UNIVERSITY OF MORATUWA, SRI LANKA

Faculty of Engineering

Department of Electronic and Telecommunication Engineering

Semester 5 (Intake 2020)

EN3160 - Image Processing and Machine Vision

Project - Diabetic Retinopathy Severity Grading

Samaraweera D. T. - 200564J

Wanigathunga W. A. S. S. - 200693D

October 23 , 2023

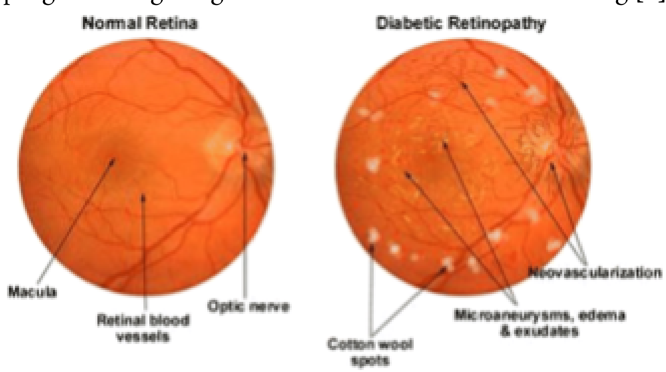
Abstract

Diabetic Retinopathy (DR) is a prevalent eye condition that poses a significant threat to vision, particularly in individuals with diabetes. The accurate grading of DR severity is vital for timely intervention and treatment. This report presents a deep learning-based approach utilizing the EfficientNet B7 architecture for DR severity grading. The study focuses on addressing challenges such as dataset preprocessing, image quality improvement, and model complexity. The evaluation is based on the Resized Diabetic Retinopathy Kaggle competition dataset, employing performance metrics including accuracy, precision, recall, F1-score, and quadratic weighted kappa. The results demonstrate competitive performance, highlighting the potential of deep learning in automating DR severity grading. This work also discusses future directions for image processing enhancement, utilization of patient information, model optimization, generalization, and possible extensions to other medical imaging applications. `https://github.com/SasiniWanigathunga/Diabetic-Retinopathy-Severity-Grading.git`

Contents

1. Introduction

The problem is Diabetic Retinopathy (DR), which is a leading cause of blindness in the working-age population in the developed world, affecting over 93 million people. In the US, approximately 29.1 million people have diabetes, and worldwide, the number is estimated at 347 million. DR is associated with long-standing diabetes, and early detection is critical to slow its progression. Earlier, DR detection was a time-consuming and manual process performed by trained clinicians, resulting in delayed results and potential loss of follow-up. The need for an automated DR screening method is recognized, and previous efforts have made progress using image classification and machine learning [7].



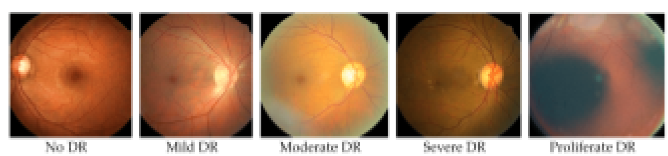
Normal and abnormal (Diabetic retinopathy) retinal image

1.1 A brief outline of the problem

Diabetic retinopathy (DR) is a serious eye condition caused by diabetes that can lead to vision loss if not detected and treated early. DR affects the blood vessels in the retina, the light-sensitive tissue at the back of the eye, and causes them to leak, swell, or grow abnormally [1].

The problem involves the accurate grading of the severity of DR, which is essential for determining the appropriate treatment and intervention. The severity of DR is usually assessed by examining the fundus images of the retina, which can be captured by digital cameras or scanners. The severity grading is based on the presence and extent of various lesions, such as microaneurysms, hemorrhages, exudates, cotton wool spots, intraretinal microvascular abnormalities, venous beading, and neovascularization [2]. The most widely used grading system is the Early Treatment Diabetic Retinopathy Study (ETDRS) scale, which divides DR into five levels: no DR, mild NPDR, moderate NPDR, severe NPDR, and PDR [3].

