XL CILAMCE
IBERO-LATIN AMERICAN
CONGRESS ON
COMPUTATIONAL
METHODS IN
ENGINEERING

NOVEMBER
11-14, 2019
Praiamar Natal Hotel & Convention
Natal, RN-BRAZIL

PREDICT STAGE OF DIABETIC RETINOPATHY USING DEEP LEARNING

Matheus Silva Santos
Sérgio Nery Simões
Cassius Zanetti Resende
Daniel Cruz Cavalieri
matheus.s.es@gmail.com
sergionery@gmail.com
cassius@ifes.edu.br
daniel.cavalieri@ifes.edu.br
Federal Institute of Espirito Santo (IFES)
ES-010 Km-6,5 - Manguinhos, 29173-087, Espírito Santo/Serra, Brazil
Gustavo Carreiro Pinasco
Vinícius Araújo Santos
Willer França Fiorotti
Bruno de Freitas Valbon
gustavo.pinasco@emescam.br
dr.vini1984@gmail.com
willerfiorotti@gmail.com
brunovalbon@gmail.com
Escola Superior de Ciências da Santa Casa de Misericórdia (EMESCAM)
Av. Nossa Sra. da Penha - Santa Luíza, 2190, 29045-402, Espírito Santo/Vitória, Brazil

Abstract. The diabetic retinopathy (DR) is one of main complications of diabetes mellitus (DM), presenting in about 40% of diabetics and being the leading cause of blindness in 16 to 64 years old people.The DR is caused by damage to the blood vessels. The diagnosis is given through analysis of images generated by retinoscopy, that is an easy operating electronic device. Despite this, in the Brazilian National Unified Health System (SUS), the access to this type of medical assessment specialized is still poor, resulting in long queues until an ophthalmological consultation. The latest Brazilian Diabetic Society Guideline recommends annual evaluations for DR in diabetic patients, a non-realistic scenario for SUS patients. In view of this, a screening tool that could predict the RD and that could be used by primary care practitioners could be an awesome public health solution. Then, we presenting a deep learning framework (DLF) to classify DR stages. In this paper, we will perform the entire DLF cycle: data collection, data preparation, model creation and validation. The main novelty is the use of siamese convolutional neural network (SCNN), which will receive input pairs of eye fundus images.The purpose of using this set of tools is to use the layers to extract the main characteristics of the inputs of each neural network, so the weights of layers are shared and the similarity degree between neural networks outputs are measure. We train this network using a high-end graphics processor unit (GPU) on the publicly available Messidor dataset and, with this approach, we expect better results in predicting the DR stage when compared to others works.

Keywords: Deep learning; Diabetic retinopathy; Siamese convolutional neural network.

Introduction

According World Health Organization [1], diabetes mellitus is a chronic disease that occurs when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Wild et al. [2] and the World Health Organization (WHO) estimated that the global prevalence of diabetes would increase from 171 million in 2000 to 366 million in 2030. As a chronic disease it is one of the causes of 16 million premature deaths each year. This increase will occur especially due to factors such as population growth and aging, inadequate diets, obesity and physical inactivity, according to Mathers and Loncar [3].

According Avila, Lavinsky and Moreira [4], DM is the most frequent cause of blindness in the industrialized countries among active populations, 2.6% of global blindness can be attributed to diabetes. Bourne et al. [5] report that the most common eye changes that can lead to blindness in DM are: diabetic retinopathy - in 70% of cases - cataract, glaucoma and neuro ophthalmopathy. Visual loss is associated with significant morbidity, including increased falls, hip fracture and a fourfold increase in mortality.

The exam to detect DR is the eye fundus. However, this examination, as well as his analysis, is done only by the ophthalmologist. The Ministry of Health [6] advises that this professional, within the Unified Health System (SUS), is only accessed by the patient when he complains of visual disturbances. On the other hand, damage to ocular vessels indicates a possible onset of other vascular lesions, especially cerebral and renal ones. Thus, early detection of DR can help not only to prevent visual diseases, but also to prevent stroke and renal failure.

Technology is a great opportunity for the disease to have a more efficient diagnosis, higher productivity and lower cost, without the need for the user to rely directly on an expert doctor (ophthalmologist), as the deep learning algorithm will be trained to provide this support to the Family Health Doctor (FHD) in the screening process and referral to the specialist doctor, usually ophthalmologist and / or endocrinologist. Given this scenario, there are already several health-related studies that use deep learning and machine learning techniques in medical imaging to provide a predicted diagnosis. These diagnoses are made through image classification [7] [8], image segmentation [9] [10] or content-based image retrieval (CBMIR) [11] [12] [13] [14] [15] .

However, previous work using the Convolutional Neural Network (CNN) architecture, machine learning techniques for image classification such as: Support Vector Machine (SVM), KNN (k-nearest neighbors) among others, and even Siamese Convolutional Neural Network (SCNN), the architecture we use, yielded results that could be improved [16] [17] [18].

Thus, as mentioned above, we use the SCNN architecture, using the Messidor database, pre-process and prepare the data, as we will see in next sections, and refer to the energy-based models (EBM), which have the objective to capture the dependencies between variables by associating a scalar energy for each variable configuration. This approach can be used to answer questions about unknown variable values given the known variable values. The prediction is to define the observable values and find values of the other variables that minimize the energy between them. The energy function is that the minor energies are correct, that is, they are similar and of the same classification, and the largest energies are incorrect, different and they do not belong to the same category. EBMs provide more flexibility in model architect, so SCNN was based on this concept [19].

The SCNN architecture is composed of two equal convolutional networks, where each network receives an input image and, at the end of both networks, the degree of similarity or energy between the eye fundus images pairs are measured. In this proposal, we will form pairs of eye fundus image of the same and different categories to train the SCNN model, extract the main features and measure the energy or degree of similarity between them. After training, the model will be able to predict the image according to similarity. That is, whether an eye fundus image (unknown variable) have a lower degree of similarity or energy with a known variable (the categorized images from the Messidor dataset).

Experiments demonstrate that the SCNN model approach in this paper obtain the better results then the CNN model with the same database.

2 Related Work

Recently, deep neural networks have been adopted in medical image learning tasks and yield the state-of-the-art performance in many medical imaging problems [11]. Using deep neural networks allows automatic feature extraction and general, expressive representation learning for different computer vision tasks [20].

After Krizhevsky et al.[21] yielded a breakthrough performance using deep convolutional neural network for ImageNet challenge [22], supervised learning with CNN architecture has become a general structure for visual tasks. For medical image, researchers mainly use CNN, stacked autoencoder [23], and restricted Boltzmann machine [24] for different tasks such as classification [7] [8], segmentation [9][10], image generation and synthesis [25][26] and image captioning [27][28].

