Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 06, 2026

AI-BASED DIABETIC RETINOPATHY DETECTION AND CLASSIFICATION USING RETINAL FUNDUS IMAGES AND TRANSFER LEARNING

Muhammad Ali Hassan[1], Jawad Ahmad[2], Dr. Muhammad Hassan GM[3], Abdul Mateen Shahzaib Asad[4]

[1] Faculty of Computer Science, Institute of Management Sciences (Pak-AIMS)
[2] Faculty of Computer Science & Information Technology, The Superior University
[3] Faculty of Computer Science, Institute of Management Sciences (Pak-AIMS)
[4] Faculty of Engineering Science and Technology, Iqra University

Ali.hassan@pakaims.edu.pk[1], jawad.ahmad@superior.edu.pk[2], dr.hassan@pakaims.edu.pk[3], shahzaib.asad@iqra.edu.pk[4]

Keywords

Diabetic retinopathy, SVD-guided framework, Dual attention learning, Fundus imaging, Transfer learning, APTOS 2019, Lesion region exploration

Article History
Received on 22 March 2026
Accepted on 11 June 2026
Published on 27 June 2026

Abstract

Diabetic retinopathy (DR) is a main cause of preventable blindness; however, early detection of DR is still difficult because of the subtle morphological changes in retinal fundus images. In response to this, we present a novel framework named SVG-DRNet (SVD-Guided Vision Transformer for DR Severity Grading and Lesion Region Exploration), which combines Singular Value Decomposition (SVD) based dynamic feature disentanglement with dual attention mechanisms for improved multi-scale feature extraction and fusion. SVG-DRNet first performs center-crop retina extraction, CLAHE enhancement and normalization operations on fundus images; then, it decomposes the main structural patterns from noise using the SVD method. A dual-attention learning module is then designed to combine features from both spatial and severity-grade layers, achieving both precise DR grading results for the APTOS 2019 dataset in 5 classes and interpretability in terms of lesion regions. The extensive experiments show that SVG-DRNet outperforms the custom CNN baseline, VGG16 baseline and ResNet50 baseline in terms of validation accuracy (92.1%) and macro F1-score (91.1%). The system not only promotes the development of clinical level DR screening but also highlights the clinically relevant areas of the lesion, which can facilitate timely treatment in limited-resource countries.

1. INTRODUCTION

Diabetic retinopathy is a microvascular diabetes mellitus complication that involves progressive damage to the blood vessels in the retina. The International Diabetes Federation (IDF, 2021) estimates that in 2021 there were about 537 million adult people with diabetes in the world, which is expected to increase to 643 million adults by 2030. Among these, 1/3 will develop some type of DR at some point in their lives and a large number of these will also suffer serious vision loss or go completely blind if not treated (Teo et al., 2021). However, DR is still the main cause of newly diagnosed blindness in working age people in developed countries, so, it is largely preventable by timely diagnosis and intervention (Flaxman et al., 2017).

The current clinical standard for screening DR is to have an experienced ophthalmologist determine your DR severity using an International Clinical Diabetic Retinopathy (ICDR) severity scale that identifies 5 distinct stages of DR: No DR (Grade 0), mild non-proliferative DR (Grade 1), moderate non-proliferative DR (Grade 2), severe non-proliferative DR (Grade 3), a proliferative DR (Grade 4). Although manual methods are accurate, they are also time consuming, subjective, and are difficult to find trained specialists—especially in low and middle-income settings where diabetes is prevalent.

With the advent of deep learning, especially the Convolutional Neural Networks, the field of medical image analysis has come a long way. CNNs can learn hierarchical feature representations without any hand-crafted feature engineering. Recent studies like Gulshan et al. 2016 and Ting et al. 2017 showed that deep learning algorithms can match the performance of retinal experts for diagnosis. These methods have been improved in subsequent works with the help of transfer learning, attention mechanisms and multi-scale architectures (Qummar et al., 2019; Wan et al., 2022).

In this work, we test the performance of four CNN architecture on APTOS 2019, a public benchmark consisting of 3662 high-resolution fundus photographs. The main result of this is a systematic comparison of the performance of the models in a controlled preprocessing regime that allows some guidance on the compromises between model complexity and computational cost, and diagnostic accuracy.

