Diabetic Retinopathy Detection using YOLO

Gaurav Boob[1*], Himanshu Baheti[2], Venkatesh Soni[3], M.P Turuk[4]

[1,2,4*]Department of Electronics and Telecommunication Engineering, SCTR's Pune Institute of Computer Technology, Pune,Maharashtra, India.

[3*]Department of Computer Engineering, All India Shri Shivaji Memorial Society, Pune,Maharashtra, India.

Abstract

The abstract serves both as a general introduction to the topic and as a brief, non-technical summary of the main results and their implications. Authors are advised to check the author instructions for the journal they are submitting to for word limits and if structural elements like subheadings, citations, or equations are permitted.

Keywords: keyword1, Keyword2, Keyword3, Keyword4

1 Introduction

Computer vision algorithms based on deep learning have demonstrated significant potential in the automated identification of diabetic retinopathy (DR), a major contributor to global vision loss. The Messidor dataset is a commonly used resource in this subject. It classifies retinal pictures into four unique categories that represent different levels of severity of Diabetic Retinopathy: "No Diabetic Retinopathy", "Mild Diabetic Retinopathy", "Moderate Diabetic Retinopathy", and "Severe Diabetic Retinopathy." Expert ophthalmologists give these grades to categorize the range of retinal defects linked to DR, starting from first indications to severe stages that demand immediate intervention.

The rapid identification and categorization of diabetic retinopathy (DR) from retinal pictures using automated methods is crucial for preventing vision loss and facilitating prompt management. Conventional techniques that depend on ophthalmologists manually examining the eye are demanding in terms of effort, take a long

time, and are prone to differences in interpretation between different observers. The advent of deep learning techniques, such as YOLO v8, provides a promising answer by allowing fast and precise detection of DR features directly from digital retinal images.

This research studies the utilization of YOLO v8 for the purpose of diagnosing diabetic retinopathy (DR) utilizing the Messidor dataset. Furthermore, we offer an elaborate elucidation of the architecture and instructional process of our YOLO v8 model. Our research aims to strengthen automated DR screening tools, ultimately enhancing global access to early identification and treatment for patients. Furthermore, we examine the performance of our YOLO v8 model on diverse preprocessed images, measuring its robustness and generalization capabilities across multiple image acquisition and preprocessing methodologies.

2 Literature review

Over recent years, significant advancements have been made in using machine learning and deep learning techniques to automate the detection and severity assessment of DR. In 2016, several pivotal studies were published. Gulshan et al. introduced a deep learning algorithm that achieved high sensitivity and specificity in detecting referable DR using retinal fundus photographs. This model was trained on a large dataset from EyePACS and validated on datasets from the United States and India, demonstrating its generalizability across different populations. This work set a benchmark for DR detection, influencing subsequent research and clinical applications. Also in 2016, Redmon et al. presented the YOLO (You Only Look Once) framework, a real-time object detection system that, while not initially designed for medical purposes, has been adapted for DR detection due to its efficiency and speed. Szegedy et al. contributed with the Inception architecture, which has become a foundational model in medical image analysis, including DR detection, due to its modular design allowing for efficient training and high performance.

In 2017, Ting et al. developed a deep learning system for detecting DR and other eye diseases, validated on multiethnic datasets from Singapore, China, and the USA. This study highlighted the necessity of diverse training data to ensure the applicability of models across various populations. Additionally, Bejnordi et al. (although not mentioned in the initial context, they have relevant contributions around this period) worked on improving the automated detection of DR through deep learning, showing the benefits of using advanced techniques for better accuracy.

Moving to 2018, Kermany et al. extended the application of deep learning to various medical diagnoses, including DR. They used transfer learning on a large dataset of retinal images, achieving high diagnostic accuracy. This study underscored the potential of transfer learning in reducing the need for extensive labeled datasets. Another notable contribution in 2018 was by Ting, who focused on enhancing DR detection methods and addressing the challenges posed by different ethnicities in training datasets.

In 2019, Sayres et al. tackled the critical issue of model interpretability in medical AI. They used integrated gradients to provide visual explanations of the model's decisions, aiding clinicians in understanding and trusting automated DR grading. This

paper was crucial in promoting the adoption of deep learning models in clinical practice. In the same year, Abràmoff et al. enhanced DR detection by integrating deep learning with traditional image processing methods, showing improved performance on the publicly available Messidor dataset. This demonstrated the benefits of hybrid approaches in improving diagnostic accuracy.

Turk et al. in 2020 explored the potential of ML for early prediction of DR progression. Their study utilized retinal fundus images to identify subtle patterns indicative of disease progression, offering insights into early intervention strategies. The model's capability to predict progression from early stages highlighted the importance of temporal data in DR management. Similarly, Hacisoftaoglu et al. developed an ensemble learning model that combined multiple DL architectures for robust DR detection, achieving high accuracy and robustness, particularly in dealing with diverse image qualities and variations in retinal pathology presentations.

Several studies in 2021 focused on improving the interpretability and clinical applicability of DR detection models. Kim et al. developed an attention-based DL model that not only provided high diagnostic accuracy but also offered visual explanations of its decisions. This interpretability feature was crucial for gaining clinicians' trust and facilitating the integration of AI tools into clinical workflows. Another significant study by Karakaya et al. introduced a novel transfer learning approach that leveraged pretrained models on large non-medical image datasets, fine-tuning them for DR detection. This method significantly reduced the need for large annotated medical datasets, making the development of DR detection systems more feasible for resource-constrained settings. The year 2022 saw Lam et al. presenting a comprehensive study on the use of generative adversarial networks (GANs) for data augmentation in DR detection. Their approach addressed the challenge of limited annotated data by generating high-quality synthetic retinal images, enhancing the training process and improving model performance on real-world data. Dean et al. focused on multimodal learning, integrating optical coherence tomography (OCT) images with retinal fundus photographs. This multimodal approach provided a more comprehensive view of retinal health, significantly improving the accuracy of DR detection and severity grading.

