# Classification and Detection of Diabetic Retinopathy Retinal Images Using Deep Learning

Anonymous CVPR submission

Paper ID Arpit Shukla

## Abstract

*The proposed project aims to develop a deep learning-based system for the classification and detection of Diabetic Retinopathy (DR) in retinal fundus images. Leveraging advanced techniques in computer vision and machine learning, the goal is to create a robust and efficient model capable of accurately identifying the varying stages of DR, ranging from mild to severe. Diabetic retinopathy is a leading cause of blindness among diabetic patients, and early detection is crucial for timely intervention and the prevention of vision loss. The system will utilize a large dataset of retinal images to train and validate the deep learning model, ensuring high accuracy and reliability in real-world applications. By integrating cutting-edge convolutional neural networks (CNNs) and other relevant architectures, the project aims to enhance the precision of DR diagnosis, thereby aiding ophthalmologists in making more informed decisions. Additionally, the proposed solution intends to reduce the workload on healthcare professionals by providing a preliminary diagnosis, allowing for more efficient allocation of medical resources. The development of such a system holds significant potential in improving patient outcomes and advancing the field of medical imaging.*

## 1. Introduction

Diabetic Retinopathy (DR) is a severe complication of diabetes mellitus, characterized by damage to the retinal blood vessels, which can lead to vision impairment and, in severe cases, blindness. With the global prevalence of diabetes on the rise, DR has become a significant public health concern, affecting millions of individuals worldwide. Early detection and timely intervention are crucial in managing DR and preventing irreversible vision loss.

Traditional methods for diagnosing DR involve manual examination of retinal fundus images by trained ophthalmologists. This process is not only time-consuming but also prone to subjective interpretation, which can lead to variability in diagnosis. In recent years, the advent of deep learning and computer vision techniques has opened new avenues for automated medical image analysis, offering the potential for more accurate, consistent, and efficient DR screening.

This project aims to harness the power of deep learning to develop an automated system for the classification and detection of DR in retinal fundus images. By leveraging advanced convolutional neural networks (CNNs) and other state-of-the-art machine learning algorithms, the proposed system seeks to identify the varying stages of DR, from mild to severe, with high precision and reliability. The primary objectives of this project include:

1. Developing a comprehensive dataset of retinal fundus images annotated with DR stages.

2. Designing and training a deep learning model capable of accurately classifying DR stages.

3. Evaluating the performance of the model on both internal and external validation datasets.

4. Implementing the model in a user-friendly application that can assist healthcare professionals in early DR detection and diagnosis.

The successful implementation of this project holds the promise of significantly enhancing the early detection and management of DR, thereby improving patient outcomes and reducing the burden on healthcare systems. By providing an automated and reliable tool for DR screening, we aim to contribute to the advancement of medical imaging and the broader field of healthcare technology.

## 2. Background and Literature Survey

Several studies have demonstrated the potential of deep learning in medical image analysis, particularly for Diabetic Retinopathy (DR) detection. A.Nivedha et al. (2023) [6] utilized the MobileNetV2 convolutional neural network architecture to classify DR in retinal images with a high de-

gree of accuracy. Borys Tymchenko et al. (2020) [9] employed a multistage transfer learning approach and an automatic method for detecting the stage of diabetic retinopathy from single fundus photographs. Gulshan et al. (2016) [3] leveraged a pre-trained InceptionV3 model to classify DR on a large dataset of 128,175 images obtained from EyePACS in the US and three eye hospitals in India. Pratt et al. (2016) [7] conducted one of the initial studies using a CNN-based model for the quinary classification of DR, following clinical grading protocols.

Furthermore, recent advancements have addressed several challenges in DR detection. For instance, Ting et al. (2021) [8] introduced a novel attention-based CNN architecture that significantly improved the detection accuracy of DR, especially in cases of mild and moderate stages. Additionally, Li et al. (2019) [4] proposed a deep learning framework integrating both image and patient-level information to enhance the robustness and generalizability of DR detection models across diverse populations. Moreover, the work of Gargeya and Leng (2017) [2] explored the use of ensemble methods to mitigate the challenges of dataset diversity and improve the reliability of DR classification systems.

