Accurate Early Diabetic Retinopathy detection by Pre-Trained Deep Learning CNNs linked to a rule-based feature classification layer

Fabietti Marcos Ignacio [(1)], Fernández Elmer Andrés [(1, 2*)]

[(1)]Facultad de Ciencias Exactas, Físicas y Naturales,

Universidad Nacional de Córdoba, Argentina

[(2)] CIDIE-CONICET-UCC, Argentina

Abstract

The worldwide incidence of diabetic retinopathy is growing. However, their blindness outcome could be prevented in almost 90% of the cases by early detection, where automatic method may play a crucial role. Deep learning strategies emerges as a candidate solution for this problem, but building such approach from scratch requires high number of training data usually scarce in most eye clinical centers. To overcome this limitation, a hybrid strategy is proposed by concatenating a pre-trained Convolutional Neural Network (CNN) as a feature extractor layer, with another classification method. Three pre-trained CNN and two classification methods were evaluated trained and tested with 1642 and 862 retinal images respectively. It was found that features from the AlexNet CNN concatenated with a support vector machine or k-Top Scoring Pairs method achieves the best results reaching F-Score values of 86% and 83%. Furthermore, the kTSP use less than 1% of the CNN features since it simultaneously performs feature selection and classification providing an easily interpretable rule based output. A diabetic retinopathy diagnostic tool applying this approach is available at https://doi.org/10.5281/zenodo.1296971

Keywords: Convolutional Neural Network, Machine Learning, k-Top Scoring Pairs, Support Vector Machine.

Declarations of interest: none

(*)Corresponding author at:

Tel: +54-(0)351-4938000 int 145
Fax: +54-(0)351-4938081
Mail Address: Universidad Católica de Córdoba, Facultad de Ingeniería, **Campus Universitario,** Avenida Armada Argentina 3555, X5016DHK, Córdoba - Argentina.
Email address: efernandez@bdmg.com.ar

1. Introduction

Diabetic Retinopathy (DR), is the third cause of blindness worldwide. In Latin America, 40% of the diabetic patients have some retinopathy level (Terradella et al. 2016). According to the World Health Organization, the number of people with DR will grow from 126.6 million in 2010 to 191.0 million by 2030 (Zheng et al. 2012). However, if early detected blindness outcome could be prevented in more than 90% of the patients (Ferris, 1993). The DR pathology as defined by The International Clinical Diabetic Retinopathy Scale, establishes that DR can be described by four stages, defined upon a dilated fundus examination (Wilkinson et al 2003). Currently, the main diagnosis strategy relies on a trained ophthalmologist to diagnose onto a digital color fundus photographs of the retina. The clinical grading process consists of detection of the aforementioned eye features, a time consuming task hard to implement as screening process in a busy eye center. Thus, both, the time consuming screening process and the high rates of blindness prevention by early detection, advocates for the advantages of developing automated systems, to facilitate the patient access to screening and early diagnosis.

In the last years the emergence of Deep Learning Artificial Neural Networks (ANN) allowed the development of successful applications in the field of image recognition and classification overcoming several common steps such as handcraft segmentation and complex preprocessing prior steps. Deep learning is a set of representation-learning methods obtained by composing simple modules transforming a representation at one level into a representation at a higher, slightly, more abstract level, has led to the development of very complex neural networks to tackle several problems by means of a new representation of the functional structure of neural units and allowing the efficient training of millions of neural connections (Kumar Srivastava et al. 2015) In the field of image analysis, there are at least three main options to face classification problems: a) to develop and train an ANN from scratch (an approach recently FDA approved)[1], b) to use an ANN built for a different purpose and modify (re-train) the weights by means of a new set of images of the specific purpose (transfer learning classification) or c) use pre-trained ANN as a general feature extraction model and link them to a new layer built from another machine learning method. In the first case, due to the fact that Deep Learning ANNs usually have millions of connection weights, turns necessary to have millions of well annotated images, in the second case, in order to successfully update an existing ANN, the amount of required images are in the order of hundreds of thousands (Soekhoe et al. 2016) . These two situations are usually far from most of the clinical ophthalmology centers. The third case, takes advantage of an ANN successfully trained on huge image data sets for a different purpose, such as those Convolutional Neural Networks (CNNs) trained to solve the image challenge ImageNet Large Scale Visual Recognition Competition (ILSVRC) (Russakovsky et al. 2015) like the well-known AlexNet, GoogLeNet and VGG-16 Net, winners of the ILSVRC.

