Towards Convolutional Neural Networks for Diabetic-Retinopathy Detection

Vitor A. Oliveira[1], Helida Santos[2], Giancarlo Lucca[3], Lizandro Oliveira[3], Adenauer Yamin[1], Renata Reiser[1]

[1]Universidade Federal de Pelotas (UFPel), Pelotas-RS, Brasil

[2]Universidade Federal do Rio Grande (FURG), Rio Grande-RS, Brasil

[3]Universidade Católica de Pelotas (UCPEL), Pelotas-RS, Brasil

{vaoliveira,adenauer,reiser}@inf.ufpel.edu.br

helida@furg.br

{giancarlo.lucca,lizandro.oliveira}@ucpel.edu.br

Abstract. *This study explores the potential of two lightweight convolutional neural networks, EfficientNet-B3 and DenseNet-169, for the diagnosis of Diabetic Retinopathy (DR), a microvascular complication of chronic hyperglycemia and a leading cause of preventable blindness among working-age adults. Threatening vision in nearly a million people who live with diabetes, timely and accurate detection is crucial. We analyze related theoretical concepts and metric evaluations, demonstrating the feasibility of compact CNN backbones for real-world deployment. All results are derived from the public EyePACS dataset, with the quadratic-weighted kappa metric used for evaluation.*

1. Introduction

Diabetic Retinopathy (DR) is the leading cause of preventable blindness among working-age adults, threatening vision in nearly one-third of the world's 540 million people who live with diabetes [International Diabetes Federation 2023]. When lesions are detected early, laser photocoagulation and anti-VEGF therapies can reduce the risk of severe vision loss by up to 90% [American Academy of Ophthalmology 2020], but their success hinges on the timely detection of subtle retinal lesions. Unfortunately, manual grading of colour fundus photographs is labour-intensive, prone to inter-grader variability, and difficult to scale, especially in low-resource settings.

Artificial-intelligence (AI) systems based on deep learning techniques have reshaped how data-rich industries generate value, automating tasks that once relied solely on expert judgment [Goodfellow et al. 2016]. In healthcare, computational intelligent systems already help clinicians diagnose earlier, triage patients faster, and tailor treatments to individual needs while easing the growing workload on medical teams [Ting et al. 2019a]. Vision-centric disciplines such as radiology, dermatology, and ophthalmology have benefited from such technologies because most diagnostic evidence is encoded in images.

Within deep learning, Convolutional Neural Networks (CNNs) hold a special place because their layered filters learn increasingly abstract visual patterns: early layers detect simple edges. In contrast, deeper layers capture complex shapes and even

disease-specific signatures, fully exploiting the spatial arrangement of pixels. Since the landmark success of AlexNet in 2012 [Krizhevsky et al. 2012], successive CNN variants have set performance records across academic benchmarks and industrial applications. In medicine, CNN pipelines now underpin certified screening tools for skin cancer [Esteva et al. 2017], pneumonia [Rajpurkar et al. 2017], and DR [Gulshan et al. 2016, Asif et al. 2025], often matching or exceeding human accuracy while operating at population scale. Recent work by [Gulshan et al. 2016] demonstrated that an Inception-v3 network can reach specialist-level performance, obtaining a quadratic-weighted κ (QWK) of 0.84 on the EyePACS dataset. Follow-up meta-analyses consistently report QWK scores above 0.80 across heterogeneous populations [Ting et al. 2019a]. Yet most state-of-the-art CNN backbones, which are the foundational architectures responsible for feature extraction, remain computationally demanding, restricting their deployment on point-of-care devices and in resource-limited clinics.

As the main objective, this work investigates whether compact CNN backbones can deliver state-of-the-art accuracy while remaining lightweight enough for real-world deployment. Specifically, we benchmark EfficientNet-B3 [Silva et al. 2021, Tan and Le 2019] and DenseNet-169 [Haque et al. 2022, Huang et al. 2017] on the EyePACS dataset [Kaggle 2015] using a progressive fine-tuning protocol.

As main results, this research promotes the following contributions:
(1) Study of the Benchmark EfficientNet-B3 and DenseNet-169 with progressive fine-tuning on EyePACS, reporting QWK and confusion-matrix metrics.
(2) Analyse training dynamics and discuss the impact of limited annotated data.
(3) Release a fully reproducible codebase[1] to facilitate future extensions and comparison.

