IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

Automated Detection of Diabetic Retinopathy from the Fundus Photography using Deep Learning Method

Balaji M[1], Harish B[2], Juhaif Ahamed H[3], Prof. Arunachalam R[4]
Students, Department of Computer Science and Engineering [1,2,3]
Assistant Professor, Department of Computer Science Engineering[4]
Anjalai Ammal Mahalingam Engineering College, Thiruvarur, India

Abstract: *Diabetic patients often experience a common disease known as Diabetic Retinopathy. This condition predominantly affects the retina, which is the light-sensitive tissue located at the back of the eye, by damaging the blood vessels that supply it. While in its early stages, Diabetic Retinopathy may not show any symptoms and can gradually lead to mild vision problems. It is crucial to detect the early stages of this disease automatically to prevent damage to the eyes and avoid vision loss. Therefore, the automatic detection of this condition is vital for early screening and diagnosis, which allows for timely treatment. Fundus cameras are used to capture retinal images, which can help in detecting and diagnosing Diabetic Retinopathy .This study proposes a method that utilizes deep learning to automatically identify the progression level of Diabetic Retinopathy. Two different deep learning architectures, namely ResNet, and Swin Transformer, were utilized in the experiment. The models were evaluated in terms of accuracy and network size, and the results were visualized using metrics like confusion matrix. The findings indicate that Swin Transformer can achieve better accuracy and robustness during classification.*

Keywords: Deep Learning, Image Pre-processing, RESNET, Swin Tranformer, Hyper-parameter tuning

I. INTRODUCTION

Diabetic retinopathy(DR) is currently a major cause of vision loss, often developing without symptoms, making timely detection and treatment challenging. Manual detection methods require experienced clinicians, but with limited advanced medical resources in many areas, many diabetic patients may miss their best treatment opportunities, leading to irreversible visual damage or even blindness. More efficient and low-cost methods for early detection of DR would be highly beneficial.

The objective of the project is to use advanced computer vision technology and AI algorithms to detect and classify the severity of diabetic retinopathy in patients, based on one or several fundus photos of the patient. The purpose of this project is to reduce the number of patients who are at risk of having undetected diabetic retinopathy, which can lead to blindness if left untreated. The classification system for the diagnosis of diabetic retinopathy is based on the standard of the American Academy of Ophthalmology[1], where a diagnosis of "0" indicates no risk of diabetic retinopathy, and a diagnosis of "1" to "4" represents increasing levels of severity of the disease. The ultimate goal of the project is to reduce the number of patients at risk of undetected diabetic retinopathy in the future..

II. LITERATURE SURVEY

Recent studies demonstrate that deep neural networks have the ability to accurately detect diabetic retinopathy and diabetic macular edema in retinal fundus images with high sensitivity and specificity, even without specifying lesion-based features, by using large datasets in their training [21][2][3].In order to obtain valuable insights on data, model and training/evaluation details, we have thoroughly examined various papers in this particular area. Among the publicly available datasets that are commonly used, EyePACS[4], DRD, MESSIDOR 2 and E-Ophtha[5] are included. While some datasets like EyePACS[6] with large data volume can be employed for model training and validation, others such as MESSIDOR 2 can be utilized for testing purposes as it may have a smaller size but encompasses actual data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

