Convolutional neural networks for diabetic retinopathy detection

Redes neuronales convolucionales para detección de retinopatía diabetica

Darwin Patiño-Pérez[1,*] ⓘ, Luis Armijos-Valarezo[1] ⓘ,
Luis Chóez-Acosta[1] ⓘ, Freddy Burgos-Robalino[1] ⓘ

Abstract

The early detection of diabetic retinopathy remains a critical challenge in medical diagnostics, with deep learning techniques in artificial intelligence offering promising solutions for identifying pathological patterns in retinal images. This study evaluates and compares the performance of three convolutional neural network (CNN) architectures ResNet-18, ResNet-50, and a custom, non-pretrained CNN using a dataset of retinal images classified into five categories. The findings reveal significant differences in the models' ability to learn and generalize. The non-pretrained CNN consistently outperformed the pretrained ResNet-18 and ResNet-50 models, achieving an accuracy of 91 % and demonstrating notable classification stability. In contrast, ResNet-18 suffered severe performance degradation, with accuracy dropping from 70 % to 26 %, while ResNet-50 required extensive tuning to improve its outcomes. The non-pretrained CNN excelled in handling class imbalances and capturing complex diagnostic patterns, emphasizing the potential of tailored architectures for medical imaging tasks. These results underscore the importance of designing domain-specific architectures, demonstrating that model complexity does not necessarily guarantee better performance. Particularly in scenarios with limited datasets, well-designed custom models can surpass pre-trained architectures in diagnostic imaging applications.

Keywords: diabetic retinopathy, blindness, detection, artificial intelligence, convolutional neural networks, image analysis

Resumen

La detección temprana de la retinopatía diabética representa un desafío crítico en el diagnóstico médico, donde el aprendizaje profundo dentro del campo de la inteligencia artificial emerge como una herramienta prometedora para optimizar la identificación de patrones patológicos en imágenes retinales. Este estudio evaluó comparativamente tres arquitecturas de redes neuronales convolucionales ResNet-18, ResNet-50 y una CNN personalizada o no-preentrenada para clasificar imágenes de retinopatía diabética en un conjunto de datos de imágenes agrupadas en cinco categorías, revelando diferencias significativas en su capacidad para aprender y generalizar. Los resultados demostraron que la arquitectura de red neuronal convolucional no-preentrenada superó consistentemente a los modelos preentrenados basados en ResNet-18 y ResNet-50, alcanzando una precisión del 91 % y una notable estabilidad en la clasificación. Mientras ResNet-18 mostró limitaciones severas, degradándose de un 70 % a un 26 % de precisión, y ResNet-50 requirió ajustes para mejorar su rendimiento, la CNN no preentrenada exhibió una capacidad sobresaliente para manejar el desbalance de clases y capturar patrones diagnósticos complejos. El estudio subraya la importancia de diseñar arquitecturas específicamente adaptadas a problemas médicos, destacando que la complejidad no garantiza necesariamente un mejor desempeño, y que un diseño cuidadoso puede superar modelos preentrenados en tareas de diagnóstico por imagen cuando la cantidad de datos con que se cuenta es limitada.

Palabras clave: retinopatía diabética, ceguera, detección, inteligencia artificial, redes neuronales convolucionales, imágenes oculares.

[1,*]Universidad de Guayaquil, Guayaquil, Ecuador. Corresponding author ✉: darwin.patinop@ug.edu.ec.

1. Introduction

The retina, located at the back of the eye, is a vital layer of light-sensitive cells essential for vision. Unfortunately, it is susceptible to various diseases, among which diabetic retinopathy (DR) stands out as one of the most common and serious conditions (Figure 1). DR is an ocular complication of diabetes, characterized by damage to the blood vessels in the retina [1]. This vascular damage can lead to several pathological issues, including:

- **Obstruction of blood flow.** Blocked blood vessels hinder sufficient blood supply to the retina, potentially resulting in the death of retinal cells and subsequent vision loss.

- **Blood leakage.** Damaged blood vessels may leak blood and other fluids into the retina, causing swelling and blurred vision.

- **Growth of abnormal blood vessels.** As a response to oxygen deprivation, the retina can develop new abnormal blood vessels which may be fragile and prone to bleeding.
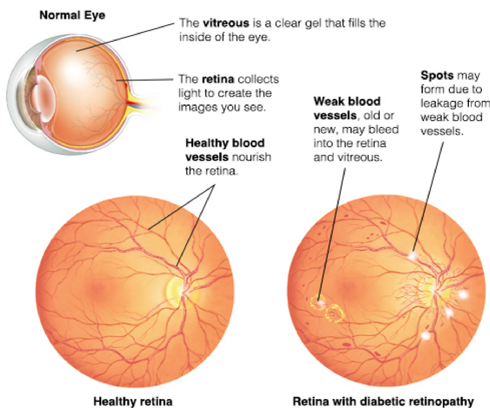

Figure 1. Diabetic retinopathy

Diabetic retinopathy (DR) is more prevalent among individuals with type 1 and type 2 diabetes, particularly those who fail to maintain adequate blood sugar control [2]. Additional risk factors include hypertension, hypercholesterolemia, smoking, overweight or obesity, and pregnancy. In its early stages, DR often presents without noticeable symptoms.

