

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING

Eduardo Pérez-Careta¹, José-Rafael Guzmán-Sepúlveda², José-Merced Lozano-García¹, Miguel-Torres Cisneros¹ y Rafael Guzmán-Cabrera¹

¹ University of Guanajuato (México)

² CINVESTAV Unidad Monterrey (México)

DOI: <http://dx.doi.org/10.6036/10117>

1. INTRODUCTION

In recent years, the use of medical imaging has become a powerful tool to aid medical diagnosis. Thanks to the availability of advanced scanners and image reconstruction software, it is possible to clearly visualize organs and tissues, as well as to obtain information that helps to characterize and quantify pathologies. The use of computer systems in the improvement of medical diagnosis is increasingly used in various areas of medicine [1]. The purpose is to provide medical experts with assistance, for example, in the interpretation of medical images to improve diagnosis; in other cases, facilitating evaluations in places where medical assistance is limited. In recent years the use of artificial intelligence (AI) tools, such as deep learning (DL), has been incorporated into the development of these systems, and they promise excellent performance [2].

In this context, image processing is widely used. For example, in [3] image processing based on texture segmentation was used to extract regions of interest potentially related to the medical diagnosis of Parkinson's. In another example with a similar idea [4], image processing, pattern recognition, and AI were used simultaneously to detect microcalcifications in mammograms. Similarly, in [5] an approach was presented uses computer-assisted mammography images to improve medical diagnosis.

Image classification refers to the task of categorizing or labeling an image based on the characteristics of its content. There are two types of classification: supervised and unsupervised. Supervised classification starts from a known set of classes, which are defined based on previous knowledge of the set of variables. On the other hand, in unsupervised classification, the classes are not established a priori; only the expected number of classes is established, and the definition of the classes is done automatically using statistical procedures. There exist in the literature many works that carry out the automatic classification of images, for example, in [6] the classification characteristics are extracted based on texture segmentation, and then standard databases were used for evaluation. In [7] they carry out the classification of images with breast cancer using thermal images.

In this work, we present a DL-based methodology to classify medical images corresponding to three conditions: tuberculosis (TB), glaucoma, and Parkinson's. The proposed methodology is based on the use of the convolutional neural network (CNN) for the extraction of the classification characteristics. Then, three standard classification methods are used in two different scenarios. The performance of the classifier is evaluated quantitatively using three evaluation metrics, which allow evaluating the feasibility of the proposed methodology. Based on the results obtained, the presented methodology can serve to strengthen the medical diagnosis of the aforementioned diseases. The rest of the paper is organized as follows: in section two, the methodology implemented is presented, beginning with the preprocessing stage and the extraction of the characteristics of the input image, which will be used as classification attributes. In section three, the experimental evaluation is presented, the databases used are described, as well as the results obtained in the different classification scenarios implemented. Finally, the conclusions and future work are presented.

2. METHODOLOGY

The main objective of this work is the classification of images, as well as the identification of the best conditions for said classification. That is, finding the most appropriate scenario and classification method. Fig. 1 shows the schematic diagram of the proposed methodology. Three databases were used, one for each disease, which are manually tagged by experts in the field. These tagged images were used to train the system; the specific details of each database are given below.

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

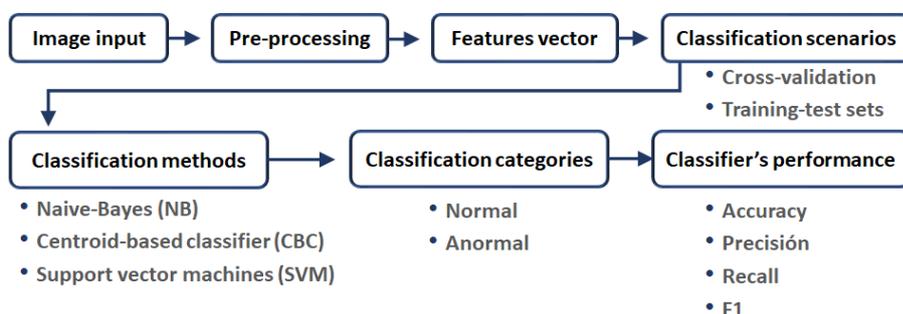


Figure 1: Schematic diagram of the stages comprising the proposed methodology.

