Tamjeed Journal of Healthcare Engineering and Science Technology (TJHEST) ISSN: 2583-7281
***Vol.3, issue** (**2**),2025, **Page** 28-41*

Leveraging DenseNet121 and Data Augmentation for High-Accuracy Diabetic Retinopathy Screening

Shihab A. Shawkat [1],* Yi Zhu[2], Shafeeq Kanaan Shakir AlDoori [3]

[1] Department of Quality Assurance and Academic Performance, University of Samarra, Salah al-Din, Iraq; shahab84ahmed@gmail.com
[2] Department of Artificial Intelligence, Nanjing University, Nanjing, Jiangsu 210003, China; 194160330@smail.cczu.edu.cn
[3] Department of Physics, College of Education, University of Samarra, Salah-AL-Din, Iraq; shafiq.k.shaker@uosamarra.edu.iq
* Correspondence: shahab84ahmed@gmail.com ; Tel.: 009647711000248

Received: 20.10.2025

Revised: 18.11.2025

Accepted: 25.12.2025

Published:27.12.2025

Abstract: This study aims to classify retinal fundus images to detect diabetic retinopathy using convolutional neural networks (CNNs). The research methodology involved developing a CNN based on the DenseNet121 architecture, augmented with a dense layer of 256 neurons and a 5-neuron SoftMax output layer. Two experiments were conducted: the first applied Deep Learning (DL) to an unbalanced dataset, while the second utilized Transfer Learning (TL) combined with Data Augmentation on a balanced dataset. Results showed that DL achieved an accuracy of 79.34% with a 20.04% loss, whereas TL significantly improved performance, yielding 97.78% accuracy and only 6% loss. The study concludes that Transfer Learning with dataset balancing produces a more precise and efficient diagnostic aid compared to standard Deep Learning, offering a reliable tool to support medical professionals in accelerating screening processes and prioritizing patient care.

Keywords: Diabetic Retinopathy, Convolutional Neural Network, Transfer Learning, Data Augmentation, Medical Image Classification, DenseNet121.

1. Introduction

Diabetes is a chronic disease that occurs in a person when their pancreas does not produce enough insulin or when it is not used efficiently, according to the International Diabetes Federation (IDF) [1]. This disease is an important cause of kidney failure, strokes, lower limb amputation, myocardial infarctions and blindness, which is the focus of the project. The number of people with diabetes, according to the World Health Organization (WHO) [1, 9] has increased from 108 million in 1980 to 422 million in 2014, a statistic obtained from the 2016 annual report of NCD-RisC [2, 10]. This indicates that a large percentage of these people may have some symptom mentioned above, and for the purposes of this project, it is sought to determine if a patient has signs of said disease [3]. Sometimes humans are careless, so an error may arise when observing, and this is where there is an opportunity with recognition and/or image processing with CNN, a tool that can provide support to a doctor's diagnosis regarding a patient who has conditions or signs of DR in the fundus examination, or have signs and not suffer from it. In order to speed up the medical process, and only as an example, automatic prioritization of waiting lists based on the state of the disease [4,5].

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The proposed research seeks to establish an aid to the medical diagnosis of a fundus examination in patients who present characteristics or symptoms of DR. For this, it makes use of the processing of these eye exam images to train convolutional neural networks and thus obtain an accuracy of 70% or higher focused on detecting whether the patient has said disease. A criterion of 70% accuracy is established as a minimum, because a lower percentage could imply that the algorithm does not have an acuity that allows demonstrating reliability in the diagnosis. This means that if the algorithm achieves 50% accuracy, it would be equivalent to flipping a coin to determine if a patient has said disease, meaning the diagnosis would be completely stochastic [6]. The articles [7-5] Physicians' Diagnostic Accuracy, Confidence, and Resource Requests shows in its results that in the easiest medical cases, diagnostic accuracy is 55.3%, while in the most difficult cases it only reaches 5.8%, therefore, an accuracy of 60% in the CNN algorithm demonstrates that the network is beginning to better recognize patterns in the images, but it is still not an acceptable range that can help a doctor's diagnosis, given that the purpose of this memoir is to provide a reliable contribution to a diagnosis, hence the idea of setting the accuracy at 70% [8].

This does not mean that if the algorithm has the mentioned accuracy or higher, it will be better than a doctor's training, but it will be a useful tool to complement the specialist's diagnosis, so that a technologist can use the algorithm and speed up the medical process. The primary objective of the work is to classify retinal fundus images through convolutional neural network processing, in order to detect diabetic retinopathy, develop an algorithm using a convolutional neural network model for image processing, under two learning methods and finally to verify the algorithm's diagnosis based on quality metrics, determined by the loss function and accuracy. The method to be used in this research is the scientific method. A method known as the persistent application of logic as a common characteristic of all reasoned knowledge. From this point of view, the scientific method is simply the way we test impressions, opinions or assumptions by examining the best evidence available for and against them as mentioned in the Figure 1.