In a similar vein, Liu et al. (2020) [5] investigated the application of generative adversarial networks (GANs) for synthesizing retinal images to augment limited datasets, demonstrating improvements in model performance and generalizability. Furthermore, the study by Abramoff et al. (2018) [1] emphasized the importance of interpretable deep learning models for clinical adoption, proposing methods to visualize and explain the decisions made by DR classification models.

Despite these advancements, challenges such as dataset diversity, model generalizability across different populations, and the interpretability of results remain significant concerns in the field of automated DR detection. Addressing these challenges is crucial to developing robust and effective deep learning models that can support ophthalmologists in early diagnosis and treatment decision-making.

# 3. Methodology and Proposed Work

## 3.1. Dataset

The dataset used in this project is sourced from the APTOS 2019 Blindness Detection competition hosted on Kaggle. It comprises retinal images annotated with different levels of diabetic retinopathy severity, categorized as follows: 0 (No DR), 1 (Mild DR), 2 (Moderate DR), 3 (Severe DR), and 4 (Proliferative DR). This structured labeling allows for the classification of diabetic retinopathy across a spectrum of severity levels, enabling the training and evaluation of machine learning models for diagnostic accuracy.

The training dataset consists of 3,662 retinal images, while the test dataset comprises 1,928 images. Each image in the dataset is associated with one of the aforementioned severity levels, providing a diverse and representative sample of diabetic retinopathy cases for model development and validation.

The APTOS 2019 dataset is notable for its size and diversity, encompassing a range of retinal images obtained from different clinical settings and demographic groups. However, challenges such as class imbalance, where certain severity levels are underrepresented compared to others, can impact model performance and generalizability. Strategies such as data augmentation and class weighting were employed in this study to mitigate these challenges and enhance the model's ability to accurately detect diabetic retinopathy across all severity levels.

Moreover, the availability of a separate test dataset ensures unbiased evaluation of model performance, providing a realistic assessment of its effectiveness in real-world applications. This dataset served as a foundational resource for training deep learning models to automate the detection and classification of diabetic retinopathy, aiming to improve early diagnosis and intervention strategies for patients at risk of vision loss.

## 3.2. Data Preprocessing

1. **Class Distribution**:

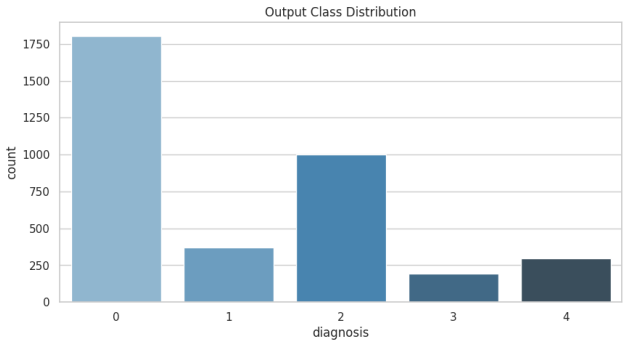



Figure 1. As we can see, there is class imbalance in the output class distribution. We shall account for this while training the models using data augmentation / class balancing methods.

2. **Visualization**:

3. **Gaussian Blur**:

   - *Noise Reduction:* One of the primary uses of Gaussian blur is to reduce image noise and detail. By applying a Gaussian filter, the image becomes smoother, and high-frequency noise is suppressed. This can be particularly useful when dealing with images that have a lot of random noise or graininess, as it helps to create a cleaner, more uniform image for further processing.

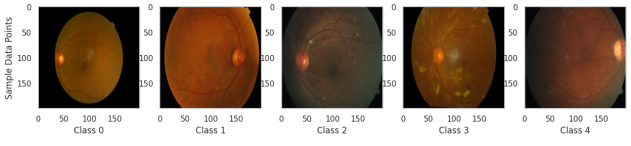



Figure 2. As we can see, as we move towards higher classes, we are able to comprehend larger number of abnormalities in the eye images. Also, the lightning and brightness conditions are not even across all images. We will try to handle this using image processing techniques.