This paper is organized as follows: Section 2 reviews background on DR and CNNs and Section 3 summarises related literature. In sequence, Section 4 details the Case Study on CNN based on Diabetic Retinopathy diseases, considering the dataset, preprocessing, and training protocol. Section 5 presents quantitative and qualitative results. Finally, Section 6 outlines our findings and future work.

2. Preliminaries

This section summarises the theoretical groundwork underlying our work. We begin with a brief overview of diabetic retinopathy (DR) and its clinical staging, followed by a concise review on CNNs, their training paradigm, and the evaluation metrics used throughout the paper.

2.1. Convolutional Neural Networks (CNNs)

CNN evolution considers the data extraction and classification steps seen in Figure 1. CNN architecture considers a family of *feed-forward* models that exploit the strong spatial locality of natural images. Instead of fully-connected layers, they use convolutional layers whose learnable kernels (filters) slide across the input while re-using the same weights – a mechanism known as *weight sharing*.

The formal expression of CNNs considers the input feature-map $\mathbf{x} \in \mathbb{R}^{H \times W}$, where H and W are its height and width, and the square kernel $\mathbf{w} \in \mathbb{R}^{(2k+1)\times(2k+1)}$ whose half-size is k (i.e., the kernel covers $(2k+1)\times(2k+1)$ pixels). The two-dimensional discrete

[1] https://github.com/Vitor-a-o/CNN_Rtinopatia_Diabetica

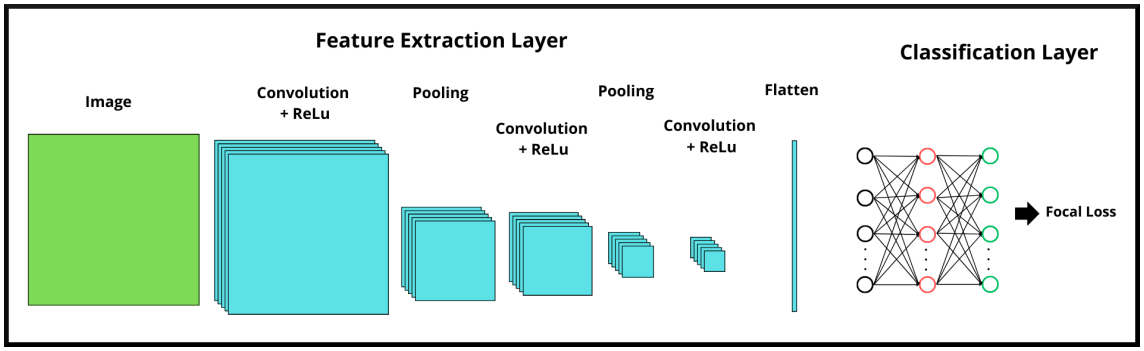

Figure 1. Convolutional Neural Network.

convolution evaluated at a pixel (u, v) is given as follows:

$$(\mathbf{w} * \mathbf{x})(u, v) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} w_{i,j}\, x_{u+i,v+j}, \tag{1}$$

The weight sharing and local connectivity yield three key properties: translation equivariance, parameter efficiency, and hierarchical feature learning. Modern blocks add ReLU, batch normalisation and spatial down-sampling; fully-connected classifiers are often replaced by global average pooling (GAP) plus soft-max.

Transfer learning and optimisation perform trained end-to-end stochastic descent gradient and its variants. Thus, given a dataset of images $\{(\mathbf{x}_n, y_n)\}_{n=1}^{N}$ and a differentiable loss $\mathcal{L}$ (e.g. cross-entropy or focal-loss), back-propagation updates all weights to minimise the empirical risk

$$\min_{\Theta} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\big(f_{\Theta}(\mathbf{x}_n),\, y_n\big). \tag{2}$$

Since large annotated medical datasets are rare, practitioners routinely *fine-tune* ImageNet-pretrained backbones. The early convolutional layers capture universal low-level patterns, which are frozen or updated with a smaller learning rate, while the final layers adapt to task-specific discriminative cues [Yosinski et al. 2014]. This strategy speeds up convergence, mitigates over-fitting, and lowers computational budget, as advantages for point-of-care deployments.

For evaluation, DR grading considers an ordered five-class problem and the quadratic-weighted Cohen's kappa (QWK) as the coefficient to measure the performance model, expressed as:

$$\text{QWK} = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}}, \tag{3}$$

where O_{ij} and E_{ij} are the observed and expected agreement matrices, respectively. And the parameters $w_{ij} = (i - j)^2/4$ penalise distant misclassifications. Additionally, we also inspect the confusion matrix to identify systematic errors across grades.

The above theoretical groundwork motivates the choice of CNN backbones.