Machine learning there are approaches by research community for medical tasks. In diabetic Retinopathy there are studies with methods like SVM(Support Vector Machine), KNN(k-nearest neighbors) to detect type of this disease in an optical image based on exudate and microaneurysm image[16]. Classifiers such as Gaussian Mixture Model(GMM), KNN, SVM and AdaBoost are analyzed for classifying retinopathy lesions from non lesions [29].

Other proposed approach elaborates the capability of deep SCNN to reduce the labeling effort by using only binary image pair information, rather than the exact multiclass labeling [18].We will use this approach.

Previous studies for DR fundus images achieved validation accuracy between 52.50% and 57.65% and test set accuracy between 49.75% and 57.25 for 4-ary classification using Messidor dataset[17]. As we use the same data set, we will determine the accuracy for 4-ary classification model and evaluate performance by comparing results to recently published research data.The others metrics as sensibility, specificity, precision, F-Score and confusion matrix will presented for evaluate the model performance.

3 Framework

While probabilistic models assign a normalized probability to every possible configuration of the variables being modeled, energy-based models (EBM) assign an unnormalized energy to those configurations [30][31].

The EBMs use pairs of inputs to perform comparisons and verify their degree of similarity or energy. This means that the lower the energy, the more similar are to the input pairs. In this way, it is possible to perform the classification of the elements referring to the energy value of the EBM output.

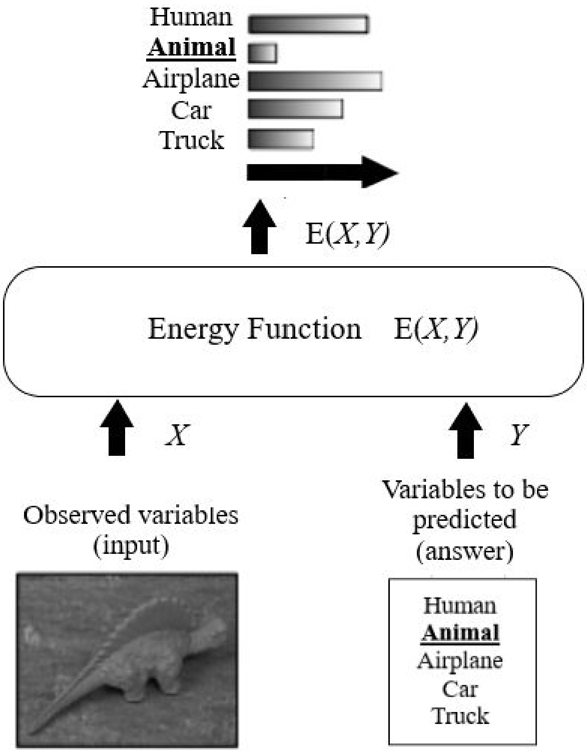


Figure 1. A model measures the compatibility between observed variables X and variables to be predicted Y using an energy function E(Y,X). For example, X could be the pixels of an image, and Y a discrete label describing the object in the image. Given X, the model produces the answer Y that minimizes the energy E [19].

We will use this type of model to perform comparisons between the categories of diabetic retinopathy. We have 4 classifications in the dataset and the model will be generated by genuine and imposities pairs, where the genuine are formed by images of eye fundus within a same classification and the imposite are formed by different categories. Therefore, those having shorter energy are more similar, and the imposities, the different ones, have larger energy. The advantage of EBMs over traditional probabilistic models, particularly generative models, is that there is no need for estimating normalized probability distributions over the input space. The absence of normalization saves us from computing partition functions that may be intractable. It also gives us considerably more freedom to choice architectures' model [31].

The learning process has performed by finding the weight in a way to adequately minimizes the loss function. The energy and loss function can result in zero, which would be a big problem for our model. But our loss function have a contrastive term to avoid that problem and to ensure the normal operation. This case will explained in the section 3.3.

3.1 Eye fundus images verification by similarity metrics

We have used a similarity model to perform resemblance metrics. A open dataset from Messidor that was composed by four categories of diabetic retinopathy[38] have generated the features for training our neural network. Our model was generated through genuine and imposities pairs and the similarity between images was measured based on energy load.

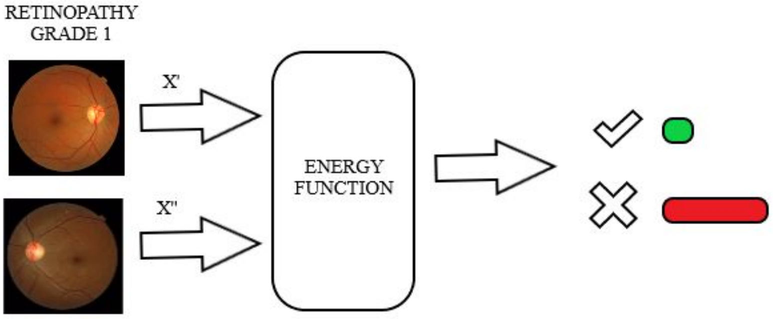


Figure 2. A model measure the similarity metric between input pair in the Retinopathy Grade 1, if the vector output has a shorter energy then it is genuine for that category, else there is a imposite and that is not of the category.

Learning of the similarity metric is realized by training a network that consists of two identical convolutional networks that share the same set of weights - a Siamese Architecture [32] (see Fig. 3).

3.2 Energy function of EBM

Figure 3 shows the architecture of the Siamese Convolutional Neural Network.

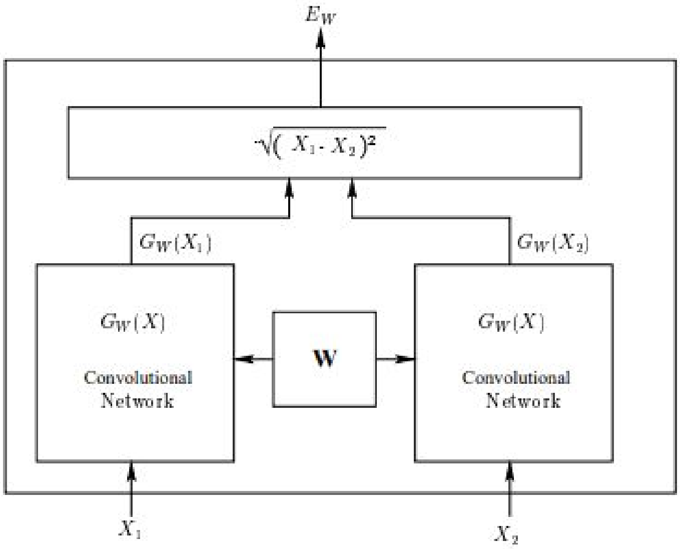


Figure 3. Siamese Architecture [33].

