Multiclass Diabetic Retinopathy Classification based on Low-Rank Adaptation

Sebastián Ferreira-Caballero, Diego P. Pinto-Roa, José Luis Vázquez Noguera
Universidad Nacional de Asunción
Facultad Politécnica
Asunción, Paraguay
sebas.ferre18@fpuna.edu.py, {dpinto,jlvazquez}@pol.una.py

and

Jordan Ayala, Pastor Pérez-Estigarribia
Universidad Nacional de Asunción,
Facultad Politécnica,
Asunción, Paraguay
jorayalag@fpuna.edu.py, peperez.estigarribia@pol.una.py

and

Pedro E. Gardel-Sotomayor
Universidad Católica de Asunción
Facultad de Ciencias y Tecnología
Hernandarías, Paraguay
pedro.gardel@uc.edu.py

Abstract

Diabetic retinopathy is a vision impairment associated with diabetes mellitus, a common medical condition. The predominant method of diagnosing this eye complication is retinographic imaging analysis of retinal fundus images. Recent advances in deep learning have shown effectiveness in detecting diabetic retinopathy, rivaling the diagnostic accuracy of human inspection. However, the success of these methods largely depends on the model's architecture. In our research, we introduce a method for training a deep learning model using a Low-Rank Adaptation (LoRA) technique to distinguish between three stages of diabetic retinopathy: no retinopathy, non-proliferative retinopathy, and proliferative retinopathy. LoRA significantly reduces the training parameters needed by employing a low-ank representation. Our experimental results on four data sets of different sizes demonstrate that the LoRA model achieves top-tier performance in classifying stages of diabetic retinopathy.

Keywords: Low-Rank Adaptation, Diabetic Retinopathy, Multiclass Classification, Deep Learning

1 Introduction

Diabetes mellitus (DM), commonly known as diabetes, is a metabolic disorder characterized by high blood sugar levels over a prolonged period, which can result in sugar presence in the urine [1]. This condition, affecting millions globally [2], leads to various complications, including Diabetic Retinopathy (DR). DR is an eye condition that occurs when high blood sugar levels cause damage to the blood vessels of the retina, potentially leading to irreversible loss of vision [3].

Exudates are apparent symptoms of DR evident in fundus images of the eye [4]. The early detection of DR can identify various lesions, including exudates [5]. Ophthalmologists utilize color fundus images to

examine multiple DR indicators, such as microa-neurysms, hemorrhages, and both hard and soft exudates, which are crucial for early diagnosis and management of the disease [6, 7].

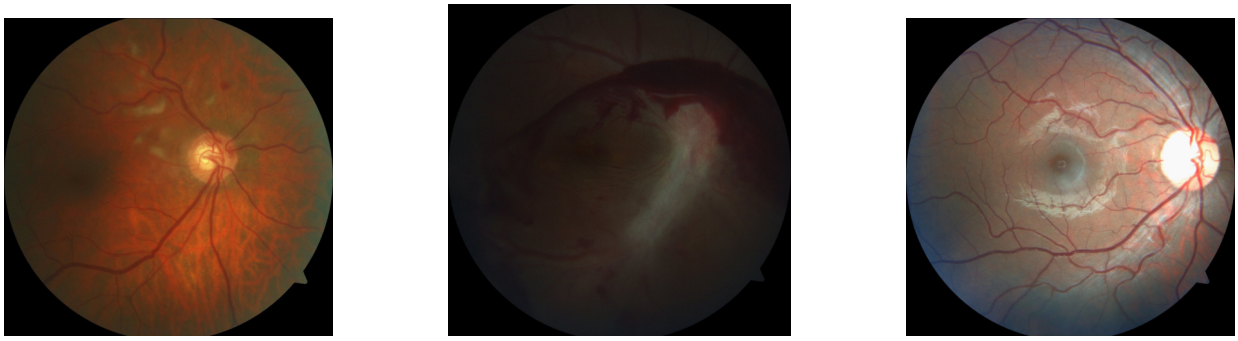
The main types of DR are [8, 9]:

- *Non-Proliferative Diabetic Retinopathy* (NPDR) is the disease's initial and less severe manifestation and may cause mild and moderate vision loss. The blood vessels are damaged at the back of the eye, resulting in hemorrhage and fluid leakage into the retina. Macular edema occurs when the macula swells, and macular ischemia arises when retinal blood vessels close [10]. Blood cannot reach the macula when macular edema occurs, affecting vision by sometimes forming exudates. An example of NPDR is shown in Figure 1.a. NPDR can be classified into three levels: *Mild*, *Moderate*, and *Severe*. At the mild level, few micro-aneurysms appear. At the Moderate level, the number of microaneurysms and dot-blot hemorrhages increases. At the severe level, the 4-2-1 rule determines four quadrants of diffuse retinal hemorrhages and micro-aneurysms, two or more quadrants of venous micro-aneurysms, and one (or more) quadrants with intraretinal microvascular abnormalities.

- *Proliferative Diabetic Retinopathy* (PDR) is the disease's most advanced manifestation. This severe stage can cause central and peripheral vision loss [11]. PDR is characterized by neovascularization of the disc or elsewhere, which is the presence of new blood vessels on the retina's surface [12]. These new blood vessels can form scar tissue, which can cause problems with the macula or lead to retinal detachment [13]. Figure 1.b shows an example of PDR.

Clinical studies advocate for periodic digital fundus imaging for diabetic patients to monitor ocular health [14]. Despite these recommendations, patients with diabetes receive regular eye exams due to factors such as limited accessibility and high costs of ophthalmological services. Over the past three decades, the demand for regular eye screenings in the diabetic population has quadrupled [15], presenting a significant challenge to healthcare providers. The traditional approach to retinal examination struggles to keep pace with the rapid increase in DR cases globally. To address this, adopting computerized automated diagnostic models has become crucial. These models facilitate DR progression control and monitoring. Modern technologies now enable retinal fundus image analysis programs, significantly enhancing the efficiency of DR diagnosis [14]. Automated DR detection methods improve diagnostic accuracy and reduce the time and cost associated with manual diagnosis, making them a valuable tool in managing this growing health concern [16].

In this paper, we extended the performance study of *Low-Rank Adaptation* (LoRA) based Deep Learning model [17] to classify three stages of DR: i) no sign of DR ii) NPDR, and iii) PDR. Using just one dataset, we applied LoRA for the first time in [18]. In this second study, we used four datasets of DR images to validate the results obtained in the first study [18].



(a) NPDR (b) PDR (c) No DR sings

Figure 1: An example image for each class [8, 18].

Adopting a low-rank representation in LoRA (Low-rank Adaptation) is driven by its efficacy in enhancing parameter and computational efficiency during model adaptation [17]. Unlike traditional fine-tuning, which struggles with the over-parameterization of pre-trained models, LoRA introduces a lightweight methodology. It integrates trainable rank decomposition matrices into each model layer while freezing the pre-trained weights. This design not only preserves the integrity of the original model but also significantly reduces the number of trainable parameters. Traditional fine-tuning methods often increase inference latency and fail to meet performance baselines [17]. LoRA addresses these shortcomings by offering a viable alternative for adapting large models, thereby minimizing memory and storage requirements and enhancing overall model efficiency.

We organize the structure of this article as follows: Section 2 reviews related works in the field. Section 3 describes the methodology proposed for the DR classification. Section 4 discusses the experimental results

obtained from applying the proposed method. Finally, Section 5 outlines the main conclusions drawn from the study and suggests directions for future research.

2 Related Work

Expert clinicians utilize retinal fundus images to detect lesions and diagnose DR [19]. Historical studies have noted the association between microaneurysms and DR since 1879 [20]. Innovative detection techniques have also been developed in [21, 22], such as the moat operator, which can detect micro-aneurysms, hemorrhages, and hard exudates. Ege et al. [23] introduced an automated tool for analyzing retinal fundus images, employing classifiers such as k-nearest neighbor, Bayesian, and Mahalanobis. The Mahalanobis classifier showed superior performance in detecting microaneurysms, hemorrhages, and exudates. Furthermore, Fujita et al. [24] have contributed to the field by researching automated detection systems for DR, focusing on micro-aneurysm-based approaches.

Techniques for detecting hemorrhages, blood vessels, and exudates are crucial in DR. These techniques generate key features that form the backbone of automatic detection systems for DR [25]. Sequential detection of these features can significantly enhance the assessment of a patient's health status affected by DR. Notably, the author of [26] developed a method to identify blood vessels using two-dimensional matched filters. This demonstrates the potential of specialized algorithms to improve diagnostic accuracy.

Exudates in the retina accumulate lipids and proteins and often appear as cream-colored, white, reflective, or bright lesions [27]. In [28], the authors developed a method for recognizing these exudates using neural networks and a hierarchical feature selection algorithm based on sensitivity analysis. The severity of DR is often indicated by the presence and number of retinal hemorrhages, which signal increased ischemia in the retina due to lack of oxygen [6]. As hemorrhages increase, indicating more severe DR, retinal vessels become more damaged and permeable, leading to further exudation of fluids, proteins, and lipids [25].