[1] https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

Here, it is shown that an efficient DR classification method can be built by using an hybrid approach concatenating a pre-trained CNN (pret-CNN), for image feature extraction then feeding a machine learning technique (a Support Vector Machine (SVM) and k–Top Scoring Pairs (kTSP)). This approach can be efficiently trained with much less data, open the chance to make them available for different medical scenarios. The developed method is made freely available through a Matlab-based decision support system application, named DiRekTS (Diabetic Retinopathy Evaluation with kTSP and SVM), which is feed with a fundus image and provides a classification as “Normal” or “Background Diabetic Retinopathy”. We assess the sensitivity, specificity, accuracy and F-score of the model and discuss these with published state-of-the-art models and clinicians’ statistics.

2. Data Description

Four different eye fundus datasets with 2476 early symptoms images were used, two for training (1642 images) and two for validation (862 images). The lesions for our early detection belong to the mild and moderate classification. Microaneurysms, which occur secondary to capillary wall outpouching due to pericyte loss; they appear as small, red dots in the superficial retinal layers. Hard exudates are caused by the breakdown of the blood-retina barrier, allowing leakage of serum proteins, lipids, and protein from the vessels; they appear as minute and yellow. Hemorrhages appear similar to microaneurysms if they are small; they occur as microaneurysms rupture in the deeper layers of the retina, such as the inner nuclear and outer plexiform layers (Watkins PJ. 2003).

2.1 Training Datasets

2.1.1 Messidor (Decencière et al. 2014)

The dataset is publicly available and has been used by other groups for benchmarking performance of algorithms for DR. It consist of 1200 eye fundus color images of the posterior pole. They were acquired by 3 ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with a 45 degree field of view. The images were captured using 8 bits per color plane at 1440*960, 2240*1488 or 2304*1536 pixels. 800 images were acquired with pupil dilation (one drop of Tropicamide at 0.5%) and 400 without dilation. Of the 1200 eye fundus, 546 presented no sign of DR and 654 presented hard exudates being considered as DR class.

2.1.2 E-ophtha (Decencière et al. 2013)

The database of color fundus images especially designed for scientific research in DR. It has been generated from the OPHDIAT© Tele-medical network for DR screening, in the framework of the ANR-TECSAN-TELEOPHTA project funded by the French Research Agency (ANR). Images were obtained with non-mydriatic retinographs: either CR-DGi (Canon, Tokyo) or TRC-NW6S (Topcon, Tokyo) retinographs. Depending on the settings of each retinograph, images with varying sizes were obtained: image sizes ranged from 1440*960 to 2544*1696 pixels. It’s composed

268 non DR and 174 with DR (hard exudates and microaneurysms or small hemorrhages).

2.2 Validation Datasets

2.2.1 DR2 dataset (Pires et al. 2014)

The dataset was captured using a TRC-NW8 retinography with a Nikon D90 camera, creating 12.2 megapixel images, which were then reduced to 867*575 pixels for accelerating computation. Of it 449 images, 300 with no sign of DR and 149 with DR. Images that only had cotton-wool spots were removed, since they belong to a more advanced stage, leaving the set with 103 DR images and 300 healthy ones.

2.2.2 IDRID challenge dataset (Porwal et al. 2018)

The fundus images were captured by a retinal specialist at an Eye Clinic located in Nanded, Maharashtra, India. Images were acquired using a Kowa VX-10 alpha digital fundus camera with 50-degree field of view (FOV), and all are centered near to the macula. The images have a resolution of 4288*2848 pixels and are stored in jpg file format. The size of each image is about 800 KB. The dataset consists of 413 images, 134 with no signs of DR and 279 with them.

3. Method

3.1 Pre-processing

Since images from the different dataset present different aspect ratio, they were preprocessed as follows: First, an Otsu threshold (Otsu 1979) was applied to standardize the color of background, followed by an opening operation. Then, columns and rows without radiograph pixels were removed. In order to standardize aspect ratios, those with 1:1 aspect were cropped to match the rest. Following this, contrast-limited adaptive histogram equalization (Zuiderveld 1994) was applied on small data regions (tiles) to enhance image contrasts. By means of this the histogram of the output region approximately matches the flat histogram set by the ‘Distribution’ parameter, using their default setting. The neighboring tiles were then combined using bilinear interpolation to eliminate artificially induced boundaries, thus resulting in a contrast limitation to avoid amplifying the noise which might be present in the image. Finally, the images were resized according to the evaluated pret-CNN, to 227x227 pixels for AlexNet and 224x224 pixels for both GoogLeNet and VGG-16 Net.

