ISSN 1870-4069

Automatic Detection of Diabetic Retinopathy Using Classification Techniques and Computer Vision

Jorge Antonio Hernández Magallanes

Universidad Autónoma de Aguascalientes,
Centro de Ciencias Básicas,
Departamento de Ciencias de la Computación,
Mexico

jhernandez.dev00@gmail.com

Abstract. One of the most important challenges in modern medicine is the timely diagnosis of chronic diseases, as early detection can make a significant difference in the patient's quality of life and better management of the health system. A clear example of this problem is diabetic retinopathy (DR), an ocular complication of diabetes that constitutes one of the main causes of blindness worldwide. In this sense, early detection is key to avoid irreversible vision damage. This paper explores the use of imaging processing techniques and artificial intelligence as tools to address this challenge, focusing specifically on the automatic analysis of retinal scans for the early detection of diabetic retinopathy. Among the techniques used to preprocess the retinographies, resampling to standardize the number of images, noise elimination, cropping of the area of interest, as well as brightness and contrast adjustment stand out. In addition, contrast Limited Adaptative Histogram Equalization (CLAHE) and gramma correction were applied to improve image quality. For classification tasks, a convolution neuronal network (CNN) based on the DenseNet121 model is employed, which has shown promising results in initial tests. Although it is still under development, the aim is to improve its accuracy and efficiency to make it a practical and reliable tool in the early diagnosis of diabetic retinopathy.

Keywords: Diabetic retinopathy, retinography, automatic detection, computer vision, convolutional neural networks.

1 Introduction

Medicine is one of mankind’s oldest branches of science, which over the centuries has been transformed for the purpose of improving people's quality of life. Thanks to advances in technology, especially in the computer science field, medicine has undergone a revolution in the way diseases are diagnosed and treated. Currently, one of the biggest challenges facing medicine is the timely diagnosis of chronic diseases, as the timing of such diagnoses is crucial to prevent a person’s health from being severely impacted.

A good example of this is diabetic retinopathy (DR), a common complication of diabetes that, if not detected early, can lead to irreversible vision loss, severely affecting

Fig. 1. Examples of types of retinopathy a) NO DR, b) MILD, c) MODERATE, d) SEVERE, d) DR PROLIFERATIVE.

the patient's quality of life. This disease, associated with diabetes mellitus [1], does not present significant symptoms in the early stages, making it difficult to detect in a timely manner.

According to the study carried out in 2021, entitled "Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045" estimates that currently, the worldwide prevalence of patients with this medical complication is situated between 22.27% among people with diabetes, and it is estimated that year 2045 this figure will increase exponentially [2]. These data are more alarming when we consider that according to another study published in the journal springer entitled "Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications-risks and mitigation" diabetic retinopathy (DR) is considered one of the leading causes of irreversible blindness in the world, especially in developing countries [3], As for Mexico, the prospect is not encouraging, since, according to a Mexican government fact sheet issued on July 22, 2018, the prevalence of diabetic retinopathy among people with diabetes is close to 31.5% [4].

Due to the nature of this disease, timely diagnosis is crucial to prevent serous complications and irreversible damage. However, traditional diagnosis based on manual inspection of retinal images is a time – consuming process and, above all, requires experience, which limits its accessibility in many regions. In this sense, image processing and the use of advanced artificial intelligence algorithms represent an opportunity to address and try to provide a solution to this problem.

By automating the retinal scan analysis process, early signs of diabetic retinopathy can be detected and help healthcare professionals make faster and more effective diagnostics, even in resource-limited settings, and enable more continuous and accessible monitoring for a wider range of patients.

With this scenario, the objective of this research is to develop an automated system that using image processing techniques and artificial intelligence models can detect early any of the stages of diabetic retinopathy, which can be seen in the image (See Fig. 1), thanks to the combination of image processing and the use of convolutional neural network can significantly improve the accuracy and efficiency of automated diagnosis compared to traditional approaches, this contributing to a useful tool especially in context with limited resources. However, the study continues to evaluate the performance of other neural networks models to identify the most suitable architecture of this task.