Over the past two decades, Computer-Aided Diagnosis (CAD) systems have seen rapid development [29]. A deeper understanding of diseases, improved sensor data accuracy, and increased processing capabilities have driven enhancements in these systems. In [30], Lee et al. demonstrated that the accuracy of computer vision systems in detecting early retinal injuries is comparable to that of human experts. Further refining diagnostic precision, Wang et al. [31] introduced a method to classify each pixel as either a lesion or non-lesion using color features in a Bayesian statistical classifier. Lee et al. [32] investigated the automated detection of Non-Proliferative Diabetic Retinopathy (NPDR), focusing on three types of lesions: hemorrhages and micro-aneurysms, cotton wool spots, and hard exudates Additionally, the authors of [33] developed a feed-forward neural network that can automatically classify stages of diabetic retinopathy, ranging from mild to proliferative, by analyzing the area and perimeter of the RGB components of blood vessels. This classification system achieved an accuracy of 84%, with sensitivity and specificity rates of 90% and 100%, respectively.

Recent advancements include the integration of Convolutional Neural Networks (CNNs) for DR diagnosis, which have demonstrated superior image classification capabilities compared to traditional manual techniques [34]. In [35], the authors developed a novel CNN model that enhances classification performance by using CNN-extracted features as inputs for various machine-learning classifiers. The authors conducted simulations with classifiers such as Support Vector Machine, Naive Bayes, Random Forest, J48, and AdaBoost, testing their model on databases including KAGGLE, IDRiD, and MESSIDOR.

Additionally, the authors of [36] introduced a new Deep Convolutional Neural Network (DCNN) that can detect DR in its early stages by identifying microaneurysms, which are the earliest detectable signs of the disease [37]. Further research by Jalli et al. [38] and Hellstedt and Immonen [39] have explored the dynamics of micro-aneurysms during various stages of fluorescein angiography, emphasizing their critical role in the early diagnosis of DR.

Ayala et al. [8] introduced a neuro-evolution approach to optimize the selection of a DL model for multiclass diagnosis of DR. This method employs Genetic Algorithms [40, 41], drawing inspiration from natural selection processes to optimize the hyperparameters of a Deep Neural Network. This approach iteratively refines the network parameters across various generations [42] through selection, crossover, and mutation. The chosen model for optimization is a pre-trained RESNET50v2 network. The fitness of the model is evaluated based on its accuracy, with the individual's chromosome encoded by four genes: (i) dropout rate, (ii) number of trainable layers, (iii) transfer-learning learning rate, and (iv) fine-tuning learning rate. The proposed strategy is enhanced by additional techniques such as Simulated Annealing, Population Re-initialization, and Ensembles, leading to robust results.

In [43], Rehman et al. used LoRA to adapt four vision foundation models—DINOv2, Swin, Deit, and ViT—for grading diabetic retinopathy (DR). They claim that their approach performs similarly to the most specialized models yet is also more resource-efficient, requiring significantly fewer parameters.

Unlike the state-of-the-art, which proposes using approaches based on classical machine learning and

traditional deep learning (CNN, DCNN), this research explores the performance of LoRA's in the multiclass DR classification problem on different public datasets. LoRA is a recently proposed and efficient technique for adapting pre-trained deep learning models without updating all their parameters.

3 Proposal: LoRA for Multiclass DR Classification

This section presents the proposed methodology. First, we report the datasets used for training and validating the proposal. Then, we explain the metrics used for the experiment. Finally, we describe the LoRA approach.

3.1 DR Datasets

To train the DL classification models, we use several datasets. First, we use a previously trained model on the ImageNet [44] dataset. This first step allows the model to learn the underlying patterns in the images from the dataset. The ImageNet dataset contains thousands of images of multiple categories.

The pre-trained model utilized in this study is a Vision Transformer (ViT) model adept at identifying standard features such as colors and edges. It was initially trained using generic images from the publicly available ImageNet dataset. The knowledge acquired during this initial training was then transferred to the model, allowing it to be repurposed. We now implement this pre-trained model for a different application by retraining it with retinal fundus images.

Afterward, in the proposal's transfer learning and fine-tuning steps, the fundus image datasets employed are DFISDR [45], APTOS [46], MESSIDOR [47], and IDRID [48]. Then, the knowledge from the pre-trained model is applied to the DR classification problem by training the model with the fundus images. Table 1 presents the number of images for the dataset and class considered.

Class Name	**DFISDR**	**APTOS**	**MESSIDOR**	**IDRID**
No signs of DR	711	1805	546	168
NPDR	465	1562	400	286
PDR	261	295	254	62

Table 1: Retinopathy dataset categories used in this paper

We partitioned each dataset into three subsets for training, validation, and testing, allocating 70%, 20%, and 10% of the images to each subset, respectively. To ensure representative sampling across all categories, we employed stratified sampling, maintaining consistent proportions of each class during the division process. During the training phase, we augmented the dataset to artificially increase its size by implementing data augmentation techniques [8,49]. These techniques included horizontal image flipping and rotating the images within a range of $[-0.4\pi, 0.4\pi]$.

3.2 Low-Rank Adaptation

Low-Rank Adaptation (LoRA), proposed by Hu et al. [17], is a parameter-efficient technique designed for training large neural networks. This method operates on the principle that weight updates during model training exhibit a low intrinsic rank. LoRA facilitates the integration of dense layers into an existing neural network. These layers utilize trainable rank decomposition matrices, allowing for the optimization of these matrices while the pre-trained weights of the original model remain frozen.

The proposal's main objective is to employ LoRA in a pre-trained model and verify if it obtains a high performance in multiclass DR diagnosis. The pre-trained model used in the proposal is a ViT-Base. Sensitivity and specificity are calculated for comparison with DR diagnostic standards [50]. Figure 2 shows how LoRA is implemented within the ViT model's care mechanism.

Figure 3 illustrates a pre-trained weight matrix $W_0 \in \mathbb{R}^{d\times d}$, we constrain its update by representing the latter with a low-rank decomposition with a given rank $r \ll d$:

$$h = W_0 x + \Delta W x, \text{with}: \quad \Delta W = B \cdot A \tag{1}$$

where x is the input and $B \in \mathbb{R}^{d\times r}$, $A \in \mathbb{R}^{r\times d}$ are the rank decomposition matrices [17]. A contains trainable parameters, while W_0 is frozen and doreceiverecieve gradient updates.

In our proposal, we employed LoRA to study its performance in diagnosing three stages of diabetic retinopathy.

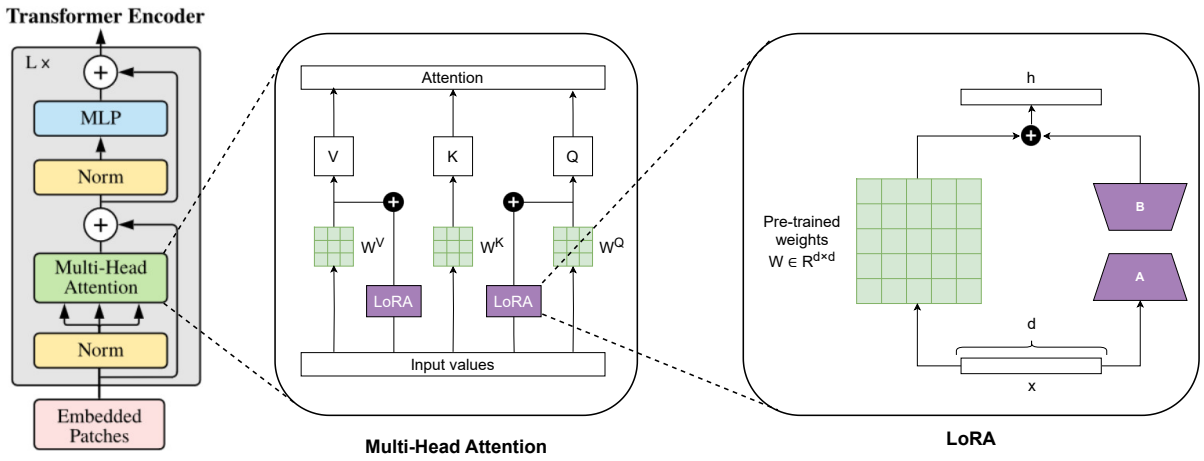


Figure 2: Structure of the attention mechanism of the ViT model with LoRA. LoRA adapts the attention matrices of the *query* and *value* vectors during the training process.

