UNIVERSITY OF SUFFOLK

DEEP LEARNING TECHNIQUES AND TOOLS

DIABETIC RETINOPATHY IMAGE CLASSIFICATION: A COMPARISON OF CONVOLUTIONARY NEURAL NETWORKS – VGG, RESNET AND ALEXNET

BY

STUDENT NUMBER: S246855

WORD COUNT: 5498 (Excluding references)

Table of Contents

ABSTRACT

This project aimed to address the image classification of diabetic retinopathy using deep learning techniques. Three pre-trained models, ResNet, VGG, and AlexNet, were utilized to classify the diabetic retinopathy images using a Gaussian filter dataset resized to a resolution of 224x224 pixels.

The results revealed notable differences in model performance. ResNet consistently achieved high precision scores across all classes, with precision values ranging from 86.19% to 96.88%. VGG exhibited excellent recall rates, with recall values ranging from 69.11% to 97.97%. AlexNet, although showing relatively lower precision and recall values, still demonstrated promising results, with precision ranging from 78.80% to 93.45% and recall ranging from 69.11% to 93.16%.

The comparison of the models highlighted their strengths and weaknesses in classifying diabetic retinopathy. The project contributes to the field by demonstrating the efficacy of deep learning models in diabetic retinopathy classification and emphasizes the importance of selecting appropriate models based on performance metrics. The limitations of using deep learning in predicting diabetic retinopathy, such as data availability and interpretability, were discussed. Possible future directions were proposed, including the exploration of advanced architectures, transfer learning, multi-modal approaches, and explainability techniques, to enhance prediction accuracy.

This project provides valuable insights into the image classification of diabetic retinopathy using deep learning, assisting in early detection and better management of the disease. The findings underscore the significance of model selection and present avenues for future research to improve the accuracy of diabetic retinopathy prediction.

INTRODUCTION

Diabetic Retinopathy (DR), a vision-related complication associated with diabetes, poses a significant threat as it can lead to blindness if left undiagnosed and untreated (NHS, 2021). High blood sugar levels in diabetes patients can cause damage to the retina, the light-sensitive layer of cells at the back of the eye responsible for converting light into electrical signals. Over time, this persistent hyperglycemia can lead to the deterioration of the retinal blood vessels, further exacerbating the condition.

The symptoms of diabetic retinopathy often remain silent or unnoticed for an extended period (Duh, Sun, & Stitt, 2017). These symptoms can affect one or both eyes and may include gradually worsening vision, sudden vision loss, the presence of floaters (shapes floating in the field of vision), blurred or patchy vision, eye pain or redness, and difficulty seeing in low-light conditions. Several factors contribute to the risk of developing diabetic retinopathy in individuals with diabetes (NHS, 2021). These factors include the duration of the disease, long-term uncontrolled blood sugar levels, high blood pressure, borderline cholesterol levels, and pregnancy in women.

Diabetic retinopathy can be classified into two categories: nonproliferative (early stage) and proliferative (advanced stage) (fatih, n.d.). Nonproliferative DR represents the initial stage of the disease, where symptoms may be mild or absent. The severity of nonproliferative DR can range from Type 0 (mild or non-existent indications) to Type 2 (moderate inflammation and fluid leakage in the macula).

Proliferative diabetic retinopathy, identified as Type 4, is the most severe form of the disease characterized by the abnormal growth of fragile blood vessels in the retina (fatih, n.d.). These new vessels are prone to leakage, leading to complications and potential blindness. It is important to note that the transition from nonproliferative to proliferative diabetic retinopathy can occur without noticeable symptoms, underscoring the necessity for regular eye examinations in diabetic patients.

The treatment of diabetic retinopathy emphasizes prevention through blood sugar control, dietary management, and increased physical activity (Hui et al., 2001). Additional treatment options, such as photocoagulation, intraocular injections, or vitrectomy, depend on the stage and extent of the disease (Duh, Sun, & Stitt, 2017).

To ensure effective treatment and management of diabetic retinopathy, early diagnosis conducted by ophthalmologists plays a vital role. These examinations involve measuring visual acuity, assessing eye muscle movement, evaluating peripheral vision, testing depth perception, and conducting cornea measurements (Duh, Sun, & Stitt, 2017). However, manual diagnosis by skilled medical professionals is resource-intensive and subject to fatigue and variations in estimation procedures. This limitation highlights the need for automated and accurate image understanding techniques that can assist medical practitioners in the diagnosis, prognosis, and treatment planning of diabetic retinopathy.

Deep learning, a subfield of machine learning, has shown great promise in image classification tasks by leveraging its ability to automatically learn complex features and make accurate predictions. By applying deep learning techniques to retinal images, the anomalies associated with diabetic retinopathy can be effectively captured and analyzed. Deep learning inherits the advantages of traditional machine learning algorithms while exhibiting superior performance, enhanced generalization capabilities, and the ability to handle complex tasks.

This project aims to utilize deep learning techniques for the classification of diabetic retinopathy images, enabling early detection and accurate prediction of the disease. By leveraging the power of deep learning, this research seeks to contribute to the development of automated and efficient systems for diabetic retinopathy diagnosis, enabling timely interventions and improving patient outcomes.

LITERATURE REVIEW

Diabetic retinopathy (DR) is a prevalent complication of diabetes and a leading cause of blindness worldwide. Early detection and accurate diagnosis of DR are crucial for timely interventions and preventing vision loss. In recent years, deep learning techniques, particularly convolutional neural networks (CNNs), have demonstrated promising results in automating the detection and prediction of DR using retinal images. This literature review aims to explore various studies that employ deep learning techniques for predicting DR and evaluate their effectiveness in improving diagnostic accuracy.

Baba and Bala (2022) proposes a CNN-based algorithm for automatic detection of diabetic retinopathy (DR) in retinal images, highlighting the time-consuming manual testing process and the potential impact on visual impairment. While the algorithm achieves a high accuracy of 99.5%, with a sensitivity of 97.6%

and specificity of 91.24%. The training and validation accuracy are reported to be 97% and 99.5%, respectively, but the paper lacks sufficient details on the dataset representativeness and network architecture. The comparison with other approaches is limited, and the future direction of the research is briefly mentioned. Figures depicting training and validation accuracy, loss graphs, and a confusion matrix are provided, but more comprehensive analysis and discussion of the results are needed. Overall, while the proposed CNN-based algorithm shows promise, further improvements and a more thorough evaluation are required to validate its effectiveness and address the limitations of the study.

Yadav, Goel and Rajeswari (2021) presents a computer vision-based technique that utilizes convolutional neural networks (CNN) for the early detection of diabetic retinopathy (DR) from retinal images. The framework involves image processing, feature extraction, and machine learning-based classification. The experimental results demonstrate the effectiveness of the CNN approach, achieving an impressive accuracy of 98.50%, outperforming support vector machines (SVM) and other unsupervised ML techniques with an accuracy of 87.40%. However, the paper would benefit from providing more detailed information on the methodology, including specific image processing techniques and CNN architecture used. Additionally, a more comprehensive evaluation, including additional metrics and a thorough comparison with related work, would strengthen the paper's findings. It is important to address the limitations and discuss potential future research directions to enhance the overall contribution of the paper. While the proposed approach shows promise, further refinement and analysis are needed for a more robust conclusion.

Hatode, Edinburgh and Jha (2022) focuses on deep learning algorithm based on the ResNet 50 architecture for the automated detection of diabetic retinopathy from retinal fundus images. The algorithm achieves an accuracy of 91.60% after training on a dataset of 3,662 images. The paper highlights the potential of deep neural networks for pattern recognition tasks and emphasizes the importance of early detection in preventing blindness caused by diabetic retinopathy. However, the paper lacks detailed methodology information, including specifics about the implementation and preprocessing techniques used. It also provides limited evaluation metrics, focusing only on accuracy. Furthermore, the dataset and its limitations are not thoroughly discussed. Addressing these shortcomings would enhance the replicability, robustness, and overall contribution of the proposed algorithm.

Shanthi and Sabeenian (2019) focuses on the classification of diabetic retinopathy (DR) fundus images based on the severity of the disease using modified AlexNet architecture. The proposed algorithm utilizes suitable layers such as Pooling, Softmax, and Rectified Linear Activation Unit (ReLU) in the CNN architecture to achieve high accuracy. The performance of the algorithm is validated using the Messidor database. The results demonstrate classification accuracies of 96.6%, 96.2%, 95.6%, and 96.6% for healthy images, stage 1 DR, stage 2 DR, and stage 3 DR, respectively.

Shahriar Maswood et al. (2020) employed a pre-trained CNN model called EfficientNet-B5 for automated DR diagnosis. Their results demonstrate the effectiveness of the proposed approach, achieving an impressive accuracy of 94.02% on the training set and 93.33% on the testing set, highlighting the potential of pre-trained models in enhancing CNN-based DR diagnosis. R, B, and P (2022) proposed a modified InceptionV3 model for classifying DR severity levels. Their model achieving an accuracy of 82.4% (quadratic weighted kappa) on a dataset of 3,724 retinal images. These studies indicate the effectiveness of EfficientNet and InceptionV3 models in improving DR diagnosis accuracy.

Bidwai et al. (2023) focuses on using CNN models, including ResNet-18, GoogLeNet, VggNet-19, and AlexNet, for the detection of diabetic retinopathy (DR) using retinal fundus images from the APTOS dataset. The proposed technique achieves a maximum accuracy of 87% for multi-class classification and 98% for binary classification. While the study presents promising results, it lacks detailed methodology information and evaluation metrics beyond accuracy. Additionally, it suggests future directions such as exploring other deep learning techniques, image augmentation, and preprocessing methods. The paper highlights the importance of explainable AI (XAI) and the integration of multimodal imaging for enhancing disease diagnosis accuracy. Further research in these areas would strengthen the study's findings and improve transparency in AI decision-making processes.