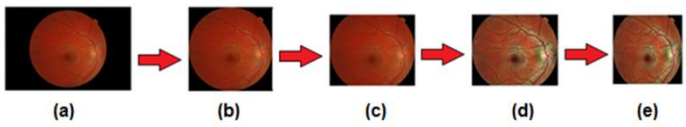
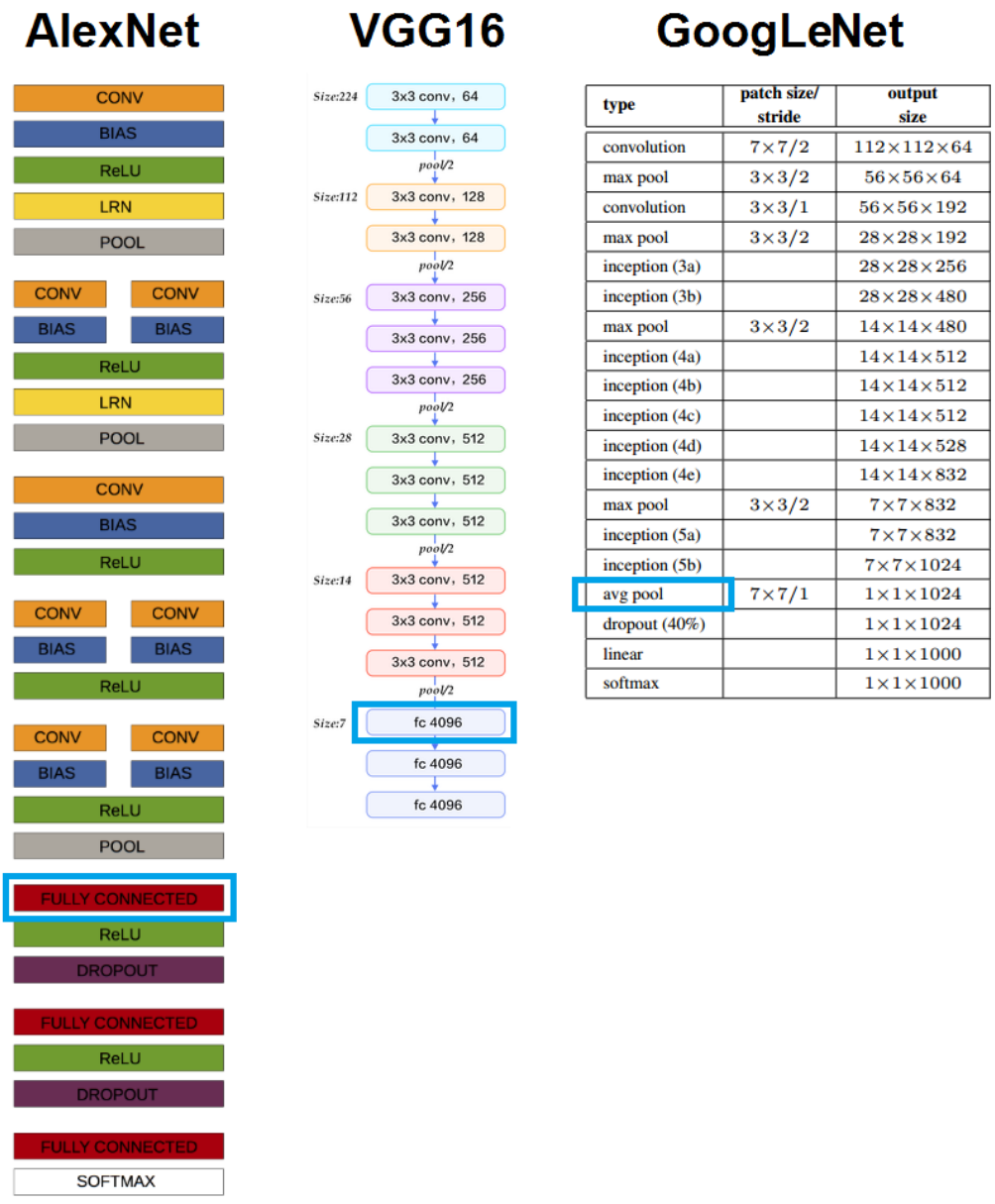


Figure 1. Illustration of the preprocessing and augmentation processes. (a) Original digital color retinal image. (b) Cropped background. (c) Standardized ratio. (d) Local color contrast enhancement. (e) Final retinal image, resized according to each CNN.

