

Deep Convolutional Neural Networks for Diabetic Retinopathy and Glaucoma Screening from Retinal Fundus Images: A Literature Review with Emphasis On Rural India

Sanika Ratanrao Kolhe¹, Dr. Abhijit Banubakode²

^{1,2}Computer Science Engineering (AI),
Pimpri Chinchwad University, Sate Maval.

Abstract

Diabetic retinopathy (DR) and glaucoma are major causes of preventable blindness, and both can be picked up on a colour fundus photograph before the patient notices any symptom. In rural India, though, ophthalmologists are few compared with the number of people at risk, so many patients are seen late or not at all. Convolutional neural networks (CNN) have been reported to grade fundus images about as well as trained graders, and a number of studies have tested them in Indian hospitals and community camps. This paper reviews that literature from a computer engineering perspective. It covers the clinical background, public datasets and preprocessing, the development of CNN architectures, published DR and glaucoma detection studies, and the practical problems of deploying such systems with smartphone-based cameras and poor connectivity. The review finds that DR detection is considerably more mature than glaucoma detection, that most of the evidence comes from well-resourced centres, and that image quality, dataset shift, explainability, joint multi-disease screening and regulatory readiness are still open problems. A framework for a rural screening pipeline and an evaluation plan are outlined to guide future work.

Keywords— Diabetic Retinopathy, Glaucoma, Retinal Fundus Images, Convolutional Neural Networks, Deep Learning, Medical Image Analysis, Rural Healthcare, India.

1. Introduction

Vision-threatening eye disease is largely a question of timing. Diabetic retinopathy damages the small blood vessels of the retina and gives no warning in its early stages. Glaucoma injures the optic nerve, usually without pain, and the vision it takes cannot be restored. For both diseases, regular examination of the retina is the main safeguard. In India the obstacle is access. A prospective study at two eye hospitals noted that more than 60 million people in the country are estimated to have diabetes and are therefore at risk of DR [1], and glaucoma was projected to affect roughly 80 million people worldwide by 2020 [2]. Specialist eye care is concentrated in cities, while a large share of the affected population lives far from it.

One response is to shift the first reading of the retinal image from a scarce specialist to a computer. Colour fundus photography is comparatively cheap, can be done by a trained technician, and has

recently been adapted to smartphone-based devices suited to camps and primary care clinics [3]. Deep learning, and CNNs in particular, supplies the missing interpretive step. In a widely cited study, a CNN trained on a large set of graded photographs reached high sensitivity and specificity for referable DR on two independent validation sets [4]. The question has since moved from whether a network can grade an image to whether it can be trusted in the field. A human-centred evaluation across eleven clinics in Thailand showed that real clinic conditions, such as poor lighting and uneven image quality, interact strongly with model behaviour and nursing workflow [5].

This review is written for an M.E. computer engineering readership and has three aims: to summarize what is established about CNN-based DR and glaucoma detection, to identify the gaps that matter most for rural India, and to indicate how a screening system could be put together to address them. Section II gives clinical background. Section III describes datasets and preprocessing. Section IV traces CNN architectures. Sections V and VI review DR and glaucoma studies. Section VII discusses deployment in low-resource settings. Section VIII analyses the gaps, Section IX outlines a proposed framework, and Section X concludes.

2. Clinical Background

A. Diabetic Retinopathy

DR is graded on an international clinical severity scale with five levels: no apparent retinopathy, mild, moderate and severe non-proliferative disease, and proliferative disease. A separate scale covers diabetic macular oedema [6]. Most screening studies reduce this to a binary decision, usually “referable” (moderate or worse, or macular oedema) against “non-referable”. For engineers the choice matters because it fixes the label space, the class balance and the cost of a missed case.

B. Glaucoma

Glaucoma is a group of optic neuropathies. On a fundus photograph the visible signs include an enlarged optic cup relative to the disc, thinning of the neuroretinal rim and changes in the nerve fibre layer. Unlike DR, there is no single accepted photographic grading scale, and a confirmed diagnosis depends on further tests such as intraocular pressure measurement and visual field testing. A fundus-image classifier therefore predicts a *suspect* label that only partly matches the final clinical diagnosis, which makes glaucoma labels noisier than DR labels.

C. Imaging in the Field

Tabletop fundus cameras give wide, evenly lit images but are costly and fixed in place. Portable and smartphone-attached devices lower the cost and make community screening possible. One study of 301 patients with type 2 diabetes used a smartphone-based fundus device with automated software and compared the output with ophthalmologist grading; it was among the first Indian studies to combine the two [3]. The price of portability is image quality, which affects every later stage.