The current challenges in DR severity grading include the high variability and subjectivity of human graders, the lack of standardized and consistent criteria, the low availability and accessibility of expert ophthalmologists, and the high cost and time required for manual grading [4]. These challenges motivate the development of automated and reliable computer-aided diagnosis (CAD) systems that can assist the clinicians in grading DR severity from fundus images. Recent advances in machine learning and deep learning have shown promising results in DR detection and classification, but there are still limitations and gaps to be addressed.



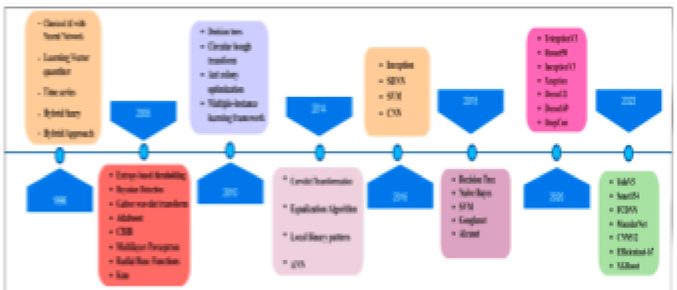
Sample images from the five different image classes showing various DR stages: No DR, Mild DR, Moderate DR, Severe DR, and Proliferate DR. [5]

1.2 Overview

The objective of this project is to propose a deep learning-based method for DR severity grading that can address some of the existing challenges and improve the performance. The proposed method utilizes the state-of-the-art deep neural network architecture, EfficientNet B7. The proposed method is evaluated on the dataset, Resized version of the Diabetic Retinopathy Kaggle competition dataset [6], and compared with the existing methods in terms of accuracy, precision, recall, F1-score and Kappa score which was the evaluation method used in the Diabetic Retinopathy Detection Kaggle competition [7]. The results show that the proposed method achieves competitive performance and provides more confidence and insight into the grading process.

2. Related work

In this section, we review the existing methods and related work in the field of Diabetic Retinopathy severity grading. We categorize the methods into three groups: traditional methods, deep learning methods, deep learning methods with uncertainty estimation, and deep learning methods with explainability [11].



Evolution of DR with artificial intelligence [14]

2.1 Traditional methods:

These methods rely on handcrafted features extracted from the fundus images, such as the number and size of lesions, the area and shape of the optic disc, and the tortuosity and caliber of the blood vessels. These features are then fed into a classifier, such as a support vector machine (SVM), a k-nearest neighbor (KNN), or a decision tree, to grade the severity of DR. Some examples of traditional methods are [1], [8] and [9]. The main limitations of these methods are that they require manual or semi-automatic segmentation of the lesions and the optic disc, which can be time-consuming and error-prone, and that they are sensitive to noise and variations in the images.

2.2 Deep learning methods:

These methods use deep neural networks (DNNs) to automatically extract features from the fundus images and classify them into different DR severity levels. DNNs are composed of multiple layers of neurons that can learn complex and hierarchical representations of the data. Some examples of deep learning methods are [9] and [10]. The main advantages of these methods are that they do not require manual or semi-automatic segmentation of the lesions and the optic disc, and that they can achieve higher accuracy and generalization than traditional methods. However, these methods also have some drawbacks, such as the need for large and diverse datasets, the lack of interpretability and explainability, and the difficulty of estimating the uncertainty of the predictions.

2.3 Deep learning methods with uncertainty estimation:

These methods are extensions of the deep learning methods that aim to address the drawback of uncertainty estimation. Uncertainty estimation is the process of quantifying the confidence and reliability of the predictions made by the DNNs. Some examples of deep learning methods with uncertainty estimation are [2] and [12] . The main benefits of these methods are that they can provide more robust and trustworthy results, and that they can help to filter out the predictions with high uncertainty and flag them for further inspection by the experts. However, these methods also have some challenges, such as the computational complexity and the lack of standard evaluation metrics.