3.2 Neural Networks

Three pret-CNN of different sizes and structures, originally trained for multiclass object classification of the ImageNet challenge, were evaluated. They were chosen not only for their renown, but because of their distinctive characteristics. The used features were the outputs of the first fully connected layer. These features were then used to feed the machine learning layer. The CNNs were compared in terms of speed, and feature layer performance on classification of DR images.



type	patch size/ stride	output size
convolution	7×7/2	112×112×64
max pool	3×3/2	56×56×64
convolution	3×3/1	56×56×192
max pool	3×3/2	28×28×192
inception (3a)		28×28×256
inception (3b)		28×28×480
max pool	3×3/2	14×14×480
inception (4a)		14×14×512
inception (4b)		14×14×512
inception (4c)		14×14×512
inception (4d)		14×14×528
inception (4e)		14×14×832
max pool	3×3/2	7×7×832
inception (5a)		7×7×832
inception (5b)		7×7×1024
avg pool	7×7/1	1×1×1024
dropout (40%)		1×1×1024
linear		1×1×1000
softmax		1×1×1000

Figure 2. Architectures of the chosen CNNs, with feature extraction layer outlined in blue

3.2.1 AlexNet

AlexNet was developed in 2012 by Alex Krizhevsky (Krizhevsky et al. 2012), is one of the most famous CNNs today. It won the ImageNet Challenge and it consists of 60 million parameters network organized as shown in Figure 1 left panel. The output of the fully connected layer fc6 were (squared box in left panel of Figure 1) were extracted consisting in 4096 features per image.

3.2.2 GoogLeNet

GoogLeNet was developed at Google Inc., This CNN won the ImageNet challenge in 2014 (Szegedy et al. 2015). It introduced the concept of Inceptions Layers and the possibility of using an ensemble classifier from the three different heads of the net. GoogLeNet consists of 6 million parameters. The output from the pooling layer fcls3_pool (squared layer in mid panel of Figure 1), providing a 1024 feature vector per image.

3.2.3 VGG-16 Net

The third pret-CNN model was VGG-16 Net. It was a runner-up at the ImageNet challenge in 2014 (Simonyan and Zisserman, 2014). It consists of 16 convolutional layers and is very appealing because of its very uniform architecture. It only performs 3×33 times 33×3 convolutions and 2×22 times 22×2 pooling all the way through. VGG-16 Net consists of 138 million parameters, from which we extract the features from the fully connected layer fc6 (squared box at right panel of Figure 1) representing 4096 feature vector per image.

3.3 Classification Layer

Each model was linked to a new machine learning layer feed by the feature vectors of the CNNs. Here, the support vector machine and the k-Top Scoring pairs (kTSP) models were compared in term of classification performance.

3.3.1 Support Vector Machines

Support vector machines are learning machines based on the statistical learning theory developed by Vapnik et al. (1995). In the classification context, SVM provides a way to find Maximal Margin Hyper plane that optimally splits the two classes, giving a specific configuration of the machine. In a general setting, the learning algorithm finds the optimum weight vector that solves the following problem: Given the set of sample $\{(x_i, y_i)\}_{i=1}^{N}\ where\ x_i \in R^p$ is the input pattern for the ith example and $y_i \in \{-1,1\}$ the desired class response, the hyper plane that separate the data should satisfy

$$y_i(w^T x_i + b) \geq 1 + \xi \quad (1)$$

where *b* is the bias and the weight vector $w \in R^p$ and the slack variable ξ allows the incorporation of some errors in the solution. The weight vector *w* can be obtained by the minimization of Lagrangian functional.

$$\Phi(\alpha) = \sum_{i=1}^{N} \alpha_i + 0.5 \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i\, \alpha_j\, y_i y_j {x_i}^T x_j \quad (2)$$

Subject to the constraints

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \ with\ 0 \leq \alpha_i \leq C\ for\ i = 1,2,\ldots,N. \quad (3)$$