collected from hospitals. The datasets mentioned above do not provide adequate information for our classification task on disease progression because they only indicate the presence or absence of the disease and do not indicate the progression level. Various preprocessing techniques were used by researchers on the datasets, such as resizing images into square shapes with sizes of 512x512 pixels[4], 299x299 pixels[2], or 224x224 pixels[7][21]. Additional techniques include normalizing image pixel values to a range of 0 and 1[4], or subtracting the mean and dividing the variance from the train image datasets[7]. Furthermore, rotation augmentation can also be applied due to the circular shape of retina fundus, which should be rotation-invariant. We can increase the size of the dataset using data augmentation methods such as randomly rotating images, cropping them to a square, and adjusting their brightness, before feeding them into our model. [4][7]. Regarding the model, many researchers have used CNN to tackle the problem at hand. Some studies developed their own customized CNN models, such as a 6-layer CNN[4]. Other models utilized lesion or red lesion detection for classification[5]. Moreover, several papers proposed transfer learning methods using pre-trained modern CNNs on ImageNET, including VGG-16, ResNet-18, GoogLeNet, DenseNet-121, and SE-BN-Inception [7][8][9][10] . The optimization functions employed varied, with examples including RMSProp[11] and SGD[7], which resulted in similar and accurate results. The learning rate used ranged from 1e-1 to 1e-4[4][5][7][11], and a specific weight decay of 4e-5 was identified as a hyperparameter[5]. Additionally, it was noted that sensitivity, specificity, accuracy, and confusion matrix are widely used and reliable evaluation metrics, which will also be evaluated in our models.

III. METHODOLOGY

In this project, Resnet and Swin Transformer are selected as our baseline models.. More detailed information about these models are provided in section B,C.

3.1 Training Methods

Our goal is to create a deep learning model capable of accurately identifying the progression level in retino images. To achieve this, we plan to use a CNN model due to its exceptional performance in image classification tasks. We will use transfer learning, as demonstrated in studies [11] and [12], which has shown better accuracy than non-transferring learning methodologies in DR image classification. Our next steps involve experimenting with different CNN model structures, including Resnet 50,andSwin Transformer, which are all well-suited for various classification tasks. We selected Resnet 50 because it has been highlighted in numerous related papers. Lastly, we also chose the modern architecture of Swin Transformer, which has performed well in other classification tasks. We will also perform hyper-parameter tuning and fine-tuning to further improve our model's performance.

3.2 Resnet

Resnet 50 is a highly prevalent neural network architecture that includes 50 layers and is primarily a convolutional neural network[14]. We utilized pretrained weights from the Imagenet dataset to ensure that the model had already learned various fundamental features, and we employed starter code to establish our pretrained Resnet50 model[13]. Resnet 50 consists of 48 convolutional layers, one max pooling layer, and one average pool layer, and its residual connection between layers helps alleviate the accuracy saturation problem. The shortcut connection introduced performs identity mappings, and the basic concept of this shortcut connection is illustrated in Figure 1. The shortcut in Resnet 50 bypasses three layers, and a 1*1 convolutional layer is included along the way. This architecture is widely utilized in numerous computer vision tasks, including image classification, object localisation, and object detection.

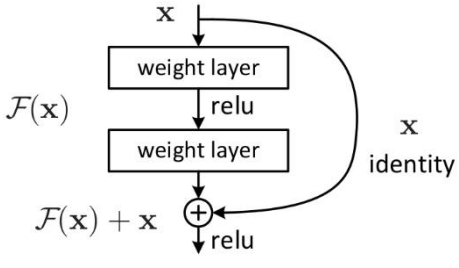
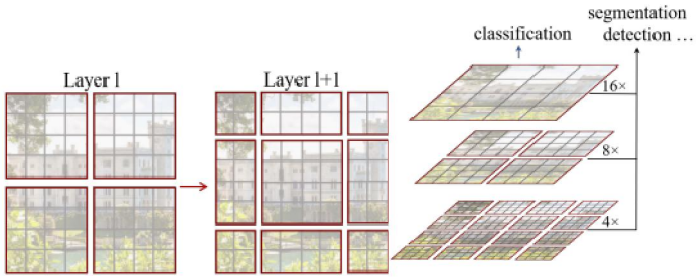


Figure 1: Basic Resnet 50 structure block

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

3.3 Swin Transformer

Recent research suggests that Vision Transformer (ViT) may not be as effective as Swin Transformer (introduced in 2021) in solving computer vision tasks [17]. This is because ViT uses fixed scale word tokens for NLP tasks, whereas computer vision tasks require elements that can vary greatly in scale and resolution. Additionally, ViT's computational complexity increases quadratically with image size, which can make it challenging to use for dense computer vision tasks. In contrast, Swin Transformer overcomes this issue by using a modified transformer that computes self-attention locally within non-overlapping windows, with a fixed number of patches in each window, resulting in complexity increasing linearly as image size increases. Swin Transformer also employs a shifted window partition between consecutive self-attention layers, which enhances the model's representation power. Lastly, Swin Transformer uses hierarchical feature maps to conveniently leverage advanced techniques for dense prediction, making it well-suited for tasks such as Diabetic Retinopathy Progression Recognition.