2.4 Deep learning methods with explainability:

These methods are extensions of the deep learning methods that aim to address the drawback of explainability. Explainability is the process of providing visual and textual explanations for the predictions made by the DNNs. Some examples of deep learning methods with explainability are [1] and [13]. The main benefits of these methods are that they can help the clinicians to understand and verify the predictions, and that they can improve the user satisfaction and acceptance of the CAD system. However, these methods also have some challenges, such as the accuracy and validity of the explanations, and the alignment with the human understanding and expectations.

3. Methodology

This section describes the discussion about diverse retinal datasets, techniques used for diagnosing DR, and performance metrics to evaluate the fundus images. The basic components involved are shown.In this work we are using a supervised learning approach using 256*256 size images and labels graded from 0 to 4 integer labels.

The given data set consist of 35126 unique images which we divided to 28000 train images and 7126 test images in approximately 4:1 ratio

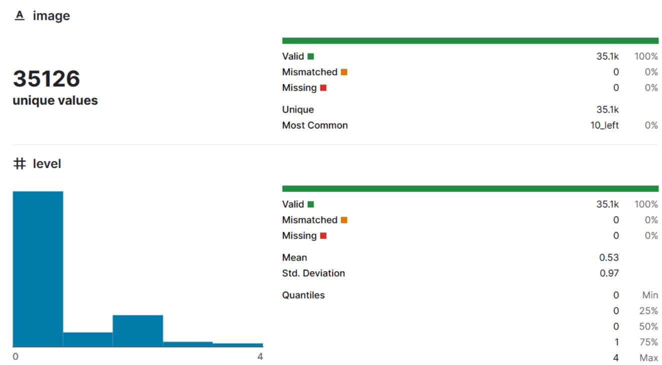
3.1 Data Acquisition

Resized version of the Diabetic Retinopathy Kaggle competition dataset:

The given dataset for the problem consists of a large set of retina images taken using fundus photography under a variety of imaging conditions. A clinician has rated each image for the severity of diabetic retinopathy on a scale of 0 to 4, according to the following scale:

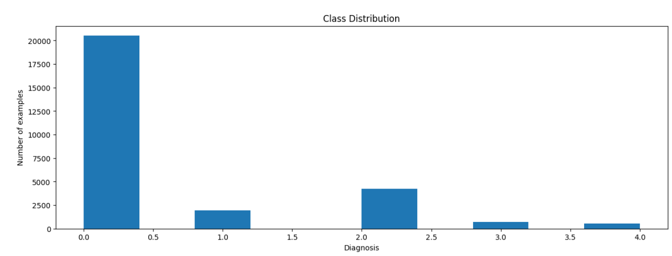
0 - No DR
1 - Mild
2 - Moderate
3 - Severe
4 - Proliferative DR

There can be noise in both the images and labels. Images may contain artifacts, be out of focus, underexposed, or overexposed. The images were gathered from multiple clinics using a variety of cameras over an extended period of time, which introduces further variation.

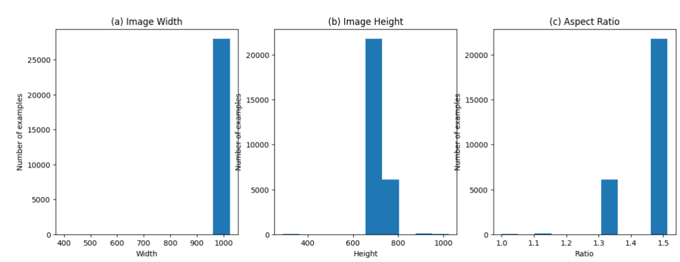


Resized Version: The original dataset resized to 1024x1024 if it is bigger than this size, else it remains the same. The distribution of the dataset among the classes is as follows

Training Set		Testing Set	
level		level	
0	20528	0	5282
2	4256	2	1036
1	1955	1	488
3	706	3	167
4	555	4	153



3.2 Data Preprocessing



To ensure the consistency and quality of our image dataset, we performed a preprocessing step to resize the images to a fixed dimension of 256*256 pixels. We noticed that some images had different heights but the same width, which could introduce unwanted distortions or artifacts in the analysis.