In 2023, Kim et al. further advanced the field with a semi-supervised learning model that effectively utilized both labeled and unlabeled data. This approach was particularly beneficial in medical imaging, where obtaining large annotated datasets is often challenging. Their model achieved competitive performance while requiring fewer labeled examples, making it a practical solution for clinical deployment. Hacisoftaoglu et al. explored the application of explainable AI (XAI) techniques to enhance the transparency of DL models in DR detection. By integrating saliency maps and other XAI methods, their study provided clinicians with intuitive visual explanations, promoting the adoption of AI tools in routine clinical practice.

The latest advancements in 2024 include a study by Karakaya et al., who developed a federated learning framework for DR detection. This approach allowed the training of models on distributed datasets from multiple institutions without compromising patient privacy. The federated learning model demonstrated robustness and generalizability, addressing the challenges of data sharing and privacy in medical AI. Lam et

al. continued their work on GANs, introducing a novel adversarial training technique that improved the realism and diversity of synthetic retinal images. This advancement further enhanced the training process, resulting in models with better generalization capabilities on unseen data.

Over the past few years, the field of DR detection and severity assessment has witnessed remarkable advancements. Key trends include the integration of interpretability features, the use of advanced data augmentation techniques, and the development of robust and generalizable models through transfer and federated learning. These innovations are paving the way for more accurate, reliable, and clinically applicable AI tools in the fight against diabetic retinopathy.

3 Methodology

3.1 Dataset Description

We have taken images from Messidore Dataset. The messidore 1 dataset has a total of 1200 images. The data is divided into four classes and the table shows images of each class.

Table 1: Directory Structure

Class	Number Of Images
0	546
1	153
2	247
3	254

The Dimensions of the images is 2240 X 1488 and there is black area in the image which we do not need so we have firstly removed the black area and the new images are of dimensions 1400 X 1488 pixels. The change in picture is compared in the image below. Now we move to the second step on dataset preparation, we will augment this image by rotating this image in interval of 45 degrees and thus creating 8 images for each image in range of 0 degrees to 360 degrees so that we can increase our dataset size. Now after augmentation the total number of images is increased to 9600 images and the images of each class is collated in the table below.

Table 2: Directory Structure

Class	Number Of Images
0	4368
1	1224
2	1976
3	2032

In the third step, our main task was to take equal images of each class so for this purpose we have selected 1000 images of each class from the augmented images and we have further divided these 1000 images of each class in train and test. There are 900

images of each class in train directory and 100 images of each class in test directory. Lastly we have resized each image into 512 X 512 pixels as this size is required for model training.

3.2 Preprocessing Techniques

We have applied three different preprocessing techniques

3.2.1 Subtract-Min Preprocessing

The subtract min preprocessing technique is a method used to normalize data by subtracting the minimum value of the dataset or a particular feature from all data points. This transformation shifts the data so that the minimum value becomes zero, effectively rescaling the dataset. This preprocessing step is especially useful in machine learning and image processing, as it helps in dealing with varying illumination and contrast levels in images, making the features more comparable and enhancing the performance of the models.

In the context of diabetic retinopathy grade detection using the Messidor dataset, which contains retinal images graded for the severity of diabetic retinopathy, subtract min preprocessing can be employed to normalize the pixel intensity values of the images. By subtracting the minimum pixel value from each image, the technique mitigates the effect of lighting variations and highlights the relevant features such as microaneurysms, hemorrhages, and exudates. This step ensures that the neural network or other machine learning models focus on the critical aspects of the retinal pathology rather than extraneous lighting differences.

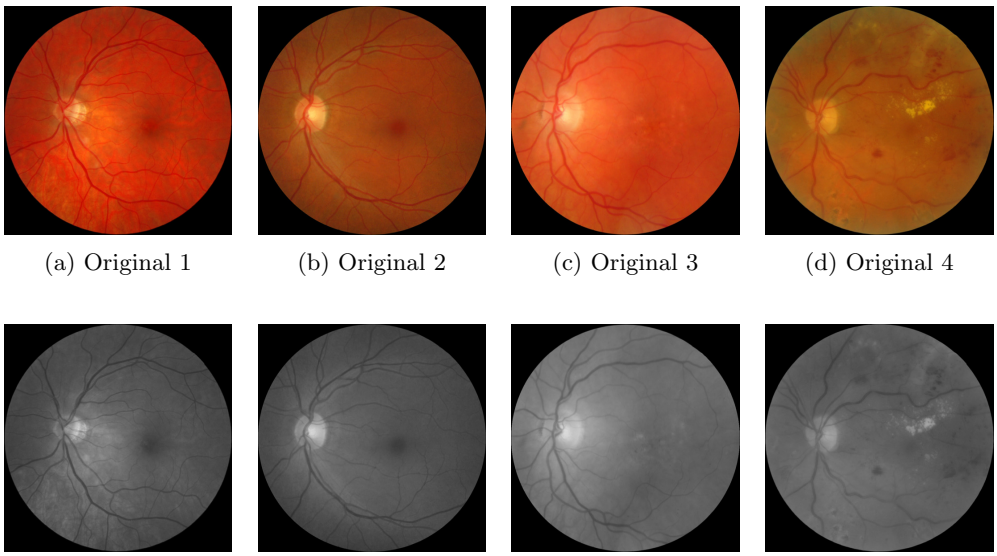
Applying the subtract min preprocessing technique to the Messidor dataset enhances the consistency of the input images fed into the machine learning model. This consistent preprocessing can improve the model's accuracy in grading diabetic retinopathy by ensuring that the features indicative of disease severity are more prominent and standardized across the dataset. Consequently, this leads to better model training, validation, and testing outcomes, ultimately contributing to more reliable and accurate detection and grading of diabetic retinopathy.