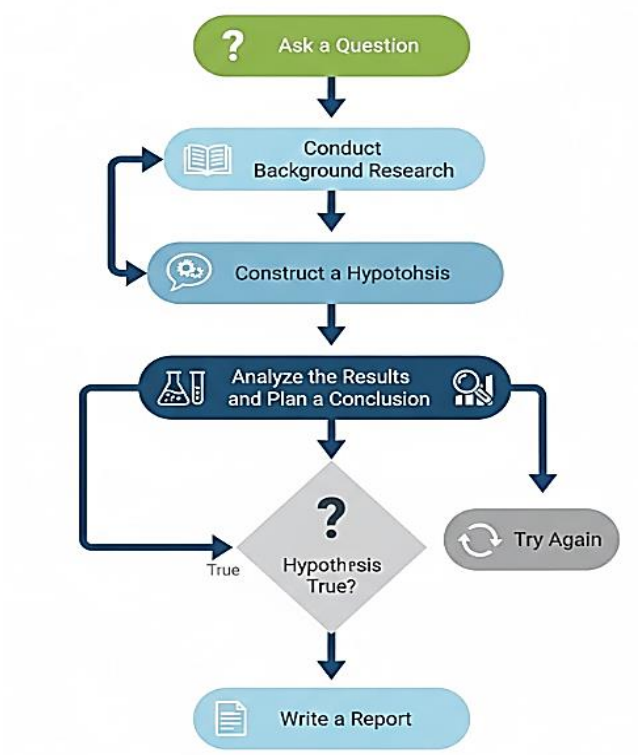

Figure 1. Illustration of the Scientific Method.

2. Theoretical Framework and Review of literature

2.1. Convolutional Neural Networks

It is known that convolutional neural networks [11] are a type of network that focuses on detecting patterns in images in order to understand a concept. To achieve this task, there are

methods that efficiently solve this detection; it must be understood that images have millions of pixels, so each of those pixels is a number representing a colour, but this number is what the neural network understands. Therefore, if you want to work with a low-quality image, assuming about 60x60 pixels, the neural network will have to work with their multiplication, resulting in 3,600 pixels. Under that premise, what happens with an image that has higher resolutions? Suppose working with a FHD (Full High Definition) image of 1920x1080 pixel resolution, the neural network will have an input to consider of 2,073,600 pixels, making the problem of working with images of an order with no apparent solution, since the time it would take to go through that image and gradually understand patterns would be immensely long, and so on with images that have a higher resolution, such as 4K images (4096x2160 = 8,847,360 pixels). To perform this task, two methods exist, mentioned in points 2.1.1 and 2.1.2.

The operation of the network, intuitively, can be seen in Figure 2, which shows a retinal fundus as input image to the network. Then the first convolution (CONV1) is applied, making the first transformation of the image. Followed by a Pooling layer to reduce dimensionality, then the process is repeated until reaching the FC (Fully-Connected) layer, which is composed of a number N of neurons for processing the features found in the previous processes, delivering this to the last process called SOFTMAX, a layer that returns a probability distribution, from which the alternative with the highest probability is chosen. This result is the classification predicted by the model [11-12].

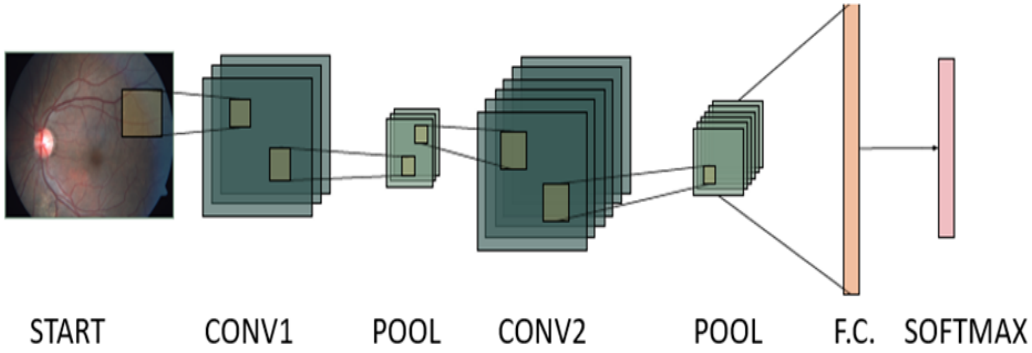

Figure 2. Representation of a CNN [12].

2.1.1. Convolution

Convolution is a method that creates a filter of a lower resolution that will divide the image into smaller images with lower resolutions, which allows the first layers of the neural network to work by understanding simpler patterns that will later transform into complex patterns, and being smaller images means that the input pixels are fewer, improving analysis performance. This process is evidenced in Figure 2, specifically in the layers called "CONV1" and "CONV2", abbreviation for Convolutional 1 and Convolutional 2, where what was mentioned in the first paragraph is appreciated, about the increase of images or known as depth [13].

2.1.2. Sub-Sampling

Subsampling, a method normally known in this field as pooling, which complements convolution, making the image "lose quality" without losing important information from the image that can be patterns. This means that if the image in the first convolution found patterns by the colour gradients in the image, such as a yellow line, the algorithm will no longer have an image of a yellow line with shades of royal yellow, duck yellow, but will only remain as yellow. These two methods can be better appreciated below with Figure 3, which shows an application of subsampling called Max Pooling, which obtains the maximum value of a set of colours in a matrix representing the image.

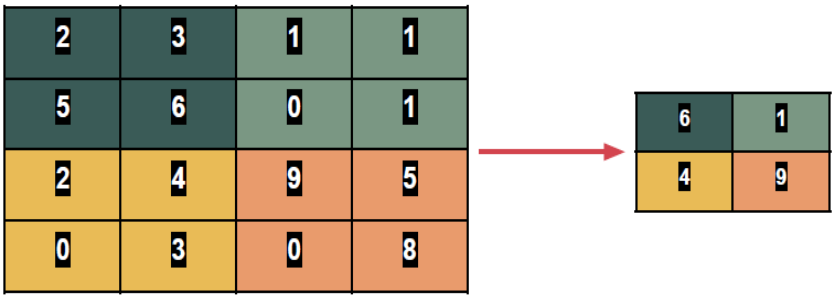

Figure 3. The matrix on the left represents the pixels of an image, with their associated value before applying Max Pooling. The matrix on the right shows the Max Pooling applied.