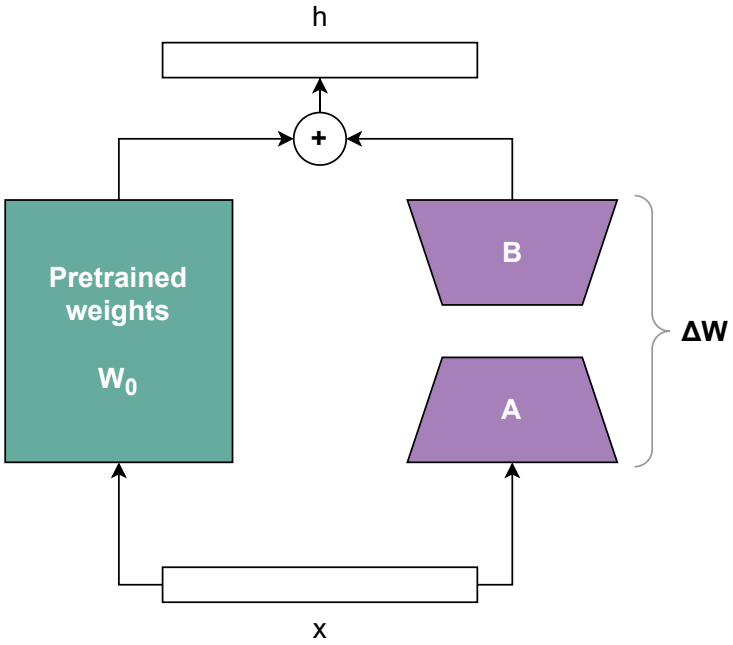


Figure 3: LoRA diagram [17].

3.3 Proposal description based on LoRA

The proposal's main objective is to employ LoRA on a pre-trained model and check if it can achieve high accuracy in the multiclass DR diagnosis. The sensitivity and specificity are compared with DR diagnosis standards [50]. Figure 4 shows the LoRA evaluation process.

We employed convolutional networks as benchmarks as proposed in [8]. The evaluation process for these networks was conducted similarly to the evaluation for LoRA. We tested four types of CNNs: RESNET50v2 [51], XCEPTION [52], RESNET101v2 [53], and MOBILENETv2 [54]. In addition to these models, we compared our results with a RESNET50v2 model that was optimized using neuro-evolution, as detailed in [8].

Afterward, we run the LoRA implementation. The main goal of our experiment here is to compare the results obtained with LoRA and the ones obtained with CNNs.

There are twelve hyperparameters in total. The default values of the convolutional models, the optimized values of RESNET50v2 obtained in [8], and our experiment uses the default values of LoRA. Table 2 shows these values. The hyperparameters are as follows:

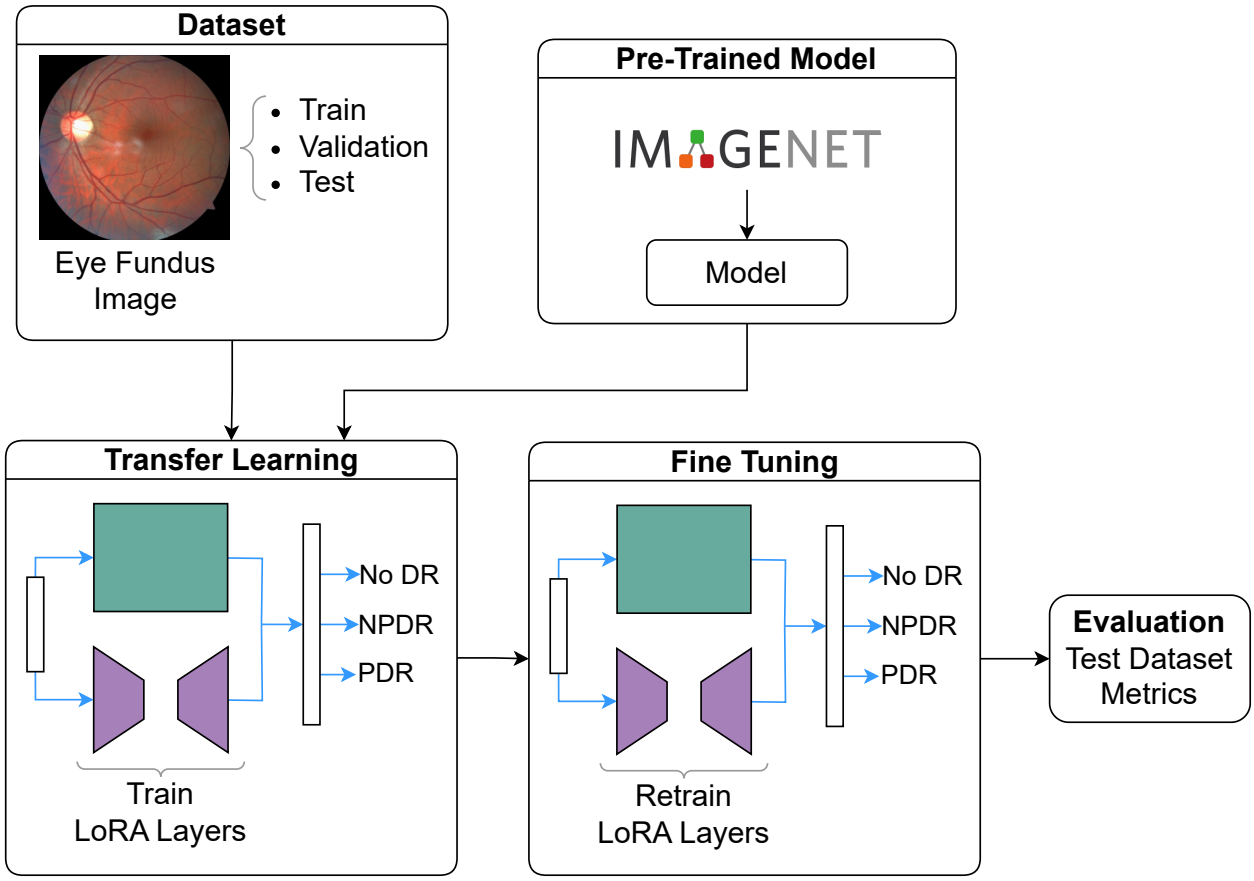


Figure 4: LoRA Evaluation Workflow [18].

1. *Dropout Rate*: according to a probability p, any neuron in a neural network is temporarily deactivated to improve the normalization capability of the network [55].
2. *Number of Trainable Layers*: indicates the number of layers retrained during fine-tuning. LoRA does not apply this hyper-parameter.
3. *Transfer Learning Learning Rate*: determines the step size at each iteration of the gradient descent during transfer learning.
4. *Fine-Tuning Learning Rate*: defines the step size at each iteration of the gradient descent during fine-tuning.
5. *Global Pooling*: defines the global pooling technique to use; three options are available: average pooling, max pooling, or no pooling at all; LoRA does not apply this hyper-parameter.
6. *Data Augmentation*: determines the use of data augmentation during training.
7. *Image Size*: defines the image dimension for the resizing process.
8. *Batch Size*: determines the number of training instances in the batch.
9. *Transfer Learning Epochs*: determines the number of iterations during training in transfer learning.
10. *Fine-Tuning Epochs*: determines the number of iterations during training in fine-tuning.
11. *Rank* (r): determines the amount of trainable parameters in the dense layers A and B that LoRA can use during training. Convolutional models do not use this hyperparameter.
12. *Alpha* (α): defines the scaling factor of the dense layers' weights over the base model ones. In other words, $\Delta W x$ is scaled by $\frac{\alpha}{r}$ [17]. Convolutional models do not use this hyperparameter.

Argument	**Symbol**	**Values CNNs**	**Values RESNET50v2**	**Values LoRA**	**Type**	**CNN**	**LoRA**
Dropout Rate	DPR	0.2	0.1	0.2	Decimal	Yes	Yes
Number of Trainable Layers	NTL	30	38	–	Integer	Yes	No
Learning Rate (Transfer Learning)	LR	10^{-4}	10^{-2}	10^{-4}	Decimal	Yes	Yes
Learning Rate (Fine-Tuning)	LR	10^{-5}	10^{-4}	10^{-5}	Decimal	Yes	Yes
Global Pooling	GP	*AVG_POOLING*	*AVG_POOLING*	–	Enum	Yes	No
Data Augmentation	DA	*True*	*True*	*True*	Boolean	Yes	Yes
Image Size	IMG	160	160	224	Integer	Yes	Yes
Batch Size	BS	32	32	32	Integer	Yes	Yes
Epochs (Transfer Learning)	E	20	20	20	Integer	Yes	Yes
Epochs (Fine-Tuning)	E	10	10	10	Integer	Yes	Yes
Rank	r	–	–	16	Integer	No	Yes
Alpha	α	–	–	16	Integer	No	Yes

Table 2: Default hyperparameters for CNNs and LoRA.

LoRA operates using default hyperparameter values, except the Number of Trainable Layers. This exclusion aligns with LoRA's approach of freezing the original weights of the pre-trained model. Instead, LoRA focuses on updating the newly added dense layers through the injected rank decomposition matrices, eliminating the need to configure multiple trainable layers.

LoRA has two exclusive parameters: Rank (r) and Alpha (α). The default values for image classification tasks were used on these parameters. Table 2 presents these values.