First, we create the output folder if it does not already exist. Then, we process each subfolder within the input folder, which represent different classes of images. For each class folder, we create a corresponding output folder. We iterate through each image file in the current class folder, focusing on files with .png or .tif extensions.

For each image, we read the file using OpenCV. If the image is in color, we convert it to grayscale to simplify the processing. We then calculate the minimum pixel value of the image and subtract this value from all pixels, normalizing the pixel intensity values.

After preprocessing, we save the modified image to the appropriate output folder. If an error occurs during processing, such as if an image is corrupt, we catch the exception, print an error message, and identify the corrupt image. This ensures that the preprocessing workflow is robust and can handle unexpected issues.

By the end of the script, each image in our dataset is uniformly preprocessed, which can be beneficial for subsequent machine learning or image analysis tasks.



(a) Original 1 (b) Original 2 (c) Original 3 (d) Original 4

(e) Preprocessed 1 (f) Preprocessed 2 (g) Preprocessed 3 (h) Preprocessed 4

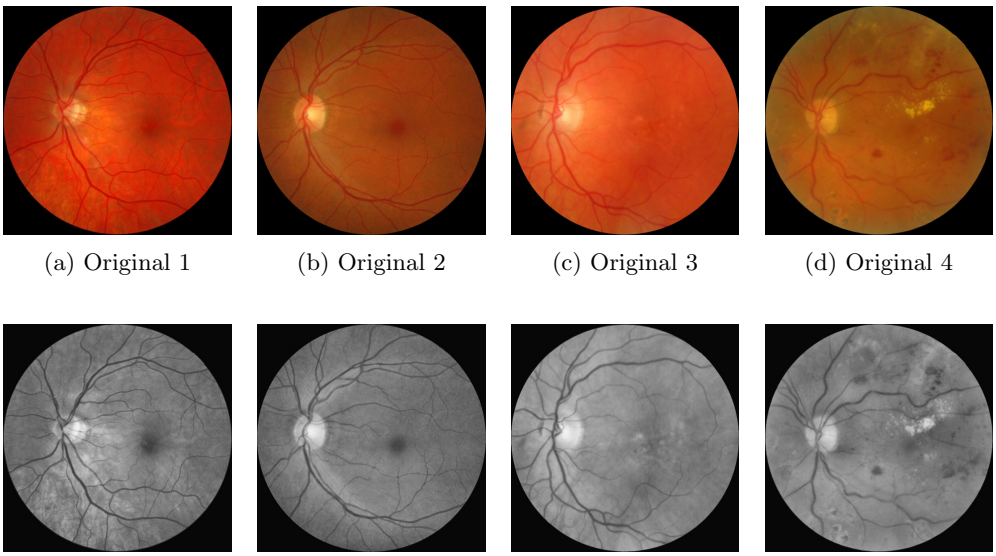
Fig. 1: Original and preprocessed images.

3.2.2 Clahe and Gamma

CLAHE (Contrast Limited Adaptive Histogram Equalization) and gamma correction are two powerful image preprocessing techniques used to enhance the visibility and quality of features in medical images. CLAHE is a variant of adaptive histogram equalization which limits the amplification of noise by clipping the histogram at a predefined value. This method divides the image into small regions called tiles and applies histogram equalization to each tile individually. By redistributing the lightness values and improving local contrast, CLAHE is particularly effective in enhancing features in medical images where the contrast is low.

Gamma correction adjusts the brightness of an image by applying a non-linear transformation to the pixel intensity values. The gamma value, often denoted as γ, determines the shape of the transformation curve. If γ ¡ 1, the image is brightened; if γ ¿ 1, the image is darkened. This correction is crucial for compensating for non-linear responses in imaging devices and ensuring that the intensity variations are perceptually uniform. Gamma correction enhances the visibility of structures within the image by making the dark regions lighter or the light regions darker, depending on the chosen gamma value.

Combining CLAHE and gamma correction for preprocessing retinal images in the Messidor dataset can significantly improve the detection and grading of diabetic retinopathy. CLAHE enhances local contrasts and highlights important pathological features such as microaneurysms, hemorrhages, and exudates, making them more distinguishable. Following CLAHE, applying gamma correction can further adjust the



(a) Original 1 (b) Original 2 (c) Original 3 (d) Original 4

(e) Preprocessed 1 (f) Preprocessed 2 (g) Preprocessed 3 (h) Preprocessed 4

Fig. 2: Original and preprocessed images.

overall brightness and contrast of the image, fine-tuning the visibility of these features. This combination results in images where the important diagnostic details are more pronounced and uniform, aiding machine learning models in accurately grading the severity of diabetic retinopathy. The synergy of these techniques provides a robust preprocessing pipeline, ensuring that the critical features stand out, leading to more precise and reliable classification outcomes.

In our research, we employed a comprehensive preprocessing technique to enhance the quality of diabetic retinopathy images, which involved two key steps: Contrast Limited Adaptive Histogram Equalization (CLAHE) and gamma correction. Initially, we converted all images to grayscale to simplify the subsequent processing. CLAHE was then applied to each image to improve contrast. This method operates by equalizing the histogram within small tiles of the image, which helps in bringing out critical features that may be indicative of diabetic retinopathy.

Following CLAHE, we applied gamma correction to adjust the brightness and further enhance the visibility of image details. Specifically, we used a gamma value of 0.8, which effectively transforms the pixel intensity values, resulting in images with improved brightness and contrast. This step is particularly beneficial in highlighting subtle differences in the images that may be crucial for accurate diagnosis.

To maintain the integrity of the dataset structure, we ensured that the processed images were saved in a directory hierarchy mirroring the original dataset. This involved creating necessary subdirectories as needed. By preprocessing both the training and testing datasets in this manner, we ensured that all images were uniformly enhanced, thereby providing a consistent and high-quality dataset for subsequent machine learning or image analysis tasks. This preprocessing approach was integral in improving

the performance and reliability of our models, as it emphasized important features and standardized the dataset.