2.1.3. Data Augmentation

A common problem in Deep Learning is the large amount of data needed for model training, which we can solve with augmented data. This is where the Data Augmentation technique comes in, which, as its name indicates, allows us to increase our dataset in two ways. The first consists of introducing perturbations into the original data, for example choosing an original cantered image, and we replicate it off-centre, inverting axes, etc [14]. The second way consists of using different distributions, for example if we want to train a model to classify high-resolution images, we add low-resolution images, with noise, or filters, always maintaining the highest proportion of high resolution.

2.1.4. Deep Learning (DL)

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have drastically improved the state of the art in speech recognition, visual object recognition, object detection and many other domains, such as drug discovery and genomics [15]. Deep learning discovers intricate structure in large datasets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional networks have produced advances in processing images, video, speech and audio, while recurrent networks have shed light on sequential data such as text and speech.

2.1.5. Transfer Learning (TL)

Transfer Learning (TL) is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned. While most machine learning algorithms are designed to address individual tasks, the development of algorithms that facilitate transfer learning is a topic of ongoing interest in the machine learning community [16].

3. Methodology

The development of an algorithm capable of processing retinal fundus images is proposed, in order to determine the existence of diabetic retinopathy, and if so, the state or degree of progression. The algorithm receives an input image of any size, to then be resized to a standard of 224×224 pixels. In this way, the neural network receives this standardized image and is processed to obtain an output probability distribution, choosing the probability that has the highest value, understanding that this is the predicted class, among the 5 existing ones. The convolutional neural network algorithm is trained with a publicly accessible dataset for research.

Algorithm1. Already trained algorithm capable of predicting the delivered sample.

```
Input: Image I
Output: Predicted disease grade and performance metrics

I ← LoadImage()
I ← Resize(I)
A ← ConvertToNumPy(I)
A ← Standardize(A)
Load A into CNN model
P ← CNN_Model(A)
if P == Expected_Grade_Type then
    Print("Predicted disease level: ", P)
end if
Print("Prediction metrics: ", Metrics(P))
```

3.1. Architecture of the Neural Network Model

The proposed algorithm consists of a base neural network model, added to a dense layer of neurons, followed by an output layer with SoftMax activation. The base neural network model is DenseNet121 or Densely Connected Convolutional Networks, which, for each layer, takes the feature maps of all previous layers and uses them as inputs, and its own feature maps are used as inputs in all subsequent layers.

When the image has been processed by this base model, the obtained parameters are captured and processed by a dense layer of 256 neurons, which deliver information to the last layer of 5 neurons, which is the output, with SoftMax activation function, which is responsible for delivering a probability distribution for the 5 classes. Finally, the ARGMAX() function returns the highest percentage obtained with the SoftMax layer, delivering an output of a single value, which is the prediction of the disease grade of the evaluated image.

3.2. Scope of The Method

The proposed method focuses solely on the automatic detection of diabetic retinopathy using prior processing of retinal fundus images to be delivered to a convolutional neural network, determining whether it has the disease or not. The classification of the degree of disease progression is not the focus of the algorithm's scope under investigation, but techniques are done to work with this problem, because it can affect accuracy. The algorithm's accuracy will be exclusively for the dataset featured in the research. Without closing the possibility of performing proof-of-concept tests with images belonging to another dataset, emphasizing that these images are not for the purpose of the study, nor are they included in the final results.

The algorithm's accuracy is assured for the type of samples used in the research, so poorly taken samples do not ensure good performance; this is explained in more detail in section 3.2.1. The fact that the algorithm can identify the disease does not mean that it can replace a doctor's work, so the prediction will always be subject to confirmation by an expert. The Legible Samples without disease Mild DR Moderate DR Severe DR Proliferative DR.

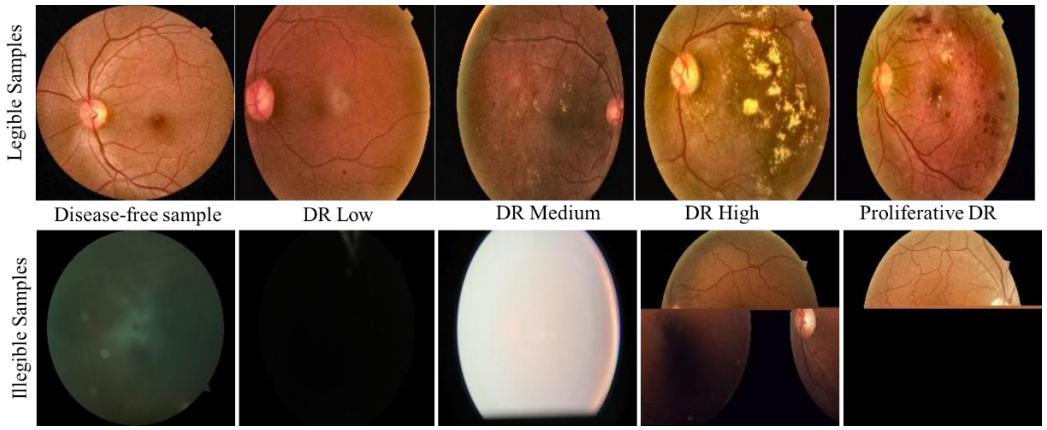

Figure 4. Legible and illegible samples of retinal fundus. Legible samples are categorized by disease grade. Illegible samples have no category because such a sample is not considered, so the patient must repeat the exam. Legible samples were pre-processed with CLAHE histogram in order to highlight the differences between each class.

