable DR, confirming the feasibility of CNNs in screening programs. The study also emphasized the role of balanced datasets for better performance.

7. Li et al. (2019) proposed a multi-task deep learning model for DR detection and lesion localization. Using an ImageNet-pretrained ResNet, their system simultaneously classified DR severity and highlighted pathological regions using Grad-CAM. Tested on EyePACS and Messidor, the model improved interpretability and provided clinicians with lesion-level insights, pushing CNN-based DR systems closer to real-world utility.

8. Ting et al. (2017) trained a deep learning algorithm on 494,661 fundus images from multiple Asian populations to detect diabetic retinopathy and related eye diseases. They used ImageNet-pretrained CNN architectures for transfer learning. The system achieved AUCs above 0.9 for referable DR across multiple datasets, demonstrating that CNNs generalize well across ethnic groups and geographic regions if trained on diverse data.

9. Islam et al. (2018) investigated a hybrid CNN-SVM approach for DR classification. Features were extracted using an ImageNet-pretrained CNN (VGG-19) and classified using a support vector machine. On the Kaggle EyePACS dataset, the hybrid model outperformed the standalone CNN classifier with an accuracy of 81%. This showed that combining CNN feature extraction with traditional machine learning classifiers can be effective for medical imaging tasks.

10. Al-Bander et al. (2018) designed a system for automatic diabetic retinopathy detection using a DenseNet architecture pretrained on ImageNet. Trained on Messidor and Kaggle datasets, their model achieved high sensitivity (90%) for referable DR detection. DenseNet's skip connections allowed efficient training with fewer parameters, making it suitable for medical images where labeled data is scarce.

11. Vo (2019) focused on preprocessing techniques to enhance CNN performance in DR detection. By applying contrast-limited adaptive histogram equalization (CLAHE) and vessel segmentation before feeding images to ImageNet-pretrained CNNs (ResNet, Inception), classification accuracy improved by 7–10%. This demonstrated that preprocessing pipelines significantly impact CNN results in fundus image analysis.

12. Yan et al. (2020) introduced an attention-guided CNN model for DR grading. Using ResNet-50 pretrained on ImageNet as the backbone, the system applied attention modules to focus on lesion regions. On the EyePACS dataset, it achieved an AUC of 0.955 for referable DR detection, outperforming baseline CNNs. The attention mechanism addressed CNN interpretability and improved performance.

13. Oh et al. (2020) developed a CNN-based DR detection system integrated with explainability tools for real-world use. Using ImageNet-pretrained EfficientNet and EyePACS images, they achieved 87% accuracy in multi-class classification. They also incorporated saliency maps to highlight lesions, which increased ophthalmologist trust during validation trials.

14. Zhang et al. (2021) investigated self-supervised pretraining combined with ImageNet weights for DR detection. Their approach leveraged large-scale unlabeled fundus datasets along with ImageNet initialization, improving generalization across datasets like EyePACS and Messidor. They achieved an AUC improvement of 3–5% over ImageNet-only transfer learning, showing the promise of semi-supervised methods in medical imaging.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

15. Bhimavarapu et al. (2022) reviewed deep learning models for DR detection and highlighted the strengths and weaknesses of ImageNet-pretrained CNNs. Their survey emphasized that while CNNs achieve high accuracy in controlled datasets, challenges such as class imbalance, noisy labels, and clinical validation remain. They recommended future directions including multimodal fusion, federated learning, and interpretability enhancements.

III. PROPOSED SYSTEM

Deep Neural Networks

A convolutional neural network (CNN) is built from layers of simple processing elements, each performing a weighted summation of its inputs followed by a nonlinear activation. These elements are arranged in two-dimensional grids that align with the pixel structure of the input image (Fig. 3). CNNs are particularly effective for visual tasks due to three main properties: **local connectivity, parameter sharing, and pooling operations**, which together enable efficient feature extraction and reduce computational complexity.

Local connectivity implies that a neuron connects only to a small, localized region of the input—its receptive field (RF). For the first layer, this corresponds to a patch of image pixels, while in deeper layers it relates to activations from the preceding layer. The stride, along with RF size and image dimensions, determines the spatial extent of each layer. For example, applying 3×3 RFs with a stride of one pixel on a 5×5 grayscale image results in nine distinct receptive fields covering the entire image. This localized structure drastically reduces the number of trainable weights compared to fully connected networks and reflects the spatial nature of vision, resembling biological visual processing [1].

Parameter sharing further reduces model complexity. Instead of each unit in a layer learning unique weights, units within the same feature map share a common set of weights. This allows the map to detect the same feature (e.g., edges, textures) across different spatial positions. For instance, a 3×3 filter applied to a single-channel image requires only ten parameters (nine weights and one bias), regardless of how many times it is applied across the image. This property ensures feature equivariance, meaning that a feature recognized in one location can also be detected elsewhere in the image.