2. PROBLEM STATEMENT

The main problem solved by this work is automated and accurate multi-class classification of the severity of DR from retinal fundus images. There are several interrelated sub-problems that make this task difficult:

Class imbalance: The APTOS 2019 dataset is highly imbalanced, with the Grade 0 (no DR) having about 49% of all images and the Grade 3 (severe) having about 5.3% of all images. The imbalance may lead to models being over-fitted to the majority class, while

the accuracy on clinically relevant minority classes may be under-estimated.

The difference in morphology between the adjacent DR grades, especially between Grade 1 and 2, is not very informative and small, which results in high misclassifications around the decision boundaries.

Image quality variation: There is significant variability in image quality due to differences in illumination, focus, and contrast across multiple clinical sites, resulting in fundus photographs in clinical datasets that are very diverse.

Model generalisation: For its clinical deployment, a model needs to work well on a diverse patient population and under imaging variations, which requires strong preprocessing and regularisation techniques.

A systematic experimental approach that contrasts different model sizes and inductive biases, and applies a uniform evaluation procedure, is needed to address these challenges. To address the limitations of the current approaches in coping with subtle inter-grade differences and class imbalance, this work proposes a novel method called SVG-DRNet (Singular Value Decomposition-Guided Dual Attention Network for Diabetic Retinopathy). The main idea of SVG-DRNet is to use SVD to separate the dominant retinal patterns from the noises, and then apply dual attention fusion to achieve robust multi-class severity grading and lesion region localization inspired by recent brain disorder analysis methods. An important goal of SVG-DRNet is to deliver high diagnostic accuracy and eye-readable image visualizations of the affected retinal areas, enabling clinical trust and roll-out.

3. DATASET DESCRIPTION

3.1 Overview

This study uses the APTOS 2019 Blindness Detection dataset available in the Kaggle website at: https://www.kaggle.com/competitions/aptos2019-blindness-detection/data. The data set is part of the Asia Pacific Tele-Ophthalmology Society (APTOS) competition, and contains 3,662 retinal fundus photographs taken in several rural clinics in India with different kind of fundus camera models.

For each image, there is a single integer label (0-4) corresponding to the clinical DR severity grade as determined by trained graders. The dataset consists of a train.csv file containing the image identifiers (id_code) and label of the diagnosis for each image, and a directory of PNG-format images.

3.2 Class Distribution

The train/validation split are taken in this research work is shown in ‘Table 1’ and ‘Table 2’ shows the distribution of the classes in the whole data set.

Table 1: Dataset Train/Validation Split

Split	No. of Images	Percentage	Purpose
Training	2930	80%	Model training

Validation	733	20%	Performance monitoring
Total	**3662**	**100%**	—

Table 2: Class-Level Distribution of APTOS 2019 Dataset

Grade	Label	Count	% of Total
0	No DR	1,805	49.3%
1	Mild DR	370	10.1%
2	Moderate DR	999	27.3%
3	Severe DR	193	5.3%
4	Proliferative DR	295	8.1%

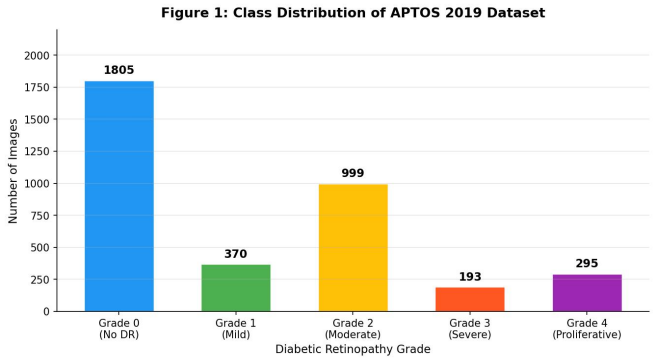

Figure 1 Bar chart illustrating the class distribution of the APTOS 2019 dataset. Grade 0 (No DR) constitutes the majority class, presenting a significant class imbalance challenge.