(a)Shifted Window (b) Hierarchical Feature Maps

Figure 2: Advanced Features of Swin Transformer [18]

The Swin Transformer model we chose to utilize is based on the Swin-B architecture, as depicted in Figure 3. To begin, an RGB image with dimensions 224x224 is divided into patches utilizing a patch splitting module. We have used a patch size of 4x4, as recommended in the original paper, resulting in each patch initially having a dimension of 48 (4x4x3). Next, this dimension is transformed linearly from 48 to c=128, which is specific to the Swin-B version. Four stages of Swin Transformer blocks are then applied to execute the feature transformation. Each stage comprises 2, 2, 18, 2 transformer blocks respectively. A patch merging layer at the beginning of each stage is utilized to generate the hierarchical representation. For instance, in stage 2, the merging layer concatenates nearby 2x2 patches and applies a linear transformation to produce an output with a shape of 28x28x256. As the neural network gets deeper, this process is repeated until stage 4 generates an output with dimensions of 7x7x1024. Finally, a global average pooling is applied to reduce the output to one dimension, which is then linearly transformed into vector scores for the five required classes, as illustrated in Figure 3.

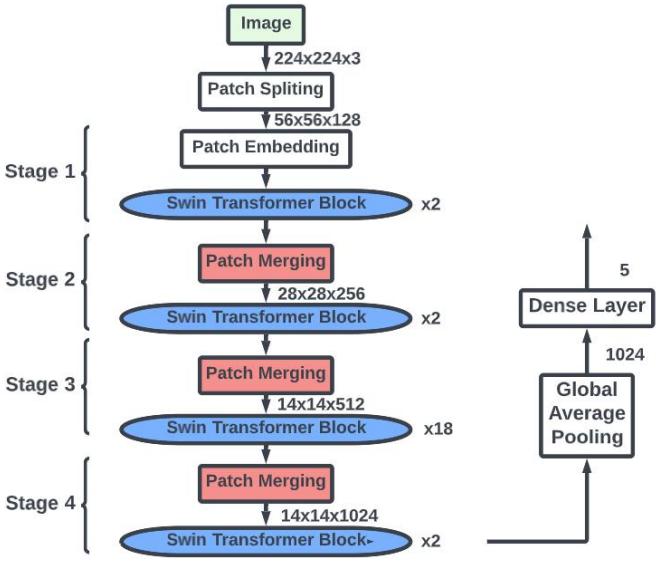


Figure 3: Architecture of Swin-B Transformer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 7.301
Volume 3, Issue 4, April 2023

IV. DATASET AND FEATURES

We opted to use public datasets from Kaggle instead of collecting our own data because the available data was sufficient for our experiment and we had limited access to clinical data. The training dataset we used was from the Diabetic Retinopathy Detection competition, while the test dataset was from the APTOS 2019 Blindness Detection competition. The dataset contained images of retinas that were classified based on the severity level of diabetic retinopathy. We found that there were more class 0 images than class 3 or 4 images in the dataset, which could result in the model only learning class 0 well and classifying most samples as class 0 while still maintaining a high accuracy. To address this issue, we generated an augmented dataset by randomly selecting 1000 images from each of level 0, level 1, and level 2 and performing data augmentation techniques such as horizontal and vertical flipping and brightness adjustments. We also included all level 3 and level 4 images in the augmented dataset. Finally, we center cropped all images to 224 * 224 pixels and obtained a dataset of around 15000 images, which we split into a training and validation set with an 8:2 ratio. Some example images in our dataset are shown below.