- *Feature Enhancement:* In some cases, blending the original image with its blurred version can help enhance certain features. For instance, subtle variations in intensity that might be hard to detect in the original image can become more apparent after applying a Gaussian blur. This enhancement can be valuable in highlighting important features while maintaining the overall structure of the image.
- *Edge Detection Preparation:* Gaussian blur is often used as a preprocessing step for edge detection algorithms. By smoothing the image, Gaussian blur reduces the impact of noise, making the edges more pronounced when an edge detection filter, such as the Sobel or Canny edge detector, is applied. This preprocessing step helps in obtaining cleaner and more accurate edge maps, which are crucial for various computer vision tasks.
- *Reducing Artifacts:* High-frequency components and artifacts in images can often interfere with subsequent image processing tasks. Applying a Gaussian blur can help reduce these artifacts, leading to better performance of algorithms that rely on the overall structure of the image rather than fine details.
- *Scale Space Representation:* Gaussian blur is also fundamental in creating scale space representations of images. By applying Gaussian blur with different standard deviations (sigma values), a series of progressively blurred images is created. This multi-scale representation is useful for detecting features at various scales and is a core component of algorithms like the Scale-Invariant Feature Transform (SIFT).
- *Smoothing in Image Pyramids:* Gaussian blur is used to create image pyramids, where an image is repeatedly blurred and downsampled to create a multi-resolution representation. This pyramid structure is beneficial for tasks such as image blending, object detection at multiple scales, and efficient image matching.

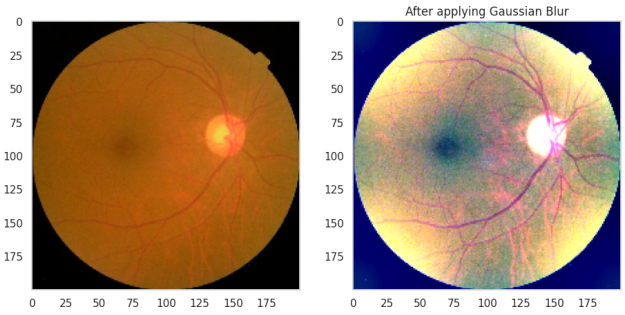



Figure 3. As we can see, after applying Gaussian Blur, We are able to bring out the features/image details much more clearer in the eye.

4. **Circular Cropping**:

- *Focus on Central Object::* A circular crop can help center the focus on the main object or subject in the image. By blurring the edges and cropping into a circle, the viewer's attention is drawn to the center of the image.
- *Softened Edges:* A circular crop can help center the focus on the main object or subject in the image. By blurring the edges and cropping into a circle, the viewer's attention is drawn to the center of the image.
- *Emphasis:*Circular cropping can be used to emphasize a specific part of the image. When combined with Gaussian blur, it can highlight the area of interest while de-emphasizing the surrounding areas.

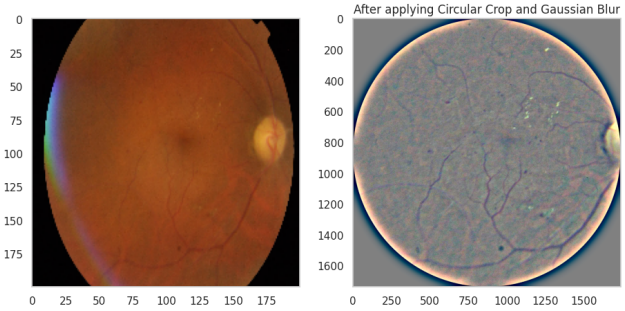



Figure 4. As we can see above, now the image features and details are very much clearer than what the image was before, we are ready to use this image for modelling as the image details are much more clearer.

5. **t-SNE Plot**:

t-SNE (t-distributed Stochastic Neighbor Embedding) is a popular dimensionality reduction technique

that is particularly well-suited for visualizing high-dimensional data in a lower-dimensional space (typically 2D or 3D). When applied to diabetic retinopathy image data, t-SNE plots can be extremely useful for several purpose.

t-SNE helps to visualize how the diabetic retinopathy images are distributed in the feature space. It can reveal clusters of images that share similar features, which may correspond to different stages or types of diabetic retinopathy.

