Early Detection of Diabetic Retinopathy Using ResNet-18

Rohan Doggalli, Aakash Deep, Rohith Naik V, Vinay B Achari, Vijaykumar Muttagi and Uday Kulkarni
School of Computer Science and Engineering, KLE Technological University, Hubballi, India

Keywords: Diabetic Retinopathy, Deep Learning, ResNet-18, Automated Screening, Retinal Image Analysis, Clinical Decision Support System.

Abstract: Diabetic retinopathy (DR) is a leading cause of preventable blindness, especially among diabetic patients. Early diagnosis is critical to halt its progression and prevent vision loss. This work leverages deep learning, specifically the ResNet-18 model, to detect DR from retinal images. Using a Kaggle dataset divided into training and validation sets, the model achieved a training accuracy of 98.57% and a validation ac- curacy of 83.49%. These findings underscore the efficacy of ResNet-18 in automating DR detection. Integrating such technology into clinical workflows has the potential to enhance early screening and treatment strategies, improving patient outcomes while optimizing healthcare re- sources.

1 INTRODUCTION

Diabetic retinopathy is a severe complication of diabetes that affects the fragile blood vessels of the retina and can cause vision loss or even blindness if left untreated. DR progresses through stages: the first one being nonproliferative di- abetic retinopathy, the earliest stage characterized by leaking and swelling blood vessels. The more advanced stage, proliferative diabetic retinopathy (PDR), is characterized by the abnormal proliferation of blood vessels, leading to detach- ment of the retina, bleeding, and irreversible vision loss.

According to the World Health Organization, more than 420 million people worldwide suffer from diabetes, and this number is expected to surge exponen- tially in the near future (Nirgude, Revathi, et al. , 2024). Diabetic retinopathy remains one of the leading causes of preventable blindness globally. In 2010 alone, it caused 0.8 million cases of blindness and 3.7 million cases of visual impairment worldwide (Bourne, Price, et al. , 2012), (Solomon, Chew, et al. , 2017). By 2030, the number of DR patients is projected to rise to 191 million, with a prevalence rate of 27% globally from 2015 to 2019 (Teo, Tham, et al. , 2021), (Yau, Rogers, et al. , 2012). These statistics underscore the urgent need for effective early detection and timely treatment to prevent vision loss.

Early diagnosis of DR is essential because the disease process can be pre- vented in its early stages if timely intervention is performed. Early DR can be controlled with laser treatment, anti-VEGF injections, and vitrectomy that stops the advancement of the disease. Traditional screening of DR has proven to be a very tedious process, laborious, and prone to human errors. The current scenario among ophthalmologists is retinal images analysis, and due to this, there have been delayed diagnoses, cases left behind, and overloads in healthcare resources. With an increase in the incidence of diabetes and the prevalence of diabetic retinopathy projected to increase, there is a dire need for automated systems that improve the screening process to give faster and more accurate diagnoses in support of early detection efforts(Abràmoff, 2020), (Cheung, Ikram, et al. , 2015).

Machine learning (ML) and deep learning (DL) have emerged as revolution- ary tools in the medical field, particularly in the automatic detection of DR. Con- volutional Neural Networks (CNNs) are a class of deep learning models that have shown impressive performance in the analysis of retinal images by autonomously learning complex features from large datasets. These models can identify even the subtlest signs of DR in its early stages, far outperforming traditional meth- ods in terms of accuracy and efficiency (Gulshan, Peng, et al. , 2016), (Krizhevsky,

Doggalli, R., Deep, A., Naik V, R., B Achari, V., Muttagi, V. and Kulkarni, U.
Early Detection of Diabetic Retinopathy Using ResNet-18.
DOI: 10.5220/0013609200004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 81-90
ISBN: 978-989-758-763-4

Sutskever, et al. , 2012). Yet, there are still challenges in applying DL in DR detection: a need for large annotated datasets to train, over- fitting issues when trained on limited data, and a lack of interpretability, which limits clinical adoption (Ward, Maselko, et al. , 2017).

This work focuses on using ResNet-18, an efficient deep learning architecture based on the residual network design, to better identify diabetic retinopathy early. The architecture of ResNet-18, better suited to the image classification task because it prevents the vanishing gradient problem with residual connec- tions, results in better training of deep networks. The model's ability to learn complex features from medical image datasets makes it a powerful tool for iden- tifying subtle patterns indicative of DR, even in its early stages (Zhang, Ren, et al., 2016). Advanced techniques like image preprocessing, data augmentation, and hyperparameter optimization are incorporated into the proposed system with the aim of enhanc- ing model robustness and generalization, thus improving its ability to detect DR across diverse populations and imaging modalities. It aims at addressing the critical problem of model transparency besides the limited annotated data to overcome the challenges this presents. This means the system develops a model that, besides providing the correct predictions, also allows interpretability us- ing techniques like Class Activation Maps (CAMs), hence gaining trust among healthcare professionals. This approach allows ophthalmologists to understand how the model comes to its conclusion, thereby facilitating better integration of AI-based tools into clinical decision-making processes (Ward, Maselko, et al. , 2017).