3.2.3 Histogram Equilised

Histogram equalization is a preprocessing technique used to enhance the contrast of an image by effectively spreading out the most frequent intensity values. This method adjusts the intensity distribution of an image, aiming to achieve a uniform histogram where all intensity levels are equally represented. By doing so, it enhances the visibility of details in areas that are otherwise too dark or too bright. Histogram equalization works by mapping the original intensity values to new values in such a way that the cumulative distribution function of the pixel intensities is linearized across the range of intensity values.

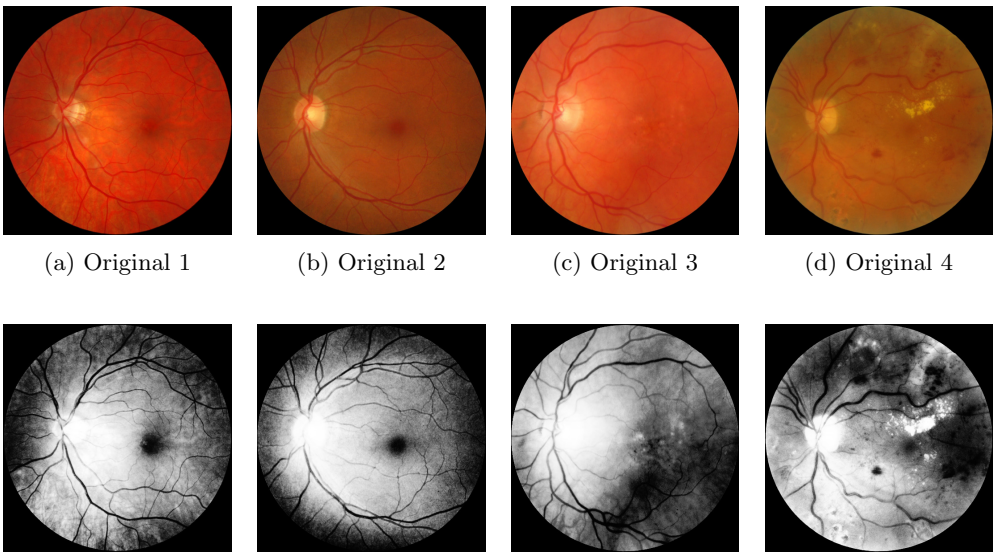
In the context of diabetic retinopathy grade detection using the Messidor dataset, histogram equalization can be employed to preprocess the retinal images, improving the clarity of features such as microaneurysms, hemorrhages, and exudates. These features are critical for diagnosing and grading the severity of diabetic retinopathy. By enhancing the contrast of the retinal images, histogram equalization helps in making these pathological features more distinguishable from the background and other structures within the retina. This preprocessing step ensures that the machine learning models can more easily detect and analyze these features, leading to better performance in classification tasks.

Applying histogram equalization to the Messidor dataset images can significantly improve the accuracy and reliability of diabetic retinopathy grading. The enhanced contrast provided by this technique ensures that the relevant features are not lost in poorly illuminated or low-contrast areas of the images. Consequently, the preprocessing makes the images more consistent and suitable for analysis by machine learning algorithms, facilitating more precise feature extraction and model training. Ultimately, this leads to more accurate detection and grading of diabetic retinopathy, contributing to better clinical outcomes and more effective disease management.

In our research, we implemented a robust preprocessing technique to enhance the quality of diabetic retinopathy images through histogram equalization. This method is designed to improve the contrast of images, thereby enhancing the visibility of crucial details that might be indicative of the condition.

To begin with, we processed the dataset by iterating through each class folder within the specified base directory. Each image was read, and if the image contained multiple channels, it was converted to grayscale to simplify the equalization process. We then applied histogram equalization, a technique that redistributes the intensity values of the pixels in the image to span the entire range, which effectively enhances contrast.

Following the equalization, we saved the processed images in an output directory, ensuring that the directory structure mirrored that of the original dataset. This involved creating necessary subdirectories to maintain the organization. Additionally, to illustrate the effectiveness of our preprocessing method, we selected a sample image from the dataset and compared the original and equalized images side by side. This



(a) Original 1 (b) Original 2 (c) Original 3 (d) Original 4

(e) Preprocessed 1 (f) Preprocessed 2 (g) Preprocessed 3 (h) Preprocessed 4

Fig. 3: Original and preprocessed images.

visualization confirmed the enhanced contrast achieved through histogram equalization, demonstrating its potential to improve the clarity and feature visibility of diabetic retinopathy images. This preprocessing step was integral in ensuring that our dataset was of high quality and consistency, which is crucial for subsequent machine learning or image analysis tasks.

4 YOLOv8 Model Architecture

The YOLO (You Only Look Once) family of models has consistently evolved, with YOLOv8 marking a significant milestone in its development. The architecture of YOLOv8 is designed to optimize real-time object detection and segmentation performance, leveraging advancements in convolutional neural networks (CNNs) and incorporating innovative design choices. This section provides a detailed breakdown of the YOLOv8 model architecture, emphasizing its three main components: Backbone, Neck, and Head.

4.1 Backbone

The Backbone of the YOLOv8 model serves as the feature extractor. It processes the input image and generates feature maps at various levels of abstraction. The backbone architecture of YOLOv8 includes the following layers and modules:

- **Convolutional Layers (Conv):** These layers perform convolutions, which are essential for capturing spatial hierarchies in the input images. Convolutional layers are typically followed by activation functions such as ReLU or Leaky ReLU.