3. Datasets and Preprocessing

A. Public Datasets

The quality of a model is bounded by the quality of its data. Table I lists the datasets that recur in the literature.

TABLE I. Frequently Used Fundus Image Datasets

Dataset	Target	Notes
EyePACS / Kaggle DR [7]	DR grading (five levels)	More than 35,000 training images; varied cameras and uneven quality
Messidor [8]	DR and macular oedema grading	1,200 images from French ophthalmology departments
IDRiD [9]	DR and oedema grading, lesion segmentation, disc and fovea localization	516 grading images; collected at an eye clinic in Maharashtra, India
REFUGE [10]	Glaucoma classification, optic disc and cup segmentation	1,200 images from a challenge; reference labels from expert assessment

Two points stand out. First, only IDRiD comes from India, so most published models were trained on populations, cameras and imaging protocols that differ from rural Indian practice. Second, glaucoma has far fewer large public datasets than DR, which partly explains why its literature is thinner.

B. Preprocessing and Augmentation

Fundus images differ in resolution, field of view, illumination and colour balance. Typical preprocessing includes cropping the black border, resizing to the network input size, normalizing colour or illumination, and enhancing contrast in the green channel, where vessels and lesions show best. Augmentation by rotation, flipping, scaling and brightness change enlarges the training set and reduces overfitting. Small lesions such as microaneurysms can disappear at low resolution, so input size is a real design decision. Higher resolution makes lesions easier to see but costs memory and inference time, which matters on cheap hardware.

4. CNN Architectures for Fundus Image Analysis

A. From Early Networks to Modern Backbones

The current period of CNN-based vision began with a deep network trained on ImageNet [11]. Deeper plain networks followed [12], and the Inception family introduced parallel convolutions of several kernel sizes, with factorization to hold down cost [13]. Residual networks made very deep models trainable through identity shortcuts [14], and dense connectivity let layers reuse earlier features [15]. EfficientNet scaled depth, width and resolution together in a balanced way [16]. For devices with limited compute, MobileNets replaced standard convolutions with depthwise separable ones [17]. Vision transformers, which treat an image as a sequence of patches and rely on self-attention [18], are the most recent addition; they generally need large pre-training sets, which is a constraint in medical imaging.

B. Transfer Learning and Segmentation

Medical datasets are small by computer-vision standards. Tajbakhsh et al. compared training from scratch with fine-tuning ImageNet-pretrained CNNs on several medical tasks and found that fine-tuned networks did at least as well and held up better when training data were scarce [20]. Broader surveys of the field [19] describe the same pattern. For lesion detection and for delineating the optic disc and cup, encoder-decoder networks such as U-Net [21] are the usual choice, and the cup-to-disc ratio computed from such masks is a clinical feature for glaucoma.

5. Diabetic Retinopathy Detection Studies

A. Algorithm Development

Early CNN work on the Kaggle data showed that five-class grading is much harder than a binary referral decision, partly because the severe classes are rare [22]. The landmark development and validation study by Gulshan et al. trained a network on a large graded set and tested it on EyePACS-1 and Messidor-2, reporting 90.3% and 87.0% sensitivity with 98.1% and 98.5% specificity at a high-specificity operating point [4]. Gargeya and Leng reported a data-driven classifier validated externally on Messidor-2 and E-Ophtha [23]. Ting et al. extended the approach to multiethnic populations and to eye diseases related to DR [24]. Quellec et al. proposed a way of mining images to localize the lesions that drive a DR decision, which brings some interpretability into the pipeline [26].

B. Prospective and Real-World Evaluation

Retrospective accuracy is not clinical performance. A pivotal prospective trial evaluated an autonomous AI system for DR in primary care offices [25]. In India, a prospective study of 3,049 patients at Aravind Eye Hospital and Sankara Nethralaya compared an automated grader with manual graders and a retina specialist. For moderate-or-worse DR the system's sensitivity was 88.9% and 92.1% at the two sites, with specificity of 92.2% and 95.2%, equal to or better than individual manual graders [1]. A community-based study used an offline smartphone AI system, which is notable because it removes the need for a network connection [27]. Together with the smartphone study of Rajalakshmi et al. [3], this work suggests that acceptable accuracy is achievable on Indian patients. The Thailand study [5] points the other way: strict image-quality thresholds can lead to many patients being referred or re-imaged when conditions are poor.

Table II condenses the DR studies discussed.