Deep learning models suffer from several major challenges. One major issue is their dependence on large annotated datasets for training. Annotated data are often either unavailable or scarce in many healthcare settings, thereby preventing such models from generalizing well to different populations and imaging modal- ities. This problem leads to overfitting, mainly when the models are trained on less or biased datasets, resulting in poor accuracy when tested with new or varied data. Further, the issue of uninterpretability is another huge challenge for deep learning models to be used in clinical practice at large scales. For medical professionals to trust and appropriately use AI-based systems, they need to be aware of how the models generate their predictions (Ward, Maselko, et al. , 2017).

Over the years, research studies have been conducted that have explored the possibility of deep learning in diabetic retinopathy detection and have shown promising results. Gulshan et al. (2016) developed a deep learning algorithm that gained diagnostic accuracy equal to an expert ophthalmologist in DR identifica- tion from retinal images (Abràmoff, 2020). Leibig et al. demonstrated in 2017 the superiority of deep learning models compared to traditional methods for screening DR (Ward, Maselko, et al. , 2017). The remaining challenges include dataset variability, high computational require- ments from deep learning models, and the lack of model transparency, which have significantly prevented the wider clinical applicability of these technologies.

2 LITERATURE WORK:

Diabetic Retinopathy (DR) has become an extremely active area of research due to its severe contribution to blindness in diabetic patients. Timely detec- tion and treatment play a crucial role in the prevention of permanent blindness in many patients, which puts emphasis on early diagnosis. Traditional methods for DR detection involved the visual inspection of retinal fundus images by us- ing techniques like thresholding, edge detection, and region growing. However, these conventional approaches are unable to deal with intrinsic variability and complexity in the retinal images. Some of this variability includes varying illu- mination, noise, and dimensions of lesions that become challenging to diagnose accurately (Yau, Rogers, et al. , 2012).

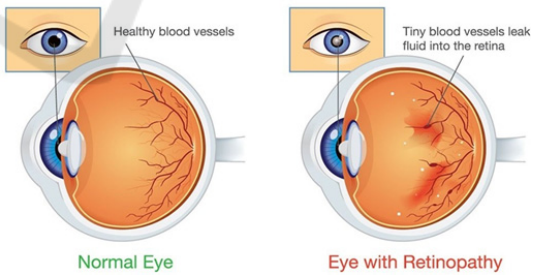

Figure 1: Healthy eye and Diabetic Retinopathy

Fig 1 presents a normal eye alongside a diabetic retinopathy-affected eye. The normal eye blood vessels look healthy and intact. Consequently, proper functioning is provided by these vessels. The DR-afflicted eye presents damaged vessels leaking fluids to the retina, leading to swelling and resulting loss of vision. These pathological changes are characteristic of DR and underlie the importance of early detection and intervention to prevent irreversible damage. This

visual representation emphasizes the need for advanced diagnostic tools, such as deep learning models, in order to identify the condition at its earliest stages to provide prompt treatment.

With the advent of ML and DL technologies, much progress has been made in automating the detection of DR. CNNs are currently identified as the best tech- niques to automatically identify DR because it can learn hierarchies of features from raw pixel data (Gulshan, Peng, et al. , 2016). Models like ResNet, Inception, and VGG, which are CNN-based, have achieved outstanding accuracy in the classification of fundus images into the different DR stages (Zhang, Ren, et al. , 2016). These models were trained and vali- dated on public datasets, including the Kaggle diabetic retinopathy competition dataset, thus making them more generalized and reproduce in real-world settings (Lu, Liu, et al. , 2018).

Among the new developments in the field, there is also the application of ResNet-18. It is a deep learning architecture, presented by He et al. (2016) for residual learning. The success of ResNet-18 has been achieved in different ap- plications, including diabetic retinopathy detection from images of the retina. By applying residual connections, the architecture is able to avoid the vanishing gradient problem, and hence train very deep networks. This innovative archi- tecture has been applied successfully to the analysis of medical images, and it is therefore a very powerful tool in the detection of subtle patterns that can indicate DR at its earliest stages (Zhang, Ren, et al. , 2016). Success of ResNet-18 in DR detection and a relatively lightweight structure of the model make it a good candidate for healthcare applications where accuracy and efficiency are equally critical.