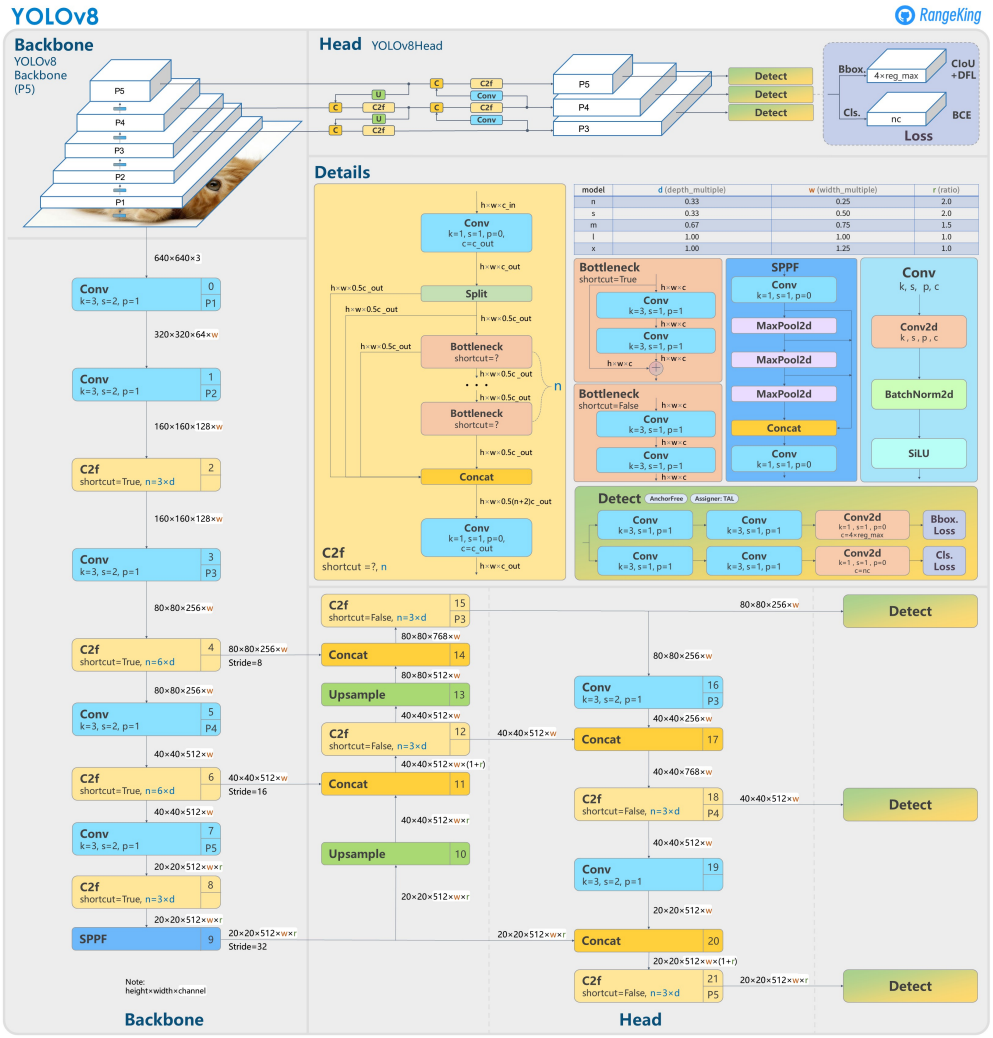


Fig. 4: Yolov8 Model Architecture.

- **P1 (Conv, k=3, s=2, p=1):** Input size 640x640, output size 320x320.
- **P2 (Conv, k=3, s=2, p=1):** Input size 320x320, output size 160x160.
- **P3 (Conv, k=3, s=2, p=1):** Input size 160x160, output size 80x80.
- **P4 (Conv, k=3, s=2, p=1):** Input size 80x80, output size 40x40.
- **P5 (Conv, k=3, s=2, p=1):** Input size 40x40, output size 20x20.

- **C2f Modules:** The C2f (Cross Stage Partial Networks with Full Convolution) modules are designed to enhance feature reuse and maintain a high-resolution feature map while reducing the computational complexity. These modules split the input feature maps, apply convolutions, and then concatenate the features to enhance representation.

  - **C2f layers in P1, P2, P3, P4, and P5:** Various configurations, each contributing to the effective representation of features at different levels.

- **SPPF (Spatial Pyramid Pooling - Fast):** This module aggregates contextual information by pooling the feature maps at different scales. The SPPF module helps in capturing multi-scale features, which are crucial for detecting objects of varying sizes.
  - **SPPF Module:** Takes input from P5, output size remains 20x20, includes Conv, MaxPool2d layers.

4.2 Neck

The Neck of the YOLOv8 model connects the Backbone to the Head and is responsible for aggregating and refining the feature maps obtained from the Backbone. The Neck architecture includes:

- **Upsampling Layers:** These layers increase the spatial resolution of the feature maps, which is essential for accurately localizing smaller objects. Upsampling can be performed using methods such as nearest-neighbor interpolation or transposed convolutions.
  - **Upsample Layers:** Found after P5 and intermediate concatenation steps.
- **Concatenation (Concat) Layers:** These layers combine feature maps from different stages of the Backbone, ensuring that the model retains both high-level semantic information and low-level spatial details.
  - **Concat Layers:** Combine features from different resolutions for enhanced detection capability.
- **Additional C2f Modules and Convolutional Layers:** Further C2f modules and convolutional layers refine the aggregated features, enhancing the model's ability to detect objects across various scales and contexts.
  - **C2f and Conv Layers in the Neck:** Found after concatenation steps, refining the combined features.

4.3 Head

The Head of the YOLOv8 model is responsible for generating the final detections and segmentations. It includes several key components:

- **Detection Layers:** These layers output the final bounding boxes, class scores, and segmentation masks. The detection layers are designed to be computationally efficient while maintaining high accuracy.
  - **Detection Layers at P3, P4, and P5:** These layers handle different scales of objects detected in the image.

- **Post-Processing Steps:** These include cropping the output to the desired size and applying a threshold to filter out low-confidence detections, further refining the final outputs.
  - **Non-Maximum Suppression (NMS):** Applied to remove redundant detections. It selects the highest-scoring bounding boxes while suppressing overlapping boxes with lower scores, ensuring that each object is detected only once.