Where *C*, the cost parameter, is a user specified positive value. In practice if all *i* < *C*, the data set can be linearly separated without error (all $\xi = 0$) and the problem is said to be a "hard-margin" problem" if not, a "soft-margin" one. The bridge between primal and dual form is given by

$$w = \sum_{i=1}^{N} \alpha_i\, y_i x_i = \sum_{i \in SV} a_i y_i x_i \quad (4)$$

where *SV* is the set of input vectors for which $a_i \neq 0$ (the support vectors). The squared norm of the weight vector can be calculated as

$$\|w\|^2 = \sum_{i=1}^{N} \alpha_i - \frac{\|\alpha\|^2}{C} \quad (5)$$

and the maximum distance between the closest patterns in the hyper-geometrical space R^p is given by

$$2\gamma = \frac{2}{\|w\|^2} \quad (6)$$

In the case of a "hard margin" solution the second term of Eq. 5 vanishes (Cohen and Fernández, 2012).

The SVM model was built using the *fitcsvm* Matlab function, standardizing the data and using a gaussian kernel. Box constraint (the cost parameter) and the kernel scale were optimized with the 'OptimizeHyperparameters' option, which uses Bayesian optimization.

3.3.2 k-Top Scoring Pairs

k–Top Scoring Pairs (Bahman et al .2015) is a classification method for high-throughput data that simultaneously select a small set of features by building pairwise rules. It was originally developed in the bioinformatics field using gene expression data to classify cancer samples. In essence, the kTSP method is a binary classifier which is composed of k elementary classifiers. Each elementary classifier is based upon a pair of input features of 'high predictive power'. This means that feature *"i"* has a prevalently higher value than feature *"j"* in one class and vice versa in the other class. An unknown sample is then classified according to the relative values of features *"i"* and *"j"* in that sample. The kTSP method simply collects the votes from the k best disjoint elementary classifiers and reports the majority vote as its decision. Each feature is allowed to participate in only one elementary classifier, because this makes the decision rule robust against systematic errors in the measurements of a single feature. On the other hand, an "indicator" feature that is very high in one group and very low in the other is most likely to be part of many high performance elementary classifiers.

In the present case, let's consider $p = (4096\ or\ 1024)$ features whose levels $x = (x_1, x_2, \ldots, x_p)$ are provided for the chosen CNN layer for each image and are regarded as random variables. Each feature profile "X_i" has a true class label in *{1, 2, ..., c}*. For simplicity, we assume *c = 2*, although the results extend to higher numbers of classes.

We focus on detecting “marker feature pairs” *(i, j)* for which there is a significant difference in the probability of $x_i < x_j$ from class 1 to class 2. Profile classification is then based on this collection of distinguished pairs. Here, the quantities of interest are $p_{i,j}(\mathrm{c}) = \mathrm{P}(x_i < x_j|\mathrm{c})$, *c = 1, 2, i.e.*, the probabilities of observing $x_i < x_j$ in each class. These probabilities are estimated by the relative frequencies of occurrences of $x_i < x_j$ within profiles and over experiments. Consequently, it is sufficient to know the ranks of the feature values within profiles on each image. The kTSP decision is based on k image features pairs, denoted $\Theta = \{(i_1, j_1), \ldots., (i_j, j_k)\}$. The particular decision rule using the k comparisons $x_{i_l} < x_{j_l}$ simply determined by the aggregate vote statistic $k = \sum_{l=1}^{k} I(x_{i_l} < x_{j_l})$ where *I* is the logical indicator function. The kTSP classification decision is based on thresholding, i.e. $\hat{y} = I\{k > \tau\}$ provided the labels $y \in \{-1,1\}$. The standard threshold is $\tau = k/2$ equivalent to majority voting.

The library SwitchBox was used in the R language. A Wilcoxon signed-rank test was applied to the pre filter features, and a range from 2 to 25 possible pairs was used. The *SWAP.Filter.Wilcoxon* function takes the phenotype factor, the predictor data, the number of feature to be returned, and a logical value to decide whether to include equal number of featured positively and negatively associated with the classes to be predicted.

3.4 Hardware and Software

The script was made in Matlab 2017a, using the AlexNet and VGG-16 Net Plug Ins and for GoogLeNet the MatCovNet (Vedaldi and Lenc, 2014) library was used, which required the compiler was changed to Microsoft Visual C++ 2013 Professional. The OS was Windows 7. k-Top Scoring Pairs was run on Rstudio and it is part of the ‘SwitchBox’ package which is freely available from Bioconductor. The hardware used was an 8 gigabyte RAM and Intel® Core™ i7- 6500U CPU @ 2.50GHz processor on an Acer TravelMate P278-MG.