3.2.1. Retinal Fundus Samples: The dataset to be used has good quality samples; in fact, there are very few images that have low quality or that could be considered illegible. Figure 4 has an extraction of samples that are part of the dataset to be used for the research, but illegible samples are also attached, although they are not part of the set, but extracted from another source. This type of illegible sample does not ensure the correct functioning of the algorithm, because it is a failure of a stage prior to the use of the algorithm. The algorithm's effectiveness is not guaranteed for a poor-quality sample, as the algorithm could not differentiate details, omitting them, or strictly the image is illegible and gives a random result. Regarding poor quality, it does not refer to the instrument used to take images being of poor quality, but the manipulation of this instrument, or some stage after saving the exam being erroneous and consequently, the image being illegible. For example, the image having flash-like lights, or the lens not being calibrated and ending up taking a photo that is too dark, obscuring the components of the retina. It should also not be confused that the image to be processed must have perfect quality, but it must be legible.

3.2.2. Technical Feasibility: For the development of the project, retinal fundus images must be available, that are healthy, as well as retinal fundus images with different degrees of the target disease. An important aspect to consider is the number of images to use, because artificial intelligence tools need a large amount of data to converge to a solution; this is why artificial neural networks provide learning capacity, as it is the training process that gives rise to the algorithm's utility. That is why we have a publicly accessible dataset for research purposes provided by the Kaggle community. This dataset provides a quantity of 3,662 images. Each one has an identification, then whether it is the left or right eye and finally the degree of the disease on a scale from 0 to 4, with 0 being a healthy eye and 4 being an eye with proliferative DR, i.e., the most severe degree of disease progression.

3.3. Data pre-Processing

This section is a very important phase for the learning of a neural network, because it is the way in which we deliver the information and with what quality we deliver it. Applied to CNNs, it is a stage that comprises transformations of the input images, with the purpose of avoiding unnecessary data that harms the algorithm's performance and the hardware's processing capacity, explained in section 2.1. That is why the phases through which an image is converted from its original structure to input data for a neural network are explained below.

3.3.1. Dataset Distribution: The dataset to be used has an original distribution observed in Figure 5, which has an imbalance between classes, meaning it has more samples of one type of class than the others. Therefore, the predominant class in the dataset is that of patients with healthy eyes (class 0). This generates inconsistency in the data to be trained, affecting the algorithm's performance,

evidenced in the experimental development chapter. For this, Data Augmentation is used, a technique used for the treatment of balanced datasets. Although the hypothesis lies in the classification of whether a retinal fundus is healthy, and not in the classification of what grade the patient is in, it is an important characteristic to consider during the development and measurement of this algorithm, as its performance is affected.

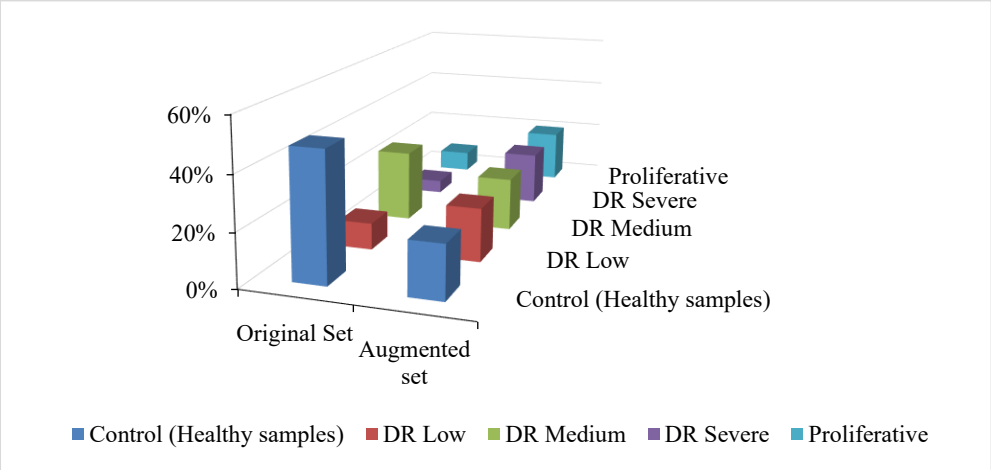

Figure 5. Distribution of the image set before and after applying Data Augmentation.

3.3.2. CLAHE Image Equalization: Contrast-Limited Adaptive Histogram Equalization (C-HLE) is a method that has proven useful for assigning intensity levels to images in medical settings. The method is designed to allow the observer to easily see, in a single image, all contrasts of interest in clinical research [17]. It calculates several histograms, each corresponding to a different section of the image, and uses them to redistribute the image's brightness values. Therefore, it is suitable for improving local contrast and edge definition in each region of an image. Applied to a fundus image, the following result is obtained, visualized in Figures 6.1 and 6.2. It can be observed that the original image shows some detail, but in a faint manner, especially when viewed by an untrained eye, such as that of someone outside the medical or ophthalmological field. Passing the image through the CLAHE equalizer reveals a change in contrast and how the details stand out, even to an untrained eye. For example, the details that show the greatest change are the haemorrhages, small red spots on the right side of the retinal fundus and the optic nerve.

Figure 6.1. Original image.

Figure 6.2. Processed image.

3.3.3. Standardization of Image Dimensions: The images in the dataset come in 4K quality, so they are of high quality and in computational terms, are quite expensive to process. On the other hand, the definition of the input size in the network must be static, so different dimensions cannot be worked with. That is why a standardization criterion is established for the dimensions, which in this case is 224 x 224 pixels. Although the definition of this parameter is not random, as it is present in the input requirements of the base neural network used in this algorithm, which is DenseNet121 [18].