However, as the disease progresses, symptoms may manifest, including blurred vision, dark spots or floaters, difficulty seeing at night, distorted vision, and even vision loss. Early detection and timely treatment are essential to prevent permanent vision impairment. Therefore, regular eye examinations are strongly recommended for individuals with diabetes, particularly those with additional risk factors, to facilitate early intervention and management.

Computer science focuses on designing and developing systems and algorithms capable of performing tasks that typically require human intelligence, such as learning, perception, reasoning, and problem-solving. This discipline forms the foundation of artificial intelligence (AI) [3]. AI integrates techniques from computer science, statistics, logic, and mathematics to create systems that can autonomously learn from data and enhance their performance in real time.

Artificial intelligence (AI) has emerged as a promising tool for DR detection. Machine Learning algorithms can analyze retinal images and identify subtle patterns indicative of the disease. This technology holds significant potential to enhance the accuracy and efficiency of DR diagnosis, enabling earlier detection and facilitating timely interventions.

DR is a severe complication of diabetes that can lead to vision loss if left untreated. Early detection and timely intervention are critical to preventing disease progression and mitigating its impact. Artificial intelligence (AI) has emerged as a promising tool to enhance the detection of DR, offering more precise and efficient diagnostic capabilities and contributing to the preservation of visual health in individuals with diabetes.

Predictive models, which provide forecasts for dichotomous outcomes (distinct yet complementary results), are widely utilized in medical applications. Figure 2 illustrates an evaluation of the most relevant models employed in this domain [4]. Deep learning, a prominent field within artificial intelligence, enables machines or computers to learn and analyze data in a manner akin to human intelligence [5]. This study examines the behavior of various deep learning-based models, highlighting their capability to leverage multiple processing layers to facilitate learning from data representations at multiple levels of abstraction [6].
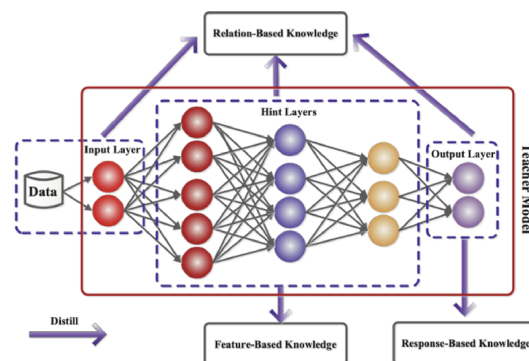

Figure 2. Deep learning model

Numerous pre-trained ResNet implementations are available across various machine learning frameworks, including TensorFlow, PyTorch, Keras, and MXNet. Each framework offers its own variants and specific

optimizations, making the selection of an appropriate pre-trained ResNet model crucial for addressing a particular task. Key factors to consider when selecting a model include the size and complexity of the dataset, the nature of the task (e.g., classification, object detection, or segmentation), and the computational resources available.

The variety of pre-trained ResNet architectures available is extensive, offering a range of options tailored to different tasks and requirements. Selecting the appropriate pre-trained ResNet model depends on the specific objectives of the project and the characteristics of the problem to be addressed. For this study, which focuses on disease recognition through the analysis of ocular images using supervised learning in a classification framework, ResNet models were chosen due to their demonstrated acceptable performance in prior studies involving other types of images. This research evaluates the performance of artificial neural network models, specifically pre-trained convolutional neural networks (CNNs) such as ResNet-18 and ResNet-50, as depicted in Figure 3.
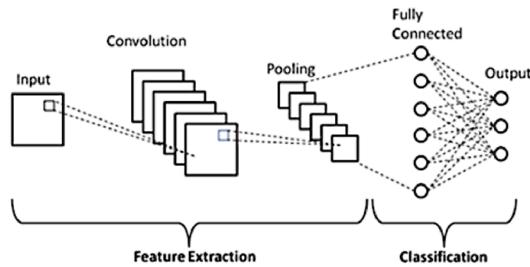
Figure 4. Analysis technique - ML

Figure 3. CNN model
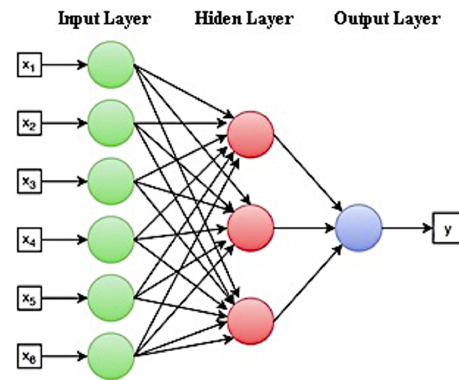
Figure 5. Analysis technique - ML