The input pairs for the learning process are X1 and X2. The output is a binary value where Y = 0 are the pairs of eye fundus images of the same category ("genuine pair") and Y = 1 are of different categories ("imposite pair"). The weights are shared between the Siamese networks by mapping pattern of the input pairs and extracting the main characteristics to be compared in the final flow of the model. We use the Euclidean distance to measure the distance between the outputs of each network, given by

$$Ew(X_1,X_2) = \sqrt{}(Gw(X_1) - Gw(X_2))^2 . \tag{1}$$

Considering the principle of equation, the value will be between 0 and 1, due to an application of the "sigmoid" activation function in the output layer of the neural network. Thus, we assume that the energy value for less than 0.5 is genuine, otherwise it is imposite, that is

$$V_{Genuine} = Ew(X_1,X_2) < 0.5 \text{, and} \tag{2}$$

$$V_{Imposite} = Ew(X_1,X_2) > 0.5 . \tag{3}$$

3.3 Contrastive Loss Function

We use the follow loss function:

$$L(W,(Y,X_1,X_2)^i) = L_I(Y(E_W(X_1,X_2)^2) + L_G((1 - Y)\ (Margin - E_W(X_1,X_2))^2), \tag{4}$$

where the W represents the weight of the neural network, Y the target, if it is 0 (genuine) or 1 (imposite), the input pairs are X_1 and X_2, i is the training sample, L represents the loss function where error minimization will occur so that the pair representation works the following way: the energy of the genuine pair becomes shorter and that imposite pair to be larger . The L_I and L_G are the partials errors for genuines and imposities pairs, so that they influences L finds a way of minimizing the error through of the follow behavior: L_I should increase monotonically and the L_G decrease monotonically. In Addition, it is necessary that this function respects 3 conditions that will give this support to find the minimum error value.

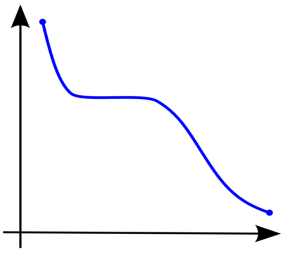


Figure 4. L_G decreasing monotonically

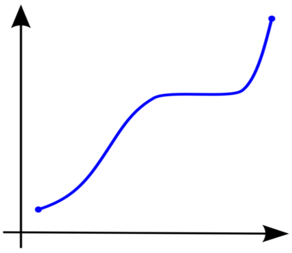


Figure 5. L_I increasing monotonically

Considering the training process of one pair, the genuine with energy Ew^{G}, the other, the imposite with energy Ew^{I}, and H that is the value of the point found on the sample space in the loss function, given by

$$H(E_W{}^{G} ,E_W{}^{I}) = L_G(E_W{}^{G}) + L_I(E_W{}^{I}) . \tag{5}$$

The first condition is that we assume that H is convex at the two inputs in relation to the weight W, and that there is a training weight of the sample satisfying this condition. The other conditions must wait for this process of the loss function H for all $E_W{}^{G}$ and $E_W{}^{I}$ values.

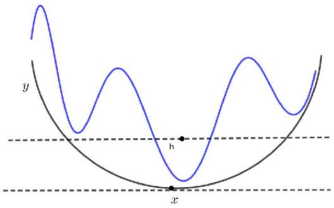


Figure 6. Representation H is in a convex space

The second condition is that the minimum result of $H(E_W{}^{G} ,E_W{}^{I})$ must be within the middle $E_W{}^{G} + Margin < E_W{}^{I}$, which we will call HP_1 and the HP_2 is the another half part, as can be seen in Fig.7. This condition ensures that H will minimize error according to weight W and machine learning is directed to the region that the solution satisfies the first condition.

The third condition is that the negative gradient of $H(E_W{}^{G} ,E_W{}^{I})$ on the margin line $E_W{}^{G} + Margin = E_W{}^{I}$ has a positive point with direction to [-1,1].

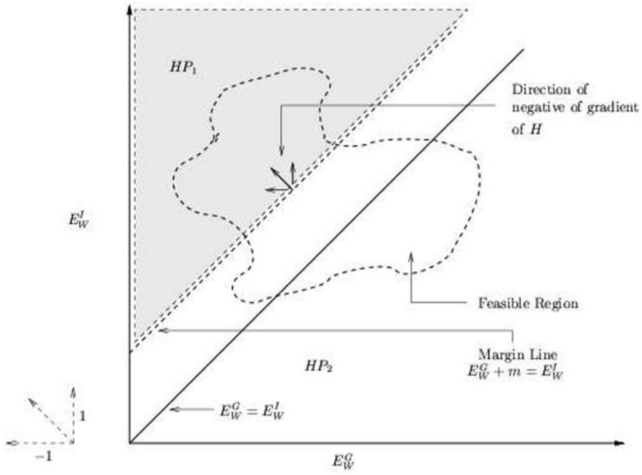


Figure 7. Represents the space where loss function flow act for minimize the error considering all conditions quote above. The feasible area in training process represent the weight values W for each sample [33].

In order to the minimum value of H to be found, which is the best scenario for the model, it is necessary to find at least one point of intersection between the feasible region and HP_1 .This point must be smaller than all points of intersection between the feasible region and HP_2 .

Given that E_G^* the point on the margin line $E_W{}^G + Margin = E_W{}^I$, where H is minimum.

It is worth noting that, as seen in the loss function equation initially, the data of the outputs of the genuine and imposite vectors are normalized by the square norm, so that the values close to zero are eliminated, and thus, the risk of failure on model is mitigated in the cases in which a possible output vector of an imposite be close to zero.

3.4 RMSProp Optimizer Function

RMSprop is adaptive learning rate method. It is as well divides the learning rate by an exponentially decaying average of squared gradients. Hinton suggests γ to be set to 0.9, while a good default value for the learning rate η is 0.001 [34].

RMSProp is Root Mean Square Propagation that attempts to solve learning rates using a moving average of the square gradient. It uses the magnitude of the recent declines to normalize the gradient. This function based on the gradient is very important for the loss function to find the minimum error value in the training. The learning rate is automatically adjusted and you choose a different learning rate for each parameter. This rate is calculated by the average exponential decay of square gradients.

We use RMSProp in the optimization function because it works well with large training steps and it is characteristic of this optimizer to converge quickly.