To reduce the noise in the image, we applied a gaussian filter with a standard deviation of 10 pixels. A gaussian filter is a type of low-pass filter that smooths the image by averaging the neighboring pixels with weights that follow a normal distribution. The standard deviation controls the size of the filter kernel and the degree of smoothing. A larger standard deviation means more blurring, but also more noise removal. The gaussian filter preserves the edges better than a simple average filter, because it gives more weight to the central pixels than to the distant ones.

3.2.1 Cropping

The function circle_crop is used to perform a circular crop around the center of an image. This is particularly useful for eye images where the subject of interest is typically located in the center.

First determines the dimensions of the image, including its height, width, and depth. It then resizes the image to be a square with sides equal to the length of the largest side of the original image. This ensures that the subject remains centered even if the original image was not a perfect square.

Next, the function calculates the center of the image and the radius of the largest possible circle that can fit within the image. A mask is created in the shape of this circle.

Finally, the function applies this circular mask to the original image. This effectively crops the image to the area within the circle, setting all pixels outside the circle to zero. The resulting circularly cropped image.

This method of cropping is advantageous for eye images as it focuses on the central area of the image where the subject is located, while eliminating potential noise and irrelevant information in the periphery of the image. This can lead to improved performance in subsequent image processing and analysis tasks.

3.2.2 Data Augmentation

The effectiveness of neural networks is largely contingent on the availability of sufficient training data. However, this requirement is often unmet in many applications, particularly in the field of medical imaging. Here, acquiring more training data means obtaining additional annotations, which can be costly due to the scarcity of skilled ophthalmologists. Another challenge is the unequal distribution of images across different disease classes.

To overcome these limitations and enhance the network's capabilities, data augmentation techniques are employed. These techniques, which include;

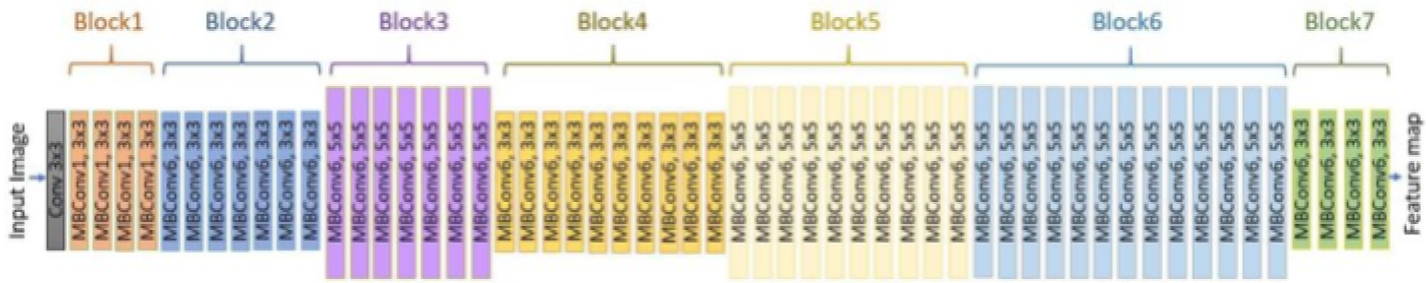
1. A random rotation to the image. The degree of rotation is chosen randomly for each image from the range -360 to 360 degrees. This helps to augment the dataset and make the model invariant to rotations.
2. A transformation that randomly flips the image horizontally and vertically. This is another form of data augmentation that helps to make the model invariant to flipping.

serve to expand the training samples and balance the class sizes.