The first stage of the proposed methodology is the preprocessing. This stage is of critical importance to be able to carry out the classification since the features need to be extracted from images with similar structure. Fig. 2 shows the schematic diagram of the preprocessing.

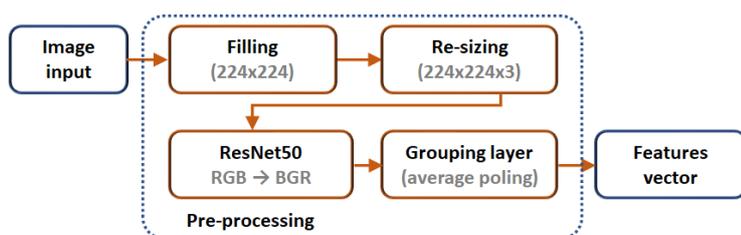


Figure 2: Preprocessing diagram to obtain the vector of characteristics.

We used the 50 layer residual network ResNet-50 to obtain the characteristic space of the input data, that is, a matrix consisting of vectors of characteristics. This network is based on residual learning, which tries to solve more complex tasks and improve the classification accuracy, two of the main problems in image classification. In residual learning, a residual can be simply understood as the subtraction of a feature learned from an input of a layer ResNet-50 operates based on a DL approach consisting of a machine learning (ML) technique, which is trained through the use of extensive labeled data sets and CNN architectures that automatically learn from the data.

For better results, it is recommended that the images fed into ResNet-50 have dimensions larger than 187 pixels and the pixels span their entire dynamic range. This stage consists of resizing via zero-padding followed by pixel scaling and then, if necessary, by layer replication. Basically, in the zero-padding operation, zeros are added to the perimeter of the original image until the desired dimensions are reached. After this operation, each layer of the image has dimensions of 224x224 pixels. Then, the pixel values of each layer were rescaled to the range of 0 to 255. Finally, it is ensured that the image has three layers (RGB), leaving a new matrix with dimensions of 224x224x3. For images that originally had a single layer, the same layer replicated in all channels. The image channels are arranged according to the requirements of the ResNet-50 network (BGR), as indicated in Fig. 2. Once the image has been properly prepared, the image enters ResNet-50 and the corresponding mean grouping layer is calculated, as described below.

Fig. 3 shows the compact architecture of ResNet-50 and its convolutional evolution. The input image is subjected to consecutive stages of convolution and sub-sampling, until reaching the middle grouping layer. In this layer, the characteristics are extracted in the form of a vector of dimensions 1x1x2048 [8]. This vector is generated for each image in the database and contains the general characteristics of the image extracted by the model, such as intensity, luminosity, saturation, among others.