Recent advances in transfer learning and attention mechanisms have further improved the performance of DR detection systems. Transfer learning, which fine-tunes pre-trained models on DR datasets, has been shown to achieve high accuracy even with limited labeled data (Lu, Liu, et al. , 2018). In addition, attention mecha- nisms such as self-attention and saliency maps allow models to focus on clinically relevant regions of retinal images, thus improving both interpretability and diag- nostic accuracy (Jia, Li, et al. , 2019). These advancements ensure that deep learning models are more capable of distinguishing between subtle features in retinal images, making them more suitable for early detection of DR. Data augmentation and multi-task learning techniques have also been used to handle imbalanced datasets and im- prove model robustness. However, several challenges remain. One major issue is the generalizability of DR detection models across different imaging de- vices, populations, and datasets. Models trained on single-source datasets tend to overfit and thus perform poorly if exposed to new or diverse data (Nirgude, Revathi, et al. , 2024). A sec- ond major problem is that deep learning models lack interpretability. Healthcare practitioners often do not embrace AI-driven tools because of opaque decision-making processes. Efforts have been made to explain these models better with XAI techniques, like Grad-CAM and saliency maps, for the sake of increased transparency of the models in this concept (Ward, Maselko, et al. , 2017)]. Another issue that still exists in this field is data imbalance, especially for underrepresented stages of diabetic retinopathy. Many datasets lack good representations of samples from early or severe stages of the disease, which makes training and evaluation difficult. Advanced preprocessing and augmentation strategies are being used to address these issues to improve model performance across different datasets(Bourne, Price, et al. , 2012).

These gaps can be filled by focusing on model generalization using robust training strategies, including advanced preprocessing, data augmentation, and hyperparameter optimization techniques in the proposed methodology.

Using ResNet-18 and transferring its ability into DR detection with high ac- curacy, it will attempt to diagnose the disease at its earliest stage. Furthermore, the incorporation of attention mechanisms is explored to improve both diagnos- tic accuracy and interpretability. This work thus contributes to a global effort aimed at reducing loss of vision due to DR by proposing an automated solution for efficient and scalable DR detection at an early stage. In summary, although deep learning has really revolutionized the detection of diabetic retinopathy, challenges like generalizability, interpretability, and data imbalance still pose se- rious issues. The gaps in these areas are likely to be bridged in the near future through improved model generalization using architectures such as ResNet-18, using explainability techniques, and enhancing the data strategies so that such models can be made clinically viable and reliable for large-scale use in healthcare systems.

3 PROPOSED WORK:

3.1 Data Collection

In the first important milestone of the project, there is the provision of data input with the sourcing of

retinal fundus images ahead of processing by further stages. For this work, the Aptos 2019 Blindness Detection dataset(Aptos, 2019) is used. This dataset includes a collection of high-resolution images of the retina with labels categorized as five types of diabetic retinopathy on a spectrum of severity: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR.in Fig 2.

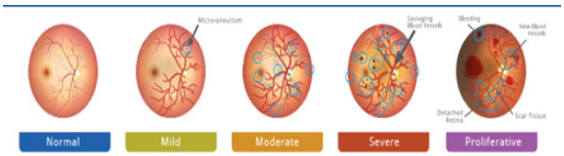

Fig. 2. Stages of Diabetic Retinopathy

All images are in PNG, ensuring high fidelity and resolution—the require- ments for a deep learning model, like ResNet-18, to correctly identify the patterns with the subtle and complex information indicative of diabetic retinopathy. Once in the system, the image passes through a series of pre-processing oper- ations that have improved consistency and quality in various operations such as standardization, cleaning of the images, normalization, and augmentation.

Presenting high-quality and standardized images, the model can obtain robust learning and generate accurate predictions for early detection and classification of diabetic retinopathy.

3.2 Data Preprocessing

Among all these steps toward preparation of retinal images for training, it is undeniable that preprocessing is one of the most important ones, especially when detecting diabetic retinopathy (DR). It actually depends on the quality of input images - hence, performance of the model. In order to optimize the dataset and to ensure that the data will be very much consistent with high quality while training, there are several preprocessing techniques, such as image cleaning, resizing, normalization, and data augmentation.

Image Cleaning: The pre-processing stage in this regard is image cleaning. This step cleans the retinal images of noise, artifacts, and distortions. Because the raw retinal images can be captured under varying conditions and on different devices, pixelation, irrelevant background noises, etc. may confuse the model. This stage enhances important features like blood vessels, hemorrhages, and microaneurysms that are important in DR, but suppresses irrelevant patterns. This is because the improvement in the signal-to-noise ratio allows the model to capture the weak changes in the retina efficiently, which may not be easy to decide and diagnose cases of early DR.