2 Related Work

During the last few years, many number of studies have tried to explore and propose various approaches and techniques to try to solve this problem, such as the study entitled "Automatic Detection of Diabetic Retinopathy Applying Computer Vision and Convolution Neural Network" [5], in which, by means of transformation of retinography to grayscale, edge detection and clipping, attempts to perform early identification of signs of retinopathy, with a degree of certainty of 91%, the model is proposed as a stable model for the detection of retinopathy. Another relevant study is "Deep learning based binary classification of diabetic retinopathy images using transfer learning approach" [6], in this work, pretrained network are used with the use of three different datasets, which are: DRD-EyePACS, IDRiD and APTOS-2019, divided into training, testing and validation sets, and with the use of preprocessing and data augmentation techniques, the most outstanding model reaches 97.33%.

In addition, a study entitled "Binary Classification of Diabetic Retinopathy Using CNN Architecture" [7] using several pretrained networks such as EfficientNet, VGG16 and MobileNet, among others, applies processing techniques such as the application of Gaussian filters, image resizing and data augmentation, to achieve an accuracy of 94.55%. However, all of them focus on binary disease classification, i.e., they determine whether an image shows signs of retinopathy.

This binary approach, although useful, limits the ability of the models to capture the complexity of diabetic retinopathy, as it does not consider the different stages of the disease or the severity of damage, which could be crucial for early diagnosis and appropriate intervention.

Other studies address this problem from a multiclass classification approach, which seeks to identify the degree of severity of diabetic retinopathy according to the different clinical stages of the disease (mild, moderate, severe and proliferative). This type of approach provides a more detailed analysis, such is the case of the study entitled "Improved Automatic Diabetic Retinopathy Severity Classification Using Deep Multimodal Fusion of UWF-CFP and OCTA Images" [8] which uses ultra-widefield fundus images (UWF-CFP) and optical coherence tomography angiography (OCTA) using ResNet50 and 3D-ResNet50 models with attention blocks, Finally, the study entitled "Transfer-Ensemble Learning based Deep Convolutional Neural Networks for Diabetic Retinopathy Classification" [9] proposes an ensemble model that combines VGG16 and Inception V3 pre-trained networks to classify DR images into five classes. Using the APTOS dataset, the model achieved an accuracy of 96.4%. However, for further studies, some papers propose the idea of using different deep learning techniques with the aim of improving the accuracy in multiclass classification of diabetic retinopathy and addressing challenges such as class imbalance and variability in image quality.

The main contribution of this article is the use of different techniques to improve the quality of images that may present unwanted variations, trying with image processing techniques that have not been used in the literature of the problem.

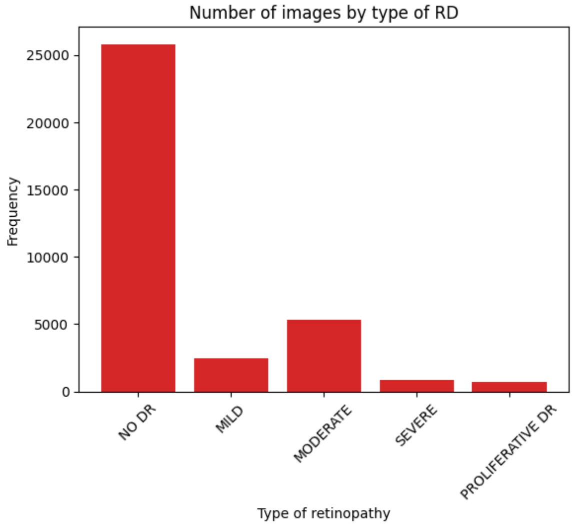

Fig. 2. Number of images per type of retinopathy are present in this dataset.

3 Methodology

The methodology of this study consists of several phases, from the acquisition of the database and the preparation of the data to the training and evaluation of the preliminary results, the process is as follows:

3.1 Image Acquisition

Preprocessing is the stage in which it is intended to repair the images obtained from any defects that may be acquired during the capture of the image, solving flaws produced or generated by the capture hardware.

To achieve the objective of building a computer vision system, it is necessary to carry out the first stage of the process, which is to obtain the dataset with which to work. In this case, the dataset used will be the “Diabetic Retinopathy Detection” [10], which was extracted from Kaggle. All the images contained in this dataset are in JPEG format with dimensions of 4752 x 3168 px. The dataset includes images covering the five stages of diabetic retinopathy, which can be seen in the image above (See Fig. 1), allowing a classification according to their severity.