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

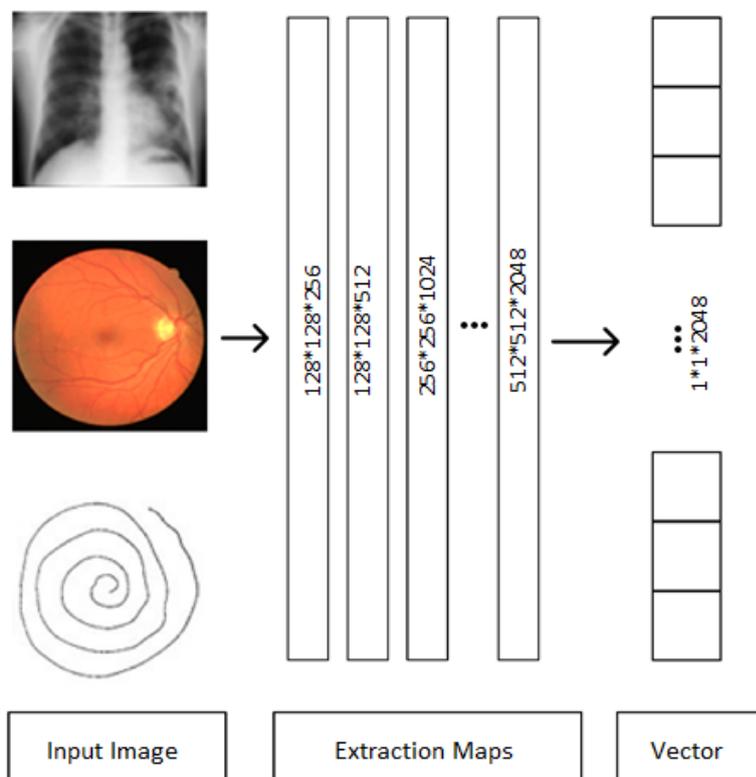


Figure 3: Architecture ResNet 50 and the convolutional evolution process.

3 EXPERIMENTAL EVALUATION

Experiment setup. As mentioned above, for each database used in the present work, two classification scenarios were implemented: 1) cross-validation (CV) and 2) training and test sets (TTS). In the case of CV, 100% of the images were used, and 10 divisions were built to carry out the validation (ten-fold CV). The same procedure was carried out in the three cases studied. The results presented correspond to the meta-average of the evaluation metrics. In the case of TTS, 80% of the images was incorporated into the training set and 20% to the test set. The same proportion was kept in all cases. It should be mentioned that this classification scenario is the most desirable when there are enough images to form both sets, since the images of the test set are never seen by the training set.

In both classification scenarios, three classification methods widely known in the state of the art were used: Support vector machines (SVM) find the optimal separating hyperplane that maximizes the margin of separation in order to minimize the risk of misclassification for both the training samples and the unseen test set. Naïve Bayes (NB) is a practical, successful ML algorithm that, based on the assumption of feature independence, allows calculating the likelihood that a test example has certain feature values. Centroid-based classifier (CBC) considers the similarity to the centroid of each class, where the similarity measure is calculated as the ones' complement of the Euclidean distance and the centroid is the vector of the average frequencies of each term across elements of a specific class.

In line with standard metrics used in data mining and information retrieval, we report the average precision (P) and recall (R) based on true positives (TP), false positives (FP), and false negatives (FN), as well as the accuracy (number of correctly predicted instances/number of instances) to evaluate the performance of the predictive models:

$$P = TP / (TP + FP) \quad (1)$$

$$R = TP / (TP + FN) \quad (2)$$

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

3.1 TUBERCULOSIS

According to the World Health Organization (WHO), TB is one of the ten leading causes of death in the world [9]. In 2018, 10 million people became ill with TB, of which 1.5 million died from the disease; among them, 251,000 people had HIV. More than 95% of TB deaths occur in third-world countries. Two tests can be performed to diagnose TB, one is by culture of a sample of sputum (phlegm that is expelled from the lungs by coughing), and the other is by chest radiographs.

Description of the database. The Montgomery database, which is available, was used. The X-ray images in this database were acquired from the TB control program of the Montgomery County, Maryland, Department of Health and Human Services in the United States. This set contains 138 posterior-anterior radiographs, of which 80 radiographs are normal and 58 radiographs are abnormal with manifestations of TB. The images in this database were manually tagged. Fig. 4 (left) shows two sample images from the database corresponds to a healthy person (bottom) and a person with TB (top).