3.3.4. Normalization of Each Pixel Value: A pixel to the human eye is a colour, but for the computer it is a number representing a colour and for most image data, pixel values are integers ranging between 0 and 255. Neural networks process inputs using small weights, and inputs with large integer values can disrupt or slow down the learning process. As such, it is good practice to normalize pixel values so that each pixel value ranges between 0 and 1. It is valid for images to have pixel values in the range 0-1 and the images can be viewed normally. This can be achieved by dividing all pixel values by the largest pixel value, which is 255. This is done on all channels, regardless of the actual range of pixel values present in the image.

3.3.5. Transformation of Numerical Variables to Categorical Type: To keep all numbers in order and work on the same scale, the disease grades (0-4) must also be worked in that style; with this transformation we avoid numbers greater than 1. To achieve this, numerical variables are converted to categorical variables, meaning their value can be 0 or 1. Figure 7 shows the format change.

Name	Grade
background_1	0
background_2	1
background_3	2
background_4	3
background_5	4

	0	1	2	3	4
Position 0	1	0	0	0	0
Position 1	0	1	0	0	0
Position 2	0	0	1	0	0
Position 3	0	0	0	1	0
Position 4	0	0	0	0	1

Figure 7. The image on the left shows the image name along with the disease grade. The image on the right shows the process of transforming a numerical variable to a categorical one.

3.3.6. Division of Data for Training and Validation: According to the small number of images present in the original distribution of the dataset, it is sought to prioritize the largest number of images for model training. Therefore, a division of the set will be set at 95% of data for training (3,479 samples) and 5% of data for validation (183 samples). For the second experiment, this distribution will also be set in order to compare the results in a more balanced way or with similar conditions, with 8,550 training images and 450 validation images. For this division, the Train_Test_Split() function provided by the Scikit Learn library will be used, together with its Stratify parameter, which allows preserving the same number of images per class, both for the training set and for the validation set. This function creates two arrays randomly, but due to a parameter called Random State, it allows reproducibility, that is, setting the initial conditions so that the random images are always the same when dividing.

3.3.7. Model Validation: For the study and validation of the values obtained by the training of these models, the loss and accuracy metrics provided by the Keras library will be used. This library in its glossary explains that the loss function calculates the error of the equalization of the network weights, with respect to the original data curve and division. Regarding the accuracy variable, Keras shows that this metric creates two local variables, True Positives and False Positives. This metric is a simple operation that divides True Positives by the sum of True Positives and False Positives. Therefore, the research must achieve over 70% accuracy to satisfy the proposed hypothesis.

4. Experimental Development

The experimental development of this memoir consists of 3 stages, 2 of them are experiments and the last one consists of the analysis of the previous stages. The experiments aim to corroborate the hypothesis based on different configurations, choosing the one that fulfils the purpose with the best results. The importance of this process is the ability to observe its behaviour based on measuring the found solution, and if it needs improvements, work on them to develop a more polished model, in the sense of minimizing errors.

Experiment 1 consists of training the neural network by means of the Deep Learning method, with the dataset in its original form as provided by the Kaggle community, therefore it presents inconsistency in the number of images per class, i.e., it presents imbalance. Experiment 2 consists of training the neural network under the Transfer Learning method and the augmented dataset, i.e., balanced for the number of images per class, by using the Data Augmentation technique, in order to reduce possible errors and biases due to the inconsistency of the original distribution.

4.1. First Experiment: Original Distribution & DL

4.1.1. Unbalanced Dataset: The data to be used in the first experiment is the original set provided by the Kaggle community [19] for educational purposes. This set has an imbalance between the classes mentioned in point 3.3.1 and its distribution can be visualized in Figure 6.1. Therefore, the results obtained will be conditioned based on that aspect. The dataset to be used comes just as a doctor would observe a colour fundus, so it has not been pre-processed with image filters, as evidenced in Figure 6.1. The only pre-processing the set has had has been the reduction of the image size to a standard of 224 pixels high by 224 pixels wide, because the neural network has those input requirements, mentioned in point 3.3.3.

4.1.2. Parameters of the First Model: For the development of this first experimental phase, the training will be with the base model of convolutional neural networks DenseNet121 added to a deep layer of 256 neurons followed by a probability distribution layer called SoftMax. The base layer DenseNet121 uses the Deep Learning technique. The learning rate configuration will be *lr = 0.00001*, which is low in order to avoid the model's learning stagnating at a local minimum and not at the global minimum, achieving more accurate results.

4.1.3. Training Results: This phase comprises the algorithm being tested under the configurations mentioned in the two previous points.

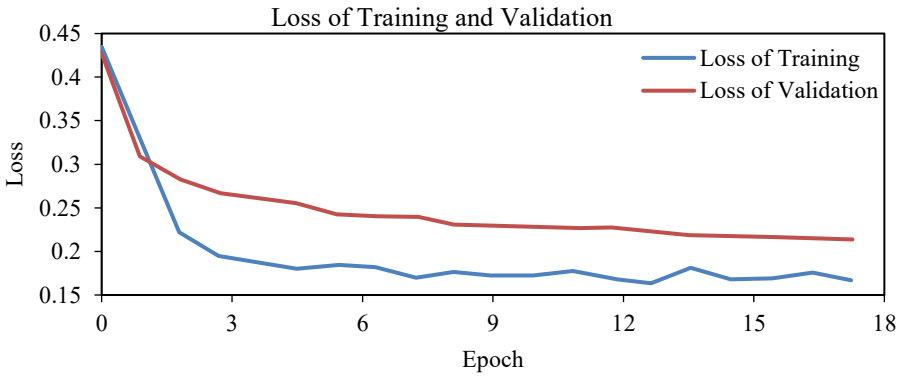

Figure 8.1. Loss function with Deep Learning.