3.5 Convolutional Networks

The CNN is a feed-forward network with the ability of extracting topological properties from the raw input image. CNNs are invariance to distortions and simple geometric transformations like translation, scaling, rotation and squeezing. They combine three architectural ideas to ensure some degree of shift, scale, and distortion invariance: local receptive fields, shared weights, and spatial or temporal sub-sampling [35].

CNN layers alternate between convolution layers with feature map $C_{k,l}^{\,i}$, given by

$$C_{k,l}^{\,i} \;=\; g(\,I_{\,k,l}^{\,i} \otimes W_{\,k,l} + B_{\,k,l}\,)\,. \tag{5}$$

And nonoverlapping sub-sampling layers with feature map $S_{k,l}^{\,i}$, given by

$$S_{k,l}^{\,i} \;=\; g(\,I_{\,k,l}^{\,i} \downarrow W_{\,k,l} + Eb_{\,k,l}\,), \tag{6}$$

where $g(x) \;=\; \tanh(x)$ is a sigmoidal activation function, B and b are the biases, W and w are the weights, $I_{\,k,l}^{\,i}$ the i'th input and $\downarrow$ the down-sampling symbol. E is a matrix whose elements are all one and denotes a two-dimensional convolution. Note that upper case letters symbolize matrices, while lower case letters present scalars [36].

A convolutional layer is used to extract features from local receptive fields in the preceding layer. In order to extract different types of local features, a convolutional layer is organized in planes of neurons called feature maps which are responsible to detect specific features. A trainable weight is assigned to each connection, but all units of one feature map share the same weights. This feature which allows reducing the number of trainable parameters is called weight sharing technique and is applied in all CNN layers. A reduction of the resolution of the feature maps is performed through the subsampling layers [37].

The convolutional networks are learning processes capable of operating at the pixels level of an image, applying filters, extracting features and recognizing patterns invariant in the space. The neural network is inspired by biological processes, it have a structure similar to the nervous system human .

In order to map the raw images to points in a low dimensional space and hence to realize a learned similarity metric, we use two identical convolutional networks [35] with a common parameter vector (see Fig. 3).

4 Proposed Baseline

In this section, we describe our baseline to predict stage of diabetic retinopathy using deep learning from the pre-processing data until model building and experiments steps .

4.1 Data Acquisition

We use the Messidor public dataset which is provided by Messidor program partners. It consists of 1200 eye fund images in the TIFF format, there are 3 packages with eye fundus images and diagnosis, one per ophthalmologic department. We use 400 eye fundus images and diagnosis of the Service Ophtalmologie Lariboisière department. Those images have dimensions of 1440x960, 2240x1488 and 2340x1536 pixels. They are classified by the criteria adopted through the rules described in the Table 1 [38] :

Table 1. Retinopathy grade [38]

Level	Criteria
0	(µA = 0) AND (H = 0)
1	(0 < µA <= 5) AND (H = 0)
2	((5 < µA < 15) OR (0 < H < 5)) AND (NV = 0)
3	(µA >= 15) OR (H >=5) OR (NV = 1)

µA: number of microaneurysms

H: number of hemorrhages

NV = 1: neovascularization

NV = 0: no neovascularization

4.2 Data Preparation

After obtaining the data, we have performed some checks, such as: data size, data distribution and directories creation for the images of each class. The images were resized to the size of 160 x 160. In the data distribution, there was an imbalance between the classes. The classes 0 and 3 had more data than the other classes, as can be see in Fig.8.

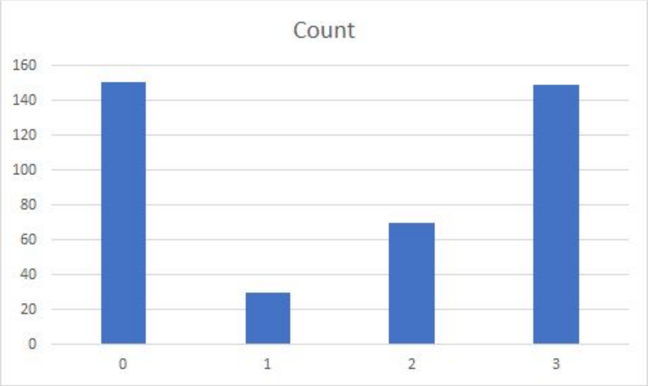


Figure 8. Display the amount of eye fundus images for each category of diabetic retinopathy (0 to 3).

In this way, data balancing was corrected with the data augmentation process, data generating with some transformations: rotate, noise and horizontal flip. Thus, all classes were balanced with the same amount of data, as can be see in Fig.9.

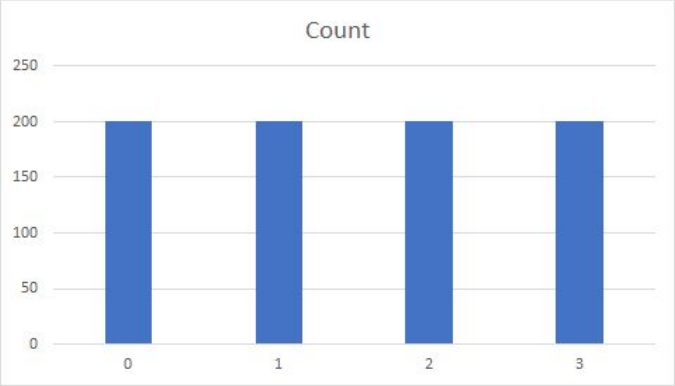


Figure 9. Display the amount of eye fundus images for each category of diabetic retinopathy after data

augmentation process.

After data balance, we have a total of 800 images to create the genuine and imposities pairs, 200 images for each diabetic retinopathy category. This would be a bad scenario for the training of a Convolutional Neural Network, however, the Siamese Convolutional Neural Network architecture does not demand much data for the model training.

We use simple combination to calculate the samples pairs. Synthetically, simple combinations account (C) for the selection of p objects from a set of n objects (with $p \leq n$), where the different orderings of the same objects do not form new possibilities, that is, in simple combinations the order in which the elements are chosen is irrelevant [39]:

$$C(n,p) = n!/p!(n-p)!, \tag{7}$$

Where p contains 2 subsets, genuine and imposite pairs for each Diabetic Retinopathy class in n(200) eye fundus images. That strategy will result in a total of 19,900 samples made up of pairs of eye fundus images (genuine and imposite) randomly selected for each category. The subsets of pairs are within the same category (genuine) and other to be different categories (imposite). Figure 10 shows a sample of genuine and imposite pairs.