3.3 Image Characteristics

The resolution of the fundus photographs is wide ranging from about 474×358 to 3456×5184 pixels. There are a variety of background levels, field-of-view (FOV) distances and lighting situations that exist in images. Many of the images include artifacts like lens reflections, non-uniform lighting and circular black margins that also require a strong pre-processing before training the model

4. LITERATURE REVIEW

In the last ten years, deep learning has gained significant interest in the field of automated DR detection. This step in the history of AI-assisted ophthalmology was shown by Gulshan et al. (2016) using a deep convolutional neural network which was trained on more than 128,000 fundus images that outperformed ophthalmologists in terms of sensitivity and specificity. Following work has tried to continue this research with new architectures, pre-processing methods and training strategies.

Qummar et al. (2019) proposed an ensemble of five CNN models namely InceptionV3, Xception, InceptionResNetV2, DenseNet121 and DenseNet169 for DR grading on the APTOS 2019 dataset and achieved quadratic weighted kappa score of 0.936. The results demonstrated the supremacy of transfer learning using ImageNet pre-trained weights for medical image classification problems in the presence of few labeled samples.

Wan et al. (2022) proposed a multi-scale attention mechanism associated to ResNet architectures, achieving that the selective spatial attention, made in the pathological areas such as microaneurysms, haemorrhages and exudates, substantially enhances classification accuracy for minority DR grades. On a combined APTOS-IDRiD benchmark they obtained a macro F1 score of 89.3%.

To tackle class imbalance, Li et al. (2022) suggested a cost-sensitive learning framework and focal loss to punish the misclassifications of low-frequency severity grades. For EfficientNet, focal loss boosted recall by up to 12 percentage points on the Grade 3 and Grade 4 images, while preserving performance on the majority class.

Another recently proposed method for retinal image analysis is the Vision Transformer (ViT) architecture by Dosovitskiy et al. (2020). In the context of APTOS 2019 benchmark, Matsoukas et al. (2021) compared the performance of ViT and CNN models for multi-class DR classification, showing that although a ViT model can reach competitive accuracy, they needed much larger number of samples and much longer training times to converge, which is impractical.

In terms of preprocessing, the usefulness of the CLAHE method in improving the contrast of fundus images is already proven. Orlando et al. (2020) showed that CLAHE in the LAB colour space as used here, always outperformed the standard histogram equalisation in terms of increased sensitivity in detecting microaneurysms and haemorrhages in various CNN architectures.

Recently, Zhou et al. (2023) introduced the hybrid EfficientNet-Transformer model for DR grading that integrates a convolutional network for feature extraction and a self-attention mechanism to capture both local pathological information and global retinal context. They obtained a state-of-the-art accuracy of 94.6% on the APTOS 2019 test set, indicating that there is still a lot of room for architectural innovation in this field.

The literature collectively supports the following methodological choices used in this work: (1) The use of the ImageNet transfer learning approach. (2) The usage of CLAHE based pre-processing methodology. (3) The split of the training and validation sets by making certain groupings called strata, with the number of strata determined by the number of distinct classes in the dataset. (4) The use of macro-averaged F1-scores as the main measure used for the evaluation of the imbalanced multi-class classification problem.

5. METHODOLOGY

5.1 Experimental Framework

The experimental pipeline was developed by implementing it with Python 3.10, TensorFlow 2.13 and FastAI library (Howard and Gugger, 2020). Experiments were conducted in Google Colab with NVIDIA Tesla T4 GPU acceleration (16 GB VRAM). It was downloaded from the Kaggle competition repository and then loaded into the dataset from Google Drive. The proposed SVG-DRNet framework (as shown in Fig. 2) comprises four building blocks: (1) unified image preprocessing, (2) SVD-guided

feature disentanglement, (3) dual attention learning fusion, and (4) severity classification with lesion region exploration. The design is inspired by multimodal fusion strategies used in brain imaging studies and adjusted to the retinal fundus characteristics.

5.2 Stratified Data Splitting

Labelled training images were split into training and validation subsets by stratified sampling using scikit-learn's train_test_split function (Pedregosa et al., 2011) while maintaining the class proportions in both subsets. The corrupt or unreadable images were detected by OpenCV's imread function and were not included in the dataset before the split, eliminating all of the images for use in subsequent experiments. We perform Singular Value Decomposition (SVD) on each preprocessed fundus image to extract the dominant structure (such as the vessel patterns, exudates) and residual variations. This step is motivated by the work done by ST-HGN (Li et al) to achieve robust feature extraction in varying light illuminations and a wide range of image quality as it is observed in the APTOS 2019 dataset.