2. Materials and methods

2.1. Methodology

Deep learning (DL), a subfield of Machine Learning (ML), represents more than a mere analysis technique (Figure 4). It is a comprehensive methodology that encompasses the entire data science pipeline, including data collection, preparation, exploration, modeling, and evaluation. This approach enables the identification of patterns, the generation of predictions, and informed decision-making. Unlike traditional statistical methods, which rely on predefined rules and static models, ML employs algorithms that learn directly from the data. These algorithms adapt to the data's complexity and evolve over time, improving performance as they are exposed to larger and more diverse datasets. This adaptability is particularly evident in models based on artificial neural networks, which comprise numerous interconnected neurons organized into layers. These networks follow a hierarchical structure, as depicted in Figure 5.
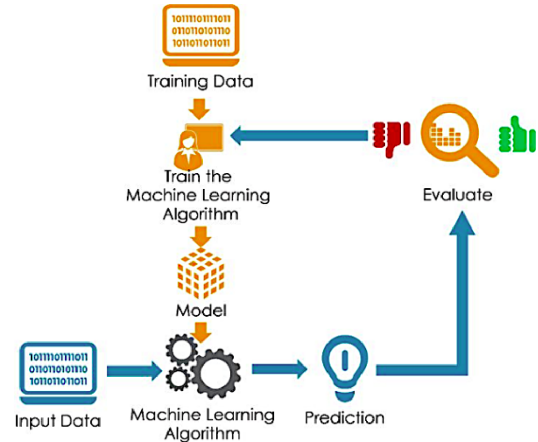
2.2. The data set

The study dataset utilized in this research comprises 3,662 retinal images sourced from the Kaggle APTOS 2019 Blindness Detection (BD) online community. These images are classified based on the severity of diabetic retinopathy, categorized as no diabetic retinopathy, mild, or severe diabetic retinopathy.

Table 1 provides an overview of the dataset, which consists of 3,662 medical images sourced from the Kaggle online community.

Table 1. Medical image segmentation

Type	Name	Number of images	Percentage
0	No DR	1805	49.29%
1	Mild	370	10.10%
2	Moderate	999	27.28%
3	Severe	193	5.27%
4	Proliferative	295	8.06%
Total		3662	100%

2.3. Treatment and adjustment of images

Due to variations in acquisition conditions and equipment, many images in the dataset display differences in retinal alignment and quality. To address these inconsistencies and enable the models to learn network properties more efficiently, an image-processing method was implemented using the OpenCV library in Python.

The preprocessing steps included Gaussian blurring and circular cropping. A contour was drawn around each image, followed by the application of a Gaussian filter. This process reduces high-frequency components, enhancing the clarity of key features in each image and improving their suitability for analysis.

2.4. Description of the variables

Figure 6 illustrates the different levels of diabetic retinopathy (DR), which are categorized as follows:

- **Level 0 (No DR).** This level indicates a non-pathological state, meaning the absence of diabetic retinopathy.

- **Level 1 (Mild).** This stage is characterized by mild non-pathological diabetic retinopathy, where microaneurysms (red spots) are present. These microaneurysms are the source of hard exudates, which appear as high-contrast yellow spots.

- **Level 2 (Moderate).** At this stage, blood vessel distortion and swelling may occur, potentially compromising their ability to transport blood effectively.

- **Level 3 (Severe).** This stage is marked by significant blockage of blood vessels, leading to impaired blood supply to the retina.

- **Level 4 (Proliferative).** This is the most advanced stage, characterized by the secretion of growth factors by the retina, stimulating the proliferation of new blood vessels. These abnormal vessels grow within the retina and extend into the vitreous gel, filling the eye.

Figure 6. Types of retinopathies

Each stage of diabetic retinopathy has distinct characteristics and properties. However, during analysis,

clinicians may overlook certain details, which could increase the likelihood of an incorrect diagnosis.

2.5. Analysis of data

Initially, the images were downloaded and uploaded to Google Drive. Subsequently, they were organized into directories corresponding to each level of diabetic retinopathy, ensuring accurate differentiation. The critical factor for verifying the correctness of the results lies in the data associated with each retinal image, which is securely stored in the cloud.

The dataset stored in the cloud was used to enable algorithms in Google Colab to access the necessary information for training. The correctness of the results is determined by comparing the output of the classifier model or algorithm with the information available in the cloud. If the results match, it can be concluded that the classification is accurate; otherwise, the result is deemed incorrect. Once the training of each algorithm is completed and the results are obtained, a comparative analysis is conducted to evaluate their performance and identify the most efficient algorithm for solving the proposed problem.