Deep learning techniques, particularly CNNs, have shown significant potential in improving the accuracy and efficiency of DR detection and diagnosis. Studies utilizing various deep learning models such as CNNs, ResNet, EfficientNet, InceptionV3, and VGG-16 have demonstrated promising results in automated DR prediction and severity classification. Further research can focus on optimizing model architectures, exploring Explainable AI techniques, and integrating multimodal imaging to enhance the interpretability and diagnostic accuracy of deep learning-based Diabetic Retinopathy detection systems.

METHODOLOGY AND EXPERIMENTAL SETUP

Google Collab Code:

https://colab.research.google.com/drive/17cLrldprgTn3-zCJSQvWRo9spx2Ijybh?usp=sharing

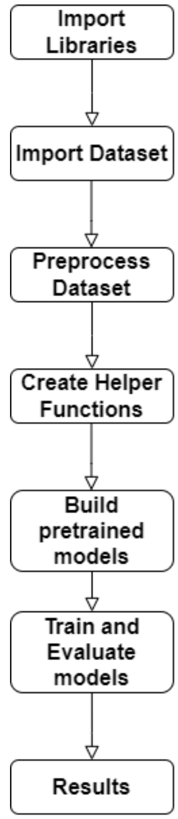

Figure 1: Proposed Methodology

CHOOSING DATASET:

These images which have been preprocessed using a Gaussian filter and resized to a resolution of 224x224 pixels are representative of the retina's condition in patients diagnosed with diabetic retinopathy. The Gaussian filter applied to the images helps in reducing noise and enhancing important features, making them suitable for analysis and classification tasks.

Each retinal image in the dataset is labelled with a corresponding class indicating the severity of diabetic retinopathy. The severity levels range from 0 to 4, representing the following classes:

0: No diabetic retinopathy

1: Mild diabetic retinopathy

2: Moderate diabetic retinopathy

3: Severe diabetic retinopathy

4: Proliferative diabetic retinopathy

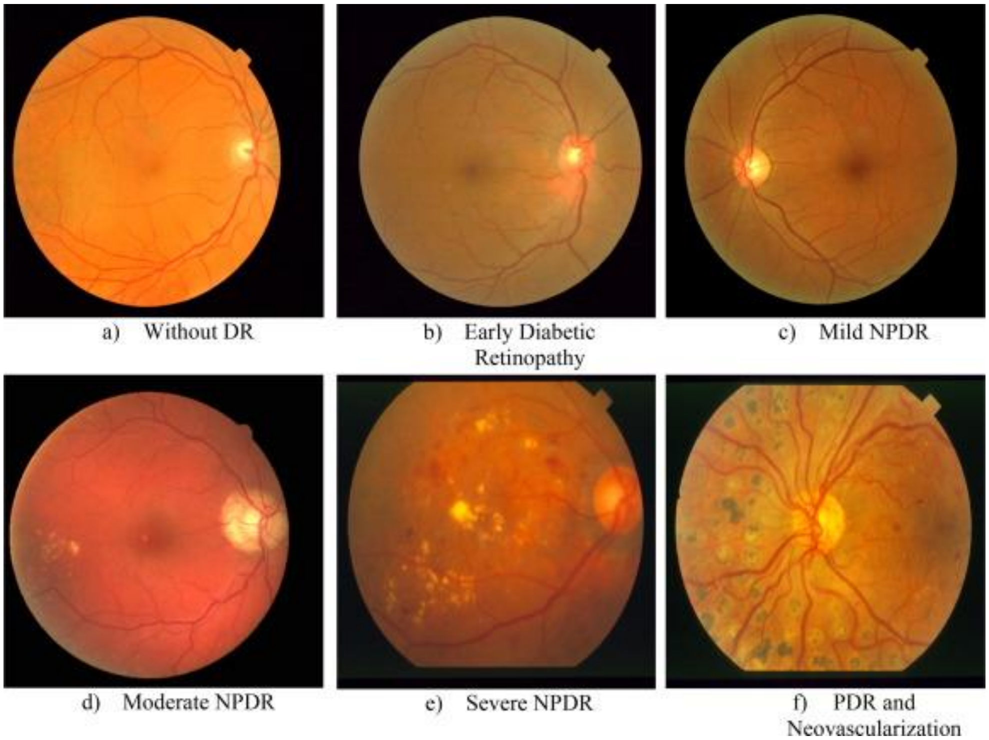

Figure 2: Diabetic Retinopathy Stages [source: (Nagpal et al., 2021)]

The dataset aims to facilitate research and development in the field of diabetic retinopathy detection, classification, and diagnosis through the application of machine learning techniques.

DEEP LEARNING FRAMEWORK:

In this project Pytorch was used. PyTorch is an open-source deep learning framework that offers a flexible and efficient platform for building neural networks. With features like tensors, automatic differentiation, pre-defined network modules, optimization algorithms, DataLoader for data handling, and GPU support, PyTorch provides a user-friendly environment for developing and training deep learning models.

DATA PREPROCESSING

The data preprocessing prepares the dataset for analysis and model development. This process involves several crucial steps to ensure data quality, organization, and transformation, facilitating accurate modelling and evaluation.

1. Balancing Data Distribution: The data distribution of the five classes was balanced by creating additional data using various transforms. These transforms included random rotation (transforms.RandomRotation(40)), random horizontal and vertical flips (transforms.RandomHorizontalFlip(), transforms.RandomVerticalFlip()), and random affine transformations (transforms.RandomAffine(0, shear=20, scale=(0.8, 1.2))). These transformations helped augment the dataset and ensure a more balanced representation of each class.

2. Splitting the Dataset: The dataset was split into training, validation, and testing sets, and distinct folders were created for each set. This allowed for proper separation and organization of the data for subsequent training, validation, and testing phases.

3. Applying Transforms: Transformations were applied to the split dataset to fit the format required by the pretrained models that were to be compared in the project, namely ResNet, AlexNet, and VGG. These transforms included normalization (tt.Normalize()), resizing the images to a specific size (tt.Resize((224, 224))), and random horizontal flipping (tt.RandomHorizontalFlip()). These transformations ensured that the dataset was appropriately pre-processed for compatibility with the selected models.

4. Converting Images to Tensors: The pre-processed images were then converted to tensors using the torchvision.transforms.ToTensor() function. This conversion enabled efficient handling and processing of the image data within the deep learning models.

5. Creating Batches: Finally, a batch size of 32 was set to load each folder of the pre-processed dataset into the data loader. This allowed for the efficient loading of the data during the training and evaluation phases, facilitating smoother model training and testing processes.

MODELS

ALEXNET:

AlexNet, a deep convolutional neural network (CNN) architecture developed by Alex Krizhevsky et al. in 2012, made a significant impact on the field of deep learning. Its performance in the ImageNet Large Scale Visual Recognition Challenge revolutionized image classification tasks and popularized deep learning methods.

AlexNet consists of eight layers, including five convolutional layers and three fully connected layers (Singh, Goyal and Chandel, 2022). The architecture incorporates innovative design choices such as rectified linear units (ReLU) as activation functions, overlapping pooling, local response normalization, and dropout regularization. These elements were specifically designed to leverage GPU acceleration for efficient training and inference.

AlexNet achieved remarkable results with a top-5 error rate of 15.3%, surpassing traditional machine learning approaches.

The impact of AlexNet extends beyond its achievements. Its design choices, including ReLU activation, overlapping pooling, and dropout regularization, became standard practices in deep learning.

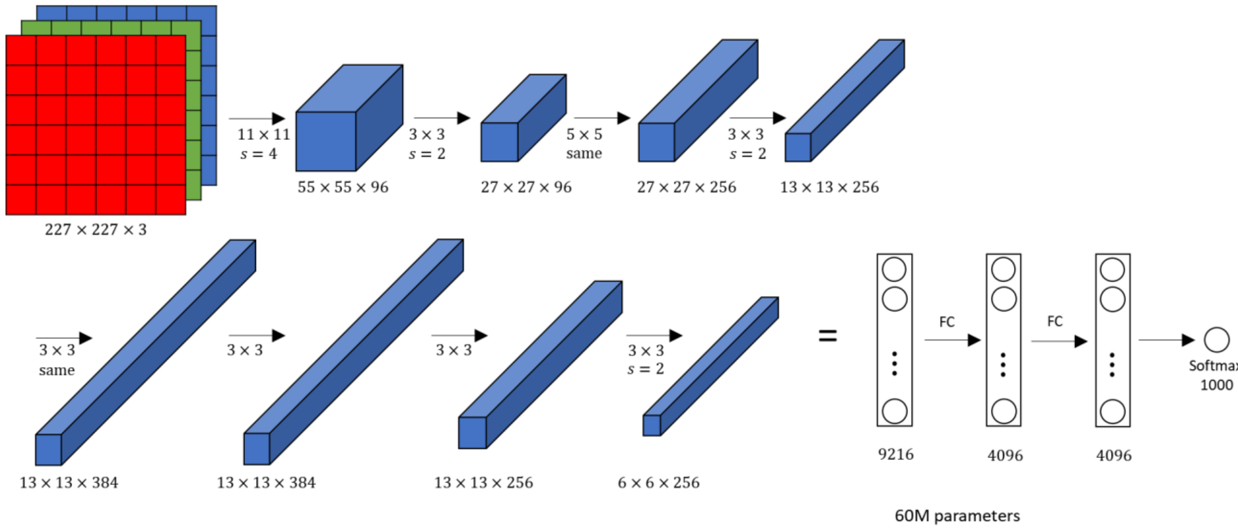

Figure 3: AlexNet architecture (source: datahacker.rs, 2018)

RESNET:

ResNet (Residual Neural Network) is a pioneering architecture in deep learning introduced by Kaiming He et al. in 2015. ResNet addressed the challenges of training very deep neural networks by utilizing residual connections. These connections, known as skip connections or shortcuts, enable information to flow directly from earlier layers to deeper layers, mitigating the vanishing gradient problem (Mujtaba, 2020). ResNet's core building blocks are residual units, composed of stacked convolutional layers with identity mappings and skip connections (Elswah, Elnakib and El-din Moustafa, 2020).