3. Related Works

Deep learning has driven rapid progress in automated screening for DR, receiving relevant contributions. The leading research, related to the development of this study, is briefly described in this section.

The timeline starts with [Gulshan et al. 2016], whose Inception-v3 *ensemble* (multiple networks determining the final answer), reached specialist-grade performance on EyePACS (QWK = 0.84), based on the metric described in Sec. 4.4, and powered the first FDA-cleared device named IDx-DR.

The next wave scaled both data and model size, as Krause *et al.* trained EfficientNet-B7 on 1.6 M images, gaining sensitivity (recall of diseased eyes) at the cost of $\sim$40 M parameters [Krause et al. 2018]. Additionally, RSG-Net added *attention blocks*—layers that dynamically weigh input elements, allowing neural networks to focus on the most relevant information for predictions [Akhtar et al. 2023].

Another parallel line of study asks whether plain ImageNet backbones are enough. Mehboob *et al.* fine-tuned DenseNet-169, VGG-16 and Inception-v3, reporting macro-accuracy (mean per-class accuracy) around 78% on EyePACS [Mehboob et al. 2023]. The results of [Lin et al. 2020] improved preprocessing for EfficientNet-B4 and reached AUC = 0.926 (area under the ROC curve) across the five DR grades. Despite being useful baselines, most papers leave open how much of the backbone must actually be fine-tuned.

Recently, the focus has shifted to models that run in the clinic, or even on a phone. In [Siahkali and Bagherian 2023], a three-stage pipeline with only 3M parameters is built, achieving 85–88% accuracy. The results in [Li et al. 2024] compressed an EfficientNet variant into a smartphone app, delivering sensitivity above 90% with 25ms latency per image. These studies show the promise of tiny networks, but they rarely compare different backbones under the same training recipe.

Moreover, two late meta-analyses echo these findings, considering CNN screeners typically score QWK/AUC above 0.80 and outperform manual grading by 10–25 percentage points [Nair et al. 2023, Ting et al. 2019b]. Both reviews, however, call out small external datasets and fuzzy train/validation splits, which prevents a fair comparison.

4. Diabetic Retinopathy - A Case Study on CNN

This paper focuses on diabetic retinopathy, a microvascular complication of chronic hyperglycaemia that progresses through five clinical grades from 0 (no DR) to 4 (proliferative DR). The main characteristic lesions include micro-aneurysms, haemorrhages, and neovascularisation, which may appear anywhere on the fundus image and vary in size and contrast [International Diabetes Federation 2023]. Early detection is critical because timely laser photocoagulation or anti-VEGF therapy can prevent up to 90% of vision loss cases [American Academy of Ophthalmology 2020].

We run a head-to-head test of two compact backbones — EfficientNet-B3 (12 M parameters) and DenseNet-169 (14 M). Both performed on identical EyePACS splits. A progressive fine-tuning schedule logs intermediate QWK and loss after each stage, clarifying when extra layers start to help. The code and weights will be available at a public repository for transparent replication. The following sections detail the data, preprocessing pipeline, and training protocol used in our simulations.

4.1. Dataset

See Figure 2 illustrating variation in sharpness, illumination, and field-of-view resulting in the large intra-classes.

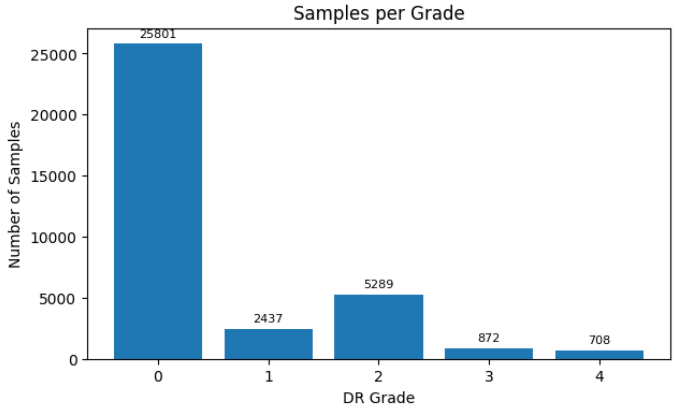

Figure 2. Class distribution.

All experiments rely on the publicly available dataset, named as *Kaggle Diabetic Retinopathy Detection* (EyePACS), considering 35,107 high-resolution colour fundus photographs labelled by certified ophthalmologists into five DR grades (0 = no DR, 4 = proliferative DR) [Kaggle 2015].