3.3 EfficientNet-B7 Model

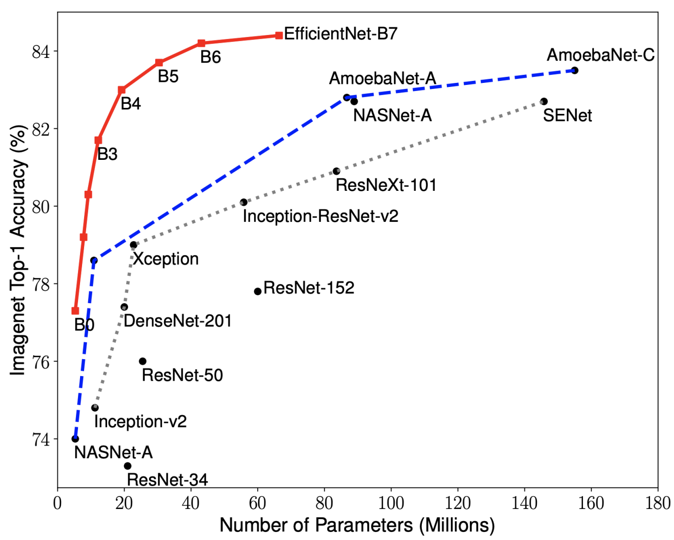
Various CNN models have been proposed for different classification tasks, such as AlexNet, VGGNet, ResNet, and Inception Architecture. The authors of this article used the EfficientNet-B7 version of the benchmark Inception network as the EfficientNet CNN model, and its variants have shown significant improvement for image classification and recognition applications.[17].

We utilized neural architecture search to develop a new foundational network, which we then scaled up to create a series of models known as EfficientNets. These models significantly outperform previous ConvNets in terms of accuracy and efficiency. Specifically, our EfficientNet-B7 achieves a record-breaking 84.3% top-1 accuracy on ImageNet.



EfficientNet B7 Architecture

Furthermore, our EfficientNets demonstrate excellent transferability, achieving unparalleled accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and three other transfer learning datasets, all while having an order of magnitude fewer parameters.[17].



EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

The EfficientNet models are built on basic and extremely efficient compounded scaling approaches. This approach allows you to scale up the ConvNet baseline to any target limited resources while retaining the model utility used for the transfer of learning datasets.

3.4 Optimizer

Adam Optimizer is a popular algorithm for gradient-based optimization of stochastic objective functions. It combines the advantages of two other extensions of stochastic gradient descent: Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). Adam Optimizer computes adaptive learning rates for each parameter, and also keeps an exponentially decaying average of past gradients and squared gradients. This helps to adjust the learning rate based on the characteristics of the data and the parameters. Adam Optimizer has been shown to work well in various settings, such as deep learning, natural language processing, computer vision, and reinforcement learning.

We have used the base parameters as 1e-3 learning rate and scheduler which multiplies the learning rate by *gamma(* $\gamma = 0.5$*)* every *step(=5)* number of epochs. This can be useful to slowly reduce the learning rate over time, which can lead to better final performance.

3.5 Evaluation

In the related competition submissions are scored based on the quadratic weighted kappa, which measures the agreement between two ratings. This metric typically varies from 0 (random agreement between raters) to 1 (complete agreement between raters). In the event that there is less agreement between the raters than expected by chance, this metric may go below 0. The quadratic weighted kappa is calculated between the scores assigned by the human rater and the predicted scores[7].

Images are assigned one of five possible ratings: 0, 1, 2, 3, 4. Each image is represented by a pair (ea, eb), where ea is the score given by Rater A (human) and eb is the score predicted by Rater B. The quadratic weighted kappa is computed as follows: Firstly, an N x N histogram matrix O is created, where Oi,j represents the count of images that received a rating i from A and a rating j from B. Then, an N x N matrix of weights, w, is computed based on the difference in scores assigned by the raters[7].

$$\omega_{i,j} = \frac{(i-j)^2}{(N-1)^2}$$

An N-by-N histogram matrix of expected ratings, E, is calculated, assuming that there is no correlation between rating scores. This is calculated as the outer product between each rater's histogram vector of ratings, normalized such that E and O have the same sum[7].