All the images are resized to the same dimension of 256x256 pixels. For a deep learning model like ResNet-18, this fixed input dimension is vital. The resolution chosen has been a balance between computational efficiency and preserving de- tails. It is large enough to preserve important retinal features so that the model could correctly identify and classify the stages of DR but not large enough to be computationally manageable.

Normalization: This step rescales pixel values of images to be within the range of 0 and 1. That way, the input to the neural network will become standardized. Without normalization, big variations in pixel values would cause instability in the training process, leading the model to converge at a slow speed. Normalizing the pixel value means that the model becomes effective at processing data with quicker convergence and better overall performance.

Data Augmentation : Techniques of data augmentation are used to make the model more robust and avoid overfitting. Techniques of data augmentation artificially increase the size of a dataset by applying transformations such as random flipping, rotation, brightness adjustment, and color jittering. Augmen- tation introduces variations in the orientation, lighting, and color of the images in a manner that it mimics real-world conditions. It helps diversify the training data but also enhances the model's ability to generalize to unseen data, which will improve the reliability of the model in clinical settings.

3.3 Data Labeling

Once all the images are pre-processed, it is the labelling procedure of data, which actually guides the model training for diabetic retinopathy (DR). Labelling in- volves associating the severity of damage caused due to this disease with the respective image of the retina. Our project uses a pre-labelled image dataset in the aptos 2019 version. These pre-labeled images would be important for supervised learning: they help the model connect input images with their classifications of severity.

Each image in the dataset is classified into one of the following severity levels: No DR, Mild DR, Moderate DR, Severe DR, or Proliferative DR. No DR is a healthy retina with no signs of diabetic retinopathy, whereas Proliferative DR is the most severe stage of the disease, which involves abnormal

blood vessel growth and significant retinal damage. These severity levels are used as ground truth for training. It can compare its prediction against the actual labels and iteratively minimize errors.

The learning process of a model heavily relies on the correctness of labeling. For supervised machine learning, a good basis is labeled data where the model learns mapping input data, such as retinal images, into output labels, which can be considered as severity levels. Quality labeling ensures the well-definition and reliability in severity for every image, enabling effective learning by the model. Poor labeling can result in lower performance and less accurate detec- tion and classification of diabetic retinopathy. Accurate labels are therefore vital in enhancing the model's ability to predict, with good performance in clinical applications.

3.4 Model Architecture

It discusses and presents a deep learning-based architecture of ResNet-18 for the detection and classification of diabetic retinopathy from retinal fundus images. Fig 3 is a representation of such architecture.

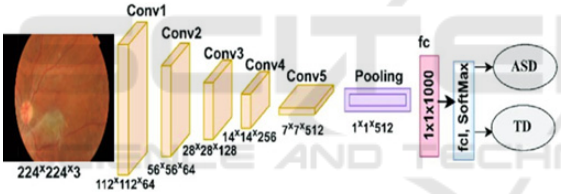

Figure 3: ResNet-18 architecture

As depicted in Fig 3, the model first accepts a 224×224×3 retinal fundus image as input. It goes through several stages:

Convolutional Layers: The initial convolutional layers (Conv1 to Conv5) ex- tract hierarchical features. Low-level features such as edges and textures are captured in the earlier layers, while deeper layers extract complex patterns, such as blood vessels, hemorrhages, and microaneurysms. The convolution operation is mathematically defined as:

$$Output(x,y) = \sum_{m=1}^{M}\sum_{n=1}^{N} 1(x+m, y+n)k(m,n) \quad (1)$$

where I(x + m, y + n) is the input pixel value, and K(m, n) is the convolutional kernel. Residual Blocks: To better learn residual connections bypass some layers such that the model is directed to learn the residual features.

$$y = F(x, Wi) + x \quad (2)$$

This framework ensures efficient training by mitigating the vanishing gradient problem. Pooling and Feature Reduction: Pooling layers reduce the spatial di- mensions, emphasizing important features. Max-pooling is computed as

$$MaxPooling = \max_{i,j}(l(x+i, y+j)) \quad (3)$$

The Final Layers: Fully Connected Layers aggregate the features that were extracted. Softmax Classification with activation for the prediction of diabetic retinopathy levels: No DR, Mild DR, Moderate DR, Severe DR and Proliferative DR.