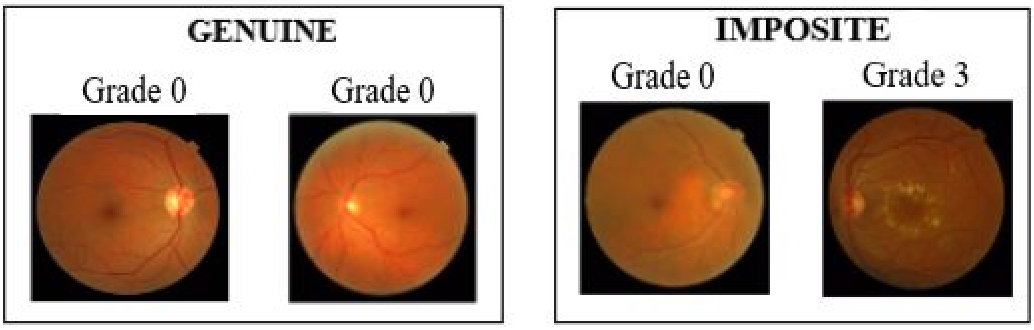


Figure 10. Sample genuine and imposite pairs.

4.3 Deep learning using Siamese Architecture

In this section, we describing the model flow to predict the stage of diabetic retinopathy. The prediction is given by the similarity degree to images pairs processed through the Siamese Convolutional Neural Network.

Siamese nets were first introduced in the early 1990s by Bromley and LeCun to solve signature verification as an image matching problem [40]. As noted at Fig. 11, the siamese neural network consists of twin networks which accept distinct inputs but are joined by an energy function at the top. This function computes some metric between the highest level feature representation on each side [41].

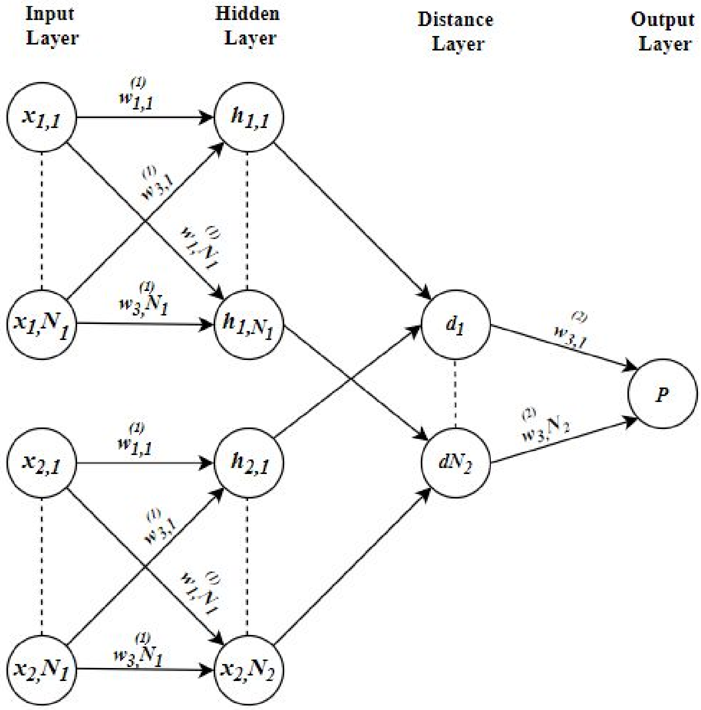


Figure 11. The structure siamese neural network

The Siamese Architecture framework comprises two identical networks and one cost module. The input to the system is a pair of images and a label. The images are passed through the sub-networks, yielding two outputs which are passed to the cost module which produces the scalar energy as discussed in chapter 3. The loss function combines the label with energy. The gradient of the loss function with respect to the parameter vector controlling both subnets is computed using back-propagation. The parameter vector is updated with a stochastic gradient method using the sum of the gradients contributed by the two subnets [33]. The weights are shared between the network, so getting the best model performance.

The deep learning flow receives as input pairs, one eye fundus image for each Convolution Neural Network, as can be see in Fig.12.

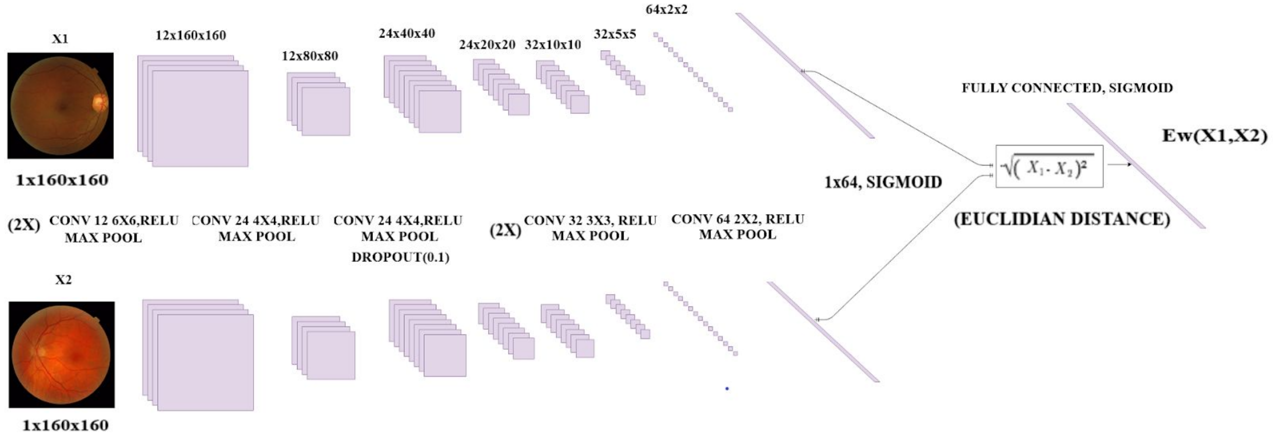
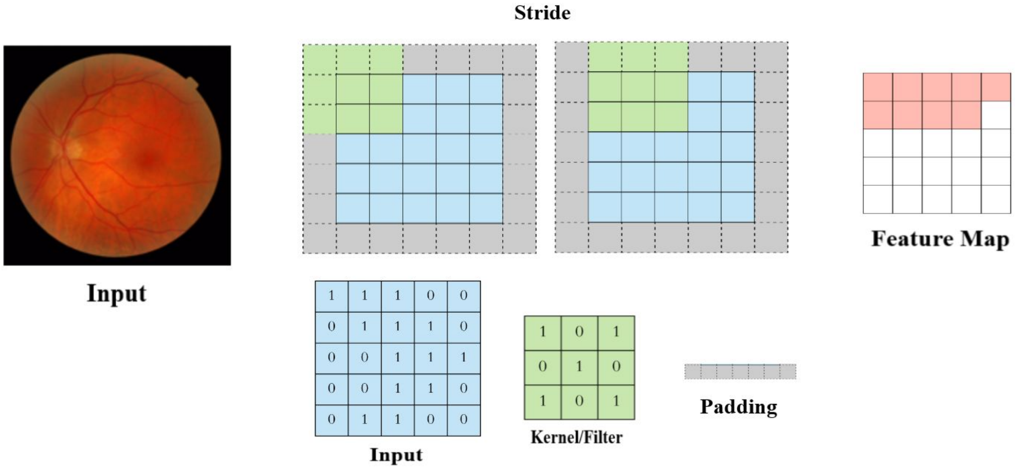


Figure 12. Siamese Network Architecture for predict stage of the retinopathy diabetic.