Before starting to apply the technique to improve the quality of the image, it is necessary to verify that the dataset is balanced.

If there are many more images of one class than another, the model may become biased and fail to correctly learn the necessary patterns, through an exploration data analysis (EDA), to understand the nature and state of the dataset, the results derived from this process can be seen in the graph above (See Fig. 2), where the distribution in the number of images between classes is evident, so a resampling technique should be applied.

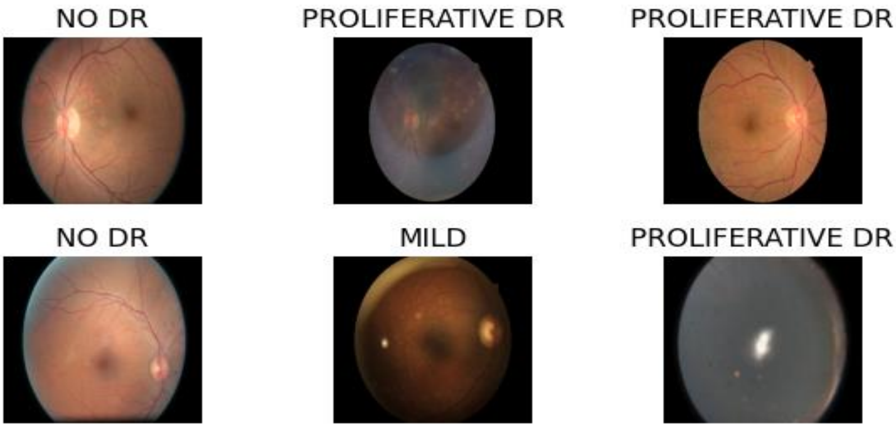

Fig. 2. Examples of retinography without preprocessing techniques

3.2 Resampling

Resampling is a technique that consists of creating a new data sample from a set of original data with some kind of bias [11], To address this problem, there are several resampling techniques, the most common being oversampling (which consists of increasing the number of samples of minority classes), under sampling (which reduces the number of samples of majority classes) and the use of class weights (which assigns a higher weight to the less represented classes during the training of the model, without the need to modify the size of the data set). The use of class weights could have led to a longer convergence time, precisely because of the large disparity in the number of images between classes.

We will then resample 700 images per class to avoid duplication of data and randomize the order of the samples, resulting in a balanced dataset of 3,500 retinography. Furthermore, recent studies in automatic medical image diagnosis, such as “Self-Supervision for Medical Image Classification: State-of-the-Art Performance with ~100 Labeled Training Samples per Class [12]”, have shown that datasets with between 500 and 1000 images per class, combined with data augmentation techniques, are sufficient to achieve robust performance metrics.

3.3 Image Preprocessing Techniques

The quality and consistency of the input images for a classification model play a crucial role in the performance of machine learning models for diabetic retinopathy detection.

In the image above (See Fig. 3), we can see the retinography of the dataset without any modifications, where we can see problems in visualizing the key factors that could help to predict the type of retinopathy in which the patient is. To address this problem, we will apply an enchantment process to optimize the quality of the images and highlight important details, allowing for a more accurate analysis, which is described below.

Denoising filter. When capturing an image, depending on multiple factors, unwanted variations may be introduced in the pixels. This is known as noise and affects the image, since it may acquire different brightness or color characteristics than the original ones [13]. They're capturing an image, depending on multiple factors, unwanted variations may be introduced in the pixels. This is known as noise and affects the image, since it may acquire different brightness or color characteristics from the original ones.

There are several types of digital noise, each with specific characteristics, which necessitates the use of different filters designed to improve image quality. One of these filters is bilateral filtering, which, unlike other types of filters, is particularly effective in reducing noise in images while keeping the edges sharp. Because of this capability, bilateral filtering is especially useful in applications where preserving fine image detail is critical, such as in medical image analysis.

Cropping. The process of cropping involves trimming the outer edges of an image to remove unnecessary or irrelevant areas, focusing the analysis on the most important central features. This technique can help eliminate noise or artifacts that may be present in the peripheral parts of the image, improving the overall quality and accuracy of the analysis. To achieve edge clipping, the following process is carried out: first, the thresholding technique is applied, which is used in image processing to generate binary images from a grayscale image. This is achieved by setting a threshold value: all pixels that exceed this value become white, while those that do not are transformed into black. Subsequently, the function find Contours of OpenCV is used, which allows to detect the internal and external contours of a binary image, this function will be used to detect the main contour, to achieve this, the results of this function will be taken, the largest contour is selected, which will serve to obtain the delimiting rectangle of the contour and finally the image is cropped.