Classification problems in machine learning are broadly divided into two main categories: binary problems and multi-class problems. The key distinction lies in the number of classes the model is required to identify within the data. In the case of binary problems, the model distinguishes between only two classes. These problems are characterized by simplicity, as binary models are generally easier to train and interpret due to the limited number of classes involved. Understanding the model's decisions is also more straightforward, as there are only two possible outcomes. In contrast, multi-class classification involves distinguishing between more than two classes, increasing the complexity of the task. These problems are more challenging to train and interpret due to the larger number of classes and the intricate relationships between them. Convolutional neural networks (CNNs) are particularly well-suited for solving multi-class problems, but interpreting the decisions of such models can be more difficult, given the wider range of potential outcomes.

2.6. Validation metrics

Confusion matrix

The confusion matrix plays a crucial role in identifying errors, enabling both descriptive and analytical evaluations of classification models. It displays the various correct and incorrect assignments made by the model [7]. Using the values provided by the confusion matrix, key evaluation metrics can be calculated to assess the model's performance, as illustrated in Figure 7.

Figure 7. Confusion matrix

The confusion matrix is an essential tool for validating neural networks, particularly in classification tasks. It offers detailed insights into the model's performance by quantifying the number of correct and incorrect predictions for each class.

Among the metrics derived from the confusion matrix and commonly applied to classification tasks in convolutional neural networks (CNNs) are accuracy and loss. These performance metrics are widely used to evaluate image classification models, both in pre-trained convolutional neural networks (pre-trained CNNs) and models developed using scikit-learn (sklearn). However, it is crucial to understand their limitations and to use them in conjunction with other metrics for a more comprehensive evaluation of the model's performance.

Accuracy

It represents the proportion of correct predictions made by the model, calculated as the number of correct predictions divided by the total number of predictions. It is an intuitive and straightforward metric to interpret; a high accuracy value indicates that the model is generally making accurate predictions. Accuracy is also a useful metric for quickly and easily comparing different models. However, accuracy is sensitive to the distribution of classes. If one class dominates the dataset, the model may achieve high accuracy by predominantly predicting the majority class, even if its performance on other classes is poor. This limitation makes accuracy less reliable in the presence of class imbalance.

Loss

It represents the average error of the model's predictions and is calculated as the sum of the individual errors for each prediction. It provides insight into the magnitude of the error, where a lower loss value indicates that the model is making predictions with smaller overall errors. Loss plays a crucial role in optimizing the model. During training, it is used to adjust the weights of the neural network to minimize error and

improve performance. The interpretation and scale of the loss depend on the specific loss function used, as different loss functions may have varying meanings and scales. However, loss can be influenced by class imbalance, which should be carefully considered when evaluating the model's performance.

For pre-trained convolutional neural networks (CNNs), the effectiveness of accuracy and loss metrics depends on the quality of the pre-trained model and its suitability for the specific classification task. Careful selection of the pre-trained network, along with appropriate hyperparameter tuning, is essential to optimize performance and ensure accurate evaluation. In models developed using sklearn, accuracy and loss metrics are directly applicable to classification tasks. However, it is crucial to account for the specific characteristics of the model and the classification problem when selecting appropriate metrics and evaluation techniques.

The effectiveness and reliability of accuracy and loss metrics depend on several factors, including the complexity of the problem, the quality of the data, the model architecture, and the additional metrics employed. It is essential to understand the limitations of these metrics and to use them responsibly in conjunction with other evaluation methods to ensure a comprehensive and robust assessment of image classification models.

2.7. Deep learning models utilized

2.7.1. Pre-trained models

The pre-trained neural networks employed in this study are based on the residual network (ResNet) architecture, which addresses the problem of gradient degradation by incorporating residual blocks. A residual block serves as a fundamental building unit in ResNets and consists of two paths within the network:

1. **Main path:** This path includes the convolutional or fully connected layers typical of a deep neural network.

2. **Direct path:** This is a direct connection that bypasses the layers in the main path, adding its output directly to the output of the main path.

This dual-path structure enables information to propagate through the network without being distorted by the transformations applied in the main path. Consequently, it simplifies the learning process and facilitates the training of much deeper neural networks compared to traditional architectures. For this study, two variants of ResNet were utilized: ResNet-18 and ResNet-50.