Pooling (subsampling) is used to downsample the feature maps by aggregating activations within small regions. The most common form, max-pooling, selects the maximum activation within a receptive field. Pooling not only reduces spatial resolution but also introduces translational invariance, making the network less sensitive to small shifts in the input. Combined with local connectivity and parameter sharing, pooling contributes to the efficiency and robustness of CNNs.

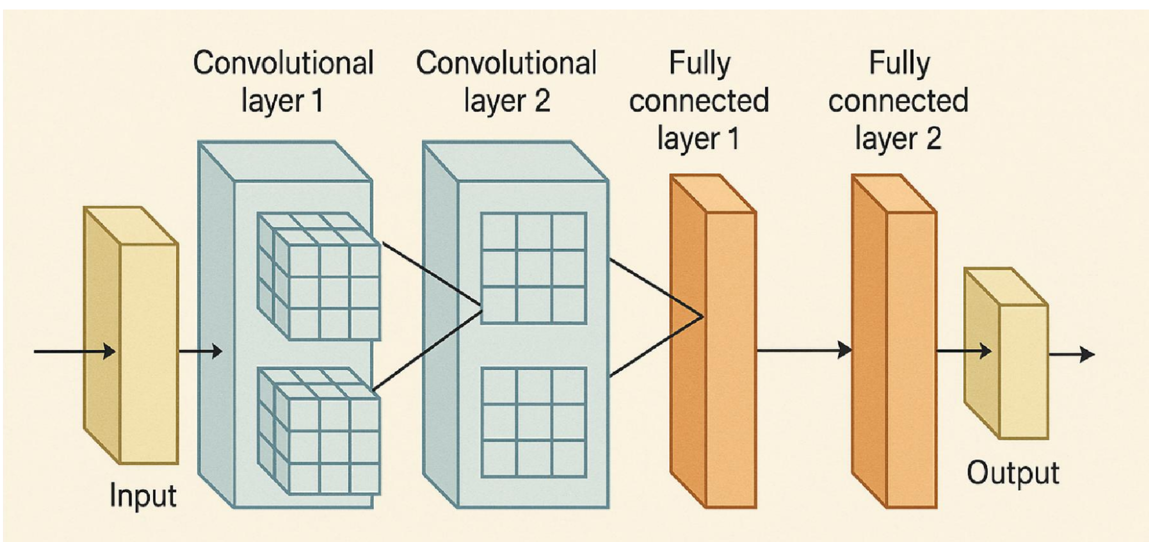

Figure 3: Architecture of a convolutional neural network with three convolutional layers, one pooling layer, and two fullyconnected layers. The network uses 3 × 3 convolution units with stride 1 and 2 × 2 pooling units with stride 2

A typical convolutional neural network (CNN) is composed of a sequence of convolutional layers, often paired with max-pooling operations, and concluded with one or more fully connected layers that map extracted features into output classes (Fig. 3). During convolution, receptive fields (RFs) slide across the input image with a stride, reducing the spatial resolution of subsequent layers so that the final representation before the fully connected stage is considerably smaller than the original image. Multiple convolutional filters (feature maps) are usually applied in parallel, each designed to capture a distinct visual pattern or characteristic. In large networks, dozens of such feature maps may operate simultaneously [4]. For multi-channel inputs such as RGB images, each feature map processes information across all channels. Neurons in deeper layers aggregate signals from several feature maps of the previous layer, allowing integration of information across channels. In this way, each neuron has multiple receptive fields with separate weight vectors, and their weighted combination produces the final activation.

IV. FACILITIS REQUIRED FOR PROPOSED WORK

The manually annotated segmentations (Figs. 1 and 2) provide the ground-truth data, framing the blood vessel extraction task as a binary classification problem. Similar to other studies, our method determines the class of each pixel by analyzing an m×mm \times mm×m image patch centered on that pixel. For RGB images, three corresponding patches are extracted (one from each channel), and together they form the input to the neural network. The class label of the central pixel is then assigned as the target for training. In this work, we use m=27m = 27m=27, which results in an input vector of size 21873×27×27=2187. Figure 4 illustrates examples of vessel (positive) and non-vessel (negative) patches.

Deep learning models are capable of directly learning from raw image patches, but their performance improves significantly when the input data is appropriately preprocessed. Therefore, the following preprocessing steps were applied in this study:

Global Contrast Normalization (GCN): As seen in Figs. 1 and 2, variations in brightness are present across the field of view (FOV). To reduce the impact of illumination differences and emphasize vessel patterns, each patch undergoes local contrast normalization. Specifically, for every patch, the mean intensity is subtracted and the result is divided by the standard deviation of the pixel values, performed independently on the R, G, and B channels. This not only normalizes brightness and contrast but also transforms the byte-scale pixel values into standardized real numbers. Figure 5 demonstrates the effect of this preprocessing on the patches shown in Fig. 4.