3.5 Statistical analysis

All the images where processed by each CNN and stored for later recall. For each CNN the image-based processing mean time was measured and compared. The extracted CNN image features of the training data sets were used to optimize the parameters of SVM with gaussian kernel and to optimize the number of rules (used pairs features) of kTSP by means of 10 fold cross validation strategy using accuracy as the optimization goal. The best models for each CNN features set was then evaluated over the validation sets by means of Accuracy, Sensitivity, Specificity and F-score, as suggested by Fernandez et al. (2005).

4. Results

Performance values for each CNN-Classification method are presented in Table 1. It can be seen that GoogLeNet was the fastest to present the image features in the fully connected layer (3.87 sec. on average), followed by AlexNet and VGG-16 Net being 3.6 and 74.32 times slower respectively than GoogLeNet.

For the SVM models, 86%, 86% and 90% of the training samples were required as support vectors for AlexNet, GoogLeNet and VGG-16 Net respectively. However, for Google net 66% of them were errors support vectors (slack variables $\xi > 0$) meanwhile the others two less than 32% yield $\xi < 0$. This suggest that the last ones provides more separable features than the GoogLeNet, and that feature reduction in the fully connected layers does not prove beneficial for the feature extraction approach (and why it may have underperformed to VGG-16 Net in ImageNet challenge in 2014).

In the case of kTSP classification layer, all the methods reach almost the same number of optimal rules, being 18, 17, and 20 for AlexNet, GoogLeNet and VGG-16 Net respectively. When comparing the whole system, i.e. the pret-CNN-(SVM or kTSP) classification method, the one using AlexNet, with 4 time more features than the GoogLeNet, yields better classification performance for both tested classifiers.

Both classification methods performs very similar at the classification layer, being the SVM just 2.5% on average more accurate with almost same F-score (80.02 and 80.04 for SVM and kTSP respectively). The best performance for both classification layer were achieved by using the AlexNet features with F-scores of 85.77% for SVM and 82.12% for kTSP. However, the later achieves such performance by using only 36 features (18 rules) against 4096 features used by the SVM. In addition, the kTSP rules behavior can be plotted as the difference between the features involved in each rule, providing a visual interpretation of the kTSP decision function as bar plots

ML	CNN	VD	F (%)	Acc (%)	Se(%)	Sp(%)	#SV(ξ > 0) or Rules	IPS (sec.)
SVM	Anet	IDRID	82.37	79.98	93.55	66.42		14.19
		DR2	84.13	83.89	85.44	82.33		
		Total	85.77	84.89	91.05	78.73	1428 (69%)	
	Gnet	IDRID	71.55	72.07	70.25	73.88		3.87
		DR2	76.45	77.57	72.82	82.33		
		Total	74.2	75.33	70.94	79.72	1427 (34%)	
	Vnet	IDRID	82.84	82.4	84.95	79.85		287.65
		DR2	79.78	77.86	87.38	68.33		
		Total	80.11	78.75	85.6	71.89	1497 (66%)	
kTSP	Anet	IDRID	78.62	74.37	94.27	54.48		14.19
		DR2	82.53	80.27	93.2	67.33		
		Total	82.12	79.76	92.93	66.59	18	
	Gnet	IDRID	79.22	76.12	91.04	61.19		3.87
		DR2	81.75	79.63	91.26	68		
		Total	80.91	78.5	91.1	65.9	17	
	Vnet	IDRID	81.83	80.27	88.89	71.64		287.65
		DR2	76.82	71.59	94.17	49		
		Total	77.09	73.15	90.31	55.99	20	

Table 1. Results of the machine learning methods. From left to right: machine learning (ML), support vector machine (SVM), k–Top Scoring Pairs (kTSP). Convolutional Neural Network (CNN), Alexnet (Anet), GoogLeNet (Gnet), VGG-16 Net (Vnet). Validation dataset (VD), F-score (F%), accuracy(Acc%), sensitivity(Se%), specificity (Sp%), number of support vectors (percentage with ξ > 0) or number of kTSP Rules (#SV(ξ > 0) or Rules), image processing speed (IPS).