Convolutional is the main part of CNN, it consists of a input data (eye fundus image), the convolution filter (kernel), and these components generate the features maps through mathematical operations of sum and multiplication of the filter applied on the input data. The convolution filter moves through the matrix at each step, the sliding value is one position (“stride” configuration). In order to the feature map have the same size as the input parameter we use the "same" setting to the "padding" configuration, as can be see in Fig.13. These parameters are in the solution.



1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

1	0	1
0	1	0
1	0	1

Figure 13. Input data and convolutional components to generate feature maps.

For any type of neural network it is necessary has non-linearity. In order to, it is necessary to have the sum of the weights through the activation function. For this, we use the Rectified Linear Activation Unit (RELU) as the activation function, which has the characteristic of returning the calculated value if it is greater than 0, otherwise it returns 0, as can be see in Fig.14.

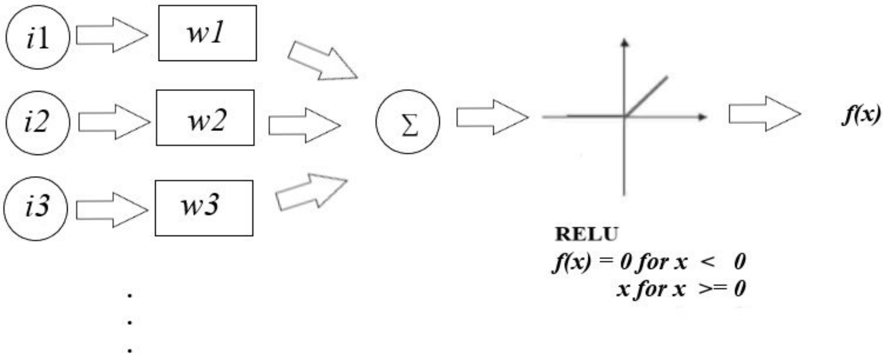
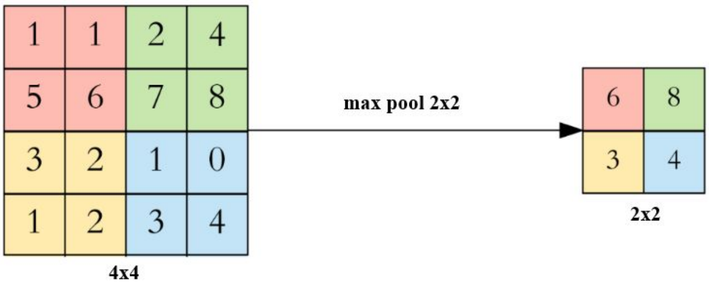


Figure 14. Sample activation function RELU.

After the convolution, we have used pooling to reduce the dimensionality for each feature map. This provides better training performance and avoid overfitting. Although a reduction in size occurs, the pooling maintains the more important characteristics in new shape. We use the value of 2x2 to reduce size by half data. As can be see in Fig.15, the example use the max pooling applying a max filter of 2x2 to reduce the size of the feature map(4x4) to a half size(2x2).



1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

6	8
3	4

Figure 15. Sample max pooling.

Dropout is used to prevent overfitting. During training, at each iteration the neuron is disabled, all the inputs and outputs of this neuron will be disabled (see example at Fig. 16). If a certain amount of neurons are disabled in the current iteration, they are included again in the next iterations and can be disable or not. We did use a 10% dropout rate on the neurons in our model.

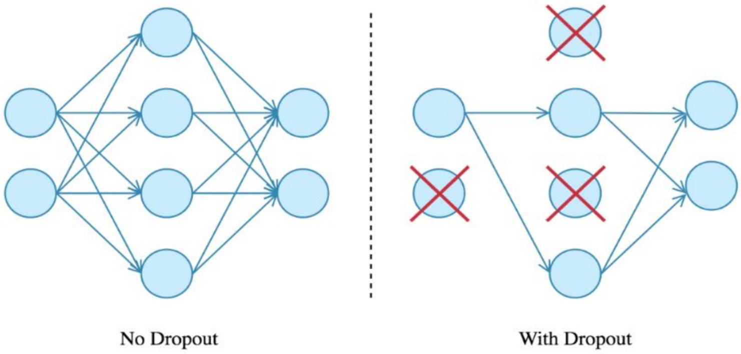


Figure 16. Sample dropout.

The fully connected layer receives the output of the convolution and pooling layers, the last layer of the neural network. The fully connected layer waits for an input vector, so the received data is transformed into vector (flatten). We did use two fully connected layers with the sigmoid activation function, in this way, the output in those layers are between 0 and 1. The first fully connected layer obtains the output data and the second output layer measures the euclidean distance between the outputs of the two neural networks to check the similarity degree them, as we can see in Fig.12.