Brightness and contrast adjustment. In image processing, brightness and contrast are two fundamental characteristics that affect the perception and quality of images. Brightness is related to the overall light intensity, while contrast measures the difference between light and dark areas, allowing them to highlight important details. To improve these parameters, a technique based on the use of the mean and the standard deviation present in the image is applied.

Contrast adjustment was performed using the inverse of the standard deviation, while brightness adjustment was centered using the median of the pixel values (50th percentile), allowing for more stable and robust corrections. Both factors were constrained with limit values to avoid overexposure or underexposure of the images.

CLAHE adjustment. CLAHE (Contrast - Limited Adaptive Histogram Equalization) is an advanced method used to enhance the contrast of images, especially those with non-uniform illumination variation.

Unlike traditional histogram equalization, which uses the range of image intensities globally, CLAHE employs an adaptive strategy, which consists of dividing the image into smaller blocks or regions, over which a local histogram is calculated for each section. The local histograms are then used to adjust the brightness and contrast of each of these regions individually. [14]

One of the main benefits of CLAHE is the ability to avoid overexposure in areas of highlight intensity, a common problem in global histogram equalization. By limiting

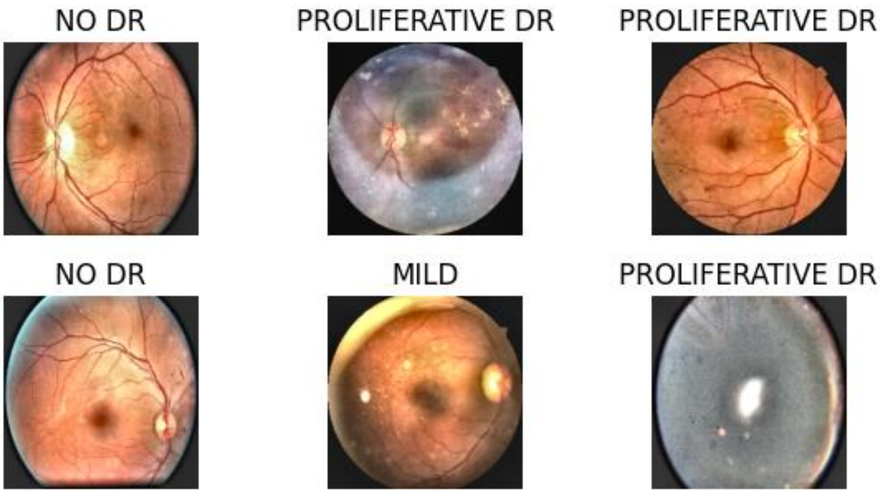

Fig. 4. Examples of retinography with preprocessing techniques applied.

the contrast of each local region by means of a threshold (which is adjusted for each image), the risk of generating unwanted changes in the image is reduced.

With these changes, image detail is more easily perceived, while preserving more subtle information in areas of low contrast.

GAMMA correction. Gamma is the relationship between the numerical value of a pixel and its actual luminance. Without this value, the tones captured by any device could not be represented in the way a person can visualize it, this correction allows them to compensate for these differences.

As a result of this process, the following image (See Fig. 4) shows a visible improvement of the details present in the image, with more balanced tones and highlighting of anatomical features of the eye necessary to perform the classification process.

To evaluate the improvement in the images two parameters will be used to evaluate them the Signal-to-Noise Ratio (SNR) and Peak Signal-to-Noise Ratio (PSNR), the SNR is a measure that compares the power of the useful signal (relevant information) with the power of the noise (unwanted interference), a high value indicates that the image is not affected by noise, while a low value indicates the presence of significant noise in the image, while the PSNR is a metric used to evaluate the quality of the processed images compared to the original ones. It quantifies the ratio between the maximum possible value of a signal (the highest pixel value) and the noise present in the image. A higher PSNR indicates that the processed image is of high quality, with little distortion compared to the original image. PSNR values above 40 dB are considered to be of excellent quality. [15]

The quantitative results suggest that the applied preprocessing has significantly improved the quality of the retinal images.

```
Average Original SNR: 0.28470235725421605
Average Processed SNR: 4.766901181346978
Average Contraste: 0.9885407340956234
Average PSNR: 58.73854062813858
```

Fig. 5. Results of the evaluation of changes in image quality.