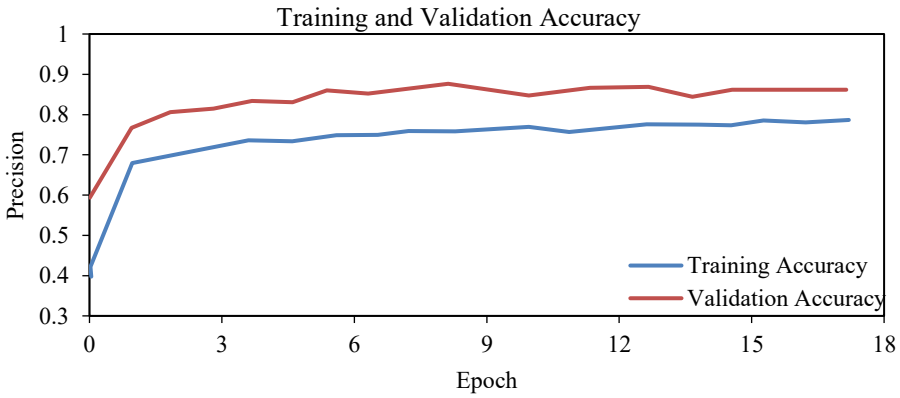

Figure 8.2. Accuracy obtained with Deep Learning.

Figures 8.1 and 8.2 show the behaviour that Deep Learning had with 20 training epochs. The loss graph shows the learning and validation path of these configurations. It can be seen that the training curve reaches a loss of 25%, being higher than the loss observed in the validation curve, which has a difference of 5%, reaching a lower level in prediction errors, with 20%. Figure 8.2 evidences the accuracy obtained from the learning of this model, as well as the validation of that learning. It is observed that the accuracy curve during training reaches a level lower than the validation curve. The training curve reaches a value slightly above 70%, while the validation curve slightly exceeds 75%.

4.2. Second Experiment: Augmented Data & TL

4.2.1. Balanced Dataset: For this second phase of experimental development, the Data Augmentation technique, explained in point 2.1.3, is used. The idea of its use lies in balancing that existing difference by classes, so the value of samples per class is approximately 1,789, making a set of 9,000 samples in total. In the original distribution, the only class that reaches that number of samples is that of healthy patients, which is why this increase in samples was made in the following 4 classes, reaching the same amount as the samples of healthy patients, generating a balance in the distribution with 20% per class. The contribution to the research that this point has is to verify whether an improvement in model prediction and accuracy is obtained.

4.2.2. Training Parameters: The second phase of experimental development, unlike the first, uses the Transfer Learning method, which should show better results than the first experiment and corroborate what is mentioned in point 2.1.5. The learning rate to be used will be the same as before lr = 0.00001, with the same purpose of not stagnating the learning at a local minimum.

4.2.3. Training Result: Figures 9.1 and 9.2 show the curves obtained during training with the Transfer Learning technique. The curves present in Figure 9.1 show that the learning loss follows a fairly similar trajectory with the validation curve. The training curve has a start above 40% error and converges at a value less than 10%, while the validation curve starts slightly above 30% error and converges at a value also less than 10% but being slightly higher than that of training. Figure 9.2 shows a similar behaviour of the training hits, as well as the validation ones. The accuracy curve of training starts at 80% while that of validation at 85%, converging both at a value close to 98%.

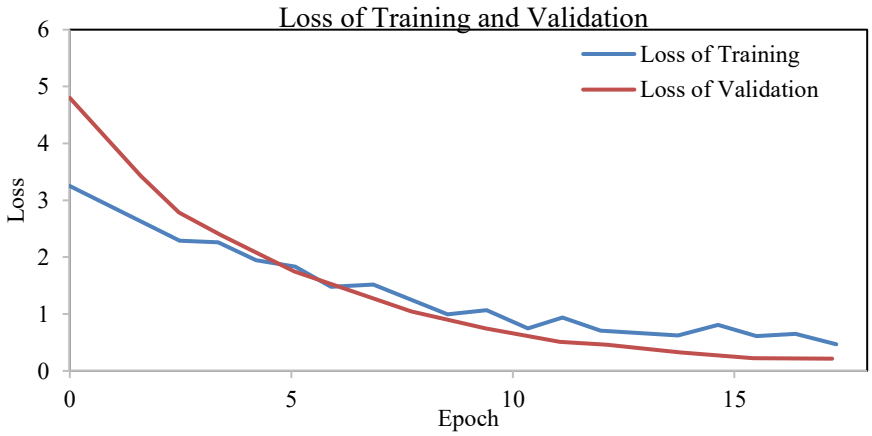

Figure 9.1. Loss under Transfer Learning.

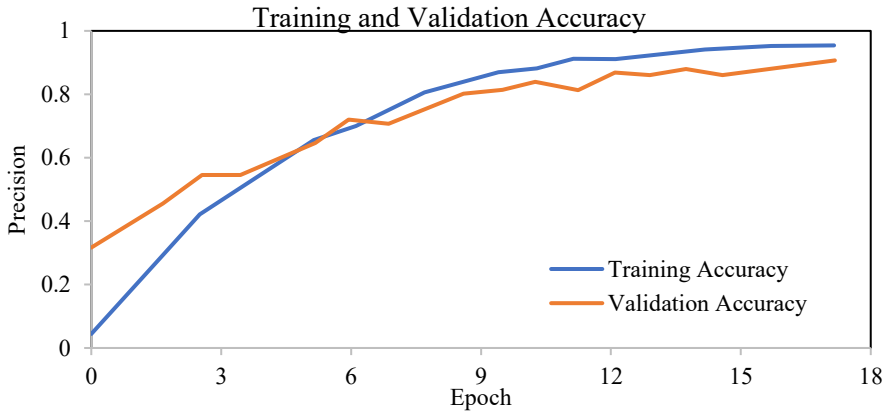

Figure 9.2. Accuracy under Transfer Learning.