ResNet-18

It is an 18-layer deep convolutional neural network, as illustrated in Figure 8. Due to its relatively shallow architecture, ResNet-18 can effectively retain low-scale features, making it appropriate for serving as a feature extractor (encoder). The ResNet-18 architecture comprises 16 convolutional layers, 2 down sampling layers, and several fully connected layers [8].
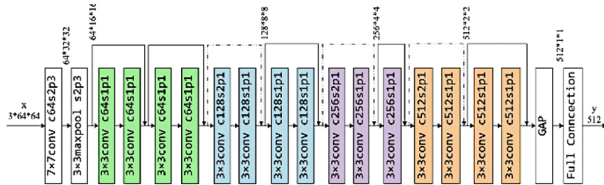

Figure 8. ResNet-18 model

ResNet-50

It is a convolutional artificial neural network with a depth of 50 layers, as depicted in Figure 9. It can utilize a pre-trained version trained on over one million images from the ImageNet database [9]. The ResNet-50 architecture consists of 48 convolutional layers, one MaxPooling layer, and one average pooling layer. It requires approximately 3.8×10^9 floating-point operations.
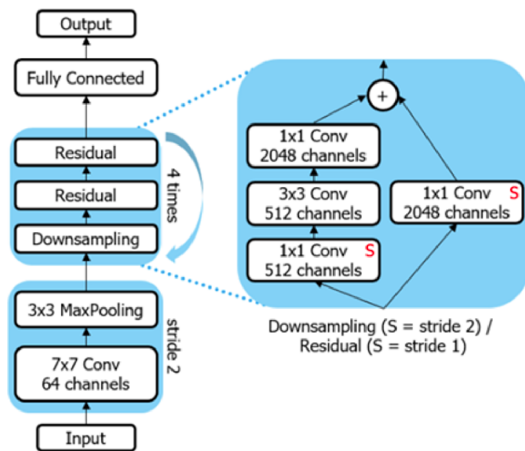

Figure 9. ResNet-50 Model

2.7.2. Non-pretrained models

The non-pretrained convolutional neural network (CNN) architecture utilized in this study consists of three 2D convolutional layers with 8, 16, and 32 filters, respectively. Each filter has a size of 3 x 3, ensuring that each convolution operation processes a 3pixel x 3 pixel region of the input. The network also includes three pooling layers, three dense layers with 64, 32,

and 3 neurons, respectively, and two dropout layers, each with a dropout rate of 15%.

Convolutional neural networks (CNNs) are a type of artificial intelligence algorithm based on multilayer neural networks. These networks are designed to learn and extract relevant features from images, as illustrated in Figure 10. CNNs are capable of performing various tasks, including object classification, detection, and segmentation [10]. They are a fundamental component of the field of deep learning [11].
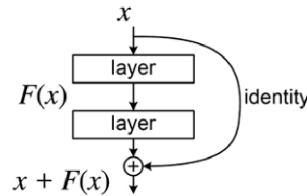

Figure 10. CNN

The Principal Component Analysis (PCA) model was also utilized in this study. PCA is a highly effective statistical technique widely applied in fields such as facial recognition and image compression. It is commonly used to identify patterns in high-dimensional data [12].

The ReLU (Rectified Linear Unit) activation function was employed in the convolutional neural networks (CNNs) used in this study. Its primary role is to enhance the nonlinear activation properties of the network without altering the receptive fields of the convolutional layers [13].

Convolutions

A convolution in an image is a pixel-by-pixel transformation achieved by applying a specific operation defined by a set of weights, commonly referred to as a filter. The convolutional layer in a neural network consists of a collection of learnable filters. Each filter is spatially small in terms of width and height but extends across the entire depth of the input volume [14].

Submapping

The pooling layer, also referred to as the subsampling layer, serves to progressively reduce the spatial dimensions of the representation, as illustrated in Figure 11. This reduction minimizes the number of parameters and computational complexity within the network [14].
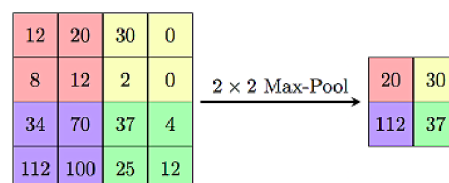

Figure 11. Submapping

Pooling layer

The pooling layer is utilized to reduce the dimensions of the feature maps, with the primary objective of decreasing processing times while preserving the most critical information. This dimensionality reduction helps mitigate overfitting in the network and introduces a degree of translation invariance [15].

Retinographies

Retinography is a diagnostic procedure that captures a non-invasive, painless color image of the fundus of the eye [16].

How CNNs work

Convolutional Neural Networks (CNNs) operate through machine learning [17] and supervised learning [18], leveraging several key components that function in an integrated manner. The core of CNNs lies in their convolutional layers, which perform convolution operations to analyze input images using small filters (kernels). These filters extract relevant features, such as edges, textures, and patterns, through matrix multiplication. By sliding across the image, the filters generate convolutional feature maps [19].

Activation functions

Following the convolution operation, a nonlinear activation function, such as the Rectified Linear Unit (ReLU), is applied. This introduces nonlinearity into the model, enabling it to capture and extract more complex features.

Pooling layers

These layers are employed to reduce the dimensionality of feature maps by summarizing the information extracted by the convolutional layers. This operation is typically performed using techniques such as max pooling or average pooling, which effectively reduce the size of the features while retaining their most relevant information.