The values for this can be seen in the image above (See Fig. 5). Overall, it can be concluded that the preprocessing has significantly reduced the noise present in the images, as evidenced by the increase in the signal-to-noise ratio (SNR) value, which indicates an improvement in the clarity of the processed images and the elevated PSNR reveals that the processed images have maintained a quality very close to the originals.

For more details on the image enhancement process, as well as access to the code and examples used, please visit the repository: https://github.com/jhernandezdev00/RDVis_DX

4 Construction of the Classifier

Classification is defined as the process by which a set of objects or elements can be grouped into different categories or classes, depending on some characteristic or factor that they share [16].

In machine learning, the choice of classification method depends largely on the type of learning used. There are three main approaches: supervised, unsupervised and semi-supervised learning, differing in the way the model works with the data.

In supervised learning, the data are labeled, that is, they have a predefined class, while in unsupervised learning the data do not have a label, so the model must group them according to patterns and similarities detected in the dataset, finally, semi-supervised learning combines the two previous ones, thus avoiding the need for exhaustive labeling [17].

4.1 Construction of the Convolutional Neural Network

A convolutional neural network (CNN) is a type of network specialized in deep pattern learning by using filters to extract relevant features from data. These networks are used to perform tasks that require object recognition or pattern identification [18].

For the structure of the model, we have chosen to use a pre-trained model as the basis for the classification model, although initially we had chosen to use EfficientNetB0 due to its optimal performance and not being a network with too much computational load, we have decided to switch to DenseNet121. This decision was based on performance comparisons reported in the literature, where DenseNet121 shows better feature extraction and recognition capability in classification tasks.

In addition, in preliminary tests it shows a better accuracy rate on the dataset that has been used, although it is important to note that, at the time of writing this paper, other techniques are still being investigated and tested to find the one that gives the best performance and results.

4.2 Network Architecture

To design the network architecture for the classification task, the input images are first resized to 256 x 256 pixels. The model uses DenseNet121, pre-trained on ImageNet, as the base. This base model is configured without its top classification layer and with its weights initially frozen to retain previously learned features. On top of the base, a Global Average Pooling 2D layer is added, followed by a sequence of dense blocks to refine the learned features. The first block includes a dense layer with 512 units and ReLU activation, followed by Batch Normalization and a 30% Dropout layer. Next, a second block includes a dense layer with 256 units and Batch Normalization. This is followed by a third block with a 128-unit dense layer, again with ReLU activation, Batch Normalization, and a 30% Dropout layer. Finally, the output layer is a dense layer with 5 units and SoftMax activation, providing the final predictions for multiclass classification.

In addition, the fine-tuning technique was applied to optimize performance. Initially, a pre-trained model was used on ImageNet, so that the base layers are frozen and only the upper layers are trained. After some epochs, some of the deeper layers are unfrozen to adjust the weights gradually on our specific dataset. In future phases of the project, other types of more comprehensive methods are planned.

4.3 Model Training

The model uses the Adam optimizer with batch of 32 images, during the first training stage, corresponding to the fine-tuning process. To dynamically adjust the learning rate, ReduceLROnPlateau is used, which reduces the value of the learning rate that allows using the validation metrics to dynamically adjust the learning value, together with the use of Early Stopping that stops the training if it does not improve over a certain number of epochs to avoid over-fitting, this first stage is executed for about 15 epochs.

For the second stage that will perform the final model adjustment, where the last 15 layers of the base model are unfrozen, the learning rate is adjusted to 0.00001, this to prevent the model from forgetting everything it learned in the first stage and Early Stopping is used again, for the final stage it is run for 30 epochs.

4.4 Model Evaluation

To test the accuracy of the training, tests were performed with random images that were not part of the training or validation, and to record the success cases of the prediction, as well as the degree of accuracy of each of the tests, and the accuracy and loss generated from the set of tests were considered as main factors.