Figure 4: Level 0 Example

Figure 5: Level 2 Example

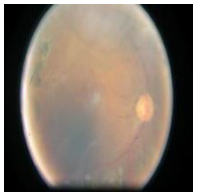


Figure 6: Level 4 Example

V. IMPLEMENTATION AND RESULTS

5.1 Resnet

Below are the hyper-parameters for the Resnet 50 model we trained. These hyper-parameters were chosen based on the results of various experiments. We initially referred to related works and then made adjustments based on our training results and conditions. Due to limited memory access, we could not set a large batch size.

We obtained the final loss and accuracy plots, as well as the confusion matrix, which are displayed below. However, the performance of Resnet 50 was not very satisfactory, as it achieved a maximum validation accuracy of only 55% before the algorithm performed early stopping to prevent overfitting to the training data.

Epoch: 30
Batch Size: 8
Warmup Epoch: 2
Learning rate: 0.0001
Warmup Learning Rate: 0.001

Table 1: Hyper-parameters in Resnet 50 training

Additionally, we plotted the diagnostic results of a small test set (details included in the Comparison section) and found that the model classified many level 1 and level 2 images as level 0. Our model did not perform satisfactorily, and we

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

attribute this to the minor variations between level 1 and level 2 retino images and healthy retino images. These differences were challenging for our model to identify, leading to its failure to capture them effectively.