Important parts of architecture like ReLU activation, batch normalization ensure efficient learning by introducing non-linearity to the ReLU function :

$$f(x) = \max(0, x) \quad (4)$$

Batch Normalization normalizes activations to accelerate training:

$$\text{x i} = \frac{\sigma 2+\epsilon}{\text{xi}-\mu} \quad (5)$$

The model is optimized using the Adam optimizer with categorical cross- entropy loss. Its equations ensure adaptive learning rates for each parameter

$$\eta_t = \frac{\eta}{\sqrt{v_t}+\epsilon} \quad (6)$$

This architecture efficiently extracts features at multiple levels of abstraction which allows for robust classification of diabetic retinopathy severity.

3.5 Training the Model

The training phase of the ResNet-18 model initiates after preprocessing and la- belling the retinal images. These images are feed into the deep convolutional net- work, and ResNet-18 automatically extracts hierarchical features such as blood vessels, hemorrhages, and microaneurysms that are pertinent for the detection of DR. The model utilizes the backpropagation algorithm, which updates its weights based on the gradients calculated from the loss function with the aim of minimizing prediction errors. It uses the Adam optimizer for weight updates, dynamically adjusting the learning rate during training, which leads to faster convergence and stable

optimization. The initial learning rate is set at 0.001 and dynamically adjusted as the training progresses.

The model uses categorical cross-entropy loss, which is the difference between the actual labels and the predicted probabilities for each class. Thus, the loss function can be defined as

$$\text{Cross Entropy Loss} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (7)$$

where C is the total number of classes, and yi is the true label, and yˆi is the predicted probability for every class.

Hyperparameter tuning is crucial for optimizing the model's performance. The three main parameters include the learning rate, dropout rate, and the number of epochs to be used. All of these are adjusted accordingly to get the best possible output. A learning rate scheduler is used to speed up convergence in the early epochs and to gradually refine the model as it approaches opti- mum performance. The dropout rate, set between 0.2 and 0.5, is used to avoid overfitting and generalization.

The ResNet-18 model was trained for 50 epochs. In this period, training and validation performance is monitored at regular time steps. After training the model obtained a training accuracy of 98.57% with the corresponding loss for training as 0.0322. Still, the validation accuracy stood at 83.49%. This implies that though the model has picked the features of interest from the training data, it underperforms a little bit on the unseen data and hence calls for further improvements in generalization. These can be achieved through methods like data augmentation, regularization, and fine-tuning.

Early stopping was applied in order to avoid overfitting and maximize the efficiency of computation. Training stopped once validation performance did not improve further, saving some computation resources and ensuring that it would not overfit on the training data.

3.6 Evaluation

The performance of the ResNet-18 model in detecting diabetic retinopathy is evaluated by considering the validation set. In the evaluation, metrics like ac- curacy, precision, recall, and F1 score are used to judge the performance of the model. These metrics provide an overview of the model's performance in detect- ing diabetic retinopathy at different stages.

Accuracy is among the evaluation measures, a ratio of correctly predicted hits to total predictions. As such, this is essentially an overall performance measure from the model:

$$\text{Accuracy} = \frac{Correct\ Prediction}{Total\ Prediction} \quad (8)$$

Precision refers to accuracy about the positive cases, i.e., true classification of images from the retinal images which is diabetic retinopathy in reality belongs to class DR. Precision is the calculation of:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

where TP denotes true positives (correctly classified DR images), and FP denotes false positives (non-DR images misclassified as DR).

Recall, in contrast, measures how well the model detects all true diabetic retinopathy cases, even those that are more challenging to detect. It is defined as

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

where FN stands for false negatives, i.e., DR images mistakenly classified as non-DR.

The F1 score is the harmonic mean of precision and recall, hence a single number that reflects both. It is also useful in case of datasets with class imbal- ance since it takes into account both the false positives and false negatives. The formula for calculating the F1 score is:

$$F1 - Score = 2.\frac{Precision.Recall}{Precision+Recall} \quad (11)$$

During the evaluation, it passes unseen data to evaluate real-time perfor- mance. Here also, validation accuracy turned out to be 83.49%whereas training accuracy was at a whopping 98.57% which shows very effective model perfor- mance on the train set though results from the validation part depict overfitting. Precision, recall, and F1 score are calculated as well to test this model in diag- nosing diabetic retinopathy stages completely.

Hyperparameter tuning: Assuming that performance isn't as it should, the learning rate, possibly the batch size, number of epochs is to be adjusted. Yet another necessary architectural adjustment would be putting more layers in or some other activation function that enhances the generality of your model. Meth- ods like cross-validation, increased diversity of the data maybe achieved using augmentation of training data to work on to improve on the validation.