From these three matrices, the quadratic weighted kappa is calculated as:

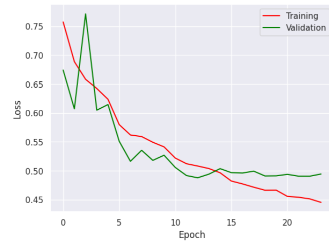
$$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}$$

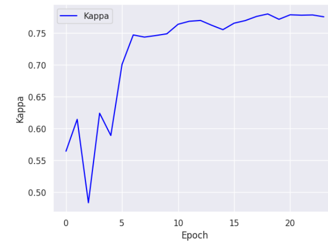
Also we calculated F1-Score which is represented as below regarding false positive (FP), false negative (FN), true negative (TN), and true positive (TP),

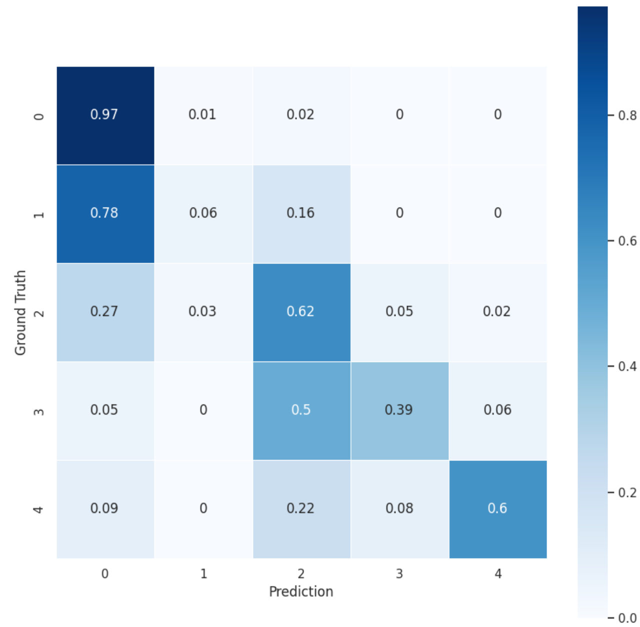
$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

4. Results

The outcomes were investigated for the epoch size of 1 to 20. The loss values and the Kappa score with respect the epochs are as follows







Because of the class imbalance, most of the images are classified under the class 1 which account for most of the data in the dataset. The least amount of F1 score has shown in the class 2 because of the lesser number of images compared to adjacent classes(488 compared to 5282 and 1036) which makes it harder for the model to differentiate.

Classification Report in Test:

	precision	recall	f1-score	support
0	0.88	0.97	0.93	5282
1	0.30	0.06	0.09	488
2	0.68	0.62	0.65	1036
3	0.49	0.39	0.43	167
4	0.75	0.60	0.67	153
accuracy			0.84	7126
macro avg	0.62	0.53	0.55	7126
weighted avg	0.80	0.84	0.81	7126

5. Comparison

In recent years, numerous studies have focused on automating Diabetic Retinopathy (DR) diagnosis. This section discusses two key areas: initial research using traditional machine learning methods for DR diagnosis, followed by recent advancements using convolutional neural network frameworks [1].

Work	Technique	Dataset	Performance analysis	Future directions
(Walter et al., 2002)	Morphological filtering techniques, Watershed transformation	Topcon TRC 50 IA retinography (own dataset)	Sen 92.8% Mean predictive value of 92.4%.	Detection of macular edema The distinction between hard and SE.
(Amin et al., 2018)	Mathematical morphology, Ensemble, Bayes Net, Naïve Bayes	DIARET DB1, DIARETDB 1, Messidor, E-ophtha and local datasets	0.99/98.9% AUC /ACC 1.00/100% on local	Detection of different types of lesions.
(Wang et al., 2020)	Deep CNN, Mathematical Morphology, Ridge-based regression feature, Random Forest	e-ophtha HEI-MED	F-score/AUC 0.8929/0.9644 0.9326/0.9323	Detection of different types of lesions.
(Prentašić and Lončarić, 2016)	CNN	DriDB	F1 measure of 0.78	Enhancing the network by using various channels, pre-processing, and post-processing steps.
(Saravanan, 2019)	Masking, Textural enhancement, and Adaptive neuro-fuzzy inference system classifier	STARE and MESSIDOR	88%	Detection of different types of lesions.