In Table 1 the differences in the values of both loss and accuracy during training, as well as in validation, that each model achieved, are observed. Table 2 shows the number of correct predictions that each model obtained with different training methods applied to the validation set, for each class of the dataset.

4.3. Analysis of the obtained results

The first experiment shows that training with an unbalanced dataset with the Deep Learning technique does solve the purpose of the proposed hypothesis; moreover, it exceeds the expected accuracy by 9%, so it is a method that works or primarily fulfils the task. It is proven that Deep Learning is a useful tool in the identification and classification of diseases, not forgetting that it will always be under the supervision of a professional, although there is a quite important detail which is the class imbalance during algorithm training. As observed in Table 2, the algorithm trained with Deep Learning is capable of correctly predicting the class of healthy patients (Class 0) with 93.33% accuracy, this is because this class has a greater number of samples than the others, so the validation image set also presents this imbalance between samples.

The number of correct predictions per class is affected based on its distribution, and as observed in Figure 5, the lower the number of samples per class, the worse the performance, demonstrated in the number of correct predictions for severe and proliferative diabetic retinopathy samples (Class 3 and 4), which are the classes with the fewest samples. Regardless that Figures 8.1 and 8.2 show that the curves have a tendency to continue decreasing, it does not mean that with a greater number of epochs greater accuracy will be obtained, although it is observed in the obtained curves, because there is a known problem called "Overlearning or Overfitting", which harms model accuracy, as the network would be learning the answers "by memory" and would not be generalizing, which is the purpose of an artificial neural network.

Another argument in favour that a greater number of epochs does not guarantee a better result is the fact that if Figure 8.2 is observed, from before the third iteration, the algorithm already begins to classify in a more correct way the original distribution dataset, so the increase in accuracy changes value more gently. This is also observed in Figure 8.1, where the loss begins to decrease gently from the second iteration.

Table 1. Loss and Accuracy Metrics for Training and Validation for Each Model.

Model	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
Deep L	24.18 %	73.57 %	20.04 %	79.34 %
Transfer L	3.44 %	98.84 %	6.02 %	97.78 %

Table 2. Accuracy of the models for prediction per class.

Model	Class 0	Class 1	Class 2	Class 3	Class 4
Deep L	93.33 %	50.00 %	86.00 %	0 %	0 %
Transfer L	100 %	84.44 %	86.67 %	96.67 %	94.50 %

On the other hand, it is observed in Figures 8.1 and 8.2 that the validation curve, both for the loss and accuracy graph, has better results than the training curve, while the common should be the opposite. This is explained in the same way with the imbalance between the classes, as it can be approached from the following two ideas. The first is that the images present in classes 3 and 4 of the training datasets have samples that are more difficult to interpret, so the neural network cannot establish the differences that exist by disease grade. The second idea is that the validation set has a greater number of easy samples to predict, which would be from the classes where there is a greater number of samples in the training, specifically class 0 and 2, given the distribution present in Figure 6.1. The only apparent solution to this problem and demonstrated in this memoir is the balancing of the dataset, making the classes that had a lower number of correct predictions, from 0% correct for the classes with the fewest samples, now exceed 94% accuracy, so the second model has a much superior performance than the first and for all classes.

There is one more element in the experimental development to explain, which is the contribution of the Transfer Learning technique. Although the results obtained with Deep Learning were not bad, they still needed perfecting, and are demonstrated in the loss and accuracy curves obtained in the second experiment, specifically Figures 9.1 and 9.2, where the error or loss of the model is 6% and the accuracy is 97.78%. It can be verified by observing the initial values of these curves, specifically in Figure 9.2, where accuracy starts from 80% for learning and from 85% in validation, finally converging in very close values around 97.5%. This is also corroborated in the similarities of Figure 8.1 and 9.1, since the loss in both methods, i.e., Deep Learning and Transfer Learning, starts from 40%. The difference between these two figures lies in that the training under Transfer Learning has a lower loss from the start for the validation curve, and for the training curve instead of stagnating at a number close to 25%, it continues decreasing until reaching 6%. A difference of 19%, which is quite significant. Therefore, it adds a plus to the main objective of the memoir, obtaining a much more adjusted model and with much less difference between prediction by classes during training.

Although the focus of this memoir is the identification of whether a patient has signs of diabetic retinopathy or not and not the classification by disease grade, which also does not belong to the algorithm's scope, it is a topic that directly affects the focus and purpose of the project, which is why the Data Augmentation technique was used. However, although the research hypothesis is fulfilled using the unbalanced set with the obtained results, it is demonstrated that for the realization of a good model, it has to be worked with the search for the least possible biases

5. Conclusions

This research comprises a decision-making process based on obtained values to determine which method best solves the proposed hypothesis. The development of an algorithm that processes eye exams in image format to be evaluated by means of a CNN was proposed, which, based on the information the image contains, delivers a result whether that sample has signs of the target disease or not. This artificial neural network is designed with a base architecture DenseNet121, along with a deep layer of 256 neurons, followed by an output layer of 5 neurons with SoftMax activation function, which is a probability distribution. The difference lies in the method implemented for the learning process, which are Deep and Transfer Learning, and in the dataset to be used per experiment. Initially, experimentation is done with the Deep Learning method along with the dataset in its original distribution, in which the difference in samples per class is observed. From

this, results that meet the expected outcome of the project are obtained, achieving an accuracy of 79.34% and an error of 20.04%, but which demonstrate a model with imperfections in its predictions due to this imbalance. That is why the second method is implemented, which has two differences compared to the previous one: the use of the Transfer Learning method, subject to a dataset with balanced distribution, as a result of applying the Data Augmentation technique. The results obtained with this change are significant compared to the first experiment; moreover, it corrects a major implicit problem, which is the improvement in predictions because it can recognize the existing differences per class, a problem that the algorithm trained with DL could not correctly solve. Therefore, it is proven that TL is capable of obtaining a more accurate result than that obtained by DL and in less time, achieving an accuracy of 97.78% and a loss of 6%.