By making use of these metrics, it ensures that the developed model is ac- curate and strong, thereby being capable to detect diabetic retinopathy in real-world applications precisely.

3.7 Deployment

Following satisfactory performance in training and evaluation, the ResNet-18 model is adopted into a clinical decision support system to detect DR through the automatic analysis of images of retinas, further classifying them on levels of DR severity and guiding healthcare providers with quicker and more accurate diagnoses. Reduction in manual assessment of images aids fast detection of DR, which further leads to better patient care through early intervention.

This is made accessible through cloud platforms or even a hospital's local network, meaning new retinal images will be processed in real-time. The model will, therefore, be retrained periodically on updated data in order to adapt to changes in imaging techniques and patterns of DR. The model is monitored in terms of its effectiveness within clinical settings to ensure continued reliability in DR diagnosis in the long term.

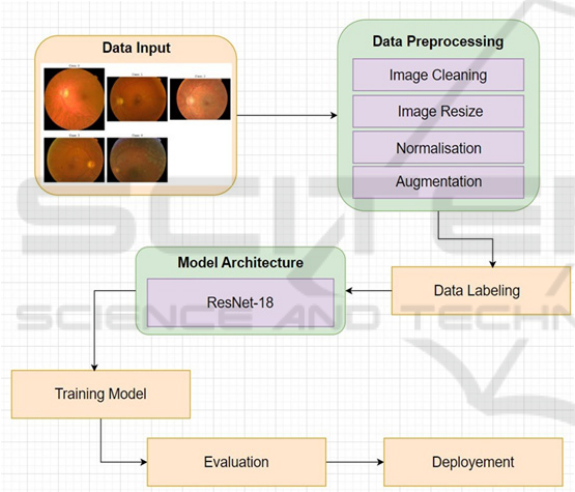

Figure 4: Flowchart illustrating machine learning workflow for Early Detection of Dia- betic Retinopathy

The critical challenges of the model are it's unable to generalize very well on unseen data; proof for this can be demonstrated with the difference in terms of training and validation accuracy. To combat the side effects, methods involving data augmentation have been adopted as ways to enhance diversity over the training set and added further dropout regularization for cutting the possibil- ity of overfitting. The learning rate scheduler has also been applied aiming to improve convergence and the generalization on both the training and validation sets.

4 RESULTS AND ANALYSIS

This study has applied a dataset labeled with a train.csv file and created to pre- dict the level of diabetic retinopathy, which had 3,662 records. Its two principal attributes include Id_Code and diagnosis. This is for the sake of Id_Code, an identifying variable in any given instance for traceability purposes. The target variable would be the diagnosis column with five levels of the severity of diabetic retinopathy: 0 is no DR, 1 is mild, 2 is moderate, 3 is severe, and 4 is prolifera- tive DR. This dataset serves as a base for training machine learning models to classify and predict the severity of diabetic retinopathy.

Table 1. Initial Records of the Dataset

Id_Code	Diagnosis
000c1434d8d7	2
001639a390f0	4
0024cdab0c1e	1
002c21358ce6	0
005b95c28852	0

Displays some sample images from the training dataset along with their re- spective labels in Fig 5.

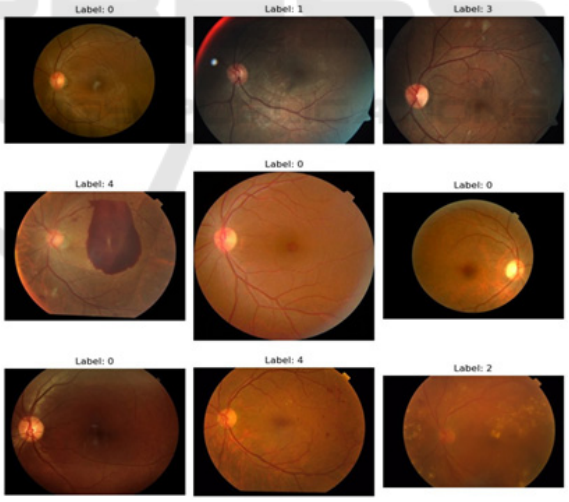

Figure 5: Sample images

The diabetic retinopathy dataset is labeled images extracted from a CSV file that removes irrelevant columns. The images are preprocessed into grayscale, resized to 256×256 pixels uniformly, and normalized pixel values into the range [0, 1]. Some representative examples are Image 90, 128, and 264, showing in the Fig 6 and content of the dataset, so one is clear about what they're looking at when it comes to analysis.

