Artificial Intelligence and Social Computing, Vol. 28, 2022, 85–91
https://doi.org/10.54941/ahfe1001450

An Analysis Model Through Machine Learning using Support Vector Machine for the Prediction of Diabetic Retinopathy

Remigio Hurtado, Janneth Matute, and Juan Boni

Universidad Politécnica Salesiana. Grupo de Investigación en Inteligencia Artificial y Tecnologías de Asistencia (GIIATA). Cuenca, EC 010101, Ecuador

ABSTRACT

Diabetic Retinopathy is a public health disease worldwide, which shows that around one percent of the population suffers from this disease. Likewise, another one percent of patients in the population suffers from this disease, but it is not diagnosed. That is why we face a great challenge, which is to predict and detect the signs of diabetic retinopathy at an early stage.For this reason, this article presents a Machine Learning model focused on the optimization of a classification method using support vector machines for the early prediction of Diabetic Retinopathy. This method has been trained using an image dataset called Messidor. In this way, the extraction and preprocessing of the data is carried out to carry out a descriptive analysis and obtain the most relevant variables through supervised learning. For the evaluation of the proposed method, quality measures have been used such as: MAE, MSE, RSME, but the most important are Accuracy, Precision, Recall and F1 for the optimization of classification problems. Therefore, to show the efficacy and effectiveness of the proposed method, we have used a public database, which has allowed us to accurately predict the signs of diabetic retinopathy. Our method has been compared with other relevant methods in classification problems, such as neural networks and genetic algorithms. The support vector machine has proven to be the best for its accuracy. This study gives way to future research related to diabetic retinopathy with the aim of conjecturing the information and thus seeking a better solution.

Keywords: Artificial intelligence, Support vector machine, Machine learning, Neural networks

INTRODUCTION

In this section, the fundamentals of Diabetic Retinopathy, of Machine Learning, the Motivation and the proposed Method of this work will be exposed.

Fundamentals of Diabetic Retinopathy

Diabetic Retinopathy has become a main ocular problem of systemic pathology, it is also one of the main reasons that leads to blindness in people

between twenty and sixty four years of age worldwide, for which epidemiological studies have been carried out, such as the one carried out in Wisconsin that have shown the presence of aneurysms and hemorrhages, this has been detected in people with type one and two diabetes, with a period of advancement of twenty years of disease, which can be said that between 90% of patients with diabetes have diabetic retinopathy. (Peris-Martínez et al. 2021).

However, the articles focused on diabetic retinopathy seek to assess different factors or variables that determine the risk of developing DR. Next, we will talk about the fundamentals of Machine Learning.

Fundamentals of Machine Learning

Machine Learning is applied in different fields such as science, engineering, medicine, management, so much so that its development has contributed considerably to image processing, pattern recognition, text mining, to voice recognition, etc. (Bensoussan et al. 2020). In this sense, Machine Learning has a great capacity to approach non-linear systems and extract important characteristics from high-dimensional data; a machine Learning algorithm can replace traditional methods, although the data used in the training of an ML algorithm contains numerical errors of approximation and rounding (Buizza et al. 2021).

Motivation and Proposed Method

Diabetic retinopathy presents different levels of severity; the patterns that reveal the existence of the disease are diverse; that is, when the disease presents a low level of severity; it is difficult to find the characteristics that reveal the existence of DR, while a high severity index is identified with the appearance of neovessels in an area of the retina and outside the optic nerve. So, this work seeks to classify the most important patterns among the Messidor images through a support vector machine that is a supervised learning algorithm for the classification of a dataset.

In short, this paper is made up of seven sections; in section II we have the related work, in section III we will talk about the proposed method, in section IV the design of experiments is presented where the dataset and quality measures are described in a general way to evaluate and compare some methods, same way in section V we will present the pertinent results and analysis, and finally in section VI we have the conclusions and future work of this research.

Next, the state of the art is exposed, where the works related to both the problem and the techniques used for the development of this project will be detailed.

STATE OF THE ART

For this section of the essay, the investigation of the works especially related to Diabetic Retinopathy and the most outstanding in both Machine Learning and Support Vector Machine and classification that seek to determine the concern that this disease has generated is proposed.

Jobs generally related to Diabetic Retinopathy

Projects related to diabetic retinopathy focus on datasets with processed images of an eve (Ocampo, 2018). To classify images, this disease has become the fundamental idea to apply strategies that help collect important information in large amounts of data (Castillo-Otí et al. 2020). For this reason, there are a variety of practices that have been developed by classifying images to identify the level of severity established in a study of retinography (Herrero, 2021).

Featured Machine Learning Works

The research work developed is focused on the classification of images, in addition to detecting, segmenting and classifying, first of all, preprocessing must be carried out, extract or select the characteristics and choose the classification method that will be used to evaluate the results. In the same way, for the extraction of characteristics, each image presents a vector with a visual characteristic such as size, shape, color, texture, intensity, etc.; this indicates that various research works are focused on the extraction of characteristics (Lucero et al. 2020).

There are also works that are focused on the recommendation of users, this refers to the suggestions that a user can give regarding products or others, with this we can say that Machine Learning can be reduced to more than health, to systems of marketing recommendation (Moreno et al. 2019).

Outstanding works in Support Vector Machines (SVM) and Classification

In article (Rojas et al. 2020) a supervised learning technique is proposed to evaluate four metaheuristics by optimizing the hyperparameters of a support vector machine, the evaluation was carried out on a database concerning Diabetic Retinopathy.

Next, the proposed method is exposed, which through a descriptive analysis aims to reveal variables that will help us predict the signs of diabetic retinopathy by means of a supervised learning technique based on the support vector machine (SVM).