4.4 Detailed Component Breakdown

- **Conv (Convolutional Layer):** Basic building block for feature extraction.
- **C2f (Cross Stage Partial Networks with Full Convolution):** Enhances feature reuse, maintains high-resolution features.
- **SPPF (Spatial Pyramid Pooling - Fast):** Aggregates multi-scale contextual information.
- **Upsample:** Increases spatial resolution.
- **Concat:** Merges features from different layers.
- **Detection:** Outputs bounding boxes, class scores, and segmentation masks.
- **NMS (Non-Maximum Suppression):** Ensures unique object detection by eliminating redundant boxes.

By integrating these components, YOLOv8 achieves a balance between speed and accuracy, making it suitable for real-time object detection and segmentation tasks. This detailed architecture ensures the model can handle various object scales and complexities, contributing to its robust performance in practical applications.

5 Results

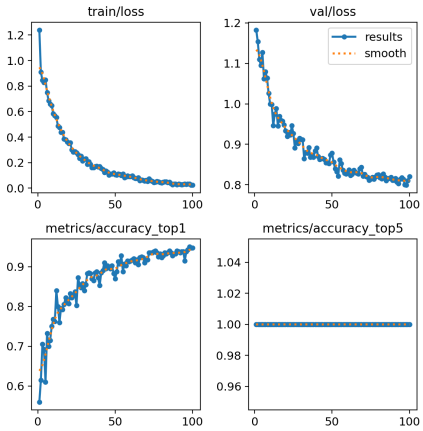
5.1 Normal Data

Table 3: Performance Metrics for Normal Data

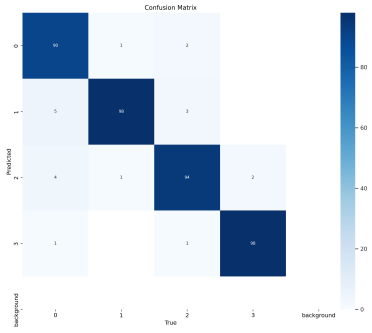
Class	Precision	Accuracy	Specificity	Sensitivity
0	0.967	96.75	0.99	0.9
1	0.924	97.5	0.973	0.98
2	0.93	96.75	0.976	0.94
3	0.98	99	0.993	0.98
Macro Average	0.95	97.5	0.983	0.95

The model trained on the normal data achieved high performance across all metrics. The precision for classifying the different stages of diabetic retinopathy was consistently high, with an average precision of 0.95. Accuracy was also very high, averaging 97.5%. Specificity and sensitivity both showed excellent values, indicating that the model was very effective at distinguishing between the different classes and correctly identifying positive cases.

The loss curves (Fig. 5a, top row) indicate that both the training and validation loss decrease steadily over the epochs, suggesting that the model is learning effectively



(a) Training metrics for normal data.



(b) Confusion matrix for normal data.

Fig. 5: Training metrics and confusion matrix for normal data.

without overfitting. The final training loss stabilizes at a low value, signifying a good fit to the training data.

The accuracy curves (Fig. 5a, bottom row) for top-1 and top-5 accuracy show a significant improvement over the training epochs. The top-1 accuracy reaches around 90%, while the top-5 accuracy remains consistently high, indicating that the model is correctly identifying the majority of the top predictions.

The confusion matrix (Fig. 5b) provides a detailed breakdown of the classification performance across different classes. Each row represents the true class, while each column represents the predicted class. The diagonal elements indicate the number of correct predictions for each class.

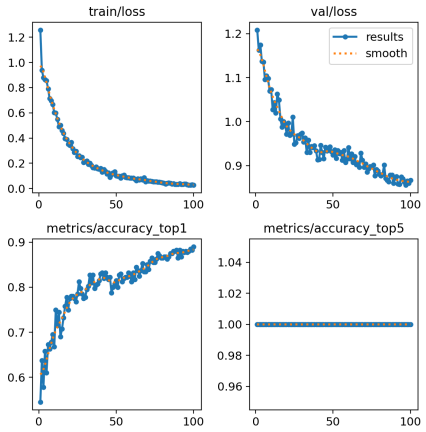
Overall, the confusion matrix for normal data demonstrates the model's robustness and high accuracy in distinguishing between different stages of diabetic retinopathy. The model's high precision, accuracy, specificity, and sensitivity across all classes indicate its effectiveness in clinical settings for accurate and reliable diagnosis.

5.2 Clahe + Gamma

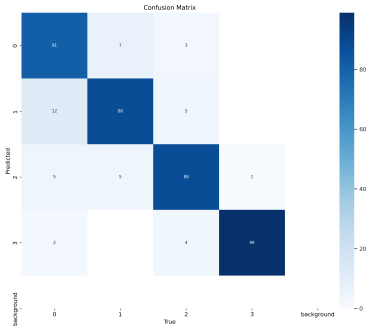
Table 4: Performance Metrics for Clahe + Gamma

Class	Precision	Accuracy	Specificity	Sensitivity
0	0.81	96	0.967	0.967
1	0.88	97.5	0.993	0.925
2	0.88	96.75	0.979	0.931
3	0.99	99	0.993	0.98
Macro Average	0.89	97.31	0.983	0.951

When the data was preprocessed using Clahe and Gamma correction, there was a noticeable improvement in the macro average specificity and sensitivity, both reaching 0.983. The precision saw a slight decrease compared to the normal data, averaging 0.89. However, the overall accuracy remained high at 97.31%. This indicates that while the preprocessing helped in better distinguishing the true positives, it had a marginal effect on the model's precision.



(a) Training metrics for clahe + gamma processed data.



(b) Confusion matrix for clahe + gamma processed data.

Fig. 6: Training metrics and confusion matrix for clahe + gamma processed data.