TABLE II. Summary of Representative DR Detection Studies

Study	Setting / data	Main contribution	Limitation for rural India
Pratt et al. [22]	Kaggle data	Early CNN for five-class grading	Weak on rare severe classes
Gulshan et al. [4]	EyePACS-1, Messidor-2	Referable DR with high sensitivity and specificity	Retrospective; non-Indian data
Gargeya and Leng [23]	EyePACS-based; external sets	Automated identification with external validation	Retrospective
Ting et al. [24]	Multiethnic cohorts	DR and related eye diseases	Tertiary-quality images
Abràmoff et al. [25]	Primary care offices	Prospective autonomous system	Specific camera protocol
Gulshan et al. [1]	Two Indian eye hospitals, 3,049 patients	Matched or exceeded manual grading	Hospital settings
Rajalakshmi et al. [3]	301 patients, smartphone device	AI on smartphone images	Small, single centre
Natarajan et al. [27]	Community screening, offline AI	Offline operation on a phone	Limited to DR

6. Glaucoma Detection Studies

Chen et al. were among the first to apply a deep CNN directly to glaucoma detection from fundus images, using two datasets [28]. Li et al. trained a deep learning system on a large collection of fundus photographs to detect glaucomatous optic neuropathy and examined its use in screening [29]. Fu et al. proposed a disc-aware ensemble network that fuses a whole-image stream, a segmentation-guided stream, a local disc stream and a polar-transformed disc stream, and reported better results than earlier methods on the SCES and SINDI datasets [30]. The REFUGE challenge supplied a shared benchmark combining classification with disc and cup segmentation, so that methods could be compared on equal terms [10].

Several limitations recur. Glaucoma ground truth is partly subjective, so inter-observer variability limits achievable accuracy. Many datasets come from clinic populations in which disease prevalence is far higher than in community screening, which can inflate apparent performance. There is also little published evidence of prospective glaucoma screening on Indian patients comparable to the DR studies in Section V, and models trained on one camera or ethnic group may transfer poorly to another.

7. Deployment In Low-Resource Settings

A. Computation and Connectivity

Rural clinics may lack stable internet access and dedicated servers. Lightweight backbones [17] and on-device inference allow screening without a network, as the offline smartphone study shows [27]. Accuracy, model size and latency have to be weighed together, and the operating threshold should reflect that a missed case costs more than an extra referral.

B. Image Quality and Human Factors

The clinic-level evidence in [5] indicates that performance in practice depends on how images are captured and by whom. A quality-assessment stage that gives the operator immediate feedback, so that a poor image can be retaken on the spot, is therefore as important as the classifier. Training and workflow design for health workers are part of the engineering problem, not an afterthought.

C. Explainability and Trust

Clinicians are more willing to act on a recommendation when they can see what drove it. Gradient-based localization maps [31] highlight the image regions that influenced a CNN decision and can be overlaid on the fundus photograph for review. Such maps are an aid, not proof of correct reasoning, and they should be checked against clinician-marked lesions.

D. Privacy, Data Sharing and Regulation

Pooling images from many centres would help generalization but raises privacy and governance concerns. Federated learning trains a shared model across sites without moving the raw data [32], which suits hospital networks, although it adds communication and heterogeneity challenges. In India, the Indian Council of Medical Research has issued ethical guidelines for the use of artificial intelligence in biomedical research and healthcare [33], which developers should treat as the baseline for consent, data protection, validation and accountability.

8. Critical Analysis and Research Gaps

The literature supports three broad conclusions. DR detection has reached prospective validation, including in India [1], [27]. Glaucoma detection from fundus images is promising but less validated [28]–[30]. Deployment evidence shows that field conditions can erode laboratory performance [5]. Table III lists the gaps most relevant to rural India.

TABLE III. Research Gaps and Their Implications

Gap	Evidence in reviewed work	Implication for a rural system
Population and camera shift	Most training data are non-Indian or from tertiary centres [4], [7], [8]	Validate and calibrate on local data and devices; test across cameras
Low-quality field images	Quality thresholds affect real clinics [5]	Include operator-guided quality checks and quality-robust training
Single-disease focus	Most systems target DR only [1], [3], [27]	Build joint DR and glaucoma screening from one photograph
Limited prospective glaucoma evidence	Retrospective or challenge-based studies [10], [28]–[30]	Run prospective community evaluation
Explainability not validated	Saliency maps used as aids [31]	Compare maps with clinician annotations
Privacy and governance	Pooled data raise consent issues [32], [33]	Consider federated training and ICMR-aligned consent
Cost-benefit unknown	Few field economic analyses in the reviewed set	Measure referral accuracy and cost per detected case

9. Proposed Framework for Rural Screening

The framework below is a proposal drawn from the gaps above and needs experimental validation.