ResNet has several architectural variants, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152. These variants offer flexibility in selecting model depth based on specific task requirements.

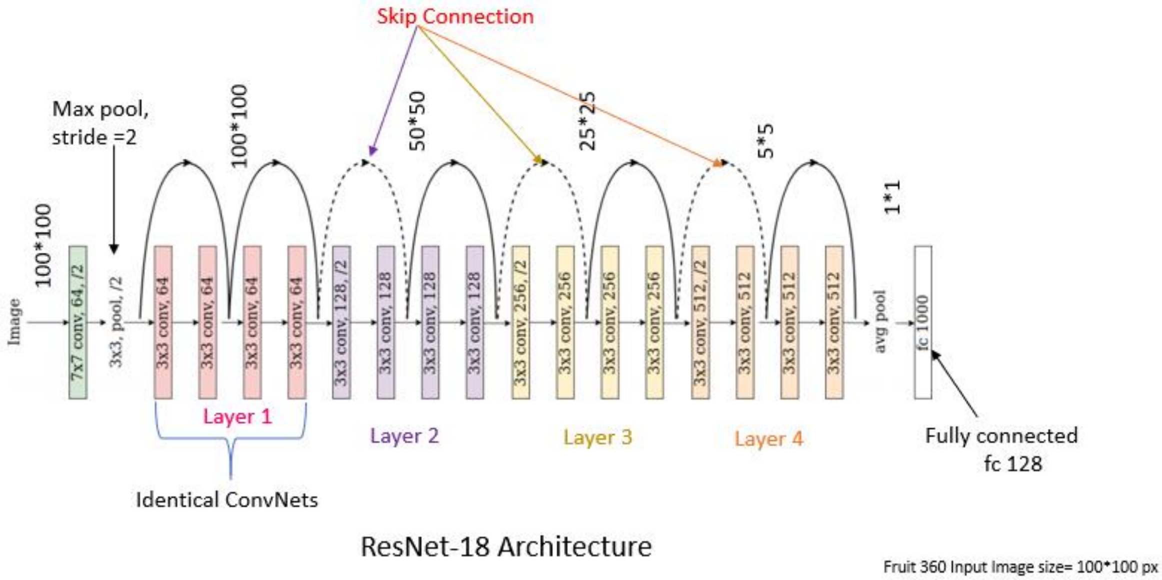

Figure 4: ResNet architecture

VGG:

Introduced by the Visual Geometry Group at the University of Oxford in 2014, VGG consists of a series of convolutional layers followed by max-pooling layers, culminating in fully connected layers for classification (Boesch, 2021). The key design principle of VGG is the use of small 3x3 filters throughout the network, allowing for a deeper architecture and learning of hierarchical features (Rakesh et al., 2023). Max-pooling layers downsample the spatial dimensions while retaining important features.

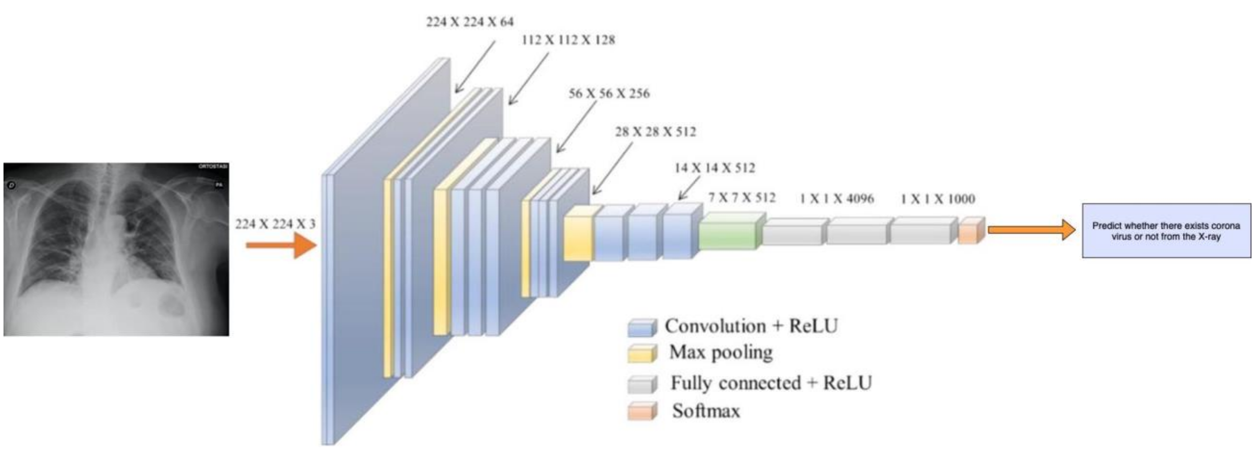

Figure 5: VGG Architecture [source: (Patel, 2020)]

TRANSFER LEARNING

Transfer learning is a powerful technique in deep learning that allows us to transfer knowledge gained from one task to another related task, rather than starting from scratch(Shao, Zhu and Li, 2015).

Medical datasets often suffer from limited availability of labelled data, making it difficult to train deep learning models from scratch. By utilizing transfer learning, we can leverage pre-trained models trained on large general datasets and adapt them to medical tasks with relatively small labeled datasets (Arbane et al., 2021). This allows us to make the most of existing knowledge and develop accurate models even with limited medical data.

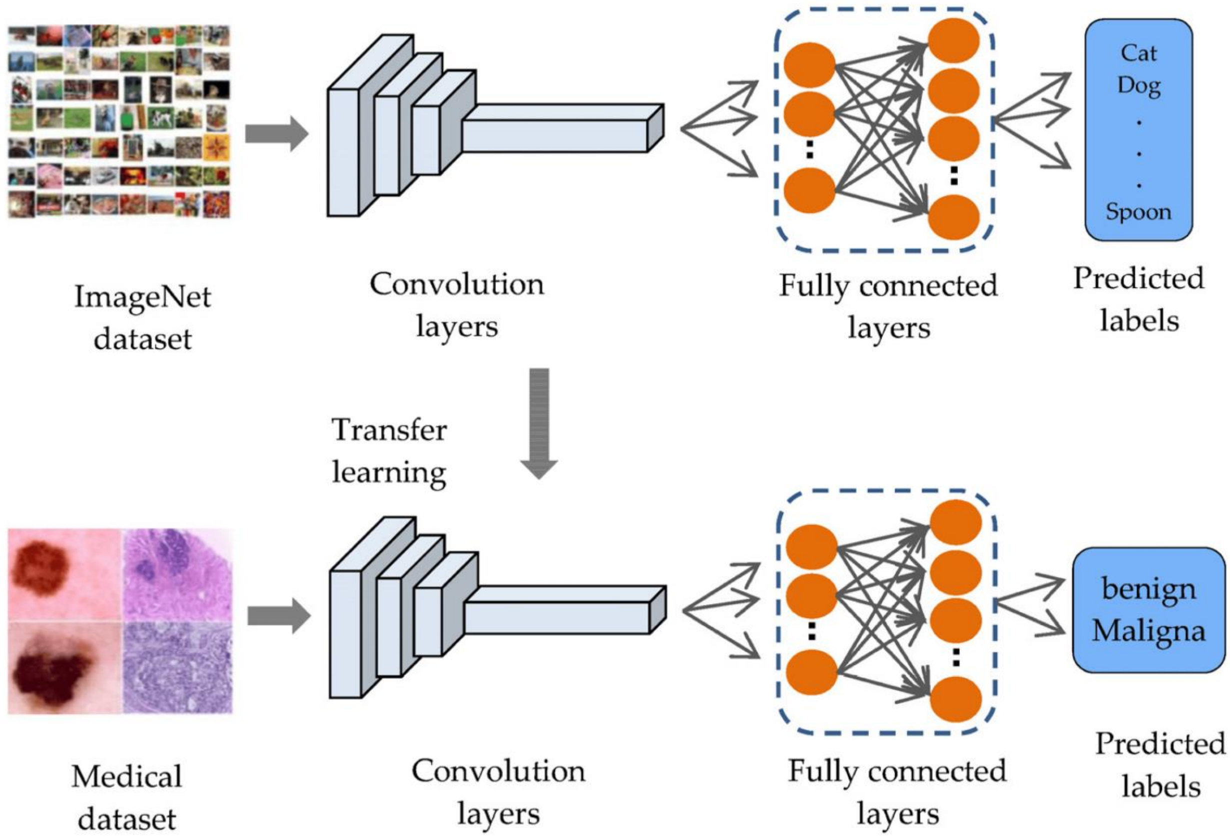

Figure 6: Transfer learning [source: (Mukhlif, Al-Khateeb and Mohammed, 2023)]

Additionally, medical datasets tend to have high dimensionality and complex feature representations. Medical images contain intricate visual patterns that can be challenging to capture using traditional machine learning techniques (Kim, Choi and Ro, 2017).