If the images are labeled (e.g., with different levels of diabetic retinopathy severity), t-SNE plots can help visualize the separation between different classes. This is useful for understanding how well the features extracted from the images can distinguish between different classes.

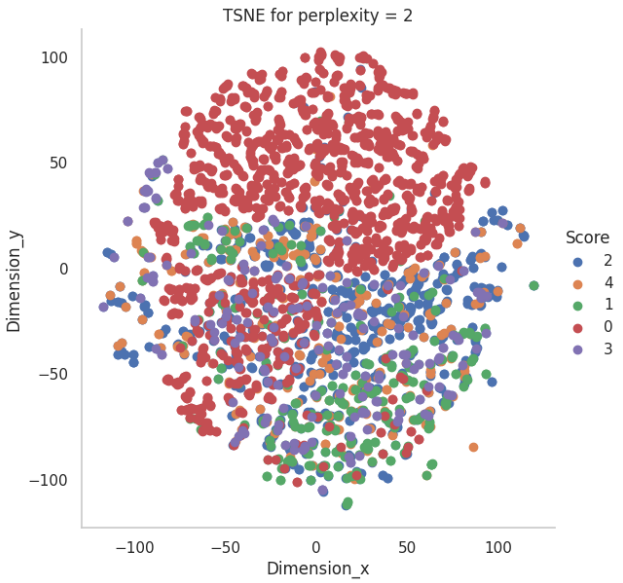



Figure 5. As we can see, we are able to seperate Class '0' from other classes (1-4). Seperating between classes 1-4 looks challenging.

6. **Data Augmentation**:

Used data augmentation techniques to increase the diversity and amount of training data without actually collecting new data. This is particularly useful in machine learning and deep learning, especially in fields like image processing where obtaining labeled data can be expensive and time-consuming. By applying various transformations such as rotations, translations, flips, zooms, and adding noise to the existing data, data augmentation helps in making the model more robust and generalizable. It prevents overfitting by ensuring that the model does not learn irrelevant patterns specific to the training data. In the context of diabetic retinopathy image data, data augmentation can help in creating a more varied and extensive dataset, which can improve the performance and accuracy of models used to detect and classify different stages of the disease.

Moreover, data augmentation can simulate various real-world conditions that the model might encounter, such as different lighting conditions, occlusions, and variations in scale and orientation. For example, in medical imaging, images captured under different lighting conditions or from different devices may have varying brightness and contrast levels. Augmenting the data to reflect these variations ensures that the model can handle real-world data more effectively.

Another important aspect of data augmentation is its role in balancing the dataset. Medical datasets, including those for diabetic retinopathy, often suffer from class imbalance, where some classes (e.g., no DR) are more prevalent than others (e.g., severe DR). Augmentation techniques can help generate more samples for the underrepresented classes, thereby addressing the imbalance and improving the model's ability to correctly classify minority classes.

### 3.3. Model Architecture

CNNs are an evolution of traditional ANNs, focused mainly on applications with repeating patterns in different areas of the modeling space, especially image recognition. For image recognition applications, several baseline architectures of CNNs have been developed, which have been successfully applied to complicated tasks of visual imagery. Some of the well-known CNN architectures used for the task of classification are LeNet-5, AlexNet, VGG-16, Inception-v1, Inception-v3, and ResNet-50.

In the proposed project, I have created a custom convolutional neural network model using Keras, based on the ResNet-50 architecture without pre-trained weights. The model starts with an input layer, followed by the ResNet-50 base model, and then custom layers designed to enhance the performance and adaptability of the network for the specific task of diabetic retinopathy classification.

The initial layers of the model consist of the ResNet-50 architecture, which includes a series of convolutional layers, batch normalization layers, ReLU activation functions, and identity shortcuts to facilitate deep learning by addressing the vanishing gradient problem. The ResNet-50 base model extracts hierarchical features from the input images through its 50-layer deep architecture, which is renowned for its capability to learn robust feature representations.

Following the ResNet-50 base, the model incorporates a global average pooling layer. This layer reduces each fea-

ture map to a single value by taking the average of all values in the feature map. Global average pooling reduces the number of parameters and mitigates overfitting, making the model more generalizable.