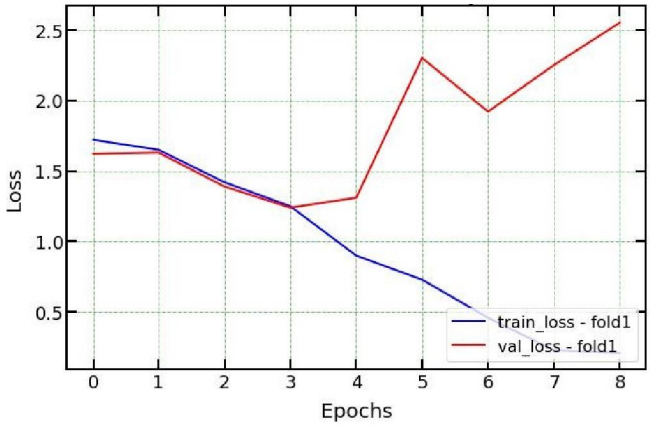


Figure 7: Loss Plot of Resnet 50 Model

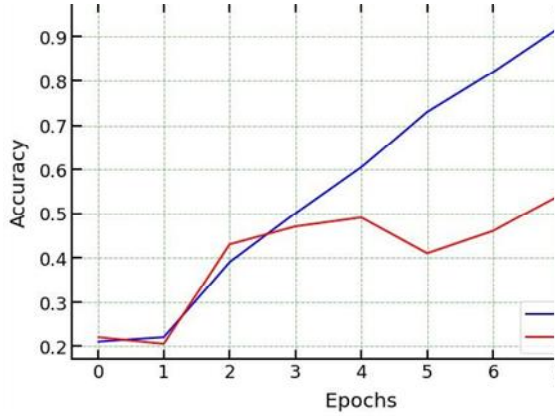


Figure 8: Accuracy Plot of Resnet 50 Model

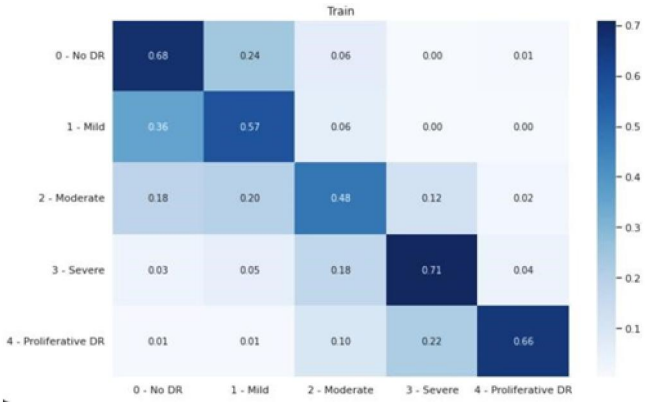


Figure 9: Confusion Matrix of Resnet 50

5.2 Swin Transformer

The table below displays the optimal hyper-parameter combination, which we selected by meticulously adjusting each parameter over a wide range of values. Due to limitations in computation power, we chose a batch size of 32. Additionally, we utilized an exponential learning rate schedule with a decay step of 100 and a decay rate of 0.95. This means that the learning rate decreases by a factor of 0.95 every 100 steps. During the training and validation processes, we observed the loss and accuracy, as shown in Figures 10 and 11. Compared to ResNet, Swin Transformer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

demonstrated significantly superior validation accuracy in the final epoch. We halted the training after 10 epochs because the validation accuracy plateaued at around 0.8. Additionally, after the seventh epoch, the validation loss began to increase, indicating that our model was overfitting to the dataset. As a result, stopping training after the tenth epoch was a wise decision.

After training, we evaluated the Swin Transformer model on the validation set and generated the confusion matrix, which is presented in Figure 12. The accuracy distribution was imbalanced, and the model had difficulty distinguishing between class 2, 3, and 4.

Epoch: 10
Batch Size: 32
Exp Decay Steps: 100
Learning rate: 0.001
Exp Decay Rate: 0.95

Table 2: Hyper-parameters in Swin Transformer training

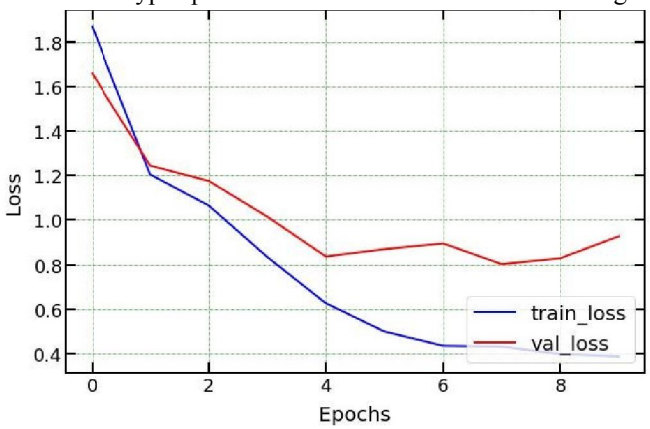


Figure 10: Loss Plot of Swin-B Transformer

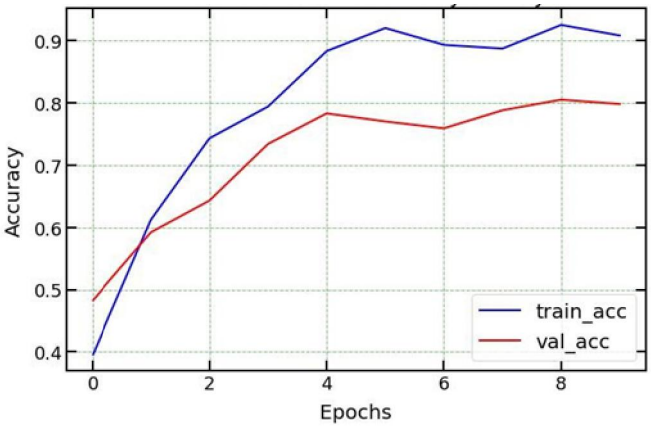


Figure 11: Accuracy Plot of Swin-B Transformer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