Fully connected layers

After passing through multiple convolutional and pooling layers, the extracted information is flattened and fed into one or more dense (fully connected) layers. These layers perform classification or regression operations to generate the final output.

Regularization

To prevent overfitting and enhance the generalization capabilities of the model, regularization techniques are employed. These include methods such as dropout, which randomly deactivates neurons during training to reduce reliance on specific features, and batch normalization, which normalizes the activations of intermediate layers.

Loss function and optimization

During training, a loss function is employed to quantify the discrepancy between the model's predictions and the actual labels. Optimization algorithms, such as stochastic gradient descent (SGD) and its variants, are then used to minimize this loss. By iteratively adjusting the weights of the neural network, these algorithms enhance the model's performance and predictive accuracy.

3. Results and discussion

ResNet (Residual Networks) addresses degradation issues in deep neural networks by introducing residual blocks. The primary differences among ResNet models lie in their depth, the size of the residual blocks, learning capacity, and computational cost. The training process was conducted in two phases, incorporating both the pre-trained ResNet models and the non-pre-trained CNN. This was performed using a dataset with imbalanced class distributions.

Phase-1

As shown in Table 2, during the training of the ResNet-18 model, the loss on the training set was observed to be 86%, while the validation loss (val_loss) was significantly higher, reaching 194%. This was accompanied by an accuracy of 60% and a validation accuracy (val_accuracy) of 70%. These results indicate potential calibration issues, early stopping, or improper training configurations caused by factors such as underfitting, excessive regularization, non-representative data, or sampling problems. For the ResNet-50 model, the training loss was 132%, and the validation loss was 126%, with training accuracy at 48% and validation accuracy at 54%. These metrics suggest challenges related to the model's learning and generalization capacity, possibly due to its increased complexity and computational requirements. In contrast, the non-pre-trained CNN demonstrated superior performance, achieving a training loss of 19% and a validation loss of 22%, with training and validation accuracies of 92% and 91%, respectively. The alignment of loss and accuracy metrics between the training and validation sets indicates that this model is generalizing well and effectively learning from the data.

As shown in Figure 12, a significant number of samples were classified into class 0 (No DR) with a count of 330. Class 1 (Mild) contained 19 samples, class 2 (Moderate) included 87 samples, class 3 (Severe) had

20 samples, and class 4 (Proliferative) comprised 35 samples.

Table 2. Training and validation phase-1

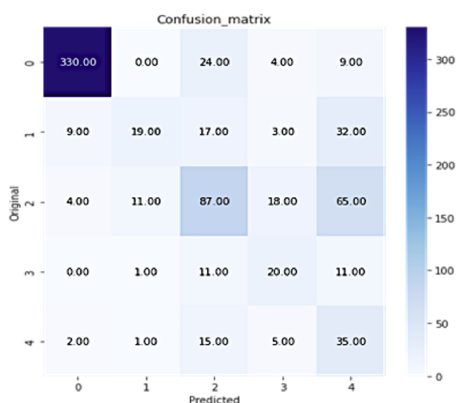|	ResNet-18	ResNet-50	CNN
Loss in training (loss)	0.86	1.32	0.19
Training precision (accuracy)	0.6	0.48	0.92
Loss in validation (val_loss)	1.94	1.26	0.22
Validation precision (val_accuracy)	0.7	0.54	0.91

Figure 12. Confusion matrix ResNet18

Phase-2

In Phase-2, a series of adjustments were made to the hyperparameter configurations of the ResNet-18, ResNet-50, and non-pretrained CNN models to develop a robust and consistent model. As indicated in Table 3, the ResNet-based models showed no notable improvements compared to the results obtained in Phase-1. In contrast, the non-pretrained CNN model demonstrated significant enhancement in performance and precision, achieving an accuracy of 94%, a validation accuracy (val_accuracy) of 93%, a loss of 18%, and a validation loss (val_loss) of 19%. These metrics indicate effective generalization of the acquired knowledge, with consistent and reliable results. The non-pretrained CNN model clearly outperformed the ResNet-based models and proved to be a superior and more suitable choice for predicting Hepatic Retinopathy.

Table 3. Training and validation phase-2

	ResNet-18	ResNet-50	CNN
Loss in training (loss)	0.83	0.065	0.18
Training precision (accuracy)	0.68	0.48	0.94
Loss in validation (val_loss)	1.88	0.2	0.19
Validation precision (val_accuracy)	0.26	0.83	0.93