4.3 Software Application

A MATLAB GUI tool (Figure 3) was developed implementing the pret-CNN and both classification layers. It was named DiRekTS (Diabetic Retinopathy Evaluation with kTSP and SVM), the software incorporates both the SVM model and kTSP pairs for AlexNet features. The Software allows loading a fundus image, which is displayed on the main panel which is then processed by the AlexNet and by both classification layers providing the positive class posterior probabilities given the SVM and the graphical representation of kTSP appears on screen, with the positive/negative rule ratio. An example for processing an image is given in Supplementary material.
The application is available at https://doi.org/10.5281/zenodo.1296971 .

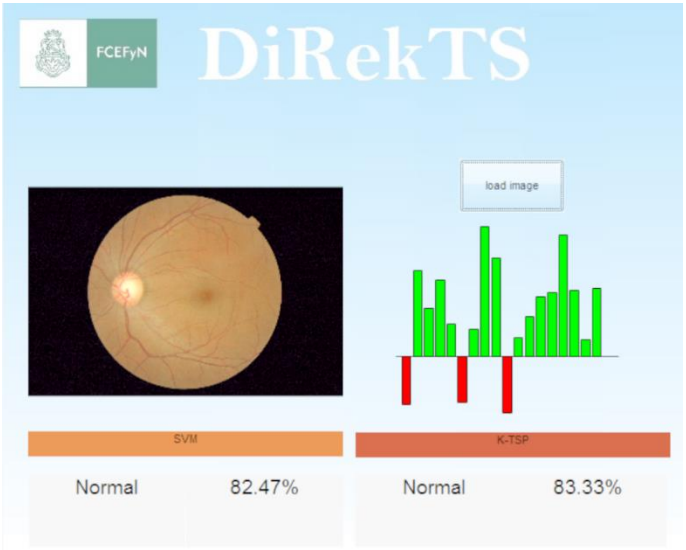


Figure 3. User interface of DiRekTS

5. Discussion and Conclusions

Here it is shown a versatile deep learning based classification approach for low sample availability, commonly found in most of the eye centers. The method has high performance compared to other already available on literature. We end with two competitive proposed models with Sensitivity higher than 91% and F-scores higher than 82%. The proposed approach is very competitive (Table 2) against other approaches that build CNN from scratch using large datasets as well as image re-annotation by ophthalmologists, and is the only one available for public use. It is worth to mention that a screening DR method should be both high sensitive and their prediction should also be highly confident i.e. high positive predictability (Fernandez et al. 2005). Thus the F-score, measuring both sensitivity and positive predictability, is a better performance score to take into account since incorporate these two good characteristics for medical decision making support systems like the highly sensitive and predictive method proposed here. In addition, to account for models with small number of parameters like the proposed kTSP model using less that 1% of all the available image features are also

a good characteristic under the information theory parsimonia concept (MacKay 2002) regarding models with less parameters but same classification power. The kTSP method, derived from the bioinformatics classification of cancer samples based on gene expression resulted to be highly competitive in this scenario. Their appropriateness relay on the fact that CNN extracted features share similar distributional characteristics that made plausible their pairwise comparison, being this a new approach for this field. Furthermore, the bar plot representation of the kTSP rules provides a visual guide allowing evaluating the decision rules, a desirable characteristics of medical diagnosis image systems unlike other deep learning approaches.

Author	VD	Se%	Sp%	Acc%	F%	Oph
trained ophthalmologist (Sundling et al. 2013)	Private	67	84	75	-	
Peng et al. (2016)	Eyepacs	97.50	93.40	-	-	yes
	Messidor	96.10	93.90	-	-	
Abramoff et al. (2016)	Private	96.80	87.00	-	-	yes
AlexNet SVM	DR2/IDRID	91.05	78.73	84.89	85.77	no
AlexNet kTSP		92.93	66.59	79.79	82.12	
Lopez et. al. (2017)	Kaggle	54.47	93.65	83.68	-	no
Pratt et al. (2016)	Kaggle	30.00	95.00	75.00	-	no
Rakhlin (2017)	Kaggle	92.00	72.00	-		no
	Messidor	99.00	71.00	-		no

Table 2. Performance comparison. From left to right: author, support vector machine (SVM), k–Top Scoring Pairs (kTSP). Validation dataset (VD), sensitivity (Se%), specificity (Sp%), Accuracy(Acc%), F-score (F%) , ophthalmologist re-validation (Oph).