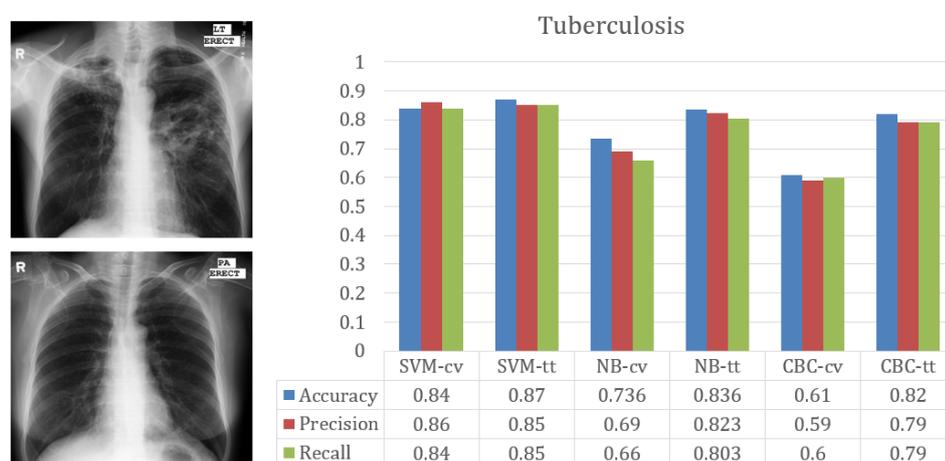


Figure 4: Left: Sample images from the Montgomery database, the top image persecutes TB while the bottom image corresponds to a healthy person. Right: Classification results of tuberculosis images using Cross Validation and Training-Test Sets.

Fig. 4 (right) shows the results obtained. In the case of CV, the best-performing classification method was SVM, with 86% precision and a recall of 84%, well above chance (50%), while the worst result was found by CBC with a 59% precision and 60% recall. For TTS, the best-performing classification method was SVM with 85% accuracy and recall while the worst result was NB with 69% accuracy and 66% recall.

3.2 Glaucoma

Glaucoma is a progressive optic neuropathy caused primarily by increased intraocular pressure in the eye and characterized by structural changes in the optic nerve. Unfortunately, there is still no cure for glaucoma. Furthermore, in most cases glaucoma is detected after loss of vision [10]. The increased pressure of the fluid in the area between the cornea and the iris, called the aqueous humor, causes damage to the optic nerve with a characteristic change in the appearance of the optic disc and visual field. Injury to the optic nerve results in loss of the visual field, while blindness appears when the latter is atrophied [11]. This disease can affect at all ages. It can be hereditary or caused by some specific issues, such as the use of steroids in eye drops, diabetes, an eye injury or surgery. The most common form of this disorder is primary open angle glaucoma. It is a disease characterized by the absence of symptoms in early stages and tends to be more frequent with age. Hence the importance of routine examinations starting at the fifth decade, particularly if there is a family history or another risk factor.

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

Description of the database. The database used was the High-Resolution Fundus (HRF) image database [11], which is publicly available. The database contains 15 images from healthy patients, and 15 images from glaucomatous. Standard vessel segmentation images are available for each image. Fig. 5 (left) shows sample images from the database from a person with glaucoma (top) and a healthy person (bottom).

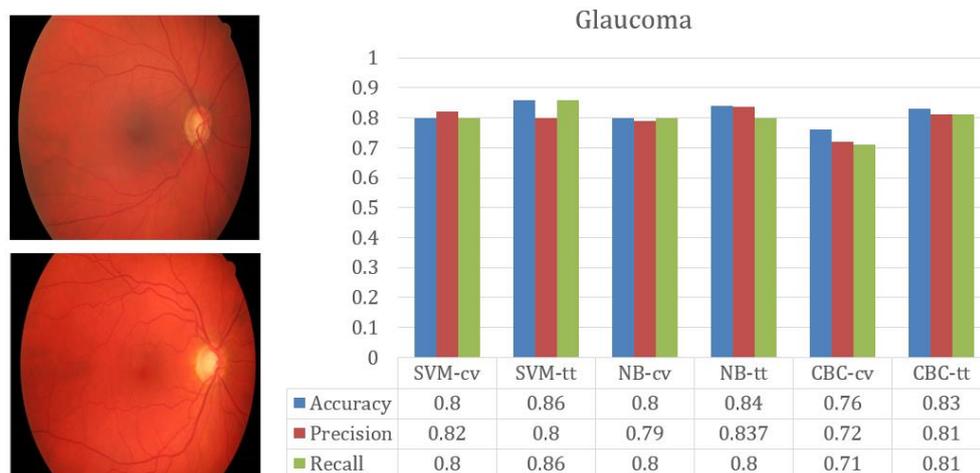


Figure 5: Left: Sample images from the High-Resolution Fundus database, the top image shows Glaucoma while the bottom image corresponds to a healthy person. Right: Classification results of glaucoma images using Cross Validation and Training-Test Sets

Fig. 5 (right) shows the results obtained for this experiment. In the case of CV, the best-performing classification method was SVM, with 82% precision and a recall of 80% while the worst result was CBC, with 72% precision and 71% recall. In the case of TTS, the best-performing classification method was NB, with 83.7% accuracy while the worst result was thrown by SVM 80% accuracy.