1. **Capture and quality check.** A trained health worker captures a fundus image with a portable or smartphone-based camera. A lightweight quality classifier flags blur, poor illumination or the wrong field and prompts a retake.
2. **Preprocessing.** Border cropping, size normalization and contrast enhancement, with resolution chosen to keep small lesions visible within the device’s memory limits.
3. **Multi-task CNN.** A shared mobile-friendly backbone [17] with transfer learning [20] feeds two heads: a DR severity or referral head and a glaucoma-suspect head, optionally supported by optic disc and cup segmentation [21].
4. **Explanation layer.** Localization maps [31] are overlaid on the image for the reviewing clinician.
5. **Referral logic.** Thresholds favour sensitivity. Uncertain or ungradable cases are referred, not silently classified.
6. **Teleophthalmology review.** Flagged images are sent, when connectivity allows, to an ophthalmologist for confirmation, and the outcome is stored as feedback for retraining.
7. **Governance.** Consent, access control and audit follow national guidance [33]; federated training [32] is considered for multi-site learning.

Evaluation plan. The primary measures are sensitivity, specificity and area under the ROC curve for referable DR and glaucoma-suspect status, reported with confidence intervals and separately by camera

and site. Secondary measures are the proportion of ungradable images, agreement with specialist grading, time per patient and referral rate. Because prevalence in community screening is lower than in hospital samples, positive predictive value should be reported at the expected local prevalence.

10. Conclusion

The reviewed studies show that CNNs can detect referable diabetic retinopathy with accuracy comparable to trained human graders, including in prospective Indian studies, and that offline smartphone-based operation is feasible. Glaucoma detection from fundus images is advancing but rests on fewer datasets and less prospective evidence. The main obstacles to rural use are no longer raw model accuracy. They are dataset and device shift, image quality in the field, explainability, joint multi-disease screening, privacy and clinical integration. A practical system should combine a quality gate, a lightweight multi-task network, explanations, conservative referral thresholds and specialist review, and should be validated on local data. Future work should concentrate on prospective community studies of joint DR and glaucoma screening on Indian populations and devices.

References

1. V. Gulshan et al., "Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India," *JAMA Ophthalmol.*, vol. 137, no. 9, pp. 987–993, Sep. 2019.
2. H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *Br. J. Ophthalmol.*, vol. 90, no. 3, pp. 262–267, Mar. 2006.
3. R. Rajalakshmi, R. Subashini, R. M. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," *Eye*, vol. 32, no. 6, pp. 1138–1144, Jun. 2018.
4. V. Gulshan 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.
5. E. Beede et al., "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy," in *Proc. CHI Conf. Human Factors in Computing Systems*, Honolulu, HI, USA, 2020, pp. 1–12.
6. C. P. Wilkinson et al., "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, Sep. 2003.
7. Kaggle, "Diabetic retinopathy detection," 2015. [Online]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection>
8. E. Decencière et al., "Feedback on a publicly distributed image database: The Messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.
9. P. Porwal et al., "Indian Diabetic Retinopathy Image Dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, Art. no. 25, 2018.
10. J. I. Orlando et al., "REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.*, vol. 59, Art. no. 101570, Jan. 2020.
11. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 25, 2012, pp. 1097–1105.

12. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learning Representations (ICLR), 2015.
13. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826.
14. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.
15. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017, pp. 2261–2269.
16. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Machine Learning (ICML), PMLR vol. 97, 2019, pp. 6105–6114.
17. A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.
18. A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.
19. G. Litjens et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.
20. N. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1299–1312, May 2016.
21. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.
22. H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," Procedia Comput. Sci., vol. 90, pp. 200–205, 2016.
23. R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," Ophthalmology, vol. 124, no. 7, pp. 962–969, Jul. 2017.
24. D. S. W. Ting et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," JAMA, vol. 318, no. 22, pp. 2211–2223, Dec. 2017.
25. M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," npj Digit. Med., vol. 1, Art. no. 39, 2018.
26. G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," Med. Image Anal., vol. 39, pp. 178–193, Jul. 2017.
27. S. Natarajan, A. Jain, R. Krishnan, A. Rogye, and S. Sivaprasad, "Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone," JAMA Ophthalmol., vol. 137, no. 10, pp. 1182–1188, Oct. 2019.
28. X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in Proc. 37th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 2015, pp. 715–718.
29. Z. Li et al., "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," Ophthalmology, vol. 125, no. 8, pp. 1199–1206, Aug. 2018.

30. H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Trans. Med. Imaging*, vol. 37, no. 11, pp. 2493–2501, Nov. 2018.
31. R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 618–626.
32. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, PMLR vol. 54, 2017, pp. 1273–1282.
33. Indian Council of Medical Research, *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare*, New Delhi, India, 2023.