The Architecture propose is SCNN, we have replicated network neural to the input data pairs. The neural network structure describe below:

- Convolutional 2 dimensions(C1). Feature maps: 12;Size: 160x160; Kernel size: 6x6; Padding: same; Activation:RELU. Fully Connected with the input.
- Max Pooling(P1). Feature maps: 12;Size: 80x80; Field of view: 2x2. Fully Connected with C1.
- Convolutional 2 dimensions(C2). Feature maps: 12;Size: 80x80; Kernel size: 6x6; Padding: same; Activation:RELU. Fully Connected with C1.
- Max Pooling(P2). Feature maps: 12;Size: 40x40; Field of view: 2x2. Fully Connected with C2.
- Convolutional 2 dimensions(C3). Feature maps: 24;Size: 40x40; Kernel size: 4x4; Padding: same; Activation:RELU. Fully Connected with P2.
- Max Pooling(P3). Feature maps: 24;Size: 40x40; Field of view: 2x2. Fully Connected with C2.
- Dropout(D1):10%. Fully Connected with P3.
- Convolutional 2 dimensions(C4). Feature maps: 24;Size: 40x40; Kernel size: 4x4; Padding: same; Activation:RELU. Fully Connected with D1.
- Max Pooling(P4). Feature maps: 24; Size: 20x20; Field of view: 2x2. Fully Connected with C4.
- Convolutional 2 dimensions(C5). Feature maps: 32;Size: 20x20; Kernel size: 3x3; Padding: same; Activation:RELU. Fully Connected with P4.
- Max Pooling(P5). Feature maps: 24; Size: 10x10; Field of view: 2x2. Fully Connected with C5.
- Convolutional 2 dimensions(C6). Feature maps: 32;Size: 10x10; Kernel size: 3x3; Padding: same; Activation:RELU. Fully Connected with P5.
- Max Pooling(P6). Feature maps: 24; Size: 5x5; Field of view: 2x2. Fully Connected with C6.
- Fully Connected Layer(F1). Input=Flatten P6; Number of units:64; Activation:Sigmoid.

- Fully Connected Layer(F2). Input=F1; Number of units:1; Activation:Sigmoid; Output:Distance Euclidean.

4.4 Model Performance

In this section, we analyze at the model performance. We split data for training and testing randomly (70% for training and 30% for testing). The hyperparameters used were batch size: 200, epochs: 100 and verbose: 2. Give this scenario, we extract the following metrics: accuracy, specificity, sensitivity, precision, F-Score, loss function values and confusion matrix.

Table 2. Metrics Model

Metric	Value
Accuracy	0.98
Specificity	0.98
Sensitivity	0.95
Precision	0.97
F-Score	0.97

As noted at Fig. 17, the loss values are decreasing in epochs, the train and test's loss value are in the same way, we did not have overfitting and it had a good generalize to new instance, because of the good performance of test loss value.

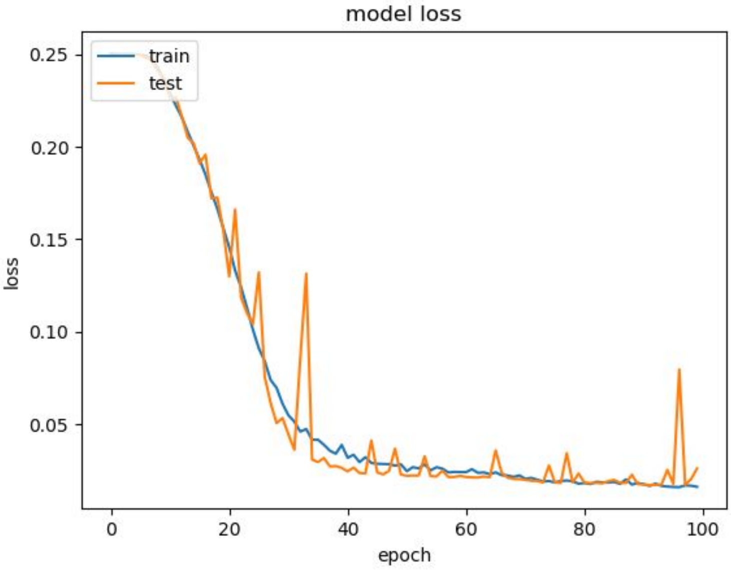


Figure 17. Loss Function

As noted at Fig. 18, confusion matrix has excellent results representing by 5878 True Positives, 62 False Positives, 259 False Negatives and 5741 True Negatives.

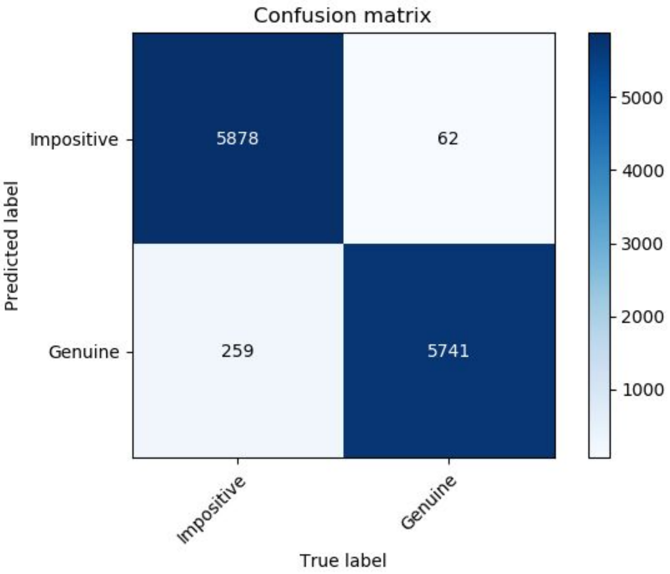


Figure 18. Confusion Matrix

We now present the final comparison results (Table 3). We borrow the baseline results from Lam et al[17] for comparison to our method SCNN. We have used the same Messidor dataset with some differences in preprocessing data step. Thus, we noticed that there was a 44.86% improvement in Validation Accuracy (98%) and 71.18% in Test Set Accuracy (98%) over the best result presented [17]. The performance of the model with the eye fundus images to predict the stage of the diabetic retinopathy demonstrate a good generalization capacity, low bias and low rate error into unknown data.

Table 3. Comparing results from others networks baselines.

Model	Architecture	Solver	Learning Rate	Policy	Validation Accuracy %	Test Set Accuracy %
4-ary	CNN	Adam	1e-4	Step Down	67.65	57.25
4-ary	CNN	SGD	1e-3	Step Down	65.07	55.25
4-ary	CNN	AdaGrad	1e-3	Exponential Decay	66.54	53.25
4-ary	CNN	NAG	1e-3	Step Down	66.18	52.75
4-ary	CNN	RMSProp	1e-4	Step Down	62.50	49.75
4-ary	SCNN	RMSProp	1e-3	Exponential Decay	98	98

Conclusion

In this paper, we present a new approach using the EBM-based SCNN architecture applied to the prediction of the diabetic retinopathy stage, which resulted in excellent metrics, with high accuracy, sensitivity and specificity, being better than the reference study that used the CNN architecture and the same dataset. Our metrics demonstrate low error rate, bias and good generalization. In addition, it can contribute to the leading cause of blindness in recent times by monitoring and preventing DR deploying this deep learning model to support FHD in DR diagnosis at primary care units.

We perform the experiments in fundus eyes images by learning an end-to-end deep SCNN with binary image pair information. This method use twin CNNs that share weights to extract the main features, in this way, the similarity is measure at the outline. For that reason, we compare the performance of our network to an existing state-of-the-art classifier with the same dataset to confirm the initial inference that the twin networks sharing weights could be work better than one CNN.