The evaluation workflows, as depicted in Figure 4, are used in this experiment for the CNNs and LoRA models, respectively. It's worth noting that both workflows are very similar, consisting of the following steps:

1. The Eye Fundus Images dataset was divided into training, validation, and testing subsets. The training and validation datasets train the model and monitor loss and accuracy metrics through each epoch.

2. During the transfer learning phase with CNNs, only the weights of the final layer are adjusted by training with the eye fundus dataset. In contrast, when using LoRA with a Vision Transformer (ViT) model, all weights of the base model are kept frozen. Instead, it introduces two dense layers to modify the query and value projection matrices. The base CNN and ViT models were pre-trained on the ImageNet dataset, leveraging this extensive dataset to expedite learning.

3. In the fine-tuning phase, the model previously adapted using the Eye Fundus Images dataset undergoes further training. However, this time, the learning rate is reduced. For CNNs, multiple neural network layers are fine-tuned, whereas for LoRA, only the newly added dense layers are fine-tuned.

4. Finally, the testing dataset is employed to evaluate the model's performance on previously unseen images.

4 Experimental Results

We implemented the proposal in Python 3.10.12; the main libraries that are required are Tensorflow [56], PyTorch [57] and Hugging Face libraries such as Transformers, Datasets, and PEFT [58, 59].

The machine used in the experiment has an NVIDIA GTX 1050 Ti GPU with 4 GB of VRAM, one Intel i3-10100F CPU, and 16 GB of RAM. It runs Windows 10 and WSL2 with Ubuntu 22.04 as the development environment.

4.1 Computational environment

The main experiment considered four pre-trained neural network models encompassing the XCEPTION [52], and MOBILENETv2 [54] architectures and the RESNET50v2 and RESNET101v2 networks [51]. The Tensorflow [60] platform was used to implement these models, while their respective weights were trained using the ImageNet dataset [44]. Afterward, it executes the transfer learning and fine-tuning processes following the CNN evaluation workflow.

This experiment also considers a RESNET50v2 model optimized with neuro-evolution demonstrated in [8]. Ayala et al. [8] selected four hyper-parameters for the optimization process: DPR, NTL, TL Learning Rate, and FT Learning Rate.

This experiment also considers a pre-trained ViT model [61]. This model implemented LoRA, and ImageNet was used to train its respective weights. Unlike the pre-trained neural networks, the developed LoRA implementation uses the PyTorch [57] and Hugging Face libraries [58, 59]. The evaluation workflow for LoRA is as follows: LoRA freezes the pre-trained weights and then proceeds to employ transfer learning and fine-tuning strategies on the injected rank decomposition matrices.

We applied hold-out random cross-validation with stratified sampling to evaluate these models [62, 63]. We executed the selected models 10 times and controlled the dataset shuffling with random states. Each run was assigned a random state to generate fixed datasets for training, validation, and testing. Thus, we ensure the selected models work on the same conditions.

This simulation considers the following classification instances in the following datasets: DFISDR, APTOS, MESSIDOR, and IDRID. Table 1 presented details of them.

4.2 Performance Metrics

The performance of LoRA and benchmark approaches was measured considering weighted sensitivity, weighted specificity, accuracy, and Cohen's kappa, which are explained below [64, 65]:

- Weighted sensitivity: This metric measures the overall performance of a multiclass classification model, considering how well it identifies each class and the actual proportion of instances of each class, which is useful in unbalanced data sets.
- Weighted specificity: This metric measures, on average, for each class, how well the model can misclassify other classes as if they were that class.
- Accuracy: This metric measures the proportion of correct predictions over the total predictions made.
- Cohen's kappa: This index is a statistical measure that calculates the degree of agreement between two classifiers, in this case, the real classification and the predicted classification.

It should be noted that sensitivity and specificity combined provide a balanced panorama of a classification algorithm's performance. Accuracy summarizes the overall performance but does not distinguish well between types of errors. Finally, the kappa index provides context as to whether the result is meaningful beyond chance.

4.3 Results

We present the results of the main experiment in a box plot for models (see Figures 5 to 8) for DFISDR, APTOS, MESSIDOR, and IDRID, respectively. The LoRA implementation was the best suited for the DR problem among all the other tested models.

For DFISDR in Figure 5, LoRA obtained an accuracy of 0.903 and a strong kappa index of 0.843, the best result achieved throughout the ten runs. This instance of LoRA also obtained 0.903 on sensitivity and 0.96 on specificity. The mean values are also consistently better than those of the other selected models, with an average accuracy, sensitivity, specificity, and kappa index ratio of 0.856, 0.856, 0.920, and 0.764, respectively. The simulations show similar results for APTOS, MESSIDOR, and IDRID in Figures 6 to 8. These simulations confirm LoRA's robustness and the partial results obtained in our previous study [18].

According to the standard criteria established by the British Diabetic Association (BDA) for DR diagnosis [50], in DFISDR and APTOS, LoRA obtained one instance with sensitivity and specificity values that fit the minimum requirements of 0.80 and 0.95, respectively. These results are promising, given that LoRA was configured with standard parameters and tested on small datasets as seen in Figures 5 and 6. On the other hand, as denoted in Figures 7 and 8, neither MESSIDOR nor IDRID approach achieved the minimum requirements of the BDA, indicating a large margin of investigation ahead for more complex datasets.

To illustrate the robustness of the results, we compare the best and worst kappa index ratios obtained from all tested models. Figures 9 to 12 show extensive statistical comparisons between models using Cohen's kappa test. The vertical gray line corresponds to the minimum acceptable interrater agreement proposed by [66]. It represents at least 0.64-0.81 of reliable data. Each model has a tag representing the procured values, with B and W being the best and the worst results obtained, respectively. For example, LoRA.B and LoRA.W correspond to the best and worst performance, respectively. LoRA is also the only model reaching the minimum acceptable agreement of 0.80 for DFISDR and APTOS. For MESSIDOR and IDRID, LoRA

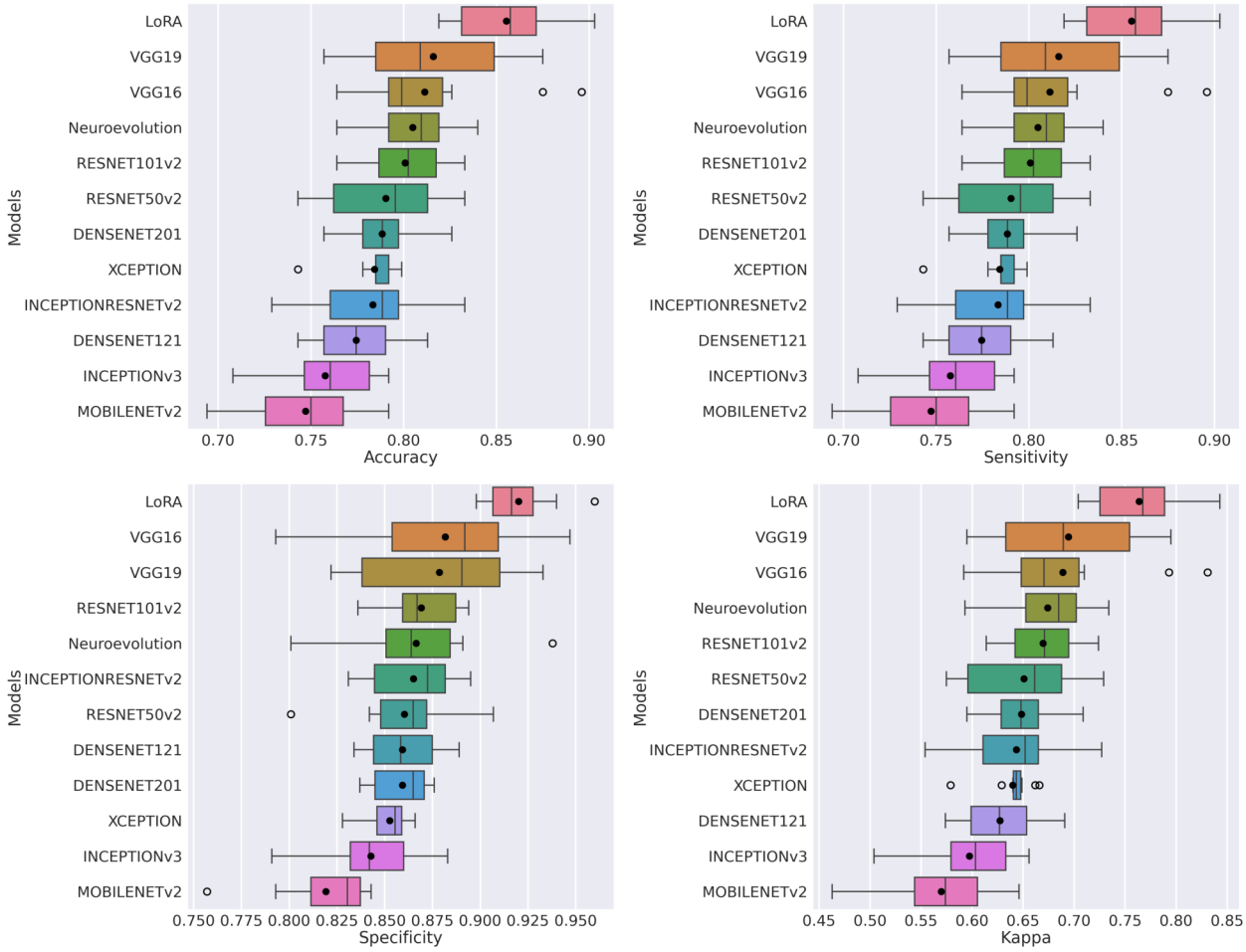


Figure 5: Box plot of the evaluation metrics for DFISDR.

gets around 0.50 and 0.60, respectively. However, the other approaches obtained less acceptable results than LoRA.