6. Discussion

In this section, the future challenges, possible extensions, and ways to improve our method and approach are discussed. The proposed method for DR severity grading has shown promising results, but there are still some challenges and limitations that need to be addressed in the future.

5.1 Future work and ways to improve our method

5.1.1 Improved Image Processing:

Although the use of Gaussian filtering has proven successful in minimizing image noise, there is potential for future research to investigate more advanced image processing methods to improve image quality. This could involve exploring techniques to manage varying light conditions in the images, or methods to emphasize the retinal features that are most relevant to the diabetic retinopathy severity grading.

5.1.2 Advanced Deep Learning Models:

The EfficientNet-B7 model has been a powerful tool for this task, but there are always new and improved deep learning models being developed. Future work could explore the use of these newer models to see if they can provide even better performance. As an example the 1st runner up team of the Diabetic Retinopathy Detection Kaggle competition [7], have proposed a novel method by adding the Regression Activation Map (RAM) after the global average pooling layer of the convolutional networks (CNN). With RAM, the proposed model can localize the discriminative regions of a retina image to show the specific region of interest in terms of its severity level [15].

5.1.3 Incorporation of Patient Information:

The current methodology relies only on the images of the retina. However, diabetic retinopathy is influenced by many factors, including the patient's overall health, lifestyle, and genetic factors. Future work could explore the incorporation of this additional patient information into the model to improve the accuracy of the severity grading [16].

5.1.4 Model complexity and efficiency:

The proposed method uses the state-of-the-art deep neural network architecture, EfficientNet B7, as a feature extractor, which is computationally expensive and requires high memory and processing power. The future work should focus on reducing the model complexity and improving the efficiency, such as by using model compression, quantization, and optimization techniques.

5.1.5 Model generalization and robustness:

The proposed method may suffer from overfitting, underfitting, or bias issues, which can affect its generalization and robustness. Therefore, future work should focus on enhancing the model generalization and robustness, such as by using cross-validation, regularization, dropout, and ensemble techniques.

5.2 Possible Extensions

The proposed method for DR severity grading can also be extended or adapted for other tasks and applications, such as:

5.2.1 Other eye diseases:

The proposed method can be applied to other eye diseases that can be detected or diagnosed from fundus images, such as glaucoma, age-related macular degeneration, cataract, and retinopathy of prematurity.

5.2.2 Other medical imaging modalities:

The proposed method can be adapted to other medical imaging modalities that can be used for DR severity grading, such as optical coherence tomography (OCT), fluorescein angiography (FA), and optical coherence tomography angiography (OCTA).

5.2.3 Other machine learning tasks:

The proposed method can be extended to other machine learning tasks that can be performed on fundus images, such as segmentation, detection, localization, and registration.

5.2.4 Other domains and applications:

The proposed method can be generalized to other domains and applications that involve image classification, uncertainty estimation, and explainability, such as natural scene understanding, face recognition, object detection, and autonomous driving.

7. Acknowledgements

This work was conducted using a Kaggle notebook, an online computational environment that allows users to run code using the browser. The Kaggle notebook provided a robust platform for running the deep learning models used in this study.

Significant computational resources were also utilized in this work. The models were trained on a NVIDIA Tesla P100 GPU. The use of this high-performance GPU significantly reduced the training time of the models and enabled the handling of large datasets.

8. References

[1] C. Bhardwaj, S. Jain, and M. Sood, "Deep Learning–Based Diabetic Retinopathy Severity Grading System Employing Quadrant Ensemble Model," *Journal of Digital Imaging*, vol. 34, no. 2, pp. 440–457, Mar. 2021, doi: https://doi.org/10.1007/s10278-021-00418-5.