PERFORMANCE METRIC

The evaluation of each neural network's performance is crucial for assessing its efficacy. Performance metrics provide quantitative measures for analyzing and comparing models. This project uses the Confusion Matrix, Accuracy, F1 Score, Recall, and Precision.

Confusion Matrix:

The confusion matrix is a tabular representation of predicted and actual class labels. It consists of four values: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These values enable a detailed analysis of model performance and form the basis for calculating other metrics (B, 2020).

Accuracy:

Accuracy measures the overall correctness of predictions, indicating the proportion of correctly classified instances among all instances. It is computed by dividing the sum of TP and TN by the total number of instances (B, 2020). The formula for accuracy is:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

F1 Score:

The F1 score combines precision and recall into a single metric, providing a balanced assessment of model performance. It considers both the ability to correctly identify positive instances (precision) and the ability to capture all positive instances (recall). The F1 score is the harmonic mean of precision and recall, calculated using the formula:

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

Recall (Sensitivity):

Recall, also known as sensitivity or true positive rate, measures the ability of a model to correctly identify positive instances. It calculates the proportion of true positives (TP) out of the sum of TP and false negatives (FN). The formula for recall is:

Recall = TP / (TP + FN)

Precision:

Precision represents the proportion of true positives (TP) out of the sum of TP and false positives (FP). It assesses the accuracy of positive predictions and is calculated using the formula:

Precision = TP / (TP + FP)

IMPROVING MODEL PERFORMANCE

Learning Rate- 0.0001:

The learning rate in deep neural networks is a hyperparameter that controls the step size of parameter updates during training. It significantly impacts model results by affecting convergence speed, the ability to find optimal solutions and generalization performance (Iiduka, 2021).

Optimizer- Adam:

In this project, Adam optimizer was used. Adam is a popular optimizer that combines the benefits of AdaGrad and RMSProp, adapting the learning rate for each parameter based on gradient moments.

RESULT

RESNET:

The ResNet model was trained using a training dataset and evaluated on a validation dataset. After 50 epochs, significant improvements were observed in the model's performance. In the first epoch, the model achieved a training accuracy of 57.3% and a validation accuracy of 65.6%. However, by the 50th epoch, the model's performance improved substantially, with a training accuracy of 98.5% and a validation accuracy of 90.6%.

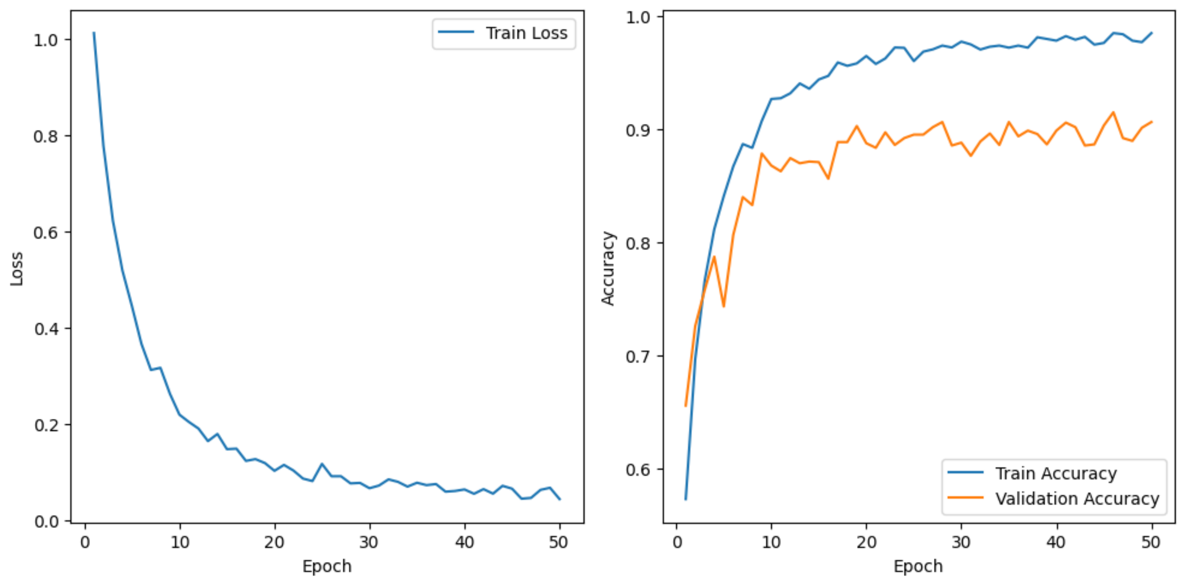

Figure 7: ResNet Loss and Accuracy curve

The above result was evaluated on the Test dataset producing the following results. The images labelled in green indicate correct prediction and the images labelled in red indicate wrong prediction.

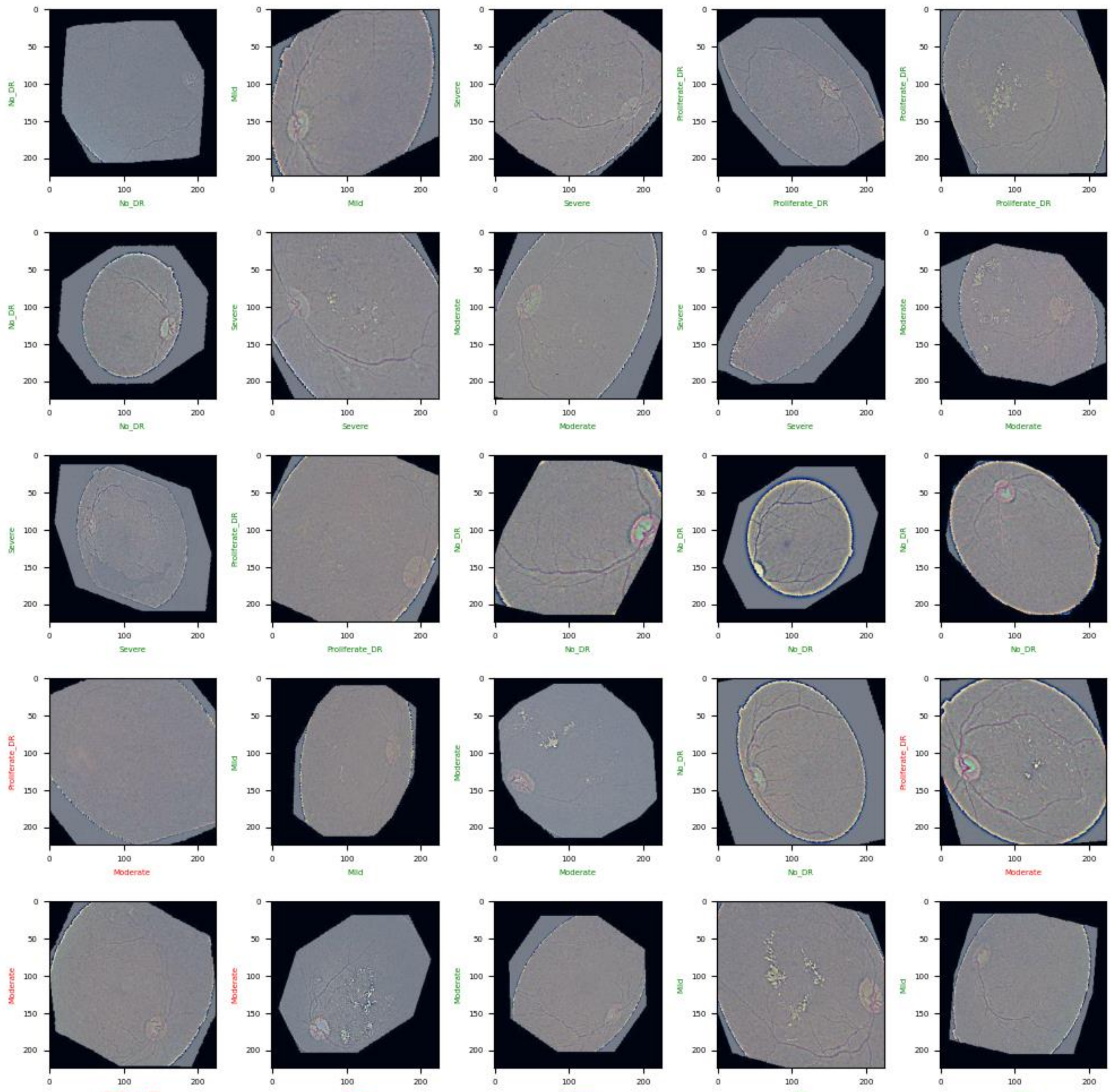

Figure 8: ResNet prediction

Confusion Matrix:

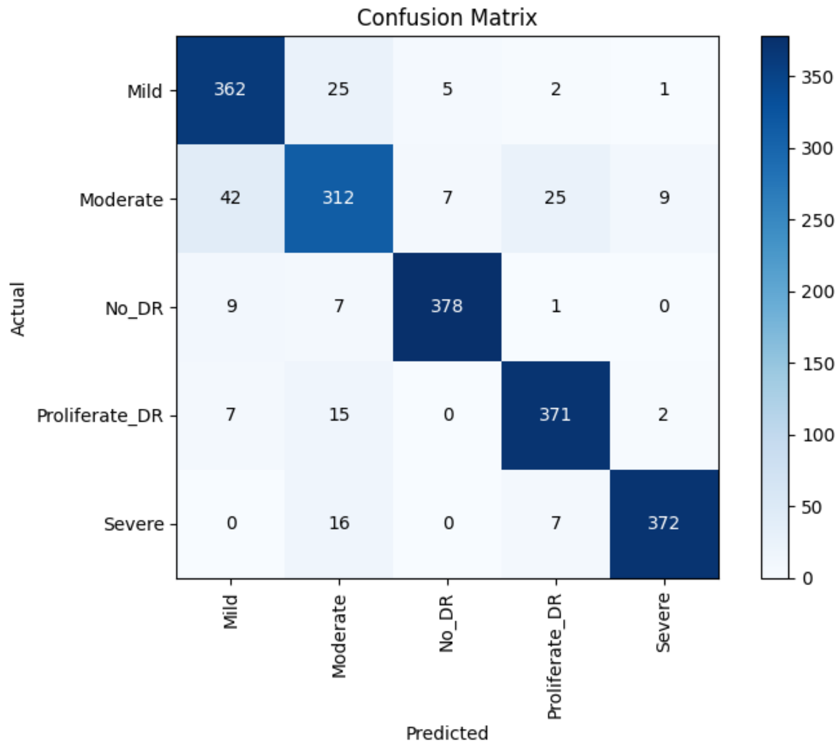

Figure 9: ResNet confusion matrix

The confusion matrix indicates that the model achieved the highest accuracy in classifying 'No Diabetic Retinopathy', with 362 true positives and a precision score of 96.9%. However, the model struggled to accurately classify the 'Mild' class, as evident from the high number of false positives and false negatives. This resulted in a lower precision score of 86.2% and recall score of 78.9% for the 'Mild' class.