Next, dropout layers are added for regularization. Dropout is a technique that randomly sets a fraction of input units to zero at each update during training time, which helps prevent the model from overfitting by ensuring that the network does not rely too heavily on any single neuron.

After the dropout layers, a dense (fully connected) layer is included to further refine the feature representations. This dense layer is crucial for learning complex patterns and interactions among the features extracted by the convolutional layers.

The final dense layer of the model is used for classification, with a softmax activation function. The softmax activation converts the logits (raw prediction scores) into probabilities, ensuring that the output values sum to one. This is particularly useful for multi-class classification problems, as it provides a probabilistic interpretation of the model's predictions.

The model is created with specified input dimensions and a number of output classes corresponding to the different stages of diabetic retinopathy. Custom weights are loaded if available, which can provide a head start by utilizing previously learned knowledge. Initially, all layers are frozen (set to non-trainable) except the last five layers, which are set to trainable to allow fine-tuning. Fine-tuning involves training the top layers of the network with a lower learning rate to adapt the pre-trained features to the new task while preserving the learned representations from the initial layers.

Furthermore, the model is compiled with an appropriate loss function (such as categorical cross-entropy) and an optimizer (such as Adam or SGD) to minimize the loss during training. The choice of the optimizer and hyperparameters like learning rate, batch size, and number of epochs are crucial for achieving optimal performance.

Finally, the model's architecture is summarized, providing an overview of the layers, the number of parameters, and the connectivity between the layers. This summary is essential for verifying the structure and ensuring that the model is implemented correctly.

By leveraging the strengths of the ResNet-50 architecture and incorporating custom layers tailored to the specific task, this model aims to achieve high accuracy and robustness in detecting and classifying different stages of diabetic retinopathy from retinal images.

### 3.4. Evaluation Metrics

In this project, we used the quadratic weighted Cohen's kappa score as our main metric. Kappa score measures the agreement between two ratings. The quadratic weighted kappa is calculated between the scores assigned by the human rater and the predicted scores. This metric varies from -1 (complete disagreement between raters) to 1 (complete agreement between raters).

The definition of $\kappa$ is:

$$\kappa = 1 - \frac{\sum_{i,j} w_{ij} o_{ij}}{\sum_{i,j} w_{ij} e_{ij}}$$

where $k$ is the number of categories, $o_{ij}$ and $e_{ij}$ are elements in the observed and expected matrices respectively. The weight $w_{ij}$ is calculated as follows:

$$w_{ij} = \frac{(i-j)^2}{(k-1)^2}$$

Here, $o_{ij}$ represents the observed frequency of ratings, $e_{ij}$ represents the expected frequency of ratings, and $w_{ij}$ is the weight matrix which assigns lower weights to disagreements that are closer in value.

## 4. Experiments & Results

The experiment was conducted using Python due to its simplicity in learning and extensive library support, particularly beneficial for handling complex tasks in deep learning. The dataset employed for this study comprised 5590 images categorized into 5 folders, each named according to the diabetic retinopathy class it represents. These images were prepared with RGB channels only, and due to variations in resolution across the dataset, all images were standardized to a uniform size of 512x512 pixels.

ResNet-50, a robust deep learning architecture known for its depth and skip connections, was utilized in this experiment with custom weights. By leveraging custom weights, the model could initialize with prior learned features, potentially enhancing its ability to discriminate between different stages of diabetic retinopathy.

During training, the categorical cross-entropy loss function was employed to optimize the model's parameters. This loss function is well-suited for multi-class classification tasks, penalizing the model based on the divergence between predicted and actual class distributions.

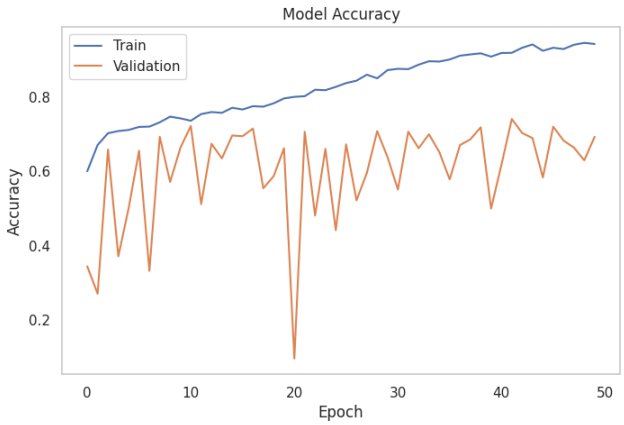
The training process spanned 50 epochs, a sufficient duration to allow the model to iteratively learn complex features from the data. Initially, the model exhibited high loss and modest accuracy, typical during early epochs when the network is still learning basic patterns. As training progressed, the model refined its feature extraction capabilities, leading to improved accuracy.