The limitation on approach is the computational high cost to train the model, however this limitation ends once the network is trained. Finally, for the future, we intend to segment the dataset to perform instance segmentation and apply autoencoder to reduce the size of the images in the pre-processing step.

References

[1] World Health Organization. Diabetes. 2017. [access in 2018 out 1]. Available in: http://www.who.int/news-room/fact-sheets/detail/diabetes

[2] S. Wild, et al. Global Prevalence Of Diabetes: Estimates For The Year 2000 And Projections For 2030. Diabetes Care. vol. 27, n. 5, pp. 1047-1053, 2004.

[3] C.D. Mathers e D. Loncar. Projections Of Global Mortality And Burden Of Disease From 2002 To 2030. Plos Medicine. vol. 3, n. 11, pp. 2011-2030, 2006.

[4] M. Ávila e J. Lavinsky e C.A. Moreira. Conselho Brasileiro de Oftalmologia. Cultura Médica, 2013-2014.

[5] R.R. Bourne, et al. 2018. Causes Of Vision Loss Worldwide, 1990-2010: a Systematic Analysis. Lancet Glob Health. vol. 1, n. 6, pp. 339-349, 2013.

[6] Ministério da Saúde. Estratégias para o cuidado da pessoa com doença crônica: diabetes mellitus. MS, 2013.

[7]A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist level classification of skin cancer with deep neural networks. Nature, 542(7639):115–118, 2017.

[8]V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Journal of the American Medical Association, 316(22):2402–2410, 2016.

[9]M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. C. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. Medical Image Analysis, 35: 18–31, 2017.

[10]Y. Guo, Y. Gao, and D. Shen. Deformable mr prostate segmentation via deep feature learning and sparse patch matching. In Deep Learning for Medical Image Analysis, pages 197–222. 2017.

[11]G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B.van Ginneken,and C.I.Sánchez. A survey on deep learning in medical image analysis. Medical Image Analysis, 42:60–88, 2017.

[12]Q. Sun, Y. Yang, J. Sun, Z. Yang, and J. Zhang. Using deep learning for content-based medical image retrieval. In SPIE Medical Imaging, 2017.

[13]Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan. Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. In SPIE Medical Imaging, 2016.

[14]X. Liu, H. R. Tizhoosh, and J. Kofman. Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In IJCNN, 2016.

[15]A. Shah, S. Conjeti, N. Navab, and A. Katouzian. Deeply learnt hashing forests for content based image retrieval in prostate mr images. In SPIE Medical Imaging, 2016.

[16]S. Aulia, S. Hadiyoso, dan D. N. Ramadan. Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma. J. ELKOMIKA -Teknik Elektro Itenas - ISSN 2338-8323, vol. 3, no. 1, pp. 75–90. 2015.

[17]Lam C, Yi D, Guo M, Lindsey T. Automated Detection of Diabetic Retinopathy using Deep Learning. AMIA Jt Summits Transl Sci Proc. 2018;2017:147–155. Published 2018 May 18.

[18]Chung Y-A, Weng W-H. Learning deep representations of medical
images using Siamese CNNs with application to content-based image
retrieval, arXiv:1711.08490v2.

[19] Y . LeCun, S. Chopra, R. Hadsell, M. Aurelio Ranzato, and F. Jie Huang. A Tutorial on Energy-Based Learning. MIT Press, 2006.

[20]Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8):1798–1828, 2013.

[21]A. Krizhevsky, I. Sutskever, and G. E.Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[22]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009.

[23]J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. Scientific Reports, 6:24454, 2016.

[24]T. Brosch, R. Tam, A. D. N. Initiative, et al. Manifold learning of brain mris by deep learning. In MICCAI, 2013.

[25]D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen. Medical image synthesis with context-aware generative adversarial networks. In MICCAI, 2017.

[26]H. Van Nguyen, K. Zhou, and R. Vemulapalli. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In MICCAI, 2015.

[27]M. Moradi, Y. Guo, Y. Gur, M. Negahdar, and T. Syeda-Mahmood. A cross-modality neural network transform for semi-automatic medical image annotation. In MICCAI, 2016.

[28]H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers. Interleaved text/image deep mining on a very large-scale radiology database. In CVPR, 2015.

[29]S. Roychowdhury, D. D. Koozekanani and K. K. Parhi, "DREAM: Diabetic Retinopathy Analysis Using Machine Learning," in IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 5, pp. 1717-1728, Sept. 2014.

[30] Y . W. Teh, M. Welling, S. Osindero, and G. E. Hinton. Energy-based models for sparse overcomplete representations. Journal of Machine Learning Research, 4:1235–1260, 2003.

[31] Y . LeCun and F. Jie Huang. Loss functions for discriminative training of energy-based models. AI-stats, 2005.

[32] J. Bromley, I. Guyon, Y . LeCun, E. Sackinger, and R. Shah. Signature verification using a siamese time delay neural network. J. Cowan and G. Tesauro (eds) Advances in Neural Information Processing Systems, 1993.

[33] Chopra, S., Hadsell, R., and LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In CVPR, pp. 539–546, Washington, DC, USA, 2005. IEEE Computer Society.

[34] Ruder, S.An overview of gradient descent optimization algorithms* . arXiv preprint arXiv:1609.04747v2, 2017.

[35] Y . LeCun, L. Bottou, Y . Bengio, and P. Haffner. Gradient based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[36]B. Fasel, Robust face analysis using convolutional neural networks, 16th Int. Conf. Pattern Recognition (2002), pp. 4043.

[37]Khalajzadeh, Hurieh & Manthouri, Mohammad & Teshnehlab, Mohammad. (2013). Hierarchical structure based convolutional neural network for face recognition. International Journal of Computational Intelligence and Applications. 12. 10.1142/S1469026813500181.

[38] E. Decenciere, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, et al. Feedback on a publicly distributed image database: the messidor database. volume 33, pages 231–234, 2014.

[39]MORGADO, A. C. de O. et al. Análise combinatória e probabilidade. Impa/vitae, 1991.
[40] Bromley, Jane, Bentz, James W, Bottou, L´eon, Guyon, Isabelle, LeCun, Yann, Moore, Cliff, S¨ackinger, Eduard, and Shah, Roopak. Signature verification using a siamese time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7 (04):669–688, 1993.
[41] G . Koch.Siamese Neural Networks for One-Shot Image Recognition.Master of Science Graduate Department of Computer Science University of Toronto, 2015.