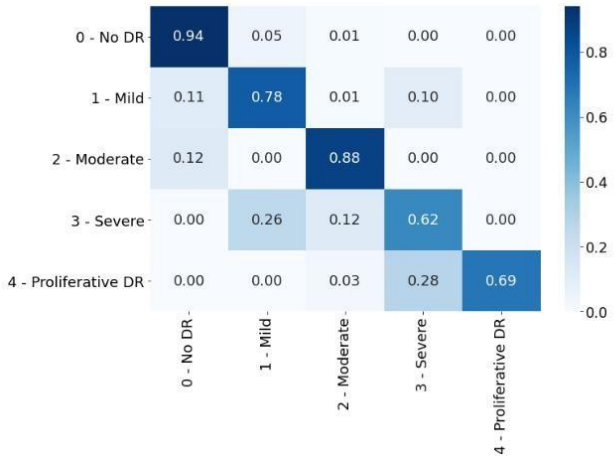


Figure 12: Confusion Matrix for Swin-B Transformer

5.3 Comparison

We evaluated our two models using three metrics: network size (parameter numbers), test accuracy, and balanced accuracy. The models were tested on a subset of 2000 samples from the APTOS-2019 Blindness Detection dataset [20], which had an imbalanced data distribution across its different classes (approximately 9:2:5:1:1). The results, presented in Table 3, revealed that Resnet 50 had the smallest network size and the fewest parameters among the two models. However, it was also the least robust to unbalanced data. On the other hand, Swin Transformer had considerably more parameters and achieved the highest accuracy among the two models. It is important to note, though, that the composition of the test dataset may have also played a role in Swin Transformer's performance advantage, as the majority of the test data came from classes 0, 1, and 2, which Swin Transformer was better at distinguishing. Conversely, classes 3 and 4 had smaller proportions in the test set, resulting in less contribution to the overall results..

Model	Parameters	Test Accuracy	Balanced Accuracy
Resnet50	27,794,320	54.30%	39.49%
Swin-B Transformer	88,109,370	76.45%	71.40%

Table 3: Comparison

In summary, while Resnet 50 is less competitive than Swin Transformer in terms of accuracy, the latter's better performance comes at a computational and time cost. Therefore, there exists a trade-off between accuracy and efficiency when choosing between the two models.

VI. CONCLUSION

Our experiment shows that Swin-B Transformer outperforms ResNet in terms of test accuracy and balanced accuracy, despite having more parameters. The additional parameters enhance the model's representation power, facilitating its ability to generalize to the training samples. Swin Transformer combines the strengths of CNNs and Transformers, using hierarchical representation for scale-invariance and self-attention to model data dependencies. Due to time constraints, we only examined two deep learning architectures. In the future, we plan to explore more advanced architectures, acquire more data, and train the models for extended periods to achieve even better performance.

VII. ACKNOWLEDGMENT

I would like to take this opportunity to express my heartfelt gratitude to all those who have supported methroughout the research project report. I am truly thankful for their unwavering guidance, invaluable constructive criticism, and friendly advice during the course of my project work. Their honest and insightful views on various project-related matters have been immensely helpful. I am also grateful to Principal Dr. S. N. Ramaswamy and the management of AAMEC for their continuous support and encouragement. My sincere indebtedness goes to my Head of Department, Dr. K. Velmurugan, and my guide, Asst. Professor Mr. R. Arunachalam, for their unwavering guidance, constant

supervision, and provision of necessary information throughout the project. I am also grateful to the review committee for their valuable suggestions and feedback. Furthermore, I extend my thanks to the laboratory staff for their valuable support. Last but not least, I sincerely appreciate the teaching and non-teaching staff from AAMEC who have contributed in various ways to myendeavor.

REFERENCES

[1]. American Academy of Ophthalmology. International Clinical Diabetic Retinopathy Disease Severity Scale DetailedTable.http://www.icoph.org/dynamic/attachments/resources/diabetic-retinopathy-detail.pdf.