We have also monitored the computational time of the runs. The CNNs have less complexity, meaning that these models are capable of training in a short amount of time. The average training time of RESNET50v2, RESNET101v2, XCEPTION, and MOBILENETv2 is 195.56, 299.89, 242.77, and 91.24 seconds, respectively. However, LoRA is time-consuming, taking 5072.31 seconds on average to fully train, i.e., the implemented LoRA in this study requires many computational resources. Considering that the LoRA implementation is the best in metrics, a trade-off between efficiency and time complexity is established.

Overall, we can observe that LoRA achieves better results in all performance metrics on all datasets compared to the other state-of-the-art approaches, but at the cost of increased computational time. We can also note that for DFISDR and APTOS, the VGG, NESNET, and Neuroevolution approaches also obtained promising results. For MESSIDOR and IDRID, all the algorithms perform poorly on all metrics, below 0.90. These results indicate that a large margin still needs to be improved in LoRA through parameter tuning techniques and improving the ViT models. Although LoRA consistently demonstrates superiority to state-of-the-art approaches, we can clearly observe that the performance of all approaches is subject to the complexity of the data set. Consequently, developing more robust and scalable approaches for the multiclass DR classification problem is necessary.

5 Conclusions and future work

In this study, we propose applying Deep Learning models with Low-Rank Adaptation (LoRA) for the Multiclass Diabetic Retinopathy (DR) classification of Eye Fundus Images. The classification considers three stages of the disease: i) no signs of DR, ii) proliferative DR, and iii) non-proliferative DR. The results obtained using LoRA for DR classification demonstrate strong performance, even without hyperparameter optimization strategies. Notably, these results were achieved using a relatively small dataset for this prob-

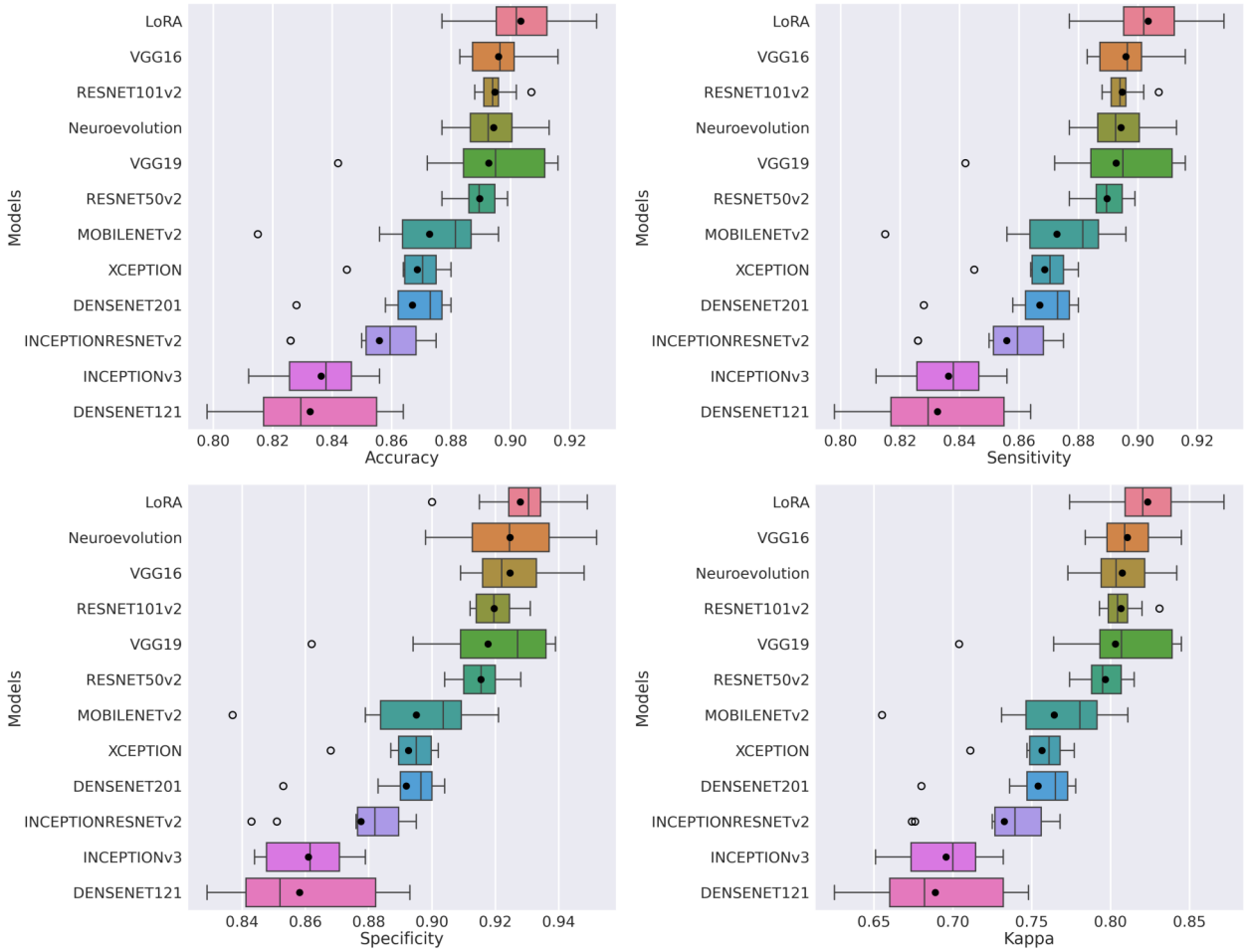


Figure 6: Box plot of the evaluation metrics for APTOS.

lem. However, for larger datasets, all models performed poorly. Nevertheless, LoRA was superior to the rest of the benchmark models considered.

The LoRA-based implementation yielded the best results overall, achieving the minimum accuracy, specificity, sensitivity, and a kappa index for small datasets. According to the criteria established by the British Diabetic Association [50], these metrics meet the minimum requirements of 80% sensitivity and 95% specificity. Furthermore, the methodology demonstrates a substantial kappa index value, indicating a reliable classification of the three DR stages.

This research opens several directions for future work. We propose further exploration into the hyperparameter optimization of LoRA, testing the approach on other datasets and benchmark algorithms, and expanding the classification task to more than three classes. Additionally, implementing other state-of-the-art DR-based Deep Learning models and optimization techniques will provide valuable comparisons, further enhancing medical image analysis with interpretable machine learning algorithms.