5 Results

Before going into detail on the results, it is important to note again that at the time of writing this article (April, 2024), the model building and testing phase is still in an early stage of development.

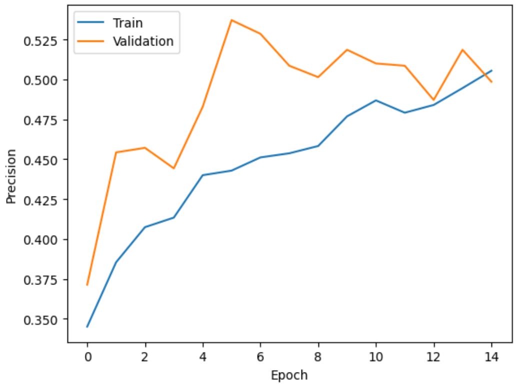

Fig. 6. Results of the first stage of model training.

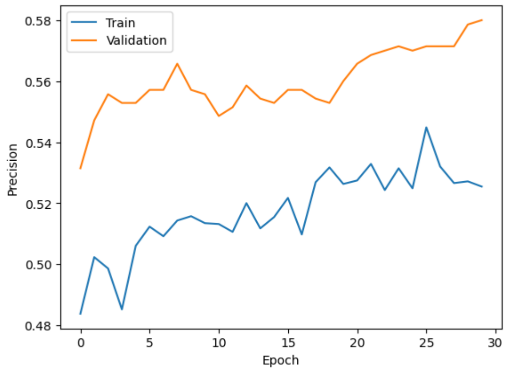

Fig. 7. Results of the second stage of model training.

```
22/22 ━━━━━━━━━━━━━━━━━━━━ 4s 173ms/step - accuracy: 0.5770 - loss: 1.0290
Loss: 1.0238395929336548
Accuracy: 0.5785714387893677
```

Fig. 8. Accuracy values obtained in tests.

To check the accuracy of the training, tests were performed with random images that were not part of the training and validation, and to record the success cases of the prediction, as well as the degree of accuracy of each of the tests. During the training it is observed that in the first phase of the fine-tuning, the training starts between 20% ~ 25%, with a loss value close to 2.15, during the 15 epochs to which the model was subjected, the value of the accuracy ended with a value close to 50% and reducing the loss value to 1.12 in the tests.

Passing these epochs in the second and last stage the model will be trained for 30 epochs or until the loss value does not improve, in order to avoid overfitting, the model starts with an accuracy value of 53% with a loss of 1.08, when completing its training the accuracy value improves to 57% and ends up reducing the loss to a value close to 0.91, these results can be seen in the graph below (See Fig 6 and 7).

As can be seen, fine tuning was a good strategy to finish adjusting and completing the model, even so, the values obtained do not reach the expected accuracy value, as can be seen in the image above, where the tests resulted in an accuracy of 57% (See Fig. 8).

Even so, these results are encouraging, since the detection of each type of retinopathy represents a very complex process, even for experienced specialists. The ability of model to differentiate between the different categories, although still at an early stage, suggests that convolutional neural networks may become a valuable diagnostic support tool in clinical settings.

6 Conclusions

Preliminary results suggest that the use of convolutional neural networks as a tool for the diagnosis of diabetic retinopathy represents a significant breakthrough in the growing integration and use of artificial intelligence in the field of medicine. Although the results do not yet reach the expected levels of accuracy, they show favorable progress in the ability of a single model to extract and recognize the most relevant patterns present in a retinography, which is essential for accurate and automated detection of diabetic retinopathy.

Furthermore, this initial breakthrough demonstrates the critical role that preprocessing plays in image standardization and enhancement, as it allows the network to learn from a more uniform and optimized dataset, which helps to improve its interpretation of the anatomical features of the eye. This stage is even more relevant considering that some studies tend to work with the dataset without applying any enhancement or modification, which can limit the performance of the model and its generalization capability.

In addition, this first phase of research has identified limitations inherent to shallow models, which usually present greater complications when generalizing the knowledge extracted from the features or biased to some of them in specific to give their prediction.

Therefore, it is recommended for these cases the use of models robust enough for this task or that integrate advanced feature extraction mechanisms, such as deep convolutional layers or attention techniques, and in cases where the number of images is reduced or the hardware capabilities are limited, the use of state-of-the-art pre-trained models, in order to significantly improve the accuracy in the detection of retinopathy, especially in its early stages where identification is more complex.