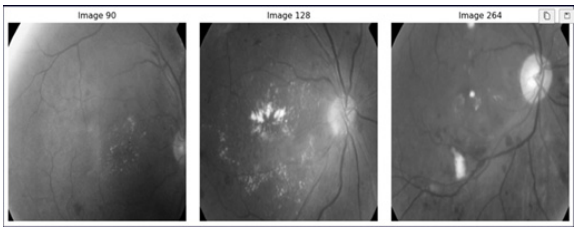

Figure 6: labeled images

As evidenced by the training results, this model's performance improves no- tably over 50 epochs. The training loss begins higher and decreases to 0.0322 by the last epoch, which means that this model is effectively minimizing errors in predictions and learning about patterns in the data set. Similarly, the accu- racy of training reaches an excellent 98.57%. The validation accuracy stands at 83.49% and shows the ability of the model to generalize for unseen data. Eval- uation metrics provide a precision of 0.70, recall of 0.52, and an F1 score of 0.58 to understand the model's efficiency in balancing false positives with false negatives.

However, the gap between training and validation accuracy still leaves room for improvement in generalization. Techniques such as further hyperparameter tuning, advanced data augmentation, or the integration of more complex archi- tectures like ResNet could be used to enhance performance. Overall, the training process highlights the model's potential but also shows areas for optimization to achieve even better results.

The confusion matrix evaluates the performance of the model on five diabetic retinopathy classes, ranging from Class 0 to Class 4. Class 0 performs well with high accuracy, correctly identifying 450 out of 540 samples as normal, effectively classifying healthy retinal images. However, it frequently confuses Class 1 and Class 2 with Class 0, highlighting challenges with class balance and overlapping features. For Class 3 and Class 4, correct predictions are sparse, suggesting the model struggles to differentiate higher severity levels due to insufficient data or distinctive features. This emphasizes the need for further model refinement. shown in Fig 7.

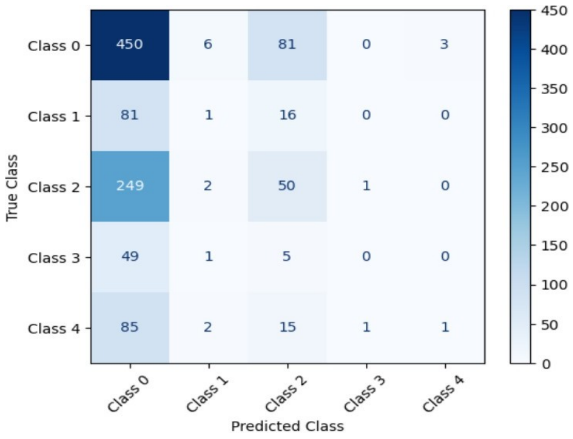

Figure 7: Normalized Confusion Matrix

The "Accuracy vs Epoch" plot shows that the model is training to a huge accuracy, well over 95%, meaning that the model is learning the patterns in the data very effectively. The steady rise in validation accuracy in the initial epochs also shows that the model is not overfitting but is rather generalizing very well for unseen data. These trends show that the model is capable of learning the data patterns successfully and has a good generalization ability, hence showing it to be able to give very good predictions on both training and validation datasets. This makes it a robust and potentially excellent model.

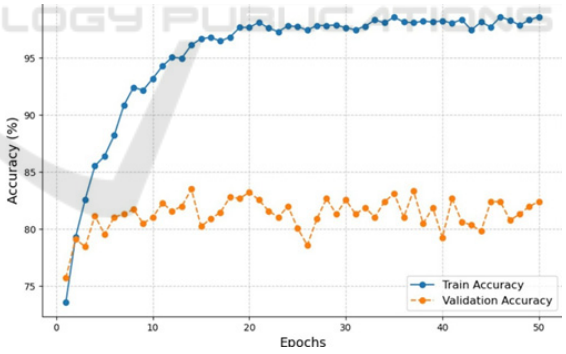

Figure 8: Accuracy vs Epoch

The "Loss vs Epoch" plot suggests that the training loss stays relatively flat, indicating proper learning and optimization of the model. Validation loss goes along a straight trend up to the first few epochs (about three to five), which again is very good generalization and stability. This pattern suggests that the model successfully captures the patterns in the training set and validation sets, as it generalizes well and maintains consistent performance.