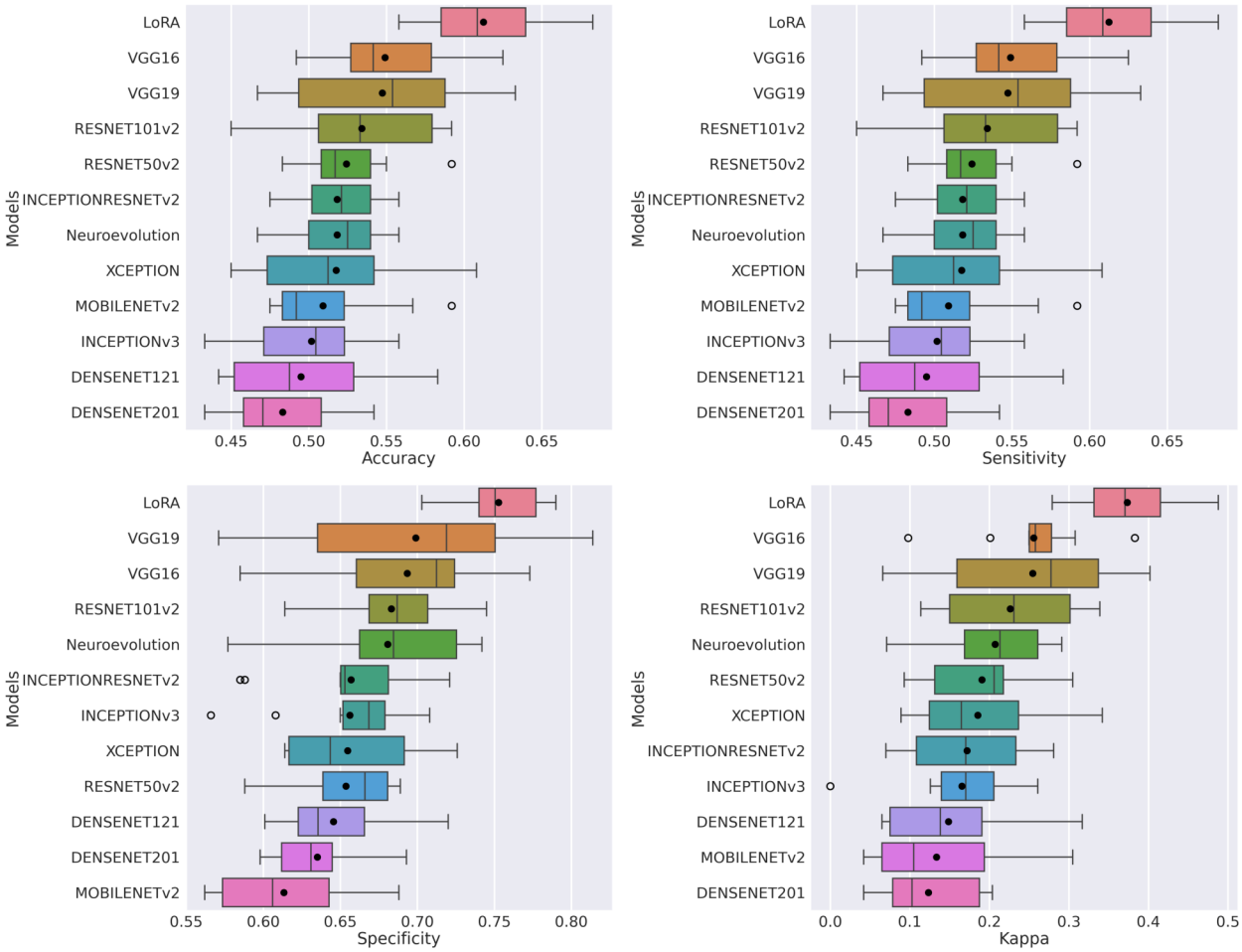


Figure 7: Box plot of the evaluation metrics for MESSIDOR.

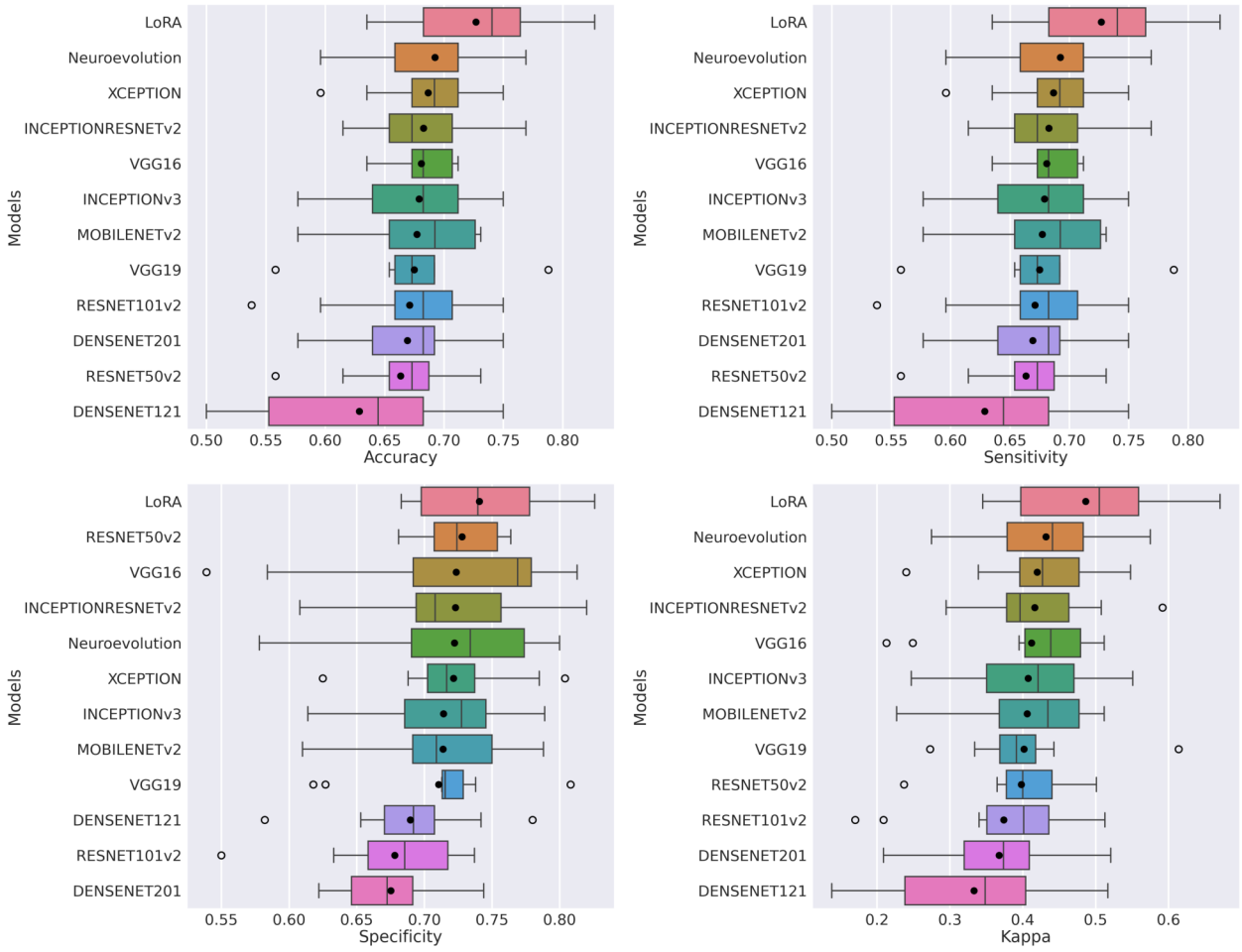


Figure 8: Box plot of the evaluation metrics for IDRID.

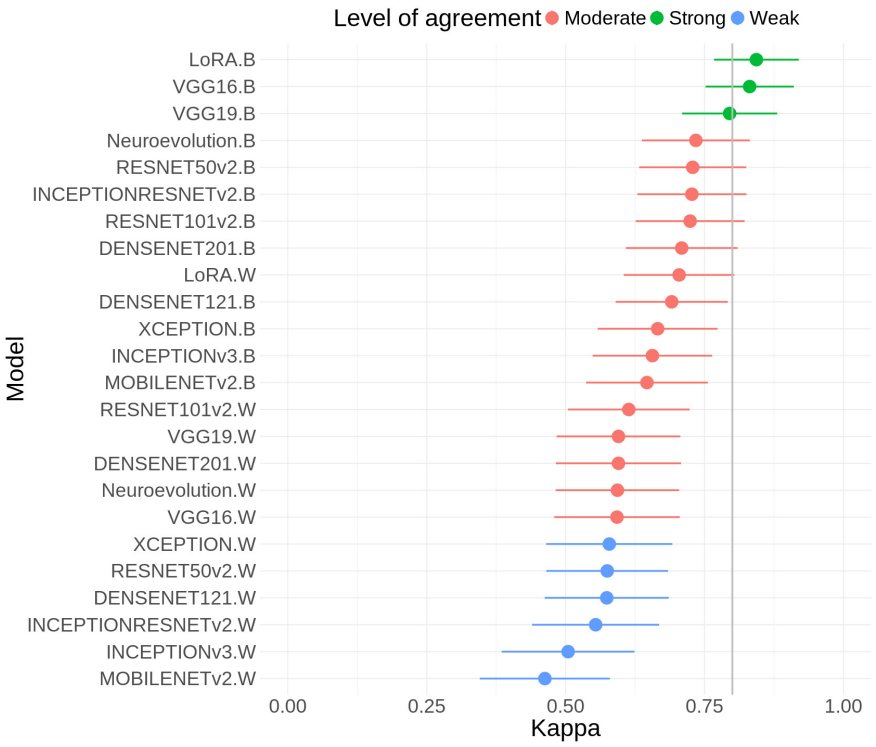


Figure 9: Cohen's kappa values in DFISDR.