3.3. PARKINSON

Parkinson's is a chronic neurodegenerative disorder caused by the destruction of dopaminergic neurons found in a region of the brain called the basal ganglia. It occurs when the nerve cells of the substantia nigra of the midbrain, the brain area that controls movement, die or suffer some deterioration. According to the WHO, Parkinson's disease affects 1 in 100 people over the age of 60. Currently, there are about 7 million people from this disease in the world and the WHO predicts that by 2030, they will become more than 12 million. Parkinson's disease manifests itself mainly through the progressive loss of the ability to coordinate movements and it is the second most frequent neurodegenerative disease after Alzheimer's [12].

In general, features in handwriting and drawing can be used for the assessment of neurodegenerative diseases. In the particular case of Parkinson's disease, it is well-known that the disease manifests itself through progressive micro-graphia and, when drawing spirals, in the speed, pen pressure, and spiral regularity. We use the Distinguishing Different Stages of Parkinson's Disease database, which is publicly available. This database was made by Kevin Mader, a software engineer, dedicated to conducting medical imaging DL programs, who did research and found a way to detect Parkinson's. A group of healthy and sick people carried out a test to draw waves and spirals, and in this way, carries out the image processing with the images of the previously labeled drawings. Spiral images were used for the present work; the database contains 102 images tagged Parkinson's and healthy. Fig. 6 (left) shows two sample images from the database.

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

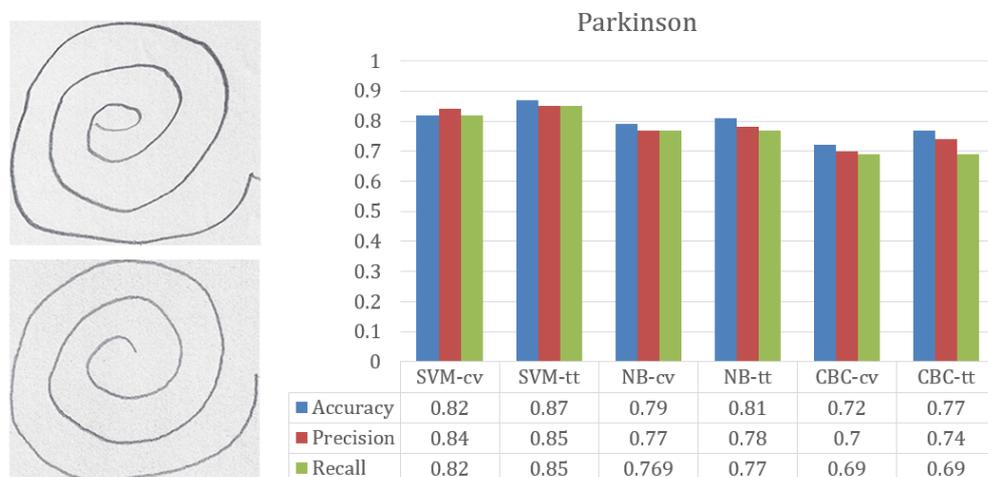


Figure 6: Left: Sample images from the so-called Distinguishing Different Stages of Parkinson's Disease database, the top image corresponds to a person with Parkinson's disease while the bottom one corresponds to a healthy person. Right: Classification results of Parkinson's disease images using Cross Validation and Training-Test Sets.