Accuracy:

The overall accuracy of the model is impressive, with an average accuracy of 96.7% across all classes. Notably, the 'No Diabetic Retinopathy' class achieved the highest accuracy of 98.5%, while the 'Mild' class exhibited a comparatively lower accuracy of 92.6%. This suggests that the model performed well in general, but struggled with accurately identifying the 'Mild' cases.

Precision:

Precision measures the model's ability to correctly identify positive instances for each class. The 'Moderate' class achieved a high precision score of 96.9%, indicating a low number of false positives. However, the 'Mild' class had a precision score of 83.2%, suggesting a higher number of false positives. This indicates a difficulty in distinguishing 'Mild' cases from other classes.

Recall:

Recall, or sensitivity, measures the model's ability to correctly detect positive instances. The 'Severe' class demonstrated the highest recall score of 94.2%, indicating a strong ability to capture true positives. However, the 'Mild' class showed a lower recall score of 78.9%, suggesting difficulty in identifying all positive instances accurately.

F1-Score:

The F1-score, a harmonic mean of precision and recall, provides an overall measure of the model's performance. The 'Moderate' class achieved the highest F1-score of 96.3%, indicating a good balance between precision and recall. In contrast, the 'Mild' class exhibited a lower F1-score of 81.0%, reflecting challenges in achieving a balance between precision and recall for this class.

Conclusion:

Overall, the pretrained ResNet model demonstrated strong performance in classifying diabetic retinopathy images into the 'Moderate', 'No Diabetic Retinopathy', 'Proliferate Diabetic Retinopathy', and 'Severe' classes. However, the model faced challenges in accurately identifying the 'Mild' cases, leading to lower precision, recall, and F1-scores for this class. Further investigation and model refinement are required to improve the performance in accurately classifying 'Mild' instances. Nonetheless, the results show promise in using a pretrained ResNet model for diabetic retinopathy image classification.

ALEXNET:

The AlexNet model underwent a training process using the training dataset and was evaluated on the validation dataset. The model was trained for 50 epochs, gradually improving its performance. In the first epoch, the model had a training loss of 1.2781 and achieved a training accuracy of 45.0%, with a validation accuracy of 51.2%. However, by the 50th epoch, the model significantly improved, with a training loss of 0.1198 and a training accuracy of 96.1%. The validation accuracy also increased to 85.7%, indicating the model's improved performance.

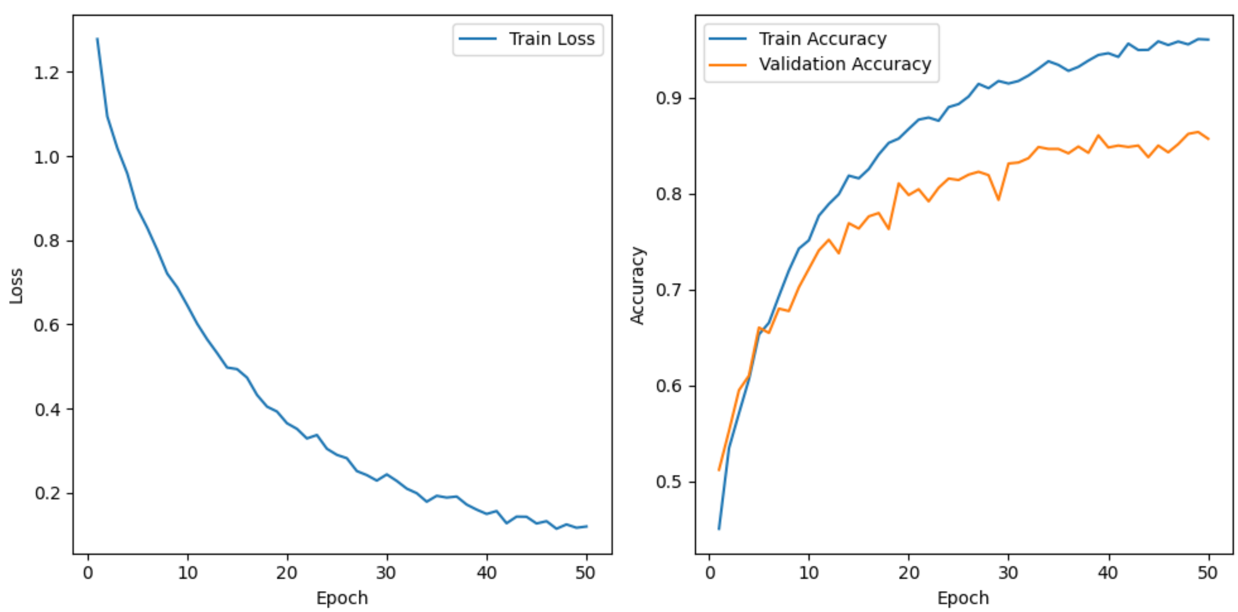

Figure 10: AlexNet Loss and Accuracy curve

The above result was evaluated on the Test dataset producing the following results. The images labelled in green indicate correct prediction and the images labelled in red indicate wrong prediction.

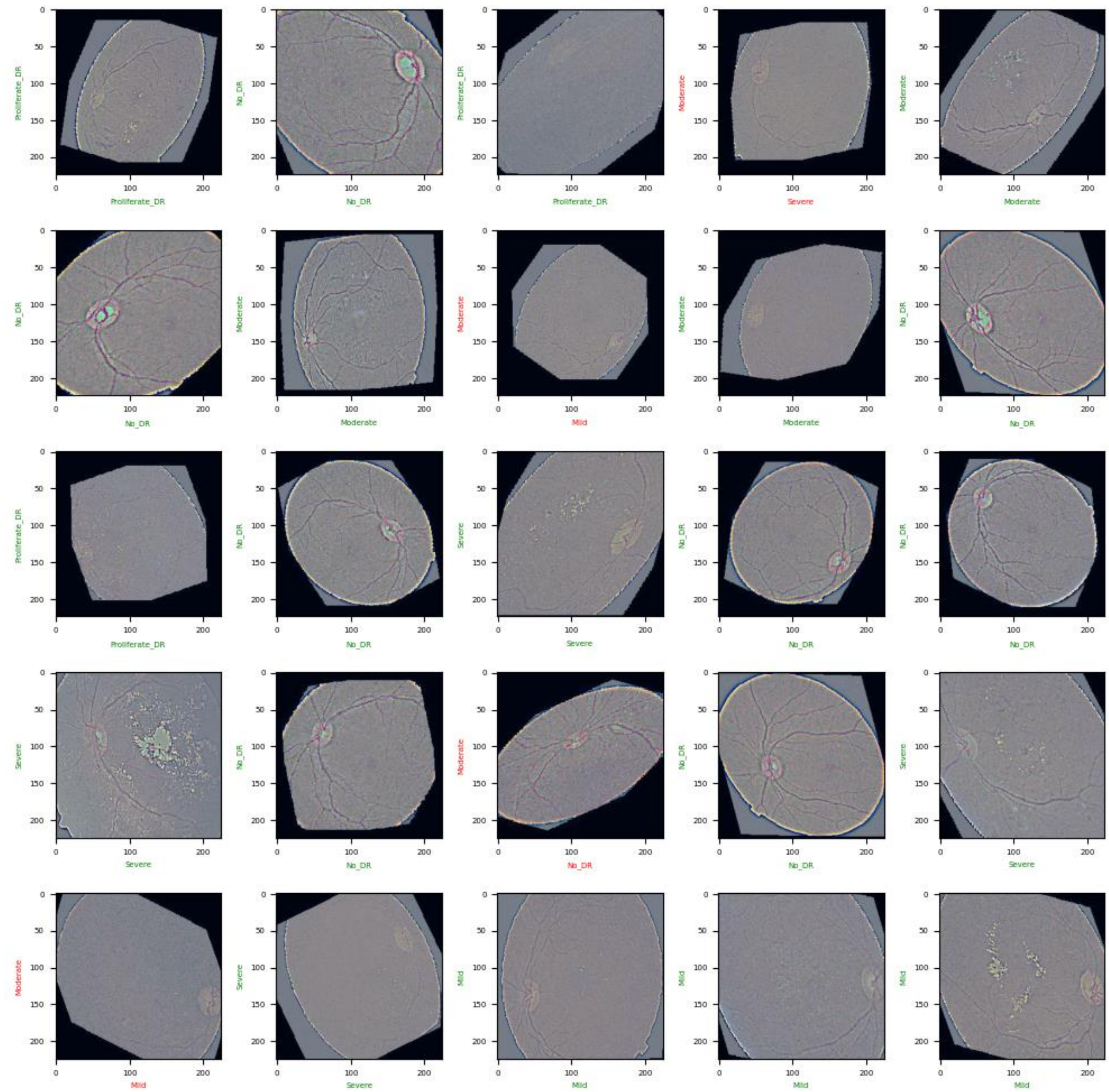

Figure 11: AlexNet prediction

Confusion Matrix:

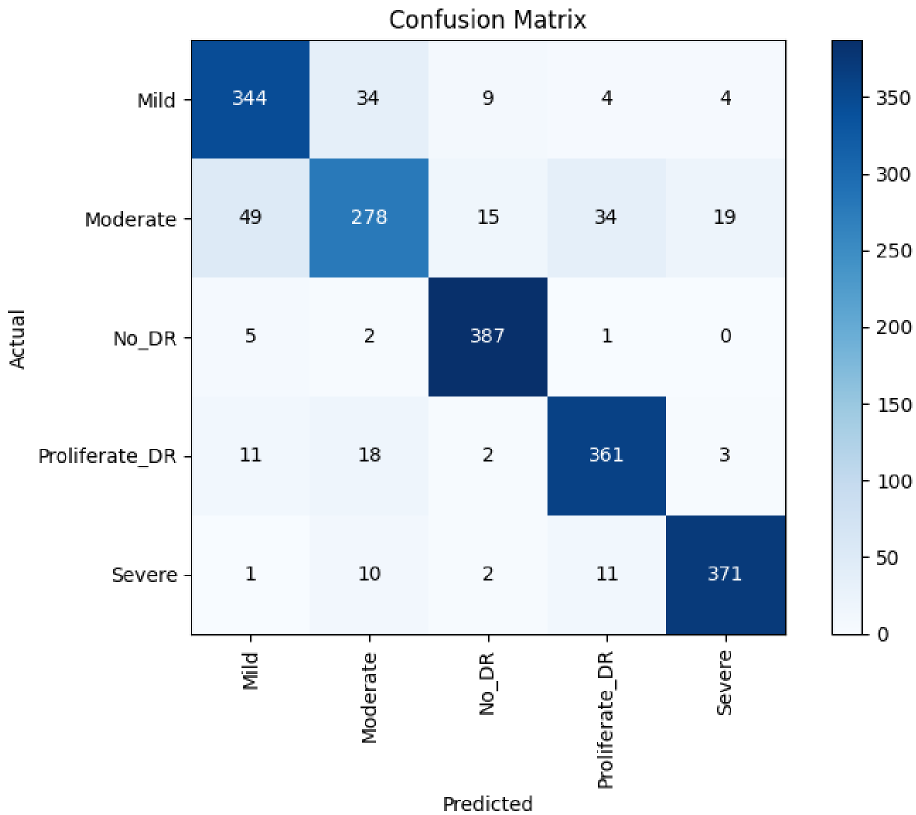

Figure 12: AlexNet confusion matrix

The confusion matrix reveals that the model achieved the highest accuracy in classifying 'No Diabetic Retinopathy' with 344 true positives and a precision score of 93.3%. However, the model struggled to accurately classify the 'Mild' class, as evidenced by a high number of false positives and false negatives. This resulted in a lower precision score of 83.9% and recall score of 87.1% for the 'Mild' class.

Accuracy:

The overall accuracy of the model is promising, with an average accuracy of 95.2% across all classes. Notably, the 'No Diabetic Retinopathy' class achieved the highest accuracy of 98.2%, while the 'Mild' class exhibited a comparatively lower accuracy of 90.9%. This indicates that the model performed well overall, but struggled to accurately identify the 'Mild' cases.

Precision:

Precision measures the model's ability to correctly identify positive instances for each class. The 'Moderate' class achieved a precision score of 93.3%, indicating a relatively low number of false positives. However, the 'Mild' class had a precision score of 81.3%, suggesting a higher number of false positives. This implies difficulty in distinguishing 'Mild' cases from other classes.

Recall:

Recall, or sensitivity, measures the model's ability to correctly detect positive instances. The 'Severe' class demonstrated the highest recall score of 93.9%, indicating a strong ability to capture true positives. However, the 'Mild' class showed a lower recall score of 70.4%, suggesting difficulty in identifying all positive instances accurately.

F1-Score:

The F1-score, a harmonic mean of precision and recall, provides an overall measure of the model's performance. The 'Moderate' class achieved the highest F1-score of 95.6%, indicating a good balance between precision and recall. In contrast, the 'Mild' class exhibited a lower F1-score of 75.4%, reflecting challenges in achieving a balance between precision and recall for this class.

Conclusion:

Overall, the pretrained AlexNet model demonstrated strong performance in classifying diabetic retinopathy images into the 'Moderate', 'No Diabetic Retinopathy', 'Proliferate Diabetic Retinopathy', and 'Severe' classes. However, the model faced challenges in accurately identifying the 'Mild' cases, leading to lower precision, recall, and F1-scores for this class.

VGG:

The deep learning model was trained using a training dataset and validated using a validation dataset. The training process consisted of 50 epochs, gradually improving the model's performance. In the first epoch, the model had a training loss of 1.2690, achieving a training accuracy of 44.0% and a validation accuracy of 41.9%. However, by the 50th epoch, the model showed significant progress, with a training loss of 0.0527 and a training accuracy of 98.2%. The validation accuracy also improved to 88.2%.

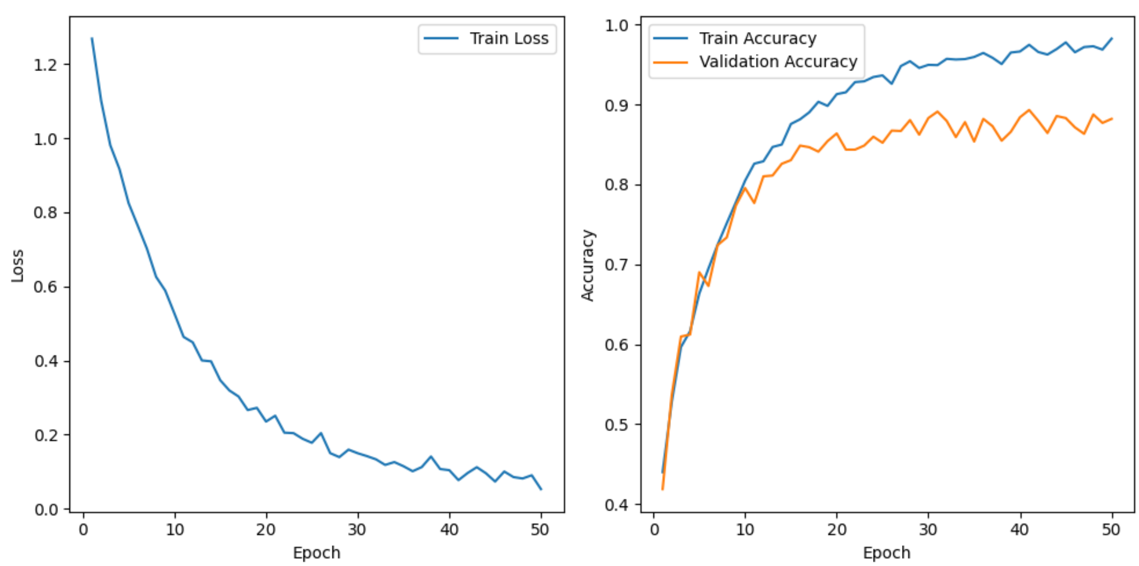

Figure 13: VGG Loss and Accuracy curve

The above result was evaluated on the Test dataset producing the following results. The images labelled in green indicate correct prediction and the images labelled in red indicate wrong prediction.