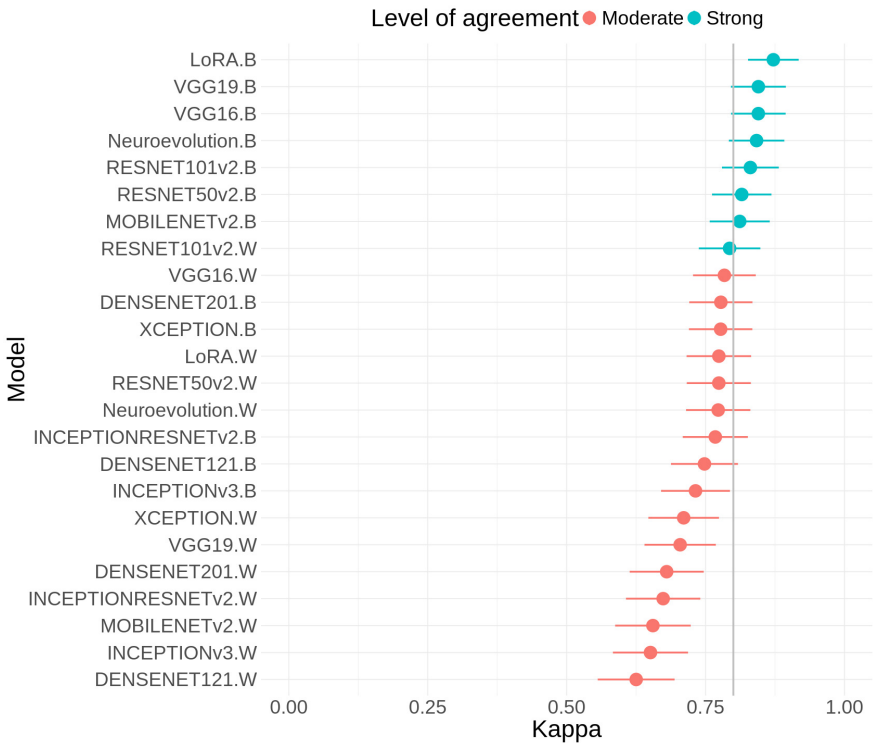


Figure 10: Cohen's kappa values in APTOS.

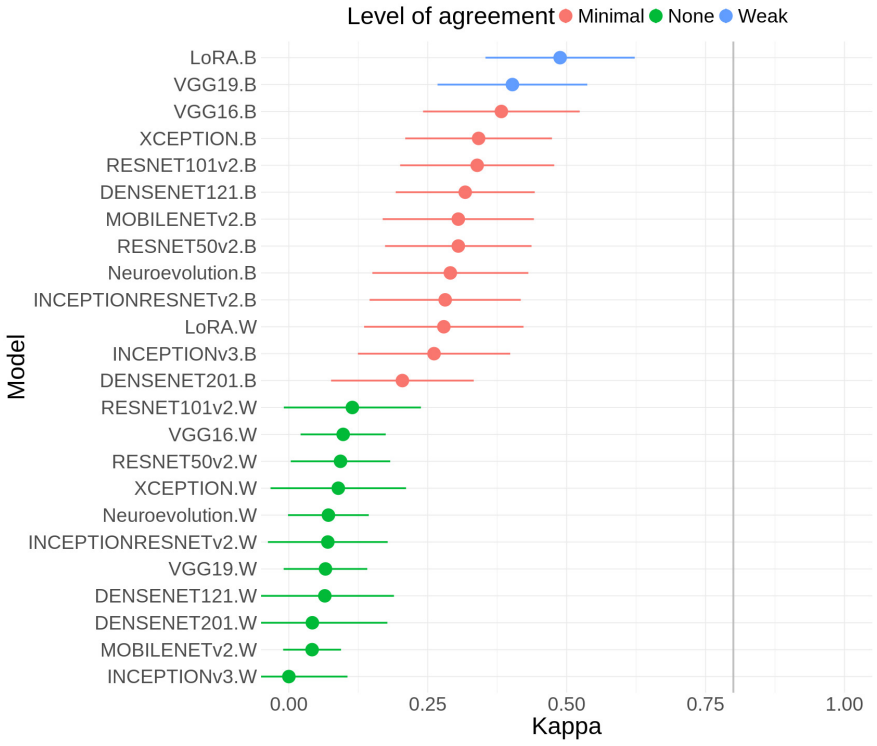


Figure 11: Cohen's kappa values in MESSIDOR.

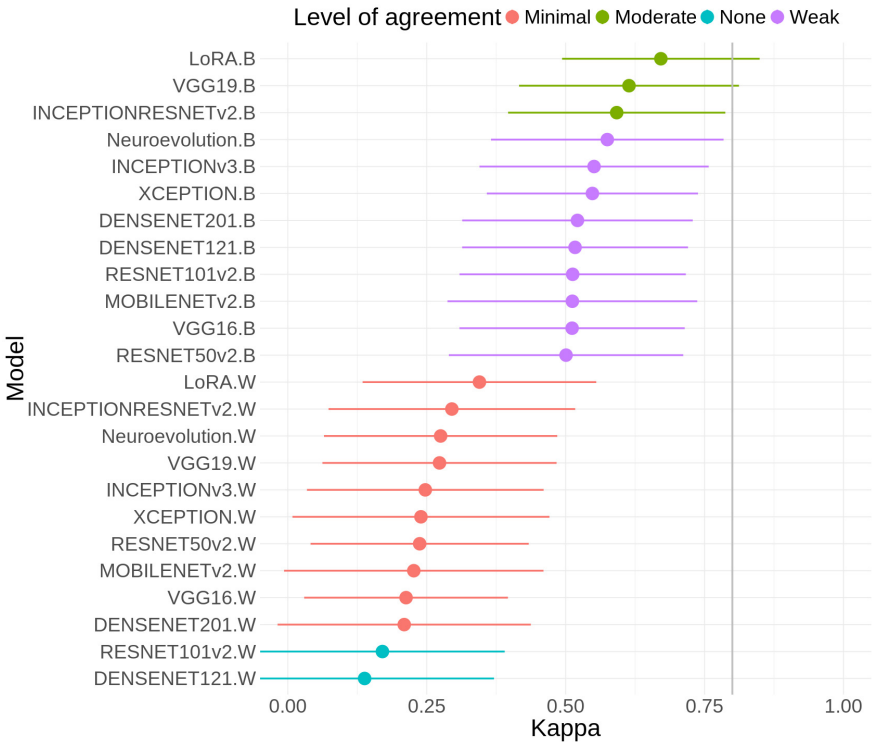


Figure 12: Cohen's kappa values in IDRID.

References

[1] S. Dutta, B. Manideep, S. M. Basha, R. D. Caytiles, and N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," *International Journal of Grid and Distributed Computing*, vol. 11, no. 1, pp. 89–106, 2018.

[2] B. Zhou, Y. Lu, K. Hajifathalian, J. Bentham, M. Di Cesare, G. Danaei, H. Bixby, M. J. Cowan, M. K. Ali, C. Taddei *et al.*, "Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4· 4 million participants," *The Lancet*, vol. 387, no. 10027, pp. 1513–1530, 2016.

[3] R. Lee, T. Y. Wong, and C. Sabanayagam, "Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss," *Eye and vision*, vol. 2, no. 1, pp. 1–25, 2015.

[4] R. J. Winder, P. J. Morrow, I. N. McRitchie, J. Bailie, and P. M. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Computerized medical imaging and graphics*, vol. 33, no. 8, pp. 608–622, 2009.

[5] V. ZELJKOVIĆ, M. Bojic, C. Tameze, and V. Valev, "Exudates detection and classification algorithm of diabetic patients'retina images," *Journal of Circuits, Systems and Computers*, vol. 22, no. 02, p. 1250087, 2013.

[6] Q. Mohamed, M. C. Gillies, and T. Y. Wong, "Management of diabetic retinopathy: a systematic review," *Jama*, vol. 298, no. 8, pp. 902–916, 2007.

[7] P. Patil, P. Shettar, P. Narayankar, and M. Patil, "An efficient method of detecting exudates in diabetic retinopathy: Using texture edge features," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2016, pp. 1188–1191.

[8] J. L. Vázquez Noguera, J. C. Mello-Román, D. P. Pinto-Roa, S. Gomez, J. Ayala, D. Aquino-Brítez, P. E. Gardel-Sotomayor, M. García-Torres, J. Facon, V. E. Castillo Benítez, I. C. Matto, and P. Pérez-Estigarribia, "Multiclass diabetic retinopathy classification of eye fundus images small datasets performance improvement – a neuroevolution approach," in *2023 XLIX Latin American Computer Conference (CLEI)*, 2023, pp. 1–10.

[9] D. K. Prasad, L. Vibha, and K. Venugopal, "Early detection of diabetic retinopathy from digital retinal fundus images," in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*. IEEE, 2015, pp. 240–245.

[10] M. Madhukar, D. Manjula, Y. Ujwal, J. Deepthi, and S. Seshacharan, "Exudate extraction: A review," *Perspectives in Communication, Embedded-systems and Signal-processing-PiCES*, vol. 1, no. 2, pp. 8–9, 2017.

[11] R. P. Danis and M. D. Davis, "Proliferative diabetic retinopathy," in *Diabetic Retinopathy*. Springer, 2008, pp. 29–65.

[12] S. S. A. Hassan, D. B. Bong, and M. Premsenthil, "Detection of neovascularization in diabetic retinopathy," *Journal of digital imaging*, vol. 25, no. 3, pp. 437–444, 2012.