As depicted in Figure 13, a large number of samples were classified into class 0 (No DR), with a count of 351. Class 1 (Mild) included 8 samples, class 2 (Moderate) comprised 162 samples, class 3 (Severe) had 25 samples, and class 4 (Proliferative) accounted for 23 samples.
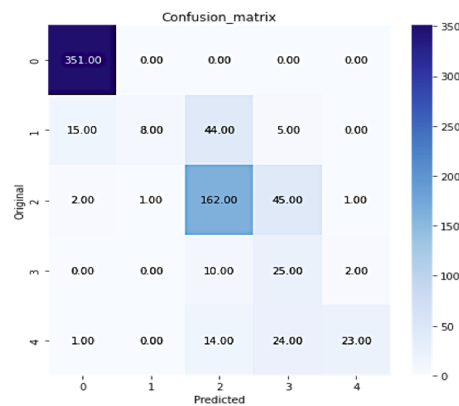

Figure 13. Confusion matrix ResNet50

The results obtained from the ResNet-based models raise several issues for discussion. The high loss observed in both training and validation phases may be attributed to the class imbalance within the dataset. Additionally, the combination of low accuracy values and high loss suggests that the models are not learning effectively from the data. This could be due to a lack of convergence or suboptimal hyperparameter configurations, leading to underfitting. Although ResNet-50 is inherently more powerful than ResNet-18 due to its greater depth and capacity, it may not be adequately suited or sufficiently tailored to the specific problem at hand.

The loss and accuracy indicators observed in Phase-1 and Phase-2 underscore the effectiveness of the non-pretrained CNN model. The high accuracy in both the training and validation sets suggests that the model successfully captures the patterns within the data and generalizes the acquired knowledge effectively. The minimal discrepancy between the training accuracy and validation accuracy (val_accuracy) is within acceptable limits and may be attributed to noise in the data or slight variations between the training and validation datasets.

The proposed approach for diabetic retinopathy detection offers significant advantages through its rigorous benchmarking of multiple neural network architectures. This process provides a comprehensive understanding of how various artificial intelligence models address a complex medical problem. The methodology is particularly notable for its ability to highlight the strengths and limitations of each architecture, demonstrating that increased model complexity does not necessarily translate into superior performance. The non-pretrained CNN emerged as a highly innovative

solution, achieving consistent accuracy exceeding 90%, robust generalization capabilities, and efficient handling of class imbalance key factors in diagnosing diseases characterized by rare but potentially severe presentations.

Despite its strengths, the proposed approach has notable limitations that warrant consideration. The reliance on a specific neural network architecture may restrict the transferability of the solution to other medical contexts, as the design is highly tailored to the dataset used in this study. Furthermore, the research highlighted the challenges faced by pre-trained models, such as ResNet-18 and ResNet-50, in adapting to medical datasets with complex and intricate features. This underscores the need for additional strategies, including advanced resampling techniques, weighted loss functions, and the augmentation of domain-specific data. These complexities introduce a more labor intensive development process, necessitating specialized expertise in both machine learning and the specific medical domain.

4. Conclusions

This analysis provided critical insights into the performance of various artificial intelligence models for the detection of diabetic retinopathy, highlighting significant variability among the evaluated neural network architectures [20].

ResNet-18 demonstrated critical limitations, with accuracy declining dramatically from an initial 70% to 26% in the final phase, underscoring its inadequacy for handling the complexity of medical image classification. In contrast, ResNet-50 exhibited a more robust learning capacity, achieving substantial improvement and reaching 83% accuracy in the final phase, emphasizing the importance of tuning and adaptation.

The non-pretrained CNN emerged as the most effective solution, consistently maintaining high levels of accuracy, nearing 91%, across both training phases and significantly outperforming the pre-trained models. This architecture achieved a training accuracy of 92% and a validation accuracy (val_accuracy) of 91% from the outset. Its stability across metrics and low validation loss (val_loss: 0.19 in Phase 2) demonstrated its capability to capture the necessary patterns for accurate image classification [21]. These results highlight that a carefully designed, simpler architecture can outperform more complex models in terms of efficiency and accuracy for specific problems.

Class imbalance was identified as a critical factor, particularly affecting the performance of the pretrained ResNet models. The non-pretrained CNN handled this challenge remarkably well, suggesting that thoughtful architectural design can overcome the structural limitations of more complex models. While the non-pretrained CNN successfully managed class imbalance, ResNet-18 and ResNet-50 struggled, particularly during the early training phases. This emphasizes the importance of implementing additional strategies, such as weighted loss functions, data augmentation, or advanced resampling techniques, to mitigate the impact of imbalance and enhance the performance of more complex models. Ensuring high-quality retinography images [2] is also crucial to avoid inconsistencies during the training phase.

Future research should focus on advanced strategies to manage class imbalance in medical datasets, addressing one of the most significant challenges identified in this study. These efforts should aim to create methodologies that ensure a more balanced representation of different image categories, particularly for minority classes that are critical to diagnosing diabetic retinopathy.

Proposed strategies include developing advanced resampling techniques, such as SMOTE, designing custom loss functions that dynamically weight classes, and creating data augmentation methods specifically tailored to medical images. These approaches aim not only to enhance model accuracy but also to improve their ability to detect rare yet clinically significant cases, representing a substantial advancement in the application of artificial intelligence to medical diagnosis.