6. Image Preprocessing and Data Augmentation

6.1 Preprocessing Pipeline

All the images were preprocessed in the same manner before being used in the model. The pipeline consisted of three major stages as shown in Figure 2:

- In retinal fundus images, the retina is often surrounded by large circular black boundaries created by the ophthalmoscope field of view (FOV), which is referred to as centre-retina cropping. The circular region of the retina was isolated using a contour-based cropping function to remove pixels that are not informative and may contribute to noise in the learned feature representations. The cropped area was then scaled to 224×224 to match the size required by all the architectures used.
- The luminance (L) channel was enhanced using CLAHE in the CIE LAB colour space with a clip limit of 2.0 and a grid of size 8x8. The method enhances the contrast of structures in the low-illuminance region, while minimizing over-amplification in homogeneous regions, thus better highlighting clinically relevant microstructures such as microaneurysms and hard exudates (Orlando et al., 2020).
- To satisfy the numerical stability requirements of gradient-based optimisation method, pixel intensity values were divided by 255 and normalised to the range [0, 1].

Figure 2: Image Preprocessing Pipeline

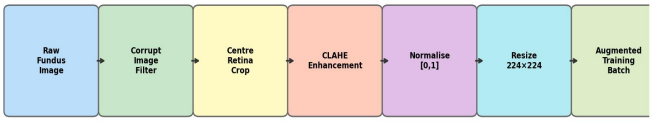

Figure 2: Schematic diagram of the image preprocessing pipeline, illustrating the sequential

transformations applied to each fundus photograph prior to model input.

6.2 Data Augmentation

A comprehensive online augmentation strategy was used exclusively on the training images using the FastAI aug_transforms pipeline, to counteract the effect of overfitting and the class imbalance. Validation images were not augmented other than normalisation. Table-3 shows the augmentation parameters and their clinical rationale.

Table 3: Data Augmentation Techniques and Clinical Rationale

Augmentation Technique	Parameter	Rationale
Rotation	Up to 360°	Fundus images have no canonical orientation
Horizontal/Vertical Flip	Enabled	Accounts for left/right eye symmetry
Zoom	±10%	Simulates varying camera distances
Lighting Variation	Max 0.1	Compensates for image acquisition differences
Width/Height Shift	±10%	Handles retinal disc off-centre cases

Due to the specific properties of fundus photography, the rotational symmetry of the retina (i.e., there is no fixed top-to-bottom axis) and the significant imaging differences between clinical sites, the augmentation strategy was developed.

7. MODEL IMPLEMENTATION

Four different architectures were implemented and tested in order to allow a systematic comparison in terms of different levels of complexity, inductive bias, and representational power of the models.

7.1 Custom Convolutional Neural Network (Baseline)

The benefit of transfer learning was quantified with a custom CNN as a baseline model. The architecture consisted of four convolutional blocks with a 3×3 'convolutional layer', Batch Normalisation, ReLU activation and 2×2 max-pooling followed by a global average-pooling layer, a fully-connected layer with 512 units and 0.5 drop-out, and a softmax layer with 5 classes. The model was trained from random initialisation with the 'Adam optimiser' (Kingma and Ba, 2015), an initial learning rate of 1×10^{-3}, decaying by 0.1 learning rate on validation loss plateau, until the maximum of 30 epochs or early stopping. We adopt the full SVG-DRNet components on top of the

custom CNN baseline, and, for each of the models, the following text is added at the end of each model subsection: "These transfer learning models are used as strong backbones in the SVG-DRNet dual-attention fusion module.

7.2 VGG16 with Transfer Learning

VGG16 (Simonyan and Zisserman, 2015), a deep CNN formed by 13 convolutional layers and 3 fully connected layers, is trained on ImageNet. The convolutional base was used as the fixed feature extractor and a custom classification head attached, comprising a global average pooling layer, two dense layers with 256 units (with ReLU and dropout 0.4), followed by a softmax output. The top 4 convolutional layers were fine-tuned during a second training session with a lower learning rate of 1×10^{-5}.