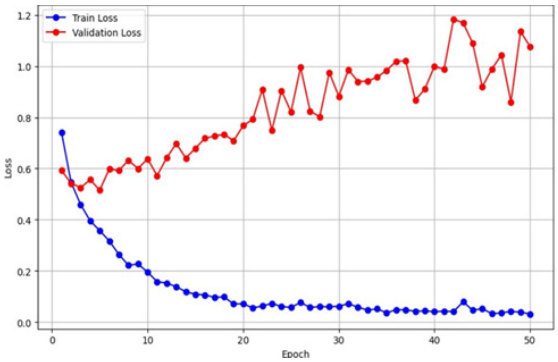

Figure 9: Loss vs Epoch

5 CONCULSION AND FEATURES

This project successfully applies deep learning in the detection of diabetic retinopa- thy using the Aptos 2019 dataset by focusing on the class-based classification of DR from the severity level in the images of the retina. All preprocessing tech- niques applied—namely, image cleaning, resizing, normalization, and augmen- tation—proved useful for improving the quality and consistency of the dataset for enhanced model performance. The ResNet-18-based model achieved a very high training accuracy of 98.57% and a validation accuracy of 83.49% after 50 epochs, demonstrating strong performance but also some room for improvement in terms of generalization.

The model employed the Adam optimizer that led to efficient training and convergence. Dropout regularization was applied, which helped prevent overfit- ting. Cross-entropy loss was used in order to optimize the model on classification tasks, therefore leading to effective learning of intricate patterns within retinal images.

Future work in improving the feature extraction ability of the model and the overall performance can be furthered by using more complex architectures such as ResNet-18. Leveraging pretrained models through transfer learning, along with hyperparameter tuning, can improve accuracy and robustness. Also, multi- modal data integration and explainable AI techniques will be critical to enhance transparency, which is paramount in clinical settings where the trust in model predictions is essential. In addition, increasing the size of the dataset to reflect different demographics and using cross-validation methods will help the model to generalize better with higher reliability and accuracy over the population of patients with diverse backgrounds.

REFERENCES

Abràmoff, M.D.: The autonomous point-of-care diabetic retinopathy examination pp. 159–178 (2020)

Aptos: Aptos 2019 blindness detection dataset. https://www.kaggle.com/c/ aptos2019-blindness-detection (2019), accessed: 2024-12-20

Bourne, R., Price, H., Stevens, G., Group, G.V.L.E., et al.: Global burden of visual impairment and blindness. Archives of ophthalmology **130**(5), 645–647 (2012)

Cheung, C.Y., Ikram, M.K., Klein, R., Wong, T.Y.: The clinical implications of recent studies on the structure and function of the retinal microvasculature in diabetes. Diabetologia **58**, 871–885 (2015)

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. jama **316**(22), 2402–2410 (2016)

He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition pp. 770–778 (2016)

Jia, W., Li, Y., Qu, R., Baranowski, T., Burke, L.E., Zhang, H., Bai, Y., Mancino, J.M., Xu, G., Mao, Z.H., et al.: Automatic food detection in egocentric images using artificial intelligence technology. Public health nutrition **22**(7), 1168–1179 (2019)

Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

Leibig, C., Allken, V., Ayhan, M.S., Berens, P., Wahl, S.: Leveraging uncertainty information from deep neural networks for disease detection. Scientific reports **7**(1), 1–14 (2017)

Lu, X., Liu, J., Wu, M., Zhang, X.: Deep learning for diabetic retinopathy: A review of recent advancements. Computers in Biology and Medicine **101**, 126–133 (2018)

Nirgude, A.S., Revathi, T., Navya, N., Naik, P.R.: "even though doctor has advised to practice foot care i have not practiced soaking feet in lukewarm water so far" self-care practices, enablers, and barriers: A mixed methods study among individ- uals with diabetes from a rural area of south india. Indian Journal of Community Medicine pp. 10–4103 (2024)

Solomon, S.D., Chew, E., Duh, E.J., Sobrin, L., Sun, J.K., VanderBeek, B.L., Wykoff, C.C., Gardner, T.W.: Diabetic retinopathy: a position statement by the american diabetes association. Diabetes care **40**(3), 412 (2017)

Teo, Z.L., Tham, Y.C., Yu, M., Chee, M.L., Rim, T.H., Cheung, N., Bikbov, M.M., Wang, Y.X., Tang, Y., Lu, Y., et al.: Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. Ophthal- mology **128**(11), 1580–1591 (2021)

Ward, C., Maselko, M., Lupfer, C., Prescott, M., Pastey, M.K.: Interaction of the human respiratory syncytial virus matrix protein with cellular adaptor protein

complex 3 plays a critical role in trafficking. PLoS One **12**(10), e0184629 (2017)

Yau, J.W., Rogers, S.L., Kawasaki, R., Lamoureux, E.L., Kowalski, J.W., Bek, T., Chen, S.J., Dekker, J.M., Fletcher, A., Grauslund, J., et al.: Global prevalence and major risk factors of diabetic retinopathy. Diabetes care **35**(3), 556–564 (2012)