Additionally, a patient-wise stratified split assigns 26,330 images (75%) to the training set and 8,777 (25%) to the validation set, preserving the natural class imbalance as can be seen in the five levels of Figure 3, outlining the image-enhancement and augmentation steps applied before training.

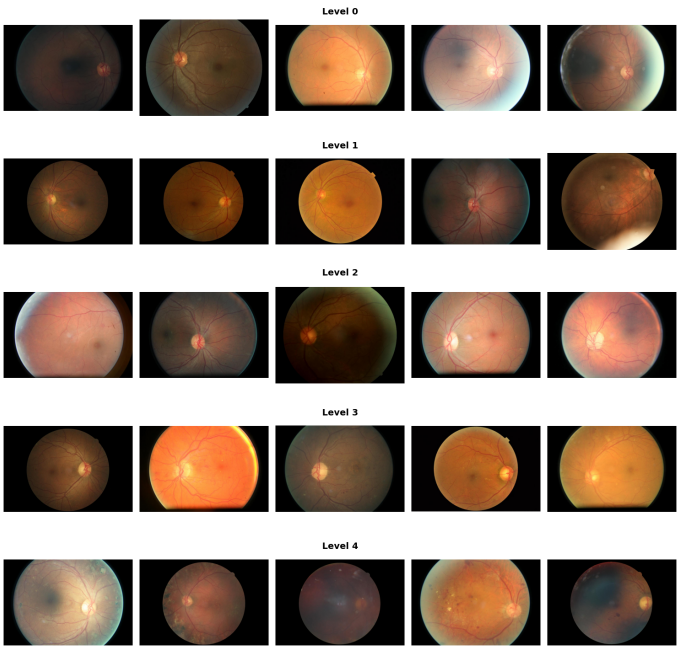

Figure 3. Grade 0–4 sample grid.

4.2. Data Preparation

We employ the Kaggle Diabetic Retinopathy Detection dataset (35,107 colour fundus photographs, five DR grades). The filename contains a patient identifier, allowing a *patient-wise*, stratified split that assigns 26,330 images (75%) to training and 8,777 (25%) to validation, thus preventing information leakage between fellow eyes.

Images are resized to each backbone's native input – 300×300 px for EfficientNet-B3 and 224×224 px for DenseNet-169 – and normalised with the corresponding ImageNet preprocessing routine.

Online augmentation related to random rotations, flips, crops, brightness and contrast jitter, enhances robustness to the variations which are illustrated in Figure 3.

4.3. Model Training

Now, the fine-tuning strategy, hyperparameters, and hardware configuration adopted for the CNN backbones are discussed:

(i) **Network Head**: Both backbones, retrained on ImageNet, are topped with a global-average pool, a dropout layer ($p = 0.2$), and a five-unit soft-max classifier.

(ii) **Progressive Fine-tuning Schedule**: Rather than running a single monolithic training session, we adopt a three-cycle progressive protocol to obtain meaningful intermediate metrics. After every warm-up (*Init.*) and fine-tuning (*FT*) stage, QWK and loss are reported, allowing to observe successive layers contributing to performance:

- **Cycle 1:** 5 epochs Init. + 5 epochs FT
- **Cycle 2:** 10 epochs Init. + 10 epochs FT
- **Cycle 3:** 10 epochs Init. + 10 epochs FT

In total, each backbone undergoes 25 warm-up and 25 fine-tuning epochs, with checkpoints saved at the end of every phase for later analysis.

(iii) **Optimisation**: During initiation, the stack's head is optimised with Adam (lr = 10^{-4}); during FT, lr = 10^{-5} and only the last 20 convolutional layers are trainable. Class imbalance is handled with inverse-frequency class weights and focal loss ($\alpha = 0.25$, $\gamma = 2$) [Lin et al. 2017]. Keras callbacks (`ModelCheckpoint`, `EarlyStopping`, `ReduceLROnPlateau`) safeguard the best weights.

4.4. Evaluation Metric and Hardware Structure

Performance is quantified with the quadratic-weighted Cohen's Kappa (QWK), the official metric adopted by Kaggle for diabetic-retinopathy grading.

During inference, predictions are generated in mini-batches to limit the memory footprint on an NVIDIA GeForce GTX TITAN X GPU. Moreover, training curves and confusion matrices are automatically exported, and every artefact – model weights, logs, and figures – is versioned, guaranteeing exact reproducibility.