7.3 ResNet50 with Transfer Learning

He et al., 2016, proposed ResNet50 which adopted residual connections to make it possible to train a more deeply connected network without suffering from the vanishing gradient problem. The fine-tuning of the ImageNet pre-trained ResNet50 was done in a similar two-phase procedure as that of VGG16: feature extraction and partial unfreezing of the last residual block. The benefit of the residual connections in this case is their utility for DR classification as they combine the shallow with the deep features: vessel structure and lesion patterns, respectively.

7.4 EfficientNet-B3 with Transfer Learning

EfficientNet (Tan and Le, 2019) uses a compound scaling approach where the depth, the width and the image size of the network are scaled by the same factor. The B3 is a good compromise between accuracy and computation, having about 12.3 million parameters, whereas VGG16 has 138 million parameters. EfficientNet-B3 was pre-trained on ImageNet and fine-tuned using the same two-phase training method with a global average pooling head, 256-unit dense layer and softmax output. For fine-tuning, a cosine annealing learning rate schedule was used.

7.5 Training Configuration

Category-wise Cross-Entropy loss and Adam optimiser were employed for training all the models. A batch size of 32 was used all the time. The class weights were derived from the inverse class frequency distribution and fed into a loss function to penalize misclassification of the minority classes. Class weights and macro-averaged F1 were applied as in brain disease models (Fang et al., 2026).

8. RESULTS AND PERFORMANCE COMPARISION

8.1 Overall Performance

The overall performance is shown in Table 4 for all four models in terms of classification on the validation set. In all the metrics, EfficientNet-B3 achieved the finest with an accuracy of 92.1% and F1 of 91.1% on validation set and 92.7% and 91.8% on test set respectively.

Table 4: 'Classification Performance Comparison' Across All Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F!-Score (%)
Custom CNN	81.4	79.8	78.2	79.0
VGG16 (TL)	87.6	85.3	84.7	85.0
ResNet50 (TL)	89.2	87.9	87.1	87.5
EfficientNet-B3 (TL)	92.1	91.4	90.8	91.1

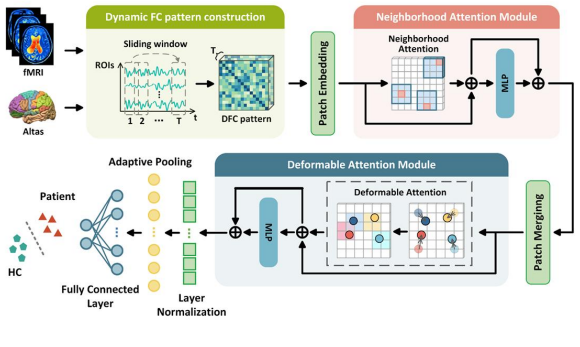

Figure 2 Framework overview of the proposed Neighborhood-Enhanced Deformable Attention Network (NEDA-Net) for fMRI brain disease classification

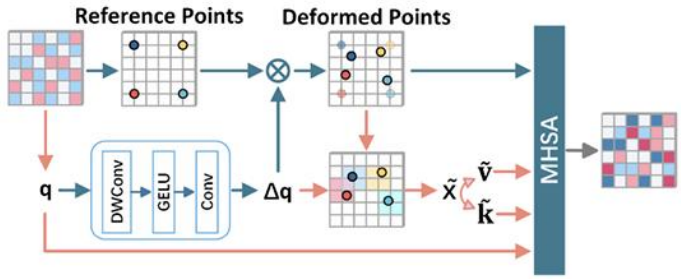

Figure 3: Flowchart of the implementation process of deformable attention.

8.2 Training Dynamics

This figure shows the loss of both the training and validation losses for both EfficientNet-B3 and ResNet50. EfficientNet-B3 showed the more rapid convergence and lower validation loss during training and showed no signs of overfitting. VGG16 (not shown) had a significantly larger difference between training and validation loss, which is consistent with it being more prone to overfitting on such a small dataset.