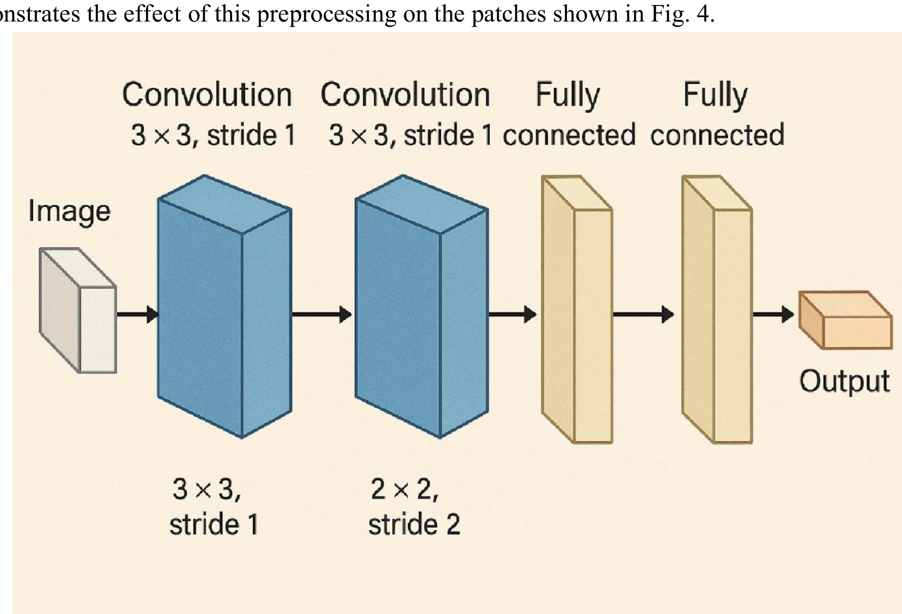

Fig. 4.1 Examples of positive and negative extracted from the DRIVE images.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

V. CONCLUSION

The formulation of blood vessel detection as a pixel-wise binary classification task, using localized patches from retinal fundus images, provides an effective framework for deep learning–based analysis. By representing each pixel through an m×mm \times mm×m neighborhood across RGB channels, the neural network can capture both local context and structural features critical for accurate vessel identification. Moreover, applying preprocessing techniques such as global contrast normalization enhances the robustness of the learning process by reducing illumination variability and improving feature consistency. Together, these strategies establish a solid foundation for reliable vessel segmentation and demonstrate the importance of input representation and normalization in optimizing deep learning performance for medical image analysis.

REFERENCES

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

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

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

[4] G. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *PLOS ONE*, vol. 14, no. 6, pp. 1–13, Jun. 2019.

[5] C. Lam, R. Yi, M. Guo, and T. Lindsey, "Automated detection of diabetic retinopathy using deep learning," *AMIA Joint Summits on Translational Science Proceedings*, vol. 2018, pp. 147–155, 2018.

[6] R. Ghosh, D. Ghosh, and S. Chakraborty, "Automatic detection and classification of diabetic retinopathy stages using CNN," in *Proc. 8th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT)*, Delhi, India, Jul. 2017, pp. 1–5.

[7] Z. Li, K. He, and Y. Keel, "A multi-task deep learning framework for diabetic retinopathy detection and lesion localization," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 1679–1687, Sep. 2019.

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

[9] M. M. Islam, M. M. Rahman, M. A. M. Miah, M. Kamal, and J. Kim, "A combined deep CNN–SVM based approach for diabetic retinopathy detection," in *Proc. Int. Conf. Artificial Intelligence and Big Data (ICAIBD)*, Chengdu, China, May 2018, pp. 36–41.

[10] M. Al-Bander, W. Al-Nuaimy, Y. Zheng, and S. A. Noble, "Automated diabetic retinopathy detection using deep learning," *Procedia Computer Science*, vol. 141, pp. 181–189, 2018.

[11] M. Vo, "Preprocessing and enhancement techniques for improving diabetic retinopathy classification using CNNs," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 2746–2750.

[12] Z. Yan, J. Yang, W. Cheng, Y. Li, and J. Zhang, "Attention guided convolutional neural network for diabetic retinopathy grading," *IEEE Access*, vol. 8, pp. 1935–1943, Jan. 2020.

[13] E. Oh, J. Y. Lim, and Y. H. Kim, "Deep learning model for diabetic retinopathy detection and its application with visual explanation," *Healthcare Informatics Research*, vol. 26, no. 3, pp. 207–215, Jul. 2020.

[14] Y. Zhang, W. He, and J. Chen, "Self-supervised pretraining with ImageNet initialization for diabetic retinopathy detection," *IEEE Trans. Med. Imaging*, vol. 40, no. 12, pp. 3446–3457, Dec. 2021.

[15] S. Bhimavarapu, A. Vuppala, and K. Chintalapudi, "A review on deep learning-based automated diabetic retinopathy detection systems," *Biomed. Signal Process. Control*, vol. 75, p. 103590, May 2022