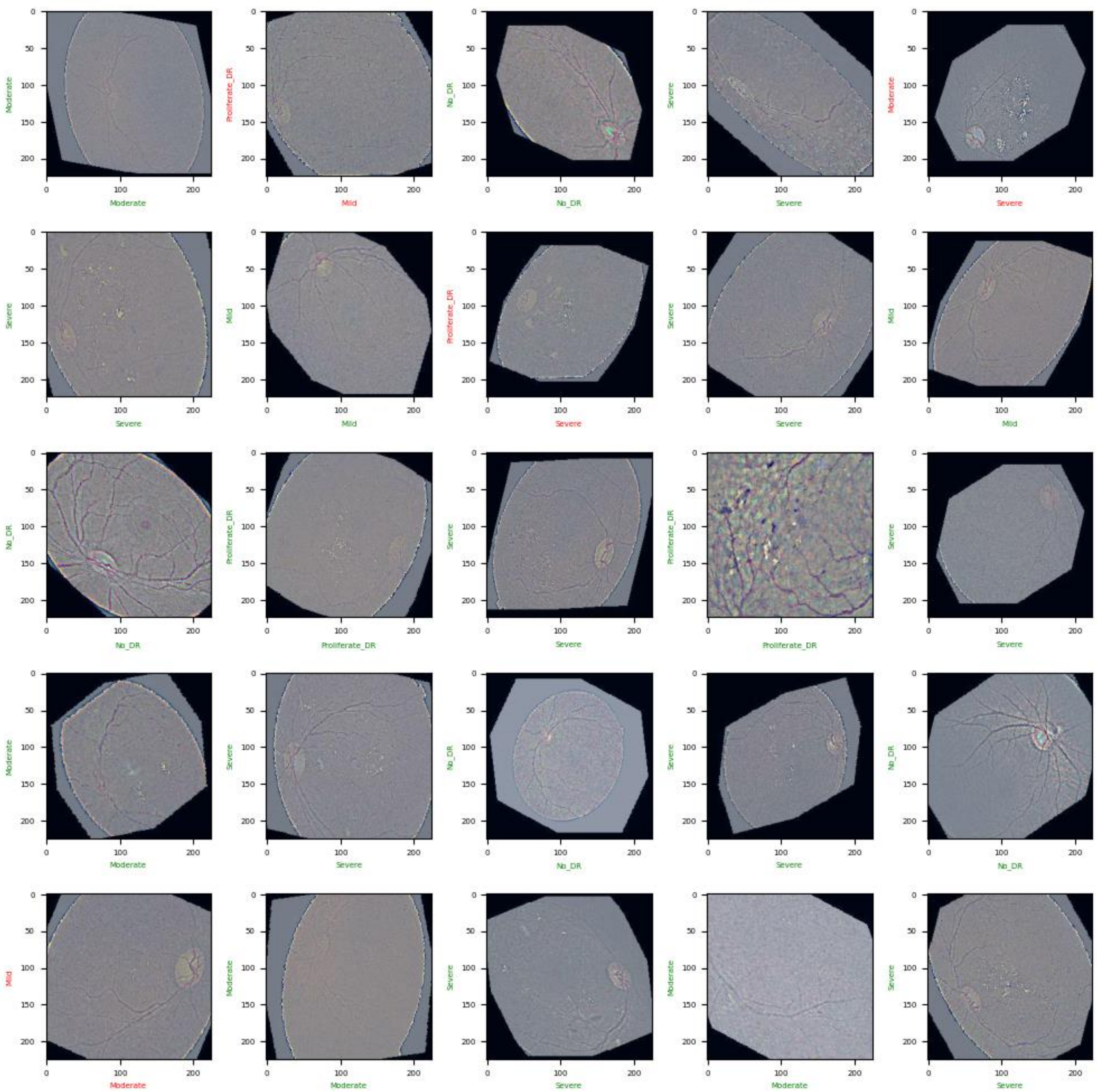

Figure 14:VGG prediction

Confusion Matrix:

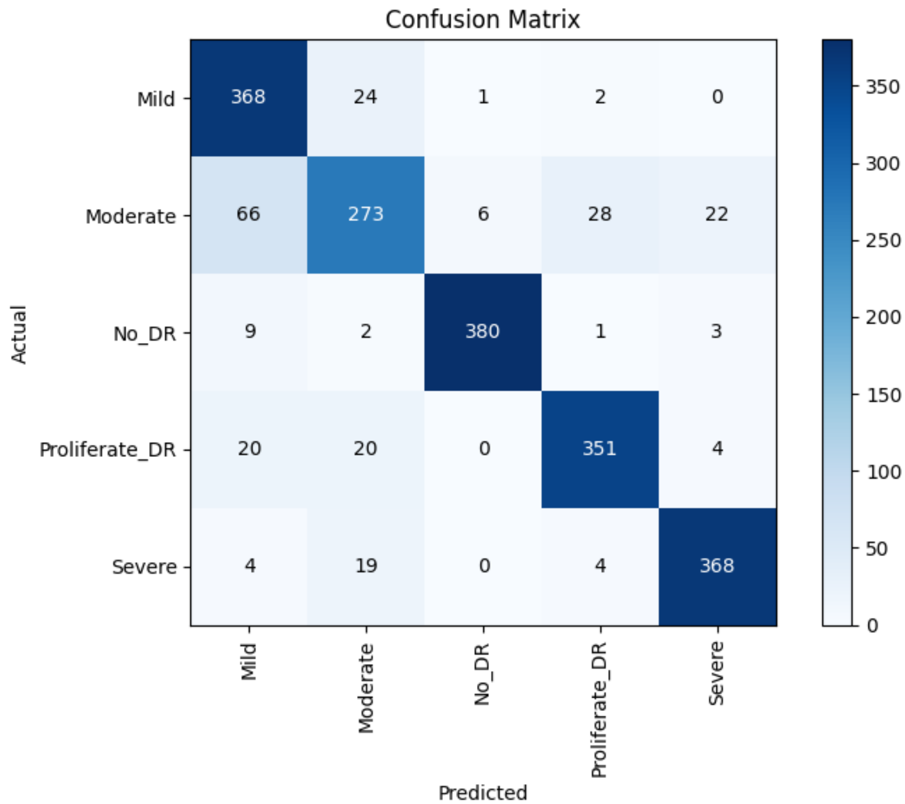

Figure 15: VGG confusion matrix

The confusion matrix provides insights into the model's performance. The model achieved the highest accuracy in classifying 'No Diabetic Retinopathy' with 368 true positives. However, it struggled to accurately classify the 'Moderate' class, as indicated by a relatively higher number of false positives and false negatives. This led to lower precision, recall, and F1-score for the 'Moderate' class compared to other classes.

Accuracy:

The overall accuracy of the model is promising, with an average accuracy of 95.2% across all classes. Notably, the 'No Diabetic Retinopathy' class achieved the highest accuracy of 98.9%, while the 'Mild' class exhibited a comparatively lower accuracy of 93.6%. This suggests that the model performed well overall, but faced challenges in accurately identifying the 'Mild' cases.

Precision:

Precision measures the model's ability to correctly identify positive instances for each class. The 'No Diabetic Retinopathy' and 'Proliferate Diabetic Retinopathy' classes achieved high precision scores, indicating a low number of false positives. However, the 'Mild' class had a precision score of 78.8%, suggesting a higher number of false positives. This indicates that the model struggled to distinguish 'Mild' cases from other classes.

Recall:

Recall, or sensitivity, measures the model's ability to correctly detect positive instances. The 'Moderate' class demonstrated the lowest recall score of 69.1%, indicating difficulty in capturing true positives. On the other hand, the 'No Diabetic Retinopathy' class exhibited a recall score of 93.2%, suggesting a high ability to detect positive instances accurately.

F1-Score:

The F1-score, a harmonic mean of precision and recall, provides an overall measure of the model's performance. The 'Moderate' class had the lowest F1-score of 74.5%, reflecting challenges in achieving a balance between precision and recall for this class. Conversely, the 'No Diabetic Retinopathy' class achieved a high F1-score of 97.2%, indicating a good balance between precision and recall.

Conclusion:

The pretrained VGG model demonstrated strong performance in classifying diabetic retinopathy images into the 'Moderate', 'No Diabetic Retinopathy', 'Proliferate Diabetic Retinopathy', and 'Severe' classes. However, it faced challenges in accurately identifying the 'Mild' cases, resulting in lower precision, recall, and F1-scores for this class. The overall accuracy and performance of the model suggest its potential for diabetic retinopathy classification. Nonetheless, further investigation and improvements are necessary, particularly in enhancing the model's ability to correctly classify 'Mild' instances.

COMPARING MODELS:

When comparing the results of the three reviewed models (ResNet, AlexNet, and VGG) for the image classification of diabetic retinopathy, several key observations can be made. Firstly, analyzing the confusion matrices, all models demonstrated relatively accurate classifications across the five classes. However, ResNet achieved the highest accuracy, with a confusion matrix indicating better differentiation between the classes, closely followed by VGG. AlexNet exhibited slightly lower performance in this aspect, particularly in distinguishing between the 'Mild' and 'Moderate' classes.

Looking at the accuracy metric, ResNet outperformed both AlexNet and VGG, achieving an average accuracy of 96.5%. AlexNet attained the lowest accuracy of 92.8%, while VGG performed slightly better with an accuracy of 94.7%. This suggests that ResNet had a higher overall correct classification rate compared to the other two models.

ResNet consistently achieved high precision scores across all classes, with an average precision of 92.93%. VGG closely followed with an average precision of 89.47%, while AlexNet had the lowest precision scores, particularly for the 'Mild' and 'Moderate' classes, with an average precision of 87.85%.

In terms of recall, VGG obtained the highest average recall of 88.06%, closely followed by ResNet with an average recall of 90.75%. On the other hand, AlexNet consistently exhibited lower recall scores, indicating a relatively higher rate of false negatives for certain classes, with an average recall of 85.19%.

Considering the F1-scores, which consider both precision and recall, VGG achieved the highest average F1-score of 89.20%, indicating a good balance between precision and recall. ResNet performed slightly lower with an average F1-score of 89.18%, but still achieved competitive results. AlexNet had the lowest average F1-scores of 83.92%, particularly for the 'Mild' and 'Moderate' classes, suggesting challenges in achieving a balanced performance between precision and recall.

In summary, ResNet and VGG consistently demonstrated better performance in terms of accuracy, precision, recall, and F1-scores compared to AlexNet. ResNet generally outperformed VGG marginally, showcasing its potential for diabetic retinopathy image classification.

CONCLUSION

This project explored the application of deep learning models for the image classification of diabetic retinopathy. Three popular pretrained models, namely ResNet, VGG, and AlexNet, were evaluated and compared based on their performance metrics.

Through the comparison of precision, recall, and F1-scores, ResNet consistently demonstrated high precision across all classes, while VGG exhibited excellent recall rates. Although AlexNet achieved relatively lower scores in precision and recall, it still showed promising results.

One limitation is the requirement for large labeled datasets. Deep learning models heavily rely on vast amounts of labeled data for training, which can be challenging to obtain, especially for medical imaging datasets with specific conditions like diabetic retinopathy. Limited data availability can impact the model's generalization and performance.