In this experiment, 51 images of patients labeled as healthy, and 51 patients labeled with Parkinson's were used. As shown in Fig. 6 (right), it was found that, in the case of CV, the method with the best results was SVM in all evaluation metrics, with an accuracy of 84% and a recall of 82% while the worst results were obtained with CBC, with 70% precision and 69% recall. In the case of TTS, the best result was obtained with SVM, with an accuracy of 85% and a Recall of 85%.

4. CONCLUSIONS

The effectiveness of the methodology proposed in this work is demonstrated by obtaining results greater than 85% in accuracy that are obtained in most cases. That is, the methodology is independent of the classification domain, and it has good performance. A relevant outcome of this work is the good results obtained in the classification scenario using a TTS. In that case, the test images are never seen by the training set. This makes it the desired classification scenario when sufficient training instances are available.

The images contained in the different databases used in this work were manually labeled by experts. Clearly, being able to automatically classify images, as demonstrated in this work, can increase throughput in medical diagnosis.

In both classification scenarios, the best classification method was SVM, which allows improving the accuracy by up to 85% compared to the other classification methods. It is very important to use computer tools that allow image classification to be carried out, in our case medical images, which may eventually help to improve a medical diagnosis

REFERENCES

- [1] Bhattacharya, S., et al., Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable cities and society, 2021. 65: p. 102589. DOI: <https://doi.org/10.1016/j.scs.2020.102589>
- [2] Tripathi, S. and N. Sharma, Computer-aided automatic approach for denoising of magnetic resonance images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2021: p. 1-10. DOI: <https://doi.org/10.1080/21681163.2021.1944914>
- [3] Guzman-Cabrera, R., et al. Parkinson's disease: Improved diagnosis using image processing. in 2017 Photonics North (PN). 2017. IEEE. DOI: <https://doi.org/10.1109/PN.2017.8090549>
- [4] Quintanilla-Domínguez, J., et al., Analysis of Mammograms Using Texture Segmentation. Advances in Language and Knowledge Engineering, 2016: p. 119. DOI: <https://doi.org/10.13053/rcs-123-1-11>
- [5] Lehman, C.D., et al., Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA internal medicine, 2015. 175(11): p. 1828-1837. DOI: <https://doi.org/10.1001/jamainternmed.2015.5231>
- [6] Wang, W., et al., Classification of focal liver lesions using deep learning with fine-tuning, in Proceedings of the 2018 International Conference on Digital Medicine Publications DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN) Tel +34 944 237 566 -- www.dyna-energia.com - email: dyna@revistadyna.com

	CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING	COMPUTER SCIENCE
COLLABORATION	Eduardo Pérez-Careta, José-Rafael Guzmán-Sepúlveda, José-Merced Lozano-García, Miguel-Torres Cisneros and Rafael Guzmán-Cabrera	Artificial Intelligence

and Image Processing. 2018. p. 56-60. DOI: <https://doi.org/10.1145/3299852.3299860>

- [7] Guzman-Cabrera, R., et al., Digital processing of thermographic images for medical applications. Revista de Chimie, 2016. 67(1): p. 53-56.
- [8] Vapnik, V., The nature of statistical learning theory. 2013: Springer science & business media.
- [9] Moreira, J.S.R., et al., La Tuberculosis y su vinculación con la pobreza. RECIAMUC, 2018. 2(2): p. 284-299.
- [10] de Paiva, L.F., et al., Metodologia Deep Features para Diagnóstico de Glaucoma/Deep Features Methodology for Glaucoma Diagnosis. Brazilian Journal of Development, 2020. 6(6): p. 41048-41060. DOI: <https://doi.org/10.34117/bjdv6n6-589>
- [11] Odstrcilik, J., et al., Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. IET Image Processing, 2013. 7(4): p. 373-383. DOI: <https://doi.org/10.1049/iet-ipr.2012.0455>
- [12] Benaiges, I.C. and C.A. Farret, Papel de la logopedia en el tratamiento de la disartria y la disfagia en la enfermedad de Parkinson. Neurol Supl, 2007. 3(7): p. 30-33.

SUPPLEMENTARY MATERIALS

https://www.revistadyna.com/documentos/pdfs/_adic/10117-1_en.pdf