The relevance of this work lies in its potential to transform AI systems capabilities for handling complex and imbalanced datasets, particularly in medical contexts where early and accurate detection is crucial for effective treatment. This direction offers promising avenues for improving diagnostic precision and addressing critical challenges in medical imaging.

References

[1] OMS, *TADDS* Instrumento para la evaluación de los sistemas de atención a la diabetes y a la retinopatía diabética.* Organización Mundial de la Salud, 2015. [Online]. Available: https://upsalesiana.ec/ing33ar8r1

[2] L. García Ferrer, M. Ramos López, Y. Molina Santana, M. Chang Hernández, E. Perera Miniet, and K. Galindo Reydmond, "Estrategias en el tratamiento de la retinopatía diabética," *Revista Cubana de Oftalmología*, vol. 31, pp. 90–99, 03 2018. [Online]. Available: https://upsalesiana.ec/ing33ar8r2

[3] F. Tablado. (2020) Inteligencia artificial en el trabajo ¿cómo afecta la ia al ámbito laboral de las empresas? Grupo Artico34. [Online]. Available: https://upsalesiana.ec/ing33ar8r3

[4] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan, "Assessing the performance of prediction models: a framework for some traditional and novel measures," *PubMed Central*, vol. 21, no. 1, pp. 128–138, 2010. [Online]. Available: https://doi.org/10.1097/EDE.0b013e3181c30fb2

[5] L. Rouhiainen, *Inteligencia Artificial: 101 cosas que debes saber hoy sobre nuestro futuro.* Editorial Alienta, 2018. [Online]. Available: https://upsalesiana.ec/ing33ar8r5

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[7] J. M. Sanchez Muñoz, "Análisis de calidad cartográfica mediante el estudio de la matriz de confusión," *Pensamiento Matematico*, vol. 6, no. 2, pp. 9–26, 2016. [Online]. Available: https://upsalesiana.ec/ing33ar8r13

[8] X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, and W. Li, "Moving object detection method via resnet-18 with encoder–decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108 152–108 160, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2931922

[9] G. Celano, "A resnet-50-based convolutional neural network model for language id identification from speech recordings," *ACL Anthology*, 2021. [Online]. Available: https://upsalesiana.ec/ing33ar8r8

[10] E. Todt and B. A. Krinski, *Convolutional Neural Network–CNN.* Universidade Federal do Paraná, 2019. [Online]. Available: https://upsalesiana.ec/ing33ar8r9

[11] J. Torres, *DEEP LEARNING Introducción práctica con Keras.* LuLu, 2018. [Online]. Available: https://upsalesiana.ec/ing33ar8r10

[12] C. Zhu, C. U. Idemudia, and W. Feng, "Improved logistic regression model for diabetes prediction by integrating pca and k-means techniques," *Informatics in Medicine Unlocked*, vol. 17, p. 100179, 2019. [Online]. Available: https://doi.org/10.1016/j.imu.2019.100179

[13] C. Bonilla Carrion, *Redes Convolucionales.* Trabajo Fin de Grado Inédito. Universidad de Sevilla, 2020. [Online]. Available: https://upsalesiana.ec/ing33ar8r12

[14] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, and S. Ridella, "The 'k' in k–fold cross validationthe 'k' in k-fold cross validation," in *European Symposium on Artificial Neural Networks, Computational Intelligenceand Machine Learning*, 2012. [Online]. Available: https://upsalesiana.ec/ing33ar8r15

[15] L. Herguedas Fenoy, *Guía práctica clínica para la realización de una retinografía.* Universidad de Valladolid, 2018. [Online]. Available: https://upsalesiana.ec/ing33ar8r17

[16] J. Chua, C. X. Y. Lim, T. Y. Wong, and C. Sabanayagam, "Diabetic retinopathy in the asia-pacific," *Asia-Pacific Journal of Ophthalmology*, vol. 7, no. 1, pp. 3–16, 2018. [Online]. Available: https://doi.org/10.22608/APO.2017511

[17] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, and L. Cilar, "Interpretability of machine learning-based prediction models in healthcare," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 5, p. e1379, 2020. [Online]. Available: https://doi.org/10.1002/widm.1379

[18] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, Jan 2018. [Online]. Available: https://doi.org/10.1093/nsr/nwx106

[19] J. Wu, *Introduction to Convolutional Neural Networks.* LAMDA Group, 2017. [Online]. Available: https://upsalesiana.ec/ing33ar8r22

[20] R. Gonzalez Gouveia, *Diferencias entre Inteligencia Artificial vs Machine Learning vs Deep Learning.* YouTube, 2021. [Online]. Available: https://upsalesiana.ec/ing33ar8r23

[21] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv*, 03 2015. [Online]. Available: https://upsalesiana.ec/ing33ar8r24