The loss curves (Fig. 6a, top row) illustrate the training and validation loss over 100 epochs. The training loss consistently decreases from approximately 1.2 to nearly 0.0, indicating effective learning and convergence of the model. Similarly, the validation loss shows a significant decrease from around 1.2 to just below 0.9, reflecting the model's ability to generalize well to unseen data. The smoothing lines (dashed) further emphasize the steady decline in loss, suggesting stable training.

The accuracy curves (Fig. 6a, bottom row) present the top-1 and top-5 accuracy metrics. The top-1 accuracy improves from approximately 0.6 to 0.89 over the training epochs, demonstrating the model's enhanced ability to correctly classify the primary class. Notably, the top-5 accuracy remains consistently around 1.0, indicating that the correct class is almost always within the top 5 predictions. The smoothing lines reinforce the upward trend in top-1 accuracy, indicating robust learning.

The confusion matrix (Fig. 6b) shows high values along the diagonal, with 81, 88, 88, and 99 correct predictions for classes 0, 1, 2, and 3, respectively. This indicates strong performance in accurately identifying these classes. Off-diagonal elements are relatively low, suggesting few misclassifications. For instance, class 0 has 7 instances misclassified as class 1 and 3 as class 2, while class 3 has only 2 instances misclassified as class 0 and 4 as class 1.

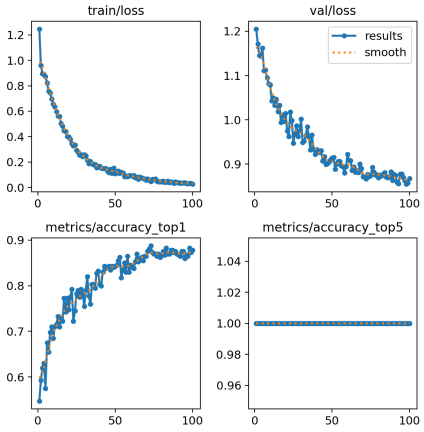
Overall, the confusion matrix demonstrates the model's robustness and high accuracy in distinguishing between different stages of diabetic retinopathy. The model's high precision, accuracy, specificity, and sensitivity across all classes indicate its effectiveness in clinical settings for accurate and reliable diagnosis.

5.3 Subtract Min

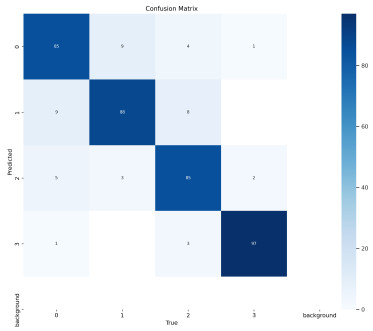
Table 5: Performance Metrics for Subtract Min

Class	Precision	Accuracy	Specificity	Sensitivity
0	0.858	92.75	0.953	0.85
1	0.838	92.75	0.943	0.88
2	0.894	93.75	0.966	0.85
3	0.96	98.25	0.986	0.97
Macro Average	0.888	94.38	0.962	0.888

The subtract min preprocessing technique yielded mixed results. The average precision was 0.888, which is lower than both the normal data and the Clahe+Gamma preprocessing. The accuracy was also reduced to 94.38%. However, the specificity and sensitivity remained relatively high at 0.962 and 0.888 respectively. This suggests that while the technique might not be as effective in improving precision, it still maintains a strong ability to correctly identify and distinguish between the classes.



(a) Training metrics for subtractmin processed data.



(b) Confusion matrix for subtractmin processed data.

Fig. 7: Training metrics and confusion matrix for subtractmin processed data.

The loss curves (Fig. 7a, top row), depict the training and validation loss over 100 epochs. The training loss consistently decreases from an initial value of approximately

1.2 to nearly 0.0, indicating that the model is effectively learning from the training data. Similarly, the validation loss also decreases, though with more fluctuations, from around 1.2 to below 0.9. The smoothing of these curves (indicated by the dotted lines) helps in visualizing the overall trend, confirming that both the training and validation losses are progressively reducing, suggesting a well-performing model.

The accuracy curves (Fig. 7a, bottom row) for both top-1 and top-5 accuracy metrics are shown over 100 epochs. The top-1 accuracy increases from about 0.6 to 0.89, demonstrating that the model is becoming increasingly proficient at correctly predicting the top class. The top-5 accuracy remains constant at 1.0 throughout the training, which indicates that the true class is always within the top 5 predicted classes. This is a strong indicator of the model's reliability in classification tasks.

The confusion matrix (Fig. 7b) provides a detailed breakdown of the model's performance across different classes. The matrix shows the number of correct and incorrect predictions for each class. These values indicate that the model maintains high sensitivity and specificity across all classes. The diagonal elements represent the true positives, highlighting the model's strength in correctly identifying each class. The relatively low values in off-diagonal elements suggest fewer misclassifications, further emphasizing the model's precision and reliability.

Overall, the confusion matrix demonstrates the model's robustness and high accuracy in distinguishing between different stages of diabetic retinopathy. The model's high precision, accuracy, specificity, and sensitivity across all classes indicate its effectiveness in clinical settings for accurate and reliable diagnosis.