[13] K. S. Kumar, D. Bhowmik, G. Harish, S. Duraivel, and B. P. Kumar, "Diabetic retinopathy-symptoms, causes, risk factors and treatment," *The Pharma Innovation*, vol. 1, no. 8, 2012.

[14] I. Qureshi, J. Ma, and Q. Abbas, "Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11 691–11 721, 2021.

[15] J.-H. Wu, T. A. Liu, W.-T. Hsu, J. H.-C. Ho, C.-C. Lee *et al.*, "Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis," *Journal of medical Internet research*, vol. 23, no. 7, p. e23863, 2021.

[16] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, p. 100377, 2020.

[17] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," 2021.

[18] S. Ferreira-Caballero, D. P. Pinto-Roa, J. L. V. Noguera, J. Ayala, P. E. Gardel-Sotomayor, and P. Pérez-Estigarribia, "Low-rank adaptation applied to multiclass diabetic retinopathy classification," in *2024 L Latin American Computer Conference (CLEI)*, 2024, pp. 1–9.

[19] K. C. Jordan, M. Menolotto, N. M. Bolster, I. A. Livingstone, and M. E. Giardini, "A review of feature-based retinal image analysis," *Expert Review of Ophthalmology*, vol. 12, no. 3, pp. 207–220, 2017.

[20] W. Mahon, M. Mitchell, J. Steinke, and M. Raben, "Microaneurysms in diabetic retinopathy," *British Medical Journal*, 1971.

[21] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic medicine*, vol. 19, no. 2, pp. 105–112, 2002.

[22] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija, "Automated screening system for diabetic retinopathy," in *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the*, vol. 2. IEEE, 2003, pp. 915–920.

[23] B. M. Ege, O. K. Hejlesen, O. V. Larsen, K. Møller, B. Jennings, D. Kerr, and D. A. Cavan, "Screening for diabetic retinopathy using computer based image analysis and statistical classification," *Computer methods and programs in biomedicine*, vol. 62, no. 3, pp. 165–175, 2000.

[24] H. J. Jelinek, M. J. Cree, D. Worsley, A. Luckie, and P. Nixon, "An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice," *Clinical and Experimental Optometry*, vol. 89, no. 5, pp. 299–305, 2006.

[25] O. Faust, R. Acharya, E. Y.-K. Ng, K.-H. Ng, and J. S. Suri, "Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review," *Journal of medical systems*, vol. 36, no. 1, pp. 145–157, 2012.

[26] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on medical imaging*, vol. 8, no. 3, pp. 263–269, 1989.

[27] M. Bolz, U. Schmidt-Erfurth, G. Deak, G. Mylonas, K. Kriechbaum, C. Scholda, D. R. R. G. Vienna *et al.*, "Optical coherence tomographic hyperreflective foci: a morphologic sign of lipid extravasation in diabetic macular edema," *Ophthalmology*, vol. 116, no. 5, pp. 914–920, 2009.

[28] A. Hunter, J. Lowell, J. Owens, L. Kennedy, and D. Steele, "Quantification of diabetic retinopathy using neural networks and sensitivity analysis," in *Artificial Neural Networks in Medicine and Biology*. Springer, 2000, pp. 81–86.

[29] H. Fujita, Y. Uchiyama, T. Nakagawa, D. Fukuoka, Y. Hatanaka, T. Hara, G. N. Lee, Y. Hayashi, Y. Ikedo, X. Gao *et al.*, "Computer-aided diagnosis: The emerging of three cad systems induced by japanese health care needs," *Computer methods and programs in biomedicine*, vol. 92, no. 3, pp. 238–248, 2008.

[30] S. C. Lee, E. T. Lee, R. M. Kingsley, Y. Wang, D. Russell, R. Klein, and A. Warn, "Comparison of diagnosis of early retinal lesions of diabetic retinopathy between a computer system and human experts," *Archives of Ophthalmology*, vol. 119, no. 4, pp. 509–515, 2001.

[31] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee, "An effective approach to detect lesions in color retinal images," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, vol. 2. IEEE, 2000, pp. 181–186.

[32] S. C. Lee, E. T. Lee, Y. Wang, R. Klein, R. M. Kingsley, and A. Warn, "Computer classification of nonproliferative diabetic retinopathy," *Archives of Ophthalmology*, vol. 123, no. 6, pp. 759–764, 2005.

[33] W. L. Yun, U. R. Acharya, Y. V. Venkatesh, C. Chee, L. C. Min, and E. Y. K. Ng, "Identification of different stages of diabetic retinopathy using retinal optical images," *Information sciences*, vol. 178, no. 1, pp. 106–121, 2008.

[34] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[35] S. Gayathri, V. P. Gopi, and P. Palanisamy, "A lightweight cnn for diabetic retinopathy classification from fundus images," *Biomedical Signal Processing and Control*, vol. 62, p. 102115, 2020.

[36] S. M. S. Islam, M. M. Hasan, and S. Abdullah, "Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images," *arXiv preprint arXiv:1812.10595*, 2018.

[37] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," *Pattern Recognition*, vol. 46, no. 1, pp. 107–116, 2013.

[38] P. Jalli, T. J. Hellstedt, and I. Immonen, "Early versus late staining of microaneurysms in fluorescein angiography." *Retina (Philadelphia, Pa.)*, vol. 17, no. 3, pp. 211–215, 1997.

[39] T. Hellstedt and I. Immonen, "Disappearance and formation rates of microaneurysms in early diabetic retinopathy." *British journal of ophthalmology*, vol. 80, no. 2, pp. 135–139, 1996.

[40] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning," *Addion wesley*, vol. 1989, no. 102, p. 36, 1989.

[41] G. V. Conroy, "Handbook of genetic algorithms by lawrence davis (ed.), chapman & hall, london, 1991, pp 385,£ 32.50." *The Knowledge Engineering Review*, vol. 6, no. 4, pp. 363–365, 1991.

[42] S. Johnson, J. Han, Y. Liu, L. Chen, and X. Wu, "Hybrid approach with improved genetic algorithm and simulated annealing for thesis sampling," *Future Internet*, vol. 10, no. 8, p. 71, 2018.

[43] N. Rehman, A. Khan, and Z. Ur Rehaman, "Parameter-efficient diabetic retinopathy grading with lora-based fine-tuning of vision foundational models," in *2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2024, pp. 284–291.

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[45] V. E. C. Benítez, I. C. Matto, J. C. M. Román, J. L. V. Noguera, M. García-Torres, J. Ayala, D. P. Pinto-Roa, P. E. Gardel-Sotomayor, J. Facon, and S. A. Grillo, "Dataset from fundus images for the study of diabetic retinopathy," *Data in Brief*, vol. 36, p. 107068, 2021.

[46] Karthik, Maggie, and S. Dane, "Aptos 2019 blindness detection," https://kaggle.com/competitions/aptos2019-blindness-detection, 2019, kaggle.

[47] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay *et al.*, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.

[48] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, 2018.

[49] D. A. Van Dyk and X.-L. Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.

[50] A. Mead, S. Burnett, and C. Davey, "Diabetic retinal screening in the uk," *Journal of the Royal Society of Medicine*, vol. 94, no. 3, pp. 127–129, 2001.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer, 2016, pp. 630–645.

[52] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[53] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[54] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[55] M.-H. Hsieh, M.-J. Hsieh, C.-M. Chen, C.-C. Hsieh, C.-M. Chao, and C.-C. Lai, "An artificial neural network model for predicting successful extubation in intensive care units," *Journal of clinical medicine*, vol. 7, no. 9, p. 240, 2018.

[56] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arxiv 2016," *arXiv preprint arXiv:1603.04467*, 2019.

[57] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," 2019.

[58] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, "Huggingface's transformers: State-of-the-art natural language processing," 2020.

[59] Q. Lhoest, A. V. del Moral, Y. Jernite, A. Thakur, P. von Platen, S. Patil, J. Chaumond, M. Drame, J. Plu, L. Tunstall, J. Davison, M. Šaško, G. Chhablani, B. Malik, S. Brandeis, T. L. Scao, V. Sanh, C. Xu, N. Patry, A. McMillan-Major, P. Schmid, S. Gugger, C. Delangue, T. Matussière, L. Debut, S. Bekman, P. Cistac, T. Goehringer, V. Mustar, F. Lagunas, A. M. Rush, and T. Wolf, "Datasets: A community library for natural language processing," 2021.

[60] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from www.tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[61] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[62] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds. Oxford: Academic Press, 2019, pp. 542–545. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012809633820349X

[63] V. L. Parsons, *Stratified Sampling*. Wiley, 2017, pp. 1–11.

[64] K. A. Naz, R. Jeena, and P. Niyas, "Deep neural network based multi-class arrhythmia classification," in *Intelligent Systems, Technologies and Applications*. Springer, 2021, pp. 223–235.

[65] A. J. Tallón-Ballesteros and J. C. Riquelme, "Data mining methods applied to a digital forensics task for supervised machine learning," in *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications*. Springer, 2014, pp. 413–428.

[66] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.