[2] T. Araújo *et al.*, "DR|GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Medical Image Analysis*, vol. 63, p. 101715, Jul. 2020, doi: https://doi.org/10.1016/j.media.2020.101715.

[3] A. Khatun and S. Hossain, "Early Detection of Diabetic Retinopathy and Severity Scale Measurement: A Progressive Review & Scopes," *arXiv (Cornell University)*, Dec. 2019, doi: https://doi.org/10.48550/arxiv.1912.12829.

[4] S. R. Sadda *et al.*, "Quantitative Assessment of the Severity of Diabetic Retinopathy," *American Journal of Ophthalmology*, vol. 218, pp. 342–352, Oct. 2020, doi: https://doi.org/10.1016/j.ajo.2020.05.021.

[5] P. Vora and S. Shrestha, "Detecting Diabetic Retinopathy Using Embedded Computer Vision," *Applied Sciences*, vol. 10, no. 20, p. 7274, Oct. 2020, doi: https://doi.org/10.3390/app10207274.

[6] "Diabetic Retinopathy (resized)," *www.kaggle.com*. https://www.kaggle.com/datasets/tanlikesmath/diabetic-retinopathy-resized

[7] Emma Dugas, Jared, Jorge, Will Cukierski. (2015). "Diabetic Retinopathy Detection," *kaggle.com*. https://kaggle.com/competitions/diabetic-retinopathy-detection.

[8] S. Chavan and N. Choubey, "An automated diabetic retinopathy of severity grade classification using transfer learning and fine-tuning for fundus images," *Multimedia Tools and Applications*, Mar. 2023, doi: https://doi.org/10.1007/s11042-023-15135-0.

[9] A. Mary Dayana and S. Emmanuel, "A Comprehensive Review of Diabetic Retinopathy Detection and Grading Based on Deep Learning and Metaheuristic Optimization Techniques," *Archives of Computational Methods in Engineering*, Jun. 2023, doi: https://doi.org/10.1007/s11831-023-09946-5.

[10] C. Bhardwaj, S. Jain, and M. Sood, "Hierarchical severity grade classification of non-proliferative diabetic retinopathy," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2649–2670, Aug. 2020, doi: https://doi.org/10.1007/s12652-020-02426-9.

[11] D. Nagpal, S. N. Panda, M. Malarvel, P. A. Pattanaik, and M. Zubair Khan, "A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends," *Journal of King Saud University - Computer and Information Sciences*, Jun. 2021, doi: https://doi.org/10.1016/j.jksuci.2021.06.006.

[12] J. Jaskari *et al.*, "Uncertainty-aware deep learning methods for robust diabetic retinopathy classification," *arXiv (Cornell University)*, Jan. 2022, doi: https://doi.org/10.48550/arxiv.2201.09042.

[13] Y. Niu, L. Gu, Y. Zhao, and F. Lu, "Explainable Diabetic Retinopathy Detection and Retinal Image Generation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 44–55, Jan. 2022, doi: https://doi.org/10.1109/jbhi.2021.3110593.

[14] P. Bidwai, S. Gite, K. Pahuja, and K. Kotecha, "A Systematic Literature Review on Diabetic Retinopathy Using an Artificial Intelligence Approach," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 152, Dec. 2022, doi: 10.3390/bdcc6040152.

[15] Z. Wang and J. Yang, "Diabetic Retinopathy Detection via Deep Convolutional Networks for Discriminative Localization and Visual Explanation." Accessed: Nov. 05, 2023. [Online]. Available: https://arxiv.org/pdf/1703.10757.pdf

[16] "Improved Automatic Grading of Diabetic Retinopathy Using Deep Learning and Principal Component Analysis," *ieeexplore.ieee.org*. https://ieeexplore.ieee.org/document/9630919/ (accessed Nov. 05, 2023).

[17] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.