5. Main Results

As described in Sections 4.2 and 4.3, the online augmentation related to random rotations, flips, crops, brightness, and contrast jitter, enhances robustness to the variations which are

illustrated in Figure 3. The validation set resulting from each training cycle is summarized in Table 1, listing the quadratic-weighted Cohen's κ (QWK). Observing that higher values indicate closer agreement with the ophthalmologist's ground truth.

Table 1. Cumulative validation QWK.

Model	Cycle 1	Cycle 2	Cycle 3
EfficientNet-B3	0.393	0.432	**0.434**
DenseNet-169	0.349	0.389	0.391

- By **Early learning**, both networks gain the bulk of their performance within the first two cycles; the third adds only ≈ 0.002 QWK.
- **EfficientNet-B3 remains superior**, the application outperforms DenseNet-169 at every stage and finishes 0.043 QWK higher, despite similar relative improvements ($+10$–$12\,\%$).

Comparison with prior work. To contextualize the QWK results in Table 1, Table 2 summarizes representative related studies and, in bold, our best validation results. Because studies report different metrics (QWK, AUC, Accuracy), we keep the "Metric" column for transparency.

Table 2. Comparison with related work (metric and value).

Work	Metric	Value
Gulshan et al. (2016)	QWK	0.84
Lin et al. (2020)	AUC	0.926
Siahkali & Bagherian (2023)	Accuracy	85–88%
This work (EfficientNet-B3)	**QWK**	**0.434**
This work (DenseNet-169)	**QWK**	**0.391**

Regarding bias towards the majority class, let C be the confusion matrices pictured in Fig. 4 and Fig. 5, related to EfficientNet-B3 and DenseNet-169, respectively.

Firstly, both matrices reveal a pronounced tilt toward predicting the majority label, grade 0, which accounts for 74% of the dataset. DenseNet-169 shows a markedly more substantial effect: it places over 92% of its outputs in the first column and misroutes grade 2 to grade 0 nearly twice as often as it correctly identifies it. So, the effect is markedly more substantial in DenseNet-169: over 92% of its outputs land in the first column, while grade 2 is misrouted to grade 0 nearly twice as often as correctly identified. However, EfficientNet-B3 still favours the healthy class, but it spreads its predictions more evenly, reclaiming a noticeably higher share of true grade 2 and grade 3 cases. And, EfficientNet earns the higher QWK reported in Table 1 by softening this majority-class bias.

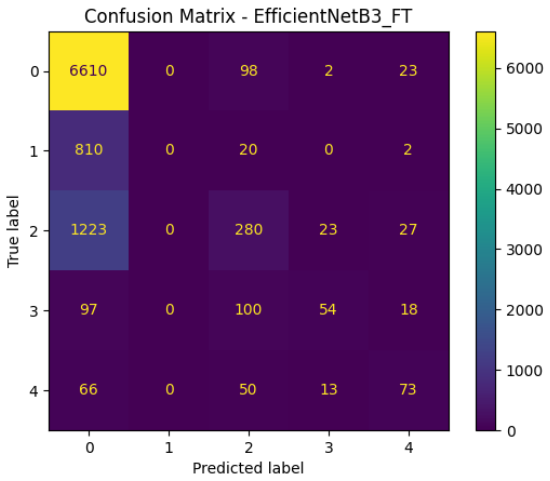

Figure 4. EfficientNet-B3 C-matrix.

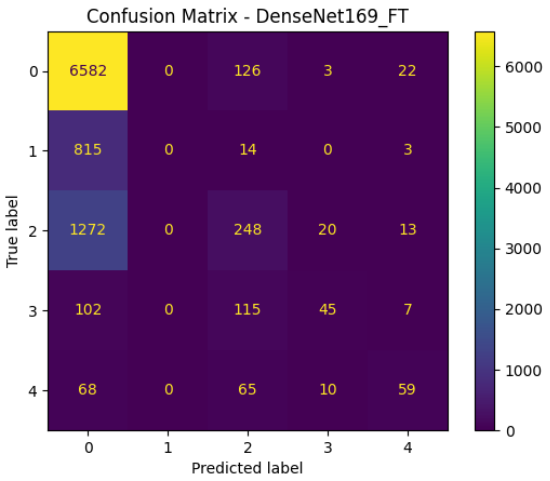

Figure 5. DenseNet-169 C-matrix.

6. Conclusion

This study showed that two lightweight convolutional networks, EfficientNet-B3 and DenseNet-169, can reliably grade diabetic-retinopathy images when they are fine-tuned for the task and supported by careful pre-processing. With only moderate training time, EfficientNet-B3 reached a quadratic-weighted kappa of 0.434 – slightly higher than DenseNet-169 – while both models kept most mistakes within neighbouring severity levels. This pattern also appears in human readings.