The contribution of this research to the medical focus is the acceleration of exam-taking processes and data organization in a company. As emphasized in several points of this memoir, the proposed solution does not replace and will not replace the capacity of medical personnel to determine and take action, but it is an aid or useful tool that can provide an instant prediction, so that the patient obtains feedback at the moment, knowing the first interpretation made by the algorithm, and does not have to wait 3 or more days until scheduling a medical visit, in order to obtain information about their condition. Emphasizing that the prediction is subject to approval by a medical specialist.

6. Future Work

As future work, it is proposed to emphasize more exhaustive preprocessing of the training image set, more specifically the quality of the samples to be used, the number of samples per class. Therefore, more than one existing dataset can be used, in order to reduce the prediction error per class. A second proposal for improving the obtained values is to study whether image filters such as CLAHE equalization, mentioned in this research, have better performance than training with unfiltered images. A third proposal for future work is the implementation of a medical-focused application, where a person who has their retinal fundus exam can upload the image to the system and it delivers the prediction, along with certain metrics. For this, the work should focus on software development and all the phases involved in software development

Funding: This research received no external funding.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. International Diabetes Federation. *IDF Diabetes Atlas*, 10th edn. International Diabetes Federation, Brussels (2021). https://diabetesatlas.org

2. Lang X, Li L, Li Y, Feng X. Effect of diabetes on wound healing: a bibliometrics and visual analysis. *J Multidiscip Healthc* **17**, 1275–1289 (2024). https://doi.org/10.2147/JMDH.S457498

3. Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, Spijker R, Hooft L, Emperador D, Dittrich S, Domen J, Horn SRA, Van den Bruel A; Cochrane COVID-19 Diagnostic Test Accuracy Group. *Cochrane Database Syst Rev* **7**, CD013665 (2020). https://doi.org/10.1002/14651858.CD013665

4. Mackenzie SC, Sainsbury CAR, Wake DJ. Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. *Diabetologia* **67**, 223–235 (2024). https://doi.org/10.1007/s00125-023-06038-8

5. Kohli M, Pandey P, Jakhmola V, et al. Revolutionizing diabetes care: the role of artificial intelligence in prevention, diagnosis, and patient care. *J Diabetes Metab Disord* **24**, 132 (2025). https://doi.org/10.1007/s40200-025-01648-y

6. Boels L, Moreno-Esteva E, Bakker A, Drijvers P. Automated gaze-based identification of students' strategies in histogram tasks through an interpretable mathematical model and a machine learning algorithm. *Int J Artif Intell Educ* **34**, 1–26 (2023). https://doi.org/10.1007/s40593-023-00368-9

7. Sarma AD, Devi M. Artificial intelligence in diabetes management: transformative potential, challenges, and opportunities in healthcare. *Hormones* **24**, 307–322 (2025). https://doi.org/10.1007/s42000-025-00644-4

8. Ryu G, Lee K, Park D, Park SH, Sagong M. A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography. *Sci Rep* **11**, 23024 (2021). https://doi.org/10.1038/s41598-021-02479-6

9. World Health Organization. *Global report on diabetes*. World Health Organization, Geneva (2016). https://www.who.int/publications/i/item/9789241565257

10. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**(10027), 1513–1530 (2016). https://doi.org/10.1016/S0140-6736(16)00618-8

11. Suedumrong C, Phongmoo S, Akarajaka T, Leksakul K. Diabetic retinopathy detection using convolutional neural networks with background removal and data augmentation. *Appl Sci* **14**(19), 8823 (2024). https://doi.org/10.3390/app14198823

12. Omer HK. Diabetic retinopathy detection using a bilayered neural network classification model with resubstitution validation. *MethodsX* **12**, 102705 (2024). https://doi.org/10.1016/j.mex.2024.102705

13. Pintelas E, Livieris IE, Pintelas PE. A convolutional autoencoder topology for classification in high-dimensional noisy image datasets. *Sensors (Basel)* **21**(22), 7731 (2021). https://doi.org/10.3390/s21227731

14. Nalepa J, Marcinkiewicz M, Kawulok M. Data augmentation for brain-tumor segmentation: a review. *Front Comput Neurosci* **13**, 83 (2019). https://doi.org/10.3389/fncom.2019.00083

15. Ahmed SF, Alam MSB, Hassan M, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif Intell Rev* **56**, 13521–13617 (2023). https://doi.org/10.1007/s10462-023-10466-8

16. Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. *J Big Data* **9**, 102 (2022). https://doi.org/10.1186/s40537-022-00652-w

17. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ter Haar Romeny B, Zimmerman JB, Zuiderveld K. Adaptive histogram equalization and its variations. *Comput Vis Graph Image Process* **39**(3), 355–368 (1987).

18. Zhang J, Wu C, Yu X, Lei X. A novel DenseNet generative adversarial network for heterogeneous low-light image enhancement. *Front Neurorobot* **15**, 700011 (2021). https://doi.org/10.3389/fnbot.2021.700011

19. Kaggle. Diabetic retinopathy detection dataset. Kaggle Inc. (accessed 2024). https://www.kaggle.com