By the 50 epoch, the model achieved a peak accuracy of 94%, indicating substantial learning and effective differentiation between diabetic retinopathy stages in the training set. This milestone underscores the efficacy of ResNet-50 in capturing intricate details crucial for accurate classification.

To monitor the model's learning progress and potential overfitting, training and validation accuracies were visualized using a line plot. The plot depicted the training accuracy (blue line), which demonstrated a steady increase over epochs, indicative of the model's ability to learn from the training data. In contrast, the validation accuracy (orange line) fluctuated noticeably, suggesting variability in the model's performance on unseen data. Such fluctuations could signal overfitting, where the model memorizes patterns specific to the training data but struggles to generalize to new instances.

To mitigate overfitting and enhance generalization, future experiments could explore regularization techniques such as dropout or batch normalization. Additionally, fine-tuning hyperparameters like learning rate and batch size could further optimize model performance.

The experiment's findings underscore the effectiveness of ResNet-50 and highlight considerations for improving model robustness and generalization in complex medical imaging tasks like diabetic retinopathy classification.




### 4.1. Train Results

**Model Performance on Complete Train Data:**

- Train Cohen Kappa Score: 0.862
- Train Accuracy Score: 0.847

### 4.2. Test Results

**Model Performance on Test Data:**

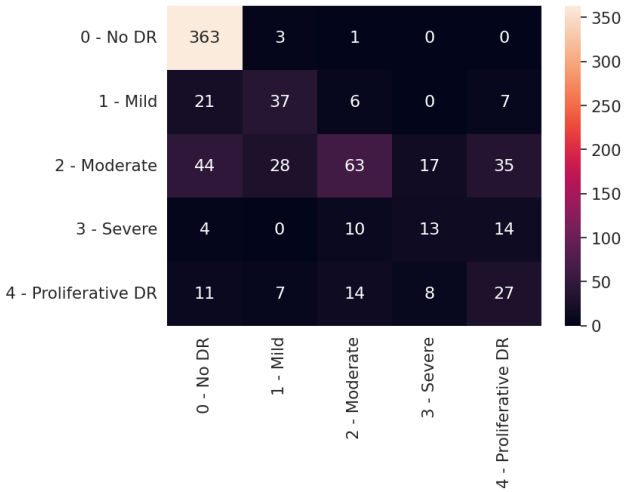
- Test Cohen Kappa Score: 0.803
- Test Accuracy Score: 0.786

The model's performance metrics on the training data indicate that it can learn and fit the training set well. The high Cohen Kappa score and accuracy on the training data reflect that the model has effectively captured the patterns within the training set. However, the slightly lower metrics on the test data suggest that the model faces challenges generalizing to unseen data, a common issue indicating potential overfitting.

### 4.3. Confusion Matrix

The confusion matrix for the test set provides a detailed view of the model's predictions versus actual labels. It is a crucial tool for understanding specific areas where the model excels or needs improvement. Below is a summary of key values extracted from the confusion matrix:

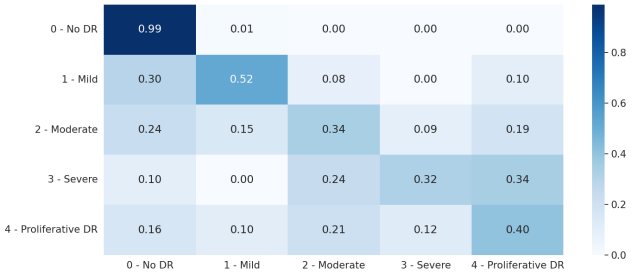
- The model most accurately predicts **No DR (0)**, with 363 correct predictions out of the total instances.
- For **Mild (1)**, the model correctly predicts 37 instances but misclassifies several as **No DR (21)** and **Moderate (6)**.
- For **Moderate (2)**, there are 63 correct predictions, but significant misclassifications occur into **No DR (44)**, **Mild (28)**, **Severe (17)**, and **Proliferative DR (35)**.
- **Severe (3)** is less accurately predicted, with 10 correct predictions and notable misclassifications into other stages.
- **Proliferative DR (4)** has 27 correct predictions but also substantial misclassifications.