Our code keeps every experiment version-controlled, logs all hyperparameters, and splits images by *patient*, reducing the risk of information leakage. Consequently, other researchers can readily replicate the tests or substitute components of the pipeline. Since all results come from the public EyePACS dataset, the single approach is strongly unbalanced and may not reflect other cameras or populations. We did not add any special, explainable ability tools beyond simple attention maps, and we have not yet checked how well the models work on truly unseen data.

Next steps include testing newer backbones (e.g. EfficientNetV2-S or ConvNeXt-Tiny), adding self-supervised or semi-supervised pre-training to use unlabelled images, and combining multiple models for higher stability. We also plan to validate the pipeline on external datasets and link saliency maps with expert feedback, making the system more transparent for clinical use.

In short, well-designed training and evaluation let classic CNNs deliver solid performance for diabetic-retinopathy screening, while leaving room for modern architectures and larger training schemes to push results even further.

Acknowledgements

This work was partially supported by the following Brazilian funding agencies: CNPq (309559/2022-7, 409696/2022-6), FAPERGS (21/2551-0002057-1, 24/2551-0000631-1, 24/2551-0001396-2), and FAPERGS/CNPq (23/2551-0000126-8).

References

Akhtar, S., Aftab, S., Kousar, S., and Saeed, A. Q. (2023). Rsg-net: A deep neural network for diabetic retinopathy severity grading. *Computers in Biology and Medicine*, 161:107081.

American Academy of Ophthalmology (2020). Diabetic retinopathy preferred practice pattern. `https://www.aao.org/ppp`. Accessed 4 May 2025.

Asif, M., Ur Rehman, F., Rashid, Z., Hussain, A., Mirza, A., and Qureshi, W. S. (2025). An insight on the timely diagnosis of diabetic retinopathy using traditional and ai-driven approaches. *IEEE Access*, 13:116869–116886.

Esteva, A., Kuprel, B., Novoa, R. A., and et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115–118.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Gulshan, V., Peng, L., Coram, M., and et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410.

Haque, M., Winston, K., and et al. (2022). Densenet-169 ensemble for robust diabetic retinopathy detection. In *IEEE International Conference on Healthcare Informatics*.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

International Diabetes Federation (2023). Idf diabetes atlas, 10th edition. `https://diabetesatlas.org`. Accessed 4 May 2025.

Kaggle (2015). Diabetic retinopathy detection dataset. `https://www.kaggle.com/c/diabetic-retinopathy-detection`. Accessed 3 May 2025.

Krause, J., Gulshan, V., Rahimy, E., Widner, K., and et al. (2018). Grader variability and the importance of reference standards for evaluating machine-learning models for diabetic retinopathy. *Ophthalmology*, 125(8):1264–1272.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.

Li, J., Wang, Y., and Chen, R. (2024). On-device diabetic retinopathy screening with a mobile efficientnet model. *IEEE Journal of Biomedical and Health Informatics*, 28(4):1553–1562.

Lin, Q., Zhang, X., and Chen, Y. (2020). Diabetic retinopathy grading with efficientnet and enhanced pre-processing. *Biomedical Signal Processing and Control*, 62:102123.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.

Mehboob, M., Raza, M. A., and Shahzad, H. (2023). Diabetic retinopathy detection using densenet-169 with transfer learning. In *International Conference on Intelligent Systems*, pages 343–350.

Nair, R., Phene, S., and Krause, J. (2023). Deep learning for diabetic retinopathy screening: A systematic review and meta-analysis. *Ophthalmology Retina*, 7(2):97–109.

Rajpurkar, P., Irvin, J., Zhu, K., and et al. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.

Siahkali, M. and Bagherian, L. (2023). A lightweight multi-stage framework for smartphone-based diabetic retinopathy screening. *Frontiers in Medicine*, 10:1184756.

Silva, R., Pereira, P., and Mendonça, L. (2021). Efficientnet-b3 for highly accurate diabetic retinopathy grading. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*.

Ting, D. S. W., Pasquale, L. R., Peng, L., and et al. (2019a). Artificial intelligence and deep learning in ophthalmology. *Nature Medicine*, 25:14–24.

Ting, D. S. W., Pasquale, L. R., Peng, L., and et al. (2019b). Artificial intelligence for diabetic retinopathy screening: A meta-analysis. *British Journal of Ophthalmology*, 103(10):167–175.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, pages 3320–3328.