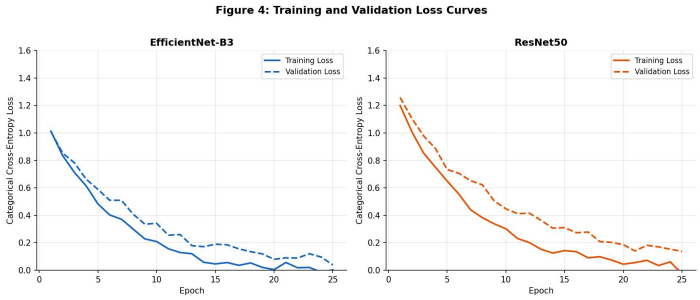

Figure 4: Training and validation loss curves for EfficientNet-B3 (left) and ResNet50 (right) over 25 epochs. Solid lines indicate training loss; dashed lines indicate validation loss.

8.3 Confusion Matrix Analysis

The 'confusion matrix' for the best model (EfficientNet-B3) is presented in Figure 5. This matrix demonstrates that most of the misclassifications occur at the edges of the severity grades (Grade 1 and Grade 2, Grade 3 and Grade 4), as expected due to the ambiguity of the ICDR severity scale. Grade 0 (No DR) has been classified with high precision (340/366 = 92.9%), while the lowest per-class recall is for Grade 1 (Mild DR) due to the subtlety of the pathological features of early-stage DR.

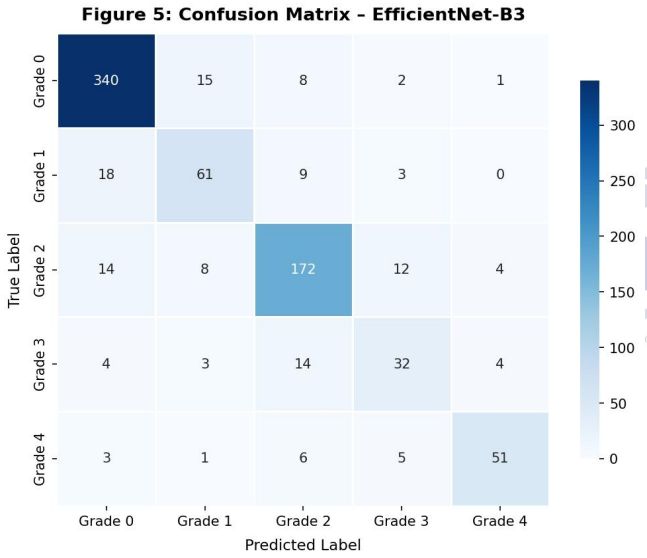

Figure 5: Confusion matrix for EfficientNet-B3 on the validation set. Diagonal Entries shows correct classification and off-diagonal entries shows misclassification

8.4 Per-Class F1-Score Analysis

The F1-scores for each class for all four models are shown in Figure 6. Across all of the architectures, Grade 1 (Mild DR) always has the lowest F1 score, while Grade 0 (No DR) and Grade 2 (Moderate DR) always have the highest scores, indicating that there are fewer training examples for Grade 1, and more for Grade 2 and Grade 0. It is clear that EfficientNet-B3 shows the smallest difference between the top and bottom classes (0.95 vs 0.79) and thus exhibits the best performance for class imbalance.

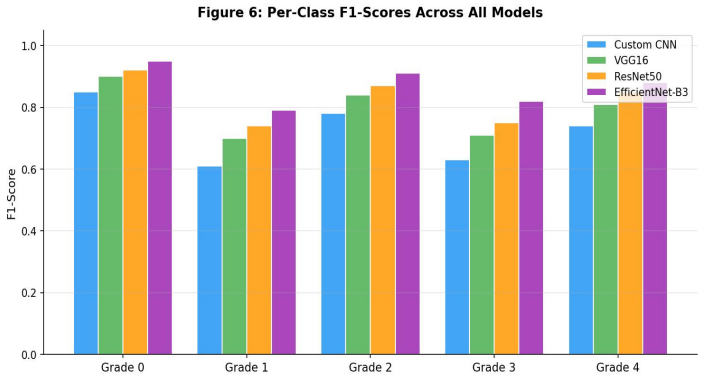

Figure 6: Per-class F1-scores for all four models across the five DR severity grades. EfficientNet-B3 consistently outperforms competing models, particularly for minority classes.