6. References

Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M. (2016); Improved Automated Detection of Diabetic Retinopathy on a Publicly

Available Dataset Through Integration of Deep Learning. Invest. Ophthalmol. Vis. Sci.; 57(13):5200-5206.
Bahman, A., Fertig, E., Donald, G., Luigi, M. (2015). “switchBox: an R package for k-Top Scoring Pairs classifier development.” Bionformatics, 31, 273–274.
Cohen, D.A., & Fernández, E.A. (2012). SVMTOCP: A Binary Tree Base SVM Approach through Optimal Multi-class Binarization. CIARP.
[Dataset] Decencière, E. et al. (2013). TeleOphta: Machine learning and image processing methods for teleophthalmology. IRBM
[Dataset] Decencière, E. et al. (2014). Feedback on a publicly distributed database: the Messidor database. Image Analysis & Stereology, v. 33, n. 3, p. 231-234. ISSN 1854-5165
Fernández, E.A., Valtuille, R., Presedo, J., Willshaw, P. (2005). "Comparison of different methods for hemodyalisis evaluations by means of ROC curves: from artificial intelligence to current methods"; Clinical Nephrology - 64(3) pp 205-213
Ferris, FL. How effective are treatments for diabetic retinopathy? (1993) JAMA. 269:1290–1291.
Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In: Proc Adv Neural Inform Process Syst. Vol. 25. Granada, Spain, pp. 1097–1105.
Kumar Srivastava, R., Greff, K., Schmidhuber, J. (2015). Training Very Deep Networks. 2015 Neural Information Processing Systems (NIPS 2015 Spotlight).
López, J., García, G., Gallardo, J., Mauricio, A., Carpio, C.D. (2017). Detection of Diabetic Retinopathy Based on a Convolutional Neural Network Using Retinal Fundus Images. ICANN.
MacKay, D. J. C. (2002). Information Theory, Inference & Learning Algorithms. Cambridge University Press, New York, NY, USA.
Otsu, N. (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9 (1): 62–66.
Peng, L., Gulshan, V., Coram, M. et al (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 316(22):2402–2410.
[Dataset] Pires, R., Jelinek, H. F., Wainer, J., Valle, E., and Rocha, A. (2014), Advancing bag-of-visual-words representations for lesion classification in retinal images,” PLoS One 9(6), e96814.
[Dataset] Porwal, P., Pachade, S., Kamble, R., Kokare, M., Girish, D., Sahasrabuddhe, V. and Meriaudeau, F. (2018). Indian Diabetic Retinopathy Image Dataset (IDRiD). IEEE Dataport.
Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., & Zheng, Y. (2016). Convolutional Neural Networks for Diabetic Retinopathy. MIUA.
Rakhlin, A., November 2017. Diabetic Retinopathy detection through integration of Deep Learning classification framework. bioRxiv 225508.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211-252.
Simonyan, K. and Zisserman, A. (2014) "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv: 1409.1556.

Soekhoe D., van der Putten P., Plaat A. (2016) On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks. In: Boström H., Knobbe A., Soares C., Papapetrou P. (eds) Advances in Intelligent Data Analysis XV. IDA 2016. Lecture Notes in Computer Science, vol 9897. Springer, Cham.

Sundling, V., Gulbrandsen, P., Straand, J. (2013) Sensitivity and specificity of Norwegian optometrists' evaluation of diabetic retinopathy in single-field retinal images - a cross-sectional experimental study. BMC health services research.13:17.

Szegedy, C. et al., "Going deeper with convolutions" (2015) IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9.

Terradella, J., Martínez Castro, F., Barría von-Bischhoshausen, F. (2016), Asociación Panamericana de Oftalmología PAAO, Actualización de la Guía clínica de Retinopatía Diabética para Latinoamérica

Vapnik, V. N. (1995). The nature of statistical learning theory. New York, NY: Springer.

Vedaldi, A. and Lenc, K. (2014). MatConvNet - Convolutional Neural Networks for MATLAB.

Watkins, PJ. (2003) Retinopathy. BMJ : British Medical Journal; 326(7395):924-926.

Wilkinson, C.P., Ferris, F.L., Klein, R.E., Lee, P.P., Agardh, C.D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., Verdaguer, J.T.(2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology.110:1677–1682

Zheng, Y., He, M., Congdon, N. (2012) The worldwide epidemic of diabetic retinopathy. Indian Journal of Ophthalmology. 60(5):428-431.

Zuiderveld, K. (1994). Contrast Limited Adaptive Histogram Equalization. Graphics Gems IV. Academic Press Professional, Inc., San Diego, CA, 1994