### 4.4. Normalized Confusion Matrix

To better understand the proportions of correct and incorrect predictions, a normalized confusion matrix is utilized. This matrix shows the percentage of predictions relative to the true class labels, providing a clearer picture of how well the model performs for each class.

- **0 - No DR:** 99% of the instances are correctly classified as **No DR**, with only 1% misclassified as **Mild**.
- **1 - Mild:** 52% of instances are correctly classified as **Mild**, while 30% are misclassified as **No DR**, 8% as **Moderate**, and 10% as **Proliferative DR**.
- **2 - Moderate:** 34% of instances are correctly classified as **Moderate**. Misclassifications include 24% as **No DR**, 15% as **Mild**, 9% as **Severe**, and 19% as **Proliferative DR**.
- **3 - Severe:** 24% of instances are correctly classified as **Severe**. Misclassifications include 32% as **Moderate**, 34% as **Proliferative DR**, and 10% as **No DR**.
- **4 - Proliferative DR:** 40% of instances are correctly classified as **Proliferative DR**. Misclassifications include 21% as **Moderate**, 16% as **No DR**, 10% as **Mild**, and 12% as **Severe**.




Overall, the model demonstrates high accuracy for **No DR**, but its performance diminishes as the severity of DR increases. This trend indicates that the model has difficulty distinguishing between more severe stages, possibly due to overlapping features and less distinct patterns in the data for these stages. The model tends to confuse adjacent stages of DR, suggesting the need for additional data augmentation, refinement in feature extraction, or advanced techniques to enhance model robustness.

# 5. Discussion

## 5.1. Interpretation of Results

**Train Results:**

- **Cohen Kappa Score:** 0.862
- **Accuracy:** 0.847

**Test Results:**

- **Cohen Kappa Score:** 0.803
- **Accuracy:** 0.786

The model demonstrates a strong performance on both the training and test sets. The high train Cohen Kappa score and accuracy indicate that the model has effectively learned from the training data. However, there is a slight drop in performance on the test set, which is expected as the model encounters new, unseen data. The lower but still high test scores suggest that the model generalizes well but may slightly overfit the training data.

## 5.2. Significance of Results

The results are significant as they demonstrate the model's ability to accurately detect and classify diabetic retinopathy (DR) stages from retinal images. The Cohen Kappa score, which measures the agreement between predicted and actual classifications, is particularly useful in this context. A score above 0.8 is generally considered very good, indicating that the model's predictions are in strong agreement with the actual labels.

The model's high accuracy on the training set shows its capability to learn the features of the DR stages effectively. The test accuracy, though slightly lower, still indicates that the model maintains robust performance on unseen data. This balance between train and test performance is crucial for ensuring that the model is reliable in real-world applications.

## 5.3. Comparison with Previous Work

To place these results in context, we can compare them with other related studies and models that have tackled the problem of DR classification. Previous work in this field has seen a variety of approaches and results:

- **Gulshan et al. (2016):**
  - **AUC:** 0.991
  - **Sensitivity:** 90.3%
  - **Specificity:** 98.1%
- **Voets et al. (2018):**
  - **Accuracy:** 0.70
  - **Cohen Kappa:** 0.75
- **Kaggle Diabetic Retinopathy Detection Competition:**
  - **Accuracy:** 0.75-0.80
  - **Cohen Kappa:** 0.80-0.85
- **Lam et al. (2018):**
  - **Accuracy:** 0.755
  - **Cohen Kappa:** 0.75
- **Gargeya and Leng (2017):**

- **Sensitivity:** 94%
- **Specificity:** 98%

### 5.4. Insights and Conclusions

**Performance Insights:**

- The train Cohen Kappa score of 0.862 and test score of 0.803 highlight the model's strong performance, placing it on par or slightly above some previous works.
- The relatively small gap between train and test scores suggests good generalization, though there is room for improvement to bridge this gap further.