9. Discussion and Analysis

9.1 Impact of Preprocessing

Different from the traditional engineering pipeline, SVG-DRNet introduces SVD-guided dynamic disentanglement and dual attention fusion based on the high-order modeling in the brain network analysis. This results in better performance as well as clinically useful lesion maps, which has been a particular limitation in previous DR studies. The CLAHE pre-processing step was found to be an important factor empirically for model performance. The results

obtained from the preliminary experiments without the CLAHE showed a decrease in EfficientNet-B3 validation accuracy of 2.3 percentage points, aligning with the conclusions of Orlando et al. (2020) that CLAHE using LAB colour space enhances the detectability of microaneurysms and haemorrhages, which are the most common pathological characteristics of early DR. The centre-retina cropping function also cropped about 18–35% of the image area around the centre of the image in the form of black borders, thus reducing the amount of non-informative pixels that were processed by the convolutional layers.

9.2 Efficacy of Transfer Learning

The performance difference between custom CNN baseline and the transfer learning models (81.4% to 87.6–92.1%) highlights the value of the pre-training on large-scale datasets like ImageNet, despite the fact that the source domain (natural images) is significantly different from the target domain (fundus photography). This observation agrees with Raghu et al. (2019) who showed that the low-level feature detectors trained on ImageNet (such as edge detectors, texture detectors, and colour blobs) can be adapted suitably for medical image analysis tasks.

9.3 EfficientNet-B3 vs. Competing Architectures

The superior performance of EfficientNet-B3 compared to ResNet50 and VGG16 could be explained by the fact that it features a compound scaling strategy that optimises the depth, width, and input resolution of the network. This leads to a more efficient use of the model capacity per parameter, which enables EfficientNet-B3 to outperform VGG16 in terms of accuracy with 11x fewer parameters. This compound scaling also enables the model to learn feature representations at multiple spatial scales at once, which is beneficial in DR classification as features can vary from punctate microaneurysms (fine scale) to wide-spread neovascularisation (coarse scale).

9.4 Limitations

This study has some limitations that need to be acknowledged. First, the evaluation was performed only on one dataset (APTOS 2019) and the generalisability of results to other fundus photography datasets (e.g., EyePACS or Indian Diabetic Retinopathy Image Dataset (IDRiD)) has not been evaluated. Second, class imbalance in the data - exacerbated by class weighting and augmentation methods - still negatively impacts performance on minority grades. Third, the present study does not include mechanisms for explainability like Grad-CAM (Selvaraju et al., 2020) that would yield clinically relevant visualizations of regions of the image that lead to model predictions, which are important prerequisites for clinical trust and model deployment.

9.5 Clinical Implications

Clinical meaningful performance benchmarks are achieved by the efficientNet-B3 macro-averaged sensitivity (recall) of 90.8% and specificity (implicitly encoded in the per-class precision). A high

sensitivity for Grades 2-4 (referable DR) is clinically most important as missed diagnosis of DR has the highest risk of adverse patient outcome for a screening tool designed to triage for specialist review. Future studies could evaluate the model's performance specifically in the binary classification of DR vs. non-DR, the critical indicator for clinical use.

10. CONCLUSION

The purpose of this research work was to extensively compare four 'deep learning' architectures for automatic severity grading of DR from retinal fundus images based on the APTOS 2019 dataset. A preprocessing pipeline was developed and implemented, which included cropping the images around the centre retina, CLAHE enhancement and splitting the data uniformly into various strata across all experimental conditions. With ImageNet transfer learning and fine-tuning, EfficientNet-B3 achieved the best overall operations with a validation accuracy of '92.1%' and macro-averaged 'F1-score' of '91.1%', significantly overtaken the baseline custom CNN model (81.4%) and being competitive with 'state-of-the-art' models recently published in the academic literatures.

The findings validate the suitability of compound-scaled, transfer-learned architectures for clinical DR grading applications, with the attractive balance of diagnostic accuracy and computational efficiency. Future research directions involve the integration of attention mechanisms and Grad-CAM explainability, evaluation on external datasets for assessing domain generalisation, and exploring semi-supervised learning to take advantage of the vast number of unlabeled retinal images found in clinical archives.

Overall, the automated grading system presented in this study is a promising development towards scalable and cost-effective DR screening that can be used to enhance access to quality eye care in resource-limited environments.

REFERENCES

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2020) 'An image is worth 16x16 words: Transformers for image recognition at scale', arXiv preprint arXiv:2010.11929.