[2]. Voets, Mike, et al. "Reproduction Study Using Public Data of: Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." PLOS ONE, Public Library of Science, https://journals.plos.org/plosone/article?id=10.13712-Fjournal.pone.0217541

[3]. "Detection of Diabetic Retinopathy Using Deep Learning Analysis." Retina Today, Bryn MawrCommunications,https://retinatoday.com/articles/2021-sept/detection-of-diabetic-retinopathy-using-deep-learning-analysis.

[4]. T, GargeyaR;Leng. "Automated Identification of Diabetic Retinopathy Using Deep Learning." Ophthalmology, U.S. National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/28359545/.

[5]. Alyoubi, Wejdan Let al. "Diabetic Retinopathy Detection through Deep Learning Techniques: A Review." Informatics in Medicine Unlocked, Elsevier, 20 June 2020, https://www.sciencedirect .com/science/article/pii/ S2352914820302069.

[6]. Bora, Ashish; "Predicting the Risk of Developing Diabetic Retinopathy Using DeepLearning." The Lancet Digital Health. https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30250-8/fulltext.

[7]. Li, Tao, et al. "Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screen-ing." Information Sciences, Elsevier, 5 June 2019, https://www.sciencedirect.com/science/article/pii/-S0020025519305377.

[8]. "Automatic Screening of Fundus Images Using a Combination of Convolutional Neural Network and Hand-Crafted Features." IEEE Xplore, https://ieeexplore.ieee.org/document/8857073.

[9]. Ayala, Angel; Figueroa, Thomas Ortiz; Fernandes, Bruno; Cruz, Francisco (2021-11). "Diabetic Retinopa-thy Improved Detection Using Deep Learning" (PDF).

[10]. Nguyen, Quang H., et al. "Diabetic Retinopathy Detection Using Deep Learning:Pro-ceedings of the 4th International Confer-enceon Machine Learning and Soft Computing." ACM Other Conferences, 1 Jan. 2020, https://dl.acm.org/doi/10.1145/3380688.3380709.

[11]. Arcadu, Filippo, et al. "Deep Learning Algorithm Pre-dicts Diabetic Retinopathy Progression in Individual Patients." Nature News, Nature Publishing Group, 20 Sept. 2019, https://www.nature.com/articles/s41746-019-0172-3.

[12]. Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. (2017-05-24). "ImageNet classification with deep convolutional neural networks" (PDF). Communica-tions of the ACM. 60 (6): 84–90. doi:10.1145/3065386. ISSN 0001-0782.

[13]. dimitreOliveira."Dimitreo-liveira/aptos2019blindnessdetection::3rd place medal: (Bronze Medal - 163rd Place Repository for the 'Aptos 2019 Blindness Detection' Kaggle Competition." GitHub, https://github.com/dimitreOliveira/APTOS2019BlindnessDetection.git.

[14]. Kaushik, Aakash. "Understanding Resnet50 Architecture." OpenGenus IQ: Computing Ex-pertise amp; Legacy, OpenGenus IQ: Com-puting Expertise amp; Legacy, 21 July 2020, https://iq.opengenus.org/resnet50-architecture/.

[15]. Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

[16]. Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., and Le, Q. V. MnasNet: Platform-aware neural architecture search for mobile. CVPR, 2019.

[17]. SrinadhBhojanapalli, Ayan Chakrabarti, Daniel Glasner, Daliang Li, Thomas Unterthiner, and Andreas Veit. Understanding Robustness of Transformers for Image Classification. CoRR. abs/2103.14586. 2021.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 4, April 2023
Impact Factor: 7.301

[18]. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer us-ing shifted windows. International Conference on Computer Vision (ICCV), 2021.

[19]. Kaggle Diabetic Retinopathy Detection compe-tition. https://www.kaggle.com/competitions/diabetic-retinopathy-detection

[20]. APTOS 2019 Blindness Detection competition. https://www.kaggle.com/competitions/aptos2019-blindness-detection/data

[21]. Zhou, Y., Wang, B., Huang, L., Cui, S., & Shao, L. (2021). A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability. IEEE Transactions on Medical Imaging, 40(3), 818–828. doi:10.1109/tmi.2020.3037771