As a next step for the research, the use of regularization techniques and strategies to increase the robustness of the model is proposed, as well as the expansion of the data set to expose the algorithm to a greater diversity of clinical cases. These actions aim to improve the generalization capacity of the system and help mitigate possible biases that the model may generate during its learning process, with the ultimate goal of moving towards a more accurate and reliable model that can be used as an effective support tool for health professionals in the early detection and timely treatment of diabetic retinopathy.

References

1. Centro de Oftalmología Barraquer 2024, Retinopatía diabética. https://www.barraquer.com/patologia/retinopatia-diabetica
2. Teo, Z.L. *et al.*: Global Prevalence of Diabetic Retinopathy and Projection of Burden Through 2045: Systematic Review and Meta-analysis, Ophthalmology, 128(11), pp. 1580–1591 (2021) doi: 10.1016/j.ophtha.2021.04.027.
3. Kropp, M. *et al.*: Diabetic Retinopathy as the Leading Cause of Blindness and Early Predictor of Cascading Complications—Risks and Mitigation. Springer Science and Business Media Deutschland GmbH, (2023) doi: 10.1007/s13167-023-00314-8.
4. Gobierno de México, Retinopatía diabética o ceguera irreversible por inadecuado control de la diabetes. https://www.gob.mx/salud/documentos/retinopatia-diabetica-o-ceguera-irreversible-por-inadecuado-control-de-la-diabetes
5. Bernal Catalán, E., De La Cruz Gámez, E., Montero V., J.A., H. Reyna, M.T.I.: Detección automática de retinopatía diabética aplicando visión artificial y redes neuronales convolucionales.
6. Saproo, D., Mahajan, A.N., Narwal, S.: Deep Learning Based Binary Classification of Diabetic Retinopathy Images Using Transfer Learning Approach. Journal of Diabetes & Metabolic Disorders, 23(2), pp. 2289–2314 (2024) doi: 10.1007/s40200-024-01497-1.
7. Khudaier, A.H., Radhi, A.M.: Binary Classification of Diabetic Retinopathy Using CNN Architecture. Iraqi Journal of Science, 65(2), pp. 963–978 (2024) doi: 10.24996/ijs.2024.65.2.31.
8. Daho, M.E.H. *et al.*: Improved Automatic Diabetic Retinopathy Severity Classification Using Deep Multimodal Fusion of UWF-CFP and OCTA Images, (2023) doi: 10.1007/978-3-031-44013-7_2.
9. Ghosh, S., Chatterjee, A.: Transfer-Ensemble Learning Based Deep Convolutional Neural Networks for Diabetic Retinopathy Classification, doi: 10.48550/arXiv.2308.00525.
10. Dugas, E., Jared, J., Cukierski, W.: Diabetic Retinopathy Detection. https://kaggle.com/competitions/diabetic-retinopathy-detection
11. Data Science. Resampling: A Method for Balancing Data. https://datascientest.com/en/resampling-a-method-for-balancing-data
12. Nielsen, M., Wenderoth, L., Sentker, T.: Self-Supervision for Medical Image Classification: State-of-the-Art Performance with ~100 Labeled Training Samples per Class, 10(8), pp. 895 (2023) doi: 10.3390/bioengineering10080895.
13. Fotolarios: El ruido digital - Qué es, por qué aparece, cómo evitarlo. https://www.fotolarios.es/2018/11/el-ruido-digital.html
14. Free-Astro Team: Contrast-Limited Adaptive Histogram Equalization (CLAHE). https://siril.readthedocs.io/es/stable/processing/clahe.html
15. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, Global Edition. 2017 https://dl.icdst.org/pdfs/files4/01c56e081202b62bd7d3b4f8545775fb.pdf
16. IBM, Tipos de clasificación. https://www.ibm.com/docs/es/msp/7.6.3?topic=classifications-classification-types
17. Zoumana, K.: Classification in Machine Learning: An Introduction. https://www.datacamp.com/blog/classification-machine-learning
18. Zoumana Keita: ¿Qué es una red neuronal convolucional (CNN)?. https://www.datacamp.com/es/tutorial/introduction-to-convolutional-neural-networks-cnns