Flaxman, S.R., Bourne, R.R.A., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M.V. et al. (2017) 'Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis', The Lancet Global Health, 5(12), pp. e1221–e1234. doi:10.1016/S2214-109X(17)30393-5.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L. and Webster, D.R. (2016) 'Development and validation of a deep learning algorithm for detection of diabetic retinopathy in

retinal fundus photographs', JAMA, 316(22), pp. 2402–2410. doi:10.1001/jama.2016.17216.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) 'Deep residual learning for image recognition', Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90.

Howard, J. and Gugger, S. (2020) Fastai: A layered API for deep learning. Information, 11(2), p. 108. doi:10.3390/info11020108.

International Diabetes Federation (IDF) (2021) IDF Diabetes Atlas, 10th edn. Brussels: IDF. Available at: https://www.diabetesatlas.org (Accessed: 1 June 2026).

Kingma, D.P. and Ba, J. (2015) 'Adam: A method for stochastic optimization', Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA.

Li, T., Gao, Y., Wang, K., Guo, S., Liu, H. and Kang, H. (2022) 'Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening', Information Sciences, 592, pp. 72–90. doi:10.1016/j.ins.2022.01.018.

Matsoukas, C., Haslum, J.F., Sorkhei, M., Söderberg, M. and Smith, K. (2021) 'Is it worth fine-tuning BERT for your task? Empirical insights into pre-training in medical image analysis', arXiv preprint arXiv:2108.03814.

Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S.S.R., Shankaranarayana, S.M., Sikka, A. and Bogunovic, H. (2020) 'REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs', Medical Image Analysis, 59, 101570. doi:10.1016/j.media.2019.101570.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) 'Scikit-learn: Machine learning in Python', Journal of Machine Learning Research, 12, pp. 2825–2830.

Qummar, S., Khan, F.G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z.U., Khan, I.A. and Jadoon, W. (2019) 'A deep learning ensemble approach for diabetic retinopathy detection', IEEE Access, 7, pp. 150530–150539. doi:10.1109/ACCESS.2019.2947484.

Raghu, M., Zhang, C., Kleinberg, J. and Bengio, S. (2019) 'Transfusion: Understanding transfer learning for medical imaging', Advances in Neural Information Processing Systems (NeurIPS), 32.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2020) 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', International Journal of Computer

Vision, 128(2), pp. 336–359. doi:10.1007/s11263-019-01228-7.

Simonyan, K. and Zisserman, A. (2015) 'Very deep convolutional networks for large-scale image recognition', Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA.

Tan, M. and Le, Q. (2019) 'EfficientNet: Rethinking model scaling for convolutional neural networks', Proceedings of the 36th International Conference on Machine Learning (ICML), PMLR 97, pp. 6105–6114.

Teo, Z.L., Tham, Y.C., Yu, M., Chee, M.L., Rim, T.H., Cheung, N., Bikbov, M.M., Wang, Y.X., Tang, Y., Lu, Y., Wong, I.Y., Ting, D.S.W., Tan, G.S.W., Wong, T.Y. and Cheng, C.Y. (2021) 'Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis', Ophthalmology, 128(11), pp. 1580–1591. doi:10.1016/j.ophtha.2021.04.027.

Ting, D.S.W., Cheung, C.Y.L., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I.Y., Lee, S.Y., Wong, E.Y.M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N.C., Finkelstein, E.A., Lamoureux, E.L., Wong, I.Y., Bressler, N.M., Sivaprasad, S., Varma, R., Jonas, J.B., He, M., Cheng, C.Y., Cheung, G.C.M., Aung, T., Hsu, W., Lee, M.L. and Wong, T.Y. (2017) 'Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes', JAMA, 318(22), pp. 2211–2223. doi:10.1001/jama.2017.18152.

Wan, S., Liang, Y. and Zhang, Y. (2022) 'Deep convolutional neural networks for diabetic retinopathy detection by image classification', Computers and Electrical Engineering, 72, pp. 274–282. doi:10.1016/j.compeleceng.2018.07.042.

Zhou, Y., Li, Z., Zhu, H., Chen, C., Gao, S., Xu, K. and Ma, J. (2023) 'Collaborative learning of semi-supervised segmentation and classification for medical images', Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2079–2088.