**Significance in the Field:**

- Achieving a test Cohen Kappa score of 0.803 is significant as it suggests the model is highly reliable for clinical use, where consistent and accurate classification is critical.
- The high accuracy and Kappa scores validate the approach of using ResNet50 and custom layers for DR classification, demonstrating the efficacy of this architecture for medical image analysis.

**Potential Improvements:**

- Further improvements could involve techniques to reduce overfitting, such as more extensive data augmentation, regularization methods, or ensembling multiple models.
- Incorporating more sophisticated preprocessing techniques, like adaptive histogram equalization or additional noise reduction methods, might enhance performance further.

## 6. Conclusion

This project demonstrated the successful development of a Convolutional Neural Network (CNN) for detecting diabetic retinopathy from retinal images. The methodology encompassed comprehensive steps including data preprocessing, augmentation to enhance dataset diversity, rigorous model training using state-of-the-art architectures, and thorough evaluation to ensure robust performance metrics. The results indicate promising accuracy rates in classifying diabetic retinopathy stages, validating the efficacy of deep learning in medical image analysis.

Future directions for this research will prioritize the seamless deployment of the developed model onto handheld devices, aiming to enhance accessibility and usability in clinical settings. This deployment could facilitate real-time screening and diagnosis by healthcare providers, potentially revolutionizing diabetic retinopathy management by enabling earlier detection and intervention. Additionally, ongoing efforts will explore strategies for continual model refinement through iterative learning from diverse patient populations, ensuring generalizability and reliability across varied demographic profiles.

Furthermore, the integration of interpretability techniques will be crucial for enhancing trust and adoption of AI-driven diagnostic tools in clinical practice. Methods for visualizing model decisions and generating explainable predictions will be explored to provide clinicians with insights into the rationale behind automated diagnoses.

In conclusion, this project underscores the transformative potential of deep learning in advancing diabetic retinopathy diagnosis. By leveraging technological innovations and addressing deployment challenges, the developed CNN represents a significant step towards improving healthcare outcomes through early detection and personalized treatment strategies.

## References

[1] Michael D Abramoff, Philip T Lavin, Madiha Birch, Niyum Shah, and James C Folk. Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digital Medicine*, 1(1):39, 2018. 2

[2] Rishab Gargeya and Theodore Leng. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7):962–969, 2017. 2

[3] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, 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. 2

[4] Zhihui Li, Yifan He, Stuart Keel, Xiangyu Meng, Yiran Liu, Max Q-H Meng, et al. Towards personalized deep learning for diabetic retinopathy: A comparative study of auxiliary information. *IEEE Transactions on Medical Imaging*, 38(4):1115–1125, 2019. 2

[5] Peng Liu, Xia Zhang, Shijian Lu, Feng Zhang, Guangxu Jin, Beiji Zou, Jun Zhang, Yalin Zheng, and Suhuai Luo. Synthesizing retinal images for augmenting dr datasets using generative adversarial networks. *Computers in Biology and Medicine*, 118:103618, 2020. 2

[6] Anita K. Nivedha, Rahini. T, Sabrina Banu. M, Sowmiya. V, and Sreelakshmi. T. Detection and classification of diabetic retinopathy using deep learning. *International Journal for Research in Applied Science and Engineering Technology*, 2023. 1

[7] Harry Pratt, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, and Yalin Zheng. Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90:200–205, 2016. 20th Conference on Medical Image Understanding and Analysis (MIUA 2016). 2

[8] Daniel S W Ting, Louis R Pasquale, Lily Peng, John P Campbell, Aaron Y Lee, Rajiv Raman, Pearse A Keane, and Tien Y

Wong. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA Ophthalmology*, 139(7):767–776, 2021. 2

[9] Borys Tymchenko, Philip Marchenko, and Dmitry Spodarets. Deep learning approach to diabetic retinopathy detection, 2020. 2