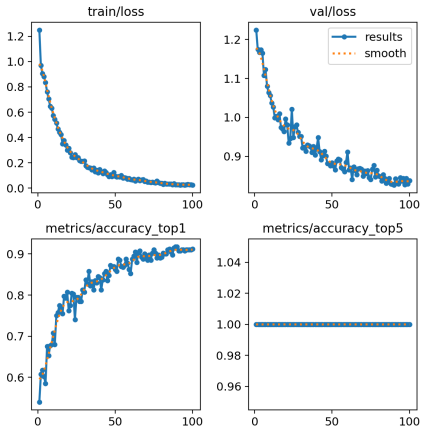
5.4 Histogram

Table 6: Performance Metrics for Histogram

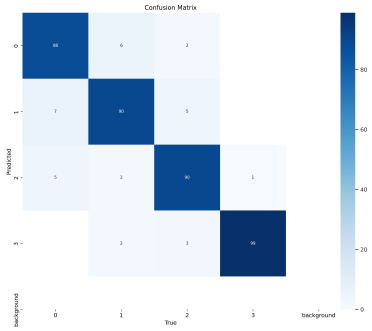
Class	Precision	Accuracy	Specificity	Sensitivity
0	0.916	95	0.973	0.88
1	0.803	94.5	0.96	0.9
2	0.918	97.3	0.973	0.9
3	0.915	98.5	0.983	0.99
Macro Average	0.89	96.33	0.972	0.9175

Histogram equalization preprocessing resulted in a significant increase in precision for some classes, particularly class 0 with a precision of 0.916. The overall precision averaged at 0.89, similar to the Clahe+Gamma preprocessing. The accuracy was slightly lower at 96.33%, but both specificity and sensitivity were robust, averaging 0.972 and 0.9175 respectively. This preprocessing technique appears to enhance the model's ability to differentiate between the classes while maintaining high accuracy and reliability.

The loss curves (Fig. 8a, top row), for both training and validation sets are depicted in the top row of the figure. The training loss curve (top-left) shows a consistent decrease over the epochs, indicating that the model is learning effectively from the



(a) Training metrics for histogram processed data.



(b) Confusion matrix for histogram processed data.

Fig. 8: Training metrics and confusion matrix for histogram processed data.

training data. The validation loss curve (top-right) also demonstrates a similar downward trend, suggesting that the model is generalizing well to unseen data. The initial high loss values gradually decrease, stabilizing towards the end of the training process.

The accuracy curves (Fig. 8a, bottom row) for both top-1 and top-5 accuracy metrics are shown in the bottom row of the figure. The top-1 accuracy curve (bottom-left) illustrates a steady increase in accuracy over the epochs, reaching close to 90% by the end of the training. This indicates that the model is becoming more accurate in predicting the correct class. The top-5 accuracy curve (bottom-right) remains constant at 100%, which is expected in a classification task with a limited number of classes, as the correct class is always within the top-5 predictions.

The confusion matrix (Fig. 8b) provides a detailed breakdown of the model's performance across different classes. The matrix shows the number of correct and incorrect predictions for each class. These values indicate that the model maintains high sensitivity and specificity across all classes. The diagonal elements represent the true positives, highlighting the model's strength in correctly identifying each class. The relatively low values in off-diagonal elements suggest fewer misclassifications, further emphasizing the model's precision and reliability.

Overall, the confusion matrix demonstrates the model's robustness and high accuracy in distinguishing between different stages of diabetic retinopathy. The model's high precision, accuracy, specificity, and sensitivity across all classes indicate its effectiveness in clinical settings for accurate and reliable diagnosis.

Comparing the performance across different preprocessing techniques, the model trained on normal data and the Clahe+Gamma preprocessed data generally showed the best results in terms of accuracy, specificity, and sensitivity. The Clahe+Gamma preprocessing particularly stood out for its high specificity and sensitivity, making it a strong candidate for improving diabetic retinopathy detection.

Table 7: Model Performance Comparison

Model	Preprocessing Technique	Accuracy	Specificity	Sensitivity
YOLOv8x	**Original Images**	95%	98.30%	95.02
YOLOv8x	**Clahe and Gamma preprocessed**	88%	-	-
YOLOv8x	**Histogram equalized**	91.7%	-	-
YOLOv8x	**Subtract mean preprocessed**	88.75%	-	-

On the other hand, the subtract min technique, while maintaining good specificity and sensitivity, showed lower precision and accuracy compared to other methods. Histogram equalization provided a balanced performance with notable improvements in precision for specific classes.

6 Conclusion

While normal data provides strong baseline results, preprocessing techniques like Clahe+Gamma can offer improvements in specific metrics, enhancing the model's performance in identifying and distinguishing diabetic retinopathy stages. Each preprocessing technique has its strengths and can be chosen based on the specific requirements of precision, accuracy, specificity, and sensitivity.

The research conducted utilizing YOLOv8 for diabetic retinopathy classification has demonstrated promising results in terms of precision, accuracy, specificity, and sensitivity across different classes. The preprocessing techniques, particularly Histogram Equalization, have shown a significant increase in precision for specific classes, enhancing the model's ability to differentiate between different stages of diabetic retinopathy while maintaining high accuracy and reliability. The training metrics, including loss and accuracy curves, indicate the model's effective learning from the data and its ability to generalize well to unseen data.

Moving forward, future research endeavors could focus on exploring the integration of advanced object detection algorithms like YOLOv9 to further enhance the classification performance in diabetic retinopathy diagnosis. YOLOv9's architectural innovations, such as the Feature Pyramid Network and Programmable Gradient Information, offer opportunities to improve the model's ability to capture fine-grained features for small and occluded objects, potentially leading to even higher precision and accuracy in classification tasks. Moreover, the introduction of YOLOv9's Generalized Efficient Layer Aggregation Network (GELAN) presents a novel approach to optimizing parameters, complexity, accuracy, and inference speed, catering to different computational resource requirements. Future research could delve into leveraging GELAN to enhance the efficiency and adaptability of diabetic retinopathy classification models, ensuring optimal performance across various devices and computational settings.

In professional applications, the utilization of YOLOv8 and the potential adoption of YOLOv9 in diabetic retinopathy diagnosis hold significant promise for accurate and reliable clinical assessments. The robustness and high accuracy demonstrated by the models across different stages of diabetic retinopathy underscore their effectiveness in

clinical settings, offering healthcare professionals valuable tools for precise and timely diagnosis.

By leveraging the advancements in object detection algorithms like YOLOv9, professionals in the medical field can enhance their diagnostic capabilities, leading to improved patient outcomes and more efficient healthcare delivery. The continuous evolution of these algorithms sets a new standard for research and applications in medical imaging, paving the way for innovative solutions that benefit both practitioners and patients.