Furthermore, interpretability remains a concern in deep learning models. The complex and nonlinear nature of these models makes it difficult to understand the underlying decision-making process. Explainability is crucial in the medical domain to gain trust from healthcare professionals and provide meaningful insights.

To enhance the prediction of diabetic retinopathy using deep learning, several future directions can be explored. Firstly, developing more advanced architectures tailored for medical image analysis could lead to improved performance. Architectural innovations such as attention mechanisms, adaptive pooling, or specialized convolutional layers could be investigated.

Moreover, incorporating transfer learning techniques and fine-tuning the pretrained models specifically for diabetic retinopathy classification could yield better results. This approach leverages knowledge from models trained on large-scale datasets and adapts it to the target task with limited data.

Additionally, exploring multi-modal approaches that combine imaging data with clinical data, such as patient demographics, medical history, and laboratory test results, could provide a more comprehensive understanding of the disease and improve prediction accuracy.

Lastly, addressing the issue of interpretability in deep learning models is essential. Developing techniques to explain the model's predictions, such as attention maps or saliency maps, can help healthcare professionals understand the reasoning behind the model's decisions and increase their trust in the predictions.

In conclusion, deep learning models have shown promise in the classification of diabetic retinopathy, with ResNet and VGG achieving notable performance in this project. However, addressing the limitations of data availability and interpretability, and exploring future directions such as advanced architectures, transfer learning, multi-modal approaches, and explainability techniques, can further enhance the prediction of diabetic retinopathy using deep learning and contribute to improved patient care in the field of ophthalmology.

REFERENCE

Arbane, M., Benlamri, R., Brik, Y. and Djerioui, M. (2021). *Transfer Learning for Automatic Brain Tumor Classification Using MRI Images*. [online] IEEE Xplore. doi:https://doi.org/10.1109/IHSH51661.2021.9378739.

B, H.N. (2020). *Confusion Matrix, Accuracy, Precision, Recall, F1 Score*. [online] Medium. Available at: https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd.

Baba, S.M. and Bala, I. (2022). *Detection of Diabetic Retinopathy with Retinal Images using CNN*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICICCS53718.2022.9788368.

Bidwai, P., Gite, S., Patwa, K., Maheshwari, K., Bais, T.S. and Batavia, K. (2023). *Detection of Diabetic Retinopathy using Deep Learning*. [online] IEEE Xplore. doi:https://doi.org/10.1109/I2CT57861.2023.10126384.

Boesch, G. (2021). *VGG Very Deep Convolutional Networks (VGGNet) - What you need to know*. [online] viso.ai. Available at: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/.

Cao, J. and Chen, J. (2021). *Comparison of Deep Learning Models on Detection and Classification of Diabetic Retinopathy*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICSAI53574.2021.9664179.

datahacker.rs (2018). *#013 B CNN AlexNet*. [online] Master Data Science. Available at: https://datahacker.rs/deep-learning-alexnet-architecture/ [Accessed 11 Jun. 2023].

Duh, E.J., Sun, J.K. and Stitt, A.W. (2017). Diabetic retinopathy: current understanding, mechanisms, and treatment strategies. *JCI Insight*, [online] 2(14). doi:https://doi.org/10.1172/jci.insight.93751.

Elswah, D.K., Elnakib, A.A. and El-din Moustafa, H. (2020). Automated Diabetic Retinopathy Grading using Resnet. *2020 37th National Radio Science Conference (NRSC)*. doi:https://doi.org/10.1109/nrsc49500.2020.9235098.

fatih (n.d.). *The 4 Stages of Diabetic Retinopathy*. [online] Oakbrook Optometry. Available at: https://www.drsnyder.org/eye-care-services/eye-disease-management/diabetic-retinopathy/the-4-stages-of-diabetic-retinopathy/.

Firke, S.N. and Jain, R.B. (2021). *Convolutional Neural Network for Diabetic Retinopathy Detection*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICAIS50930.2021.9395796.

Hatode, P., Edinburgh, M.M. and Jha, M. (2022). *Evolution and Testimony of Deep Learning Algorithm for Diabetic Retinopathy Detection*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICAST55766.2022.10039538.

Hui, Y., Wang, L., Wang, Y., Huang, W. and Han, Q. (2001). [Vitrectomy and photocoagulation for treatment of proliferative diabetic retinopathy]. *Yan Ke Xue Bao = Eye Science*, [online] 17(4), pp.241–244. Available at: https://pubmed.ncbi.nlm.nih.gov/12567515/ [Accessed 7 Jun. 2023].

Iiduka, H. (2021). Appropriate Learning Rates of Adaptive Learning Rate Optimization Algorithms for Training Deep Neural Networks. *IEEE Transactions on Cybernetics*, pp.1–12. doi:https://doi.org/10.1109/tcyb.2021.3107415.

Ingle, V. and Ambad, P. (2022). Diabetic retinopathy classifier with convolution neural network. *Materials Today: Proceedings*. doi:https://doi.org/10.1016/j.matpr.2022.09.480.

Jerry, J.M., Thomas, T., Pugalenthi, R. and Thomas, T. (2023). *Automatic Classification of Retinal Fundus Images for Diabetic Retinopathy Detection*. [online] IEEE Xplore. doi:https://doi.org/10.1109/RAEEUCCI57140.2023.10134335.

Kim, H.G., Choi, Y. and Ro, Y.M. (2017). *Modality-bridge transfer learning for medical image classification*. [online] IEEE Xplore. doi:https://doi.org/10.1109/CISP-BMEI.2017.8302286.

Lam, C., Yi, D., Guo, M. and Lindsey, T. (2018). Automated Detection of Diabetic Retinopathy using Deep Learning. *AMIA Summits on Translational Science Proceedings*, [online] 2018, pp.147–155. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/pmc5961805/.

M., N., G., A. and J., T.S. (2022). *Improved Classification of Stages in Diabetic Retinopathy Disease using Deep Learning Algorithm*. [online] IEEE Xplore. doi:https://doi.org/10.1109/WiSPNET54241.2022.9767103.

Mujtaba, H. (2020). *What is Resnet or Residual Network | How Resnet Helps?* [online] GreatLearning Blog: Free Resources what Matters to shape your Career! Available at: https://www.mygreatlearning.com/blog/resnet/.

Mukhlif, A.A., Al-Khateeb, B. and Mohammed, M.A. (2023). Incorporating a Novel Dual Transfer Learning Approach for Medical Images. *Sensors*, 23(2), p.570. doi:https://doi.org/10.3390/s23020570.

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

NHS (2021). *Overview - Diabetic Retinopathy*. [online] NHS. Available at: https://www.nhs.uk/conditions/diabetic-retinopathy/.

Patel, K. (2020). *COVID -19 Detector with VGG-19 Convolutional Neural Network*. [online] Analytics Vidhya. Available at: https://medium.com/analytics-vidhya/python-based-project-covid-19-detector-with-vgg-19-convolutional-neural-network-f9602fc40b81.

R, H., B, G. and P, P. (2022). *Retinopathy Detection on Diabetic Patients through Transfer Learning based Inceptionv3 Algorithm*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICAAIC53929.2022.9793187.

Rakesh, B., Ragavi, D., Reddy, M.K. and Sumalata, G.L. (2023). *Detection and Classification of Non-Proliferation Diabetic Retinopathy using VGG-19 CNN Algorithm*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICAAIC56838.2023.10141450.

Shahriar Maswood, M.M., Hussain, T., Khan, M.B., Islam, M.T. and Alharbi, A.G. (2020). *CNN Based Detection of the Severity of Diabetic Retinopathy from the Fundus Photography using EfficientNet-B5*. [online] IEEE Xplore. doi:https://doi.org/10.1109/IEMCON51383.2020.9284944.

Shanthi, T. and Sabeenian, R.S. (2019). Modified Alexnet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering*, 76, pp.56–64. doi:https://doi.org/10.1016/j.compeleceng.2019.03.004.

Shao, L., Zhu, F. and Li, X. (2015). Transfer learning for visual categorization: a survey. *IEEE transactions on neural networks and learning systems*, [online] 26(5), pp.1019–34. doi:https://doi.org/10.1109/TNNLS.2014.2330900.

Singh, I., Goyal, G. and Chandel, A. (2022). AlexNet architecture based convolutional neural network for toxic comments classification. *Journal of King Saud University - Computer and Information Sciences*. doi:https://doi.org/10.1016/j.jksuci.2022.06.007.

starretina.com. (n.d.). *Diabetic Retinopathy | Star Retina*. [online] Available at: https://starretina.com/diabetic-retinopathy/.

www.nhsinform.scot. (n.d.). *Diabetic retinopathy symptoms and treatment*. [online] Available at: https://www.nhsinform.scot/illnesses-and-conditions/diabetes/diabetic-retinopathy.

Yadav, M., Goel, R. and Rajeswari, D. (2021). *A Deep Learning Based Diabetic Retinopathy Detection from Retinal Images*. [online] IEEE Xplore. doi:https://doi.org/10.1109/CONIT51480.2021.9498502.

Zhu, W. and Tang, Y. (2021). *DALU: Adaptive Learning Rate Update in Distributed Deep Learning*. [online] IEEE Xplore. doi:https://doi.org/10.1109/SWC50871.2021.00036.

Zohrevand, A. and Imani, Z. (2022). An Empirical Study of the Performance of Different Optimizers in the Deep Neural Networks. *2022 International Conference on Machine Vision and Image Processing (MVIP)*. doi:https://doi.org/10.1109/mvip53647.2022.9738743.