PROPOSED METHOD

In this section, the proposed method will be exposed, which contains a table with each of the parameters that was used to develop the work, later the procedure will be described by means of a diagram.

Table with the definition of the parameters

The following table explains the parameters that were used in the SVM classification algorithm.

Analysis model process

Q1

Data extraction: Collect information to perform data analysis.

Descriptive analysis: It is a phase prior to data processing.

Data Preprocessing: Data preparation is a process that involves the

Table 1. Definition of Parameters used in the algorithm.

Variable	Description
C	Penalty parameter in an SVM.
gamma	It is the coefficient for the kernels.
kernel	Kernel function used in an SVM.

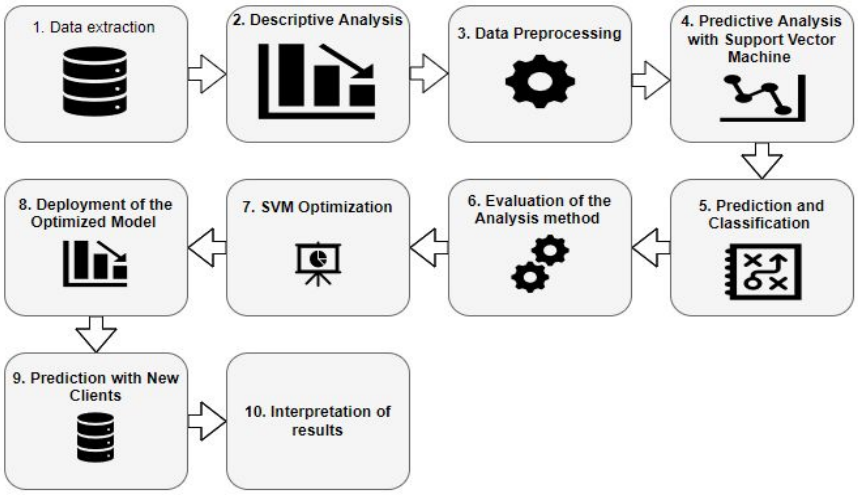

Figure 1: Analysis process diagram.

manipulation and transformation of raw data.

Predictive Analysis with Support Vector Machine: A Support Vector Machine is used for classification.

Evaluation of the analysis method: Quality measures such as precision, Recall, F1 are evaluated.

Optimization of the SVM: The values to test for each parameter are defined.

Deployment of the optimized model: add the optimized SVM classifier at the end.

Prediction with new clients: The data of another client is entered.

Interpretation of Results: The results of the data analysis are compared with those of the research.

DESIGN OF EXPERIMENTS

Section in which a description of the database we use, the methods and parameters used for the comparison will be explained.

Brief Description of the Database (Dataset)

The proposed method uses a public database with 1151 observations, contains features extracted from a group of Messidor images. The dataset is located in an Machine Learning repository: " https://archive.ics.uci.edu/

Table 2. Characteristics of the Diabetic Retinopathy database.

Data set	Diabetic retinopathy
Number of variables	20
Number of Observations	1151

Table 3. Define Parameters for the different Methods.

Method	Parameters
Neural Network	Epochs=5,10,20,50,100,300,600 batchsize=1,12,24,48,96,192 optimizer= AdaGrad, rmsprop 3 layers to use: layer 1: 2,4,8,16,32,64 layer 2: 2,4,8,16,32,64 layer 3: 1 Functions (Sigmoid, hyperbolic tangent)
SVM	gamma = 0.2, 0.5, 0.7, 1 kernel = linear, polynomial, radial C = 0.01, 0.2, 1, 5, 10, 50,100
GA	Generations=10,50,100,200 Pressure=50 Mutation=0.2 K neighbors=50,100,200,500,1000

ml/datasets/DiabeticRetinopathyDebrecenDataSet“ (Antal & Hajdu, 2014). Next, a table with the characteristics of the dataset is provided.

Use Parameters for Future Comparison

This item will define the parameters for each of the proposed methods.

In following section, the results obtained with the different methods are presented.

RESULTS AND DISCUSSION

This section will deal with the results of the proposed method and the comparison with the method used with the neural network and a genetic algorithm.

Definition of the Final Parameters of each of the Methods

This table defines the final parameters that will be used in the evaluation of each of the methods.

With respect to the genetic algorithm, it calculates the most similar K neighbors (the new population) for each user and then by means of the mode function with the K neighbors, the class of the new individual is determined.

Table 4. Definition of final parameters.

Method	Parameters
Neural Network	Epochs = 20 batchsize = 10 optimizer = rmsprop 3 layers (4,8,1) Functions (Sigmoid)
SVM	gamma = 0.5, kernel = polynomial, C= 50
GA	Generations = 100, pressure = 3, Mutation = 0.2, K neighbors = 50

Table 5. Final results.

Method	Precision
Neural Network	0.7446
SVM	0.761
GA	0.5086

Final Results of the Comparison of the three Methods

The following table shows the final results obtained when comparing the three methods.

That is, in this work we are concerned with the classification model to determine whether or not there is Diabetic Retinopathy, which is why we are interested in classification measures such as Precision, Recall, F1, Accuracy.

CONCLUSION

A classification has been created by means of an SVM, which has carried out the tuning of the hyperparameters. It can be seen that without the optimization the accuracy with cross validation is 0.70, if the optimization and adjustment of parameters is applied, the result is 0.75 with (kernel=polynomial, C=0.001 and gamma=1). Where the important phase of optimization and tuning is reflected, using parameters of machine learning models.

In addition, we can confirm that the model proposed in this article exceeds the base model in both quality measures with a considerable difference, additionally, train and test values are obtained with a considerable improvement to the base model. For the realization of this article we took the model of a duly published paper, in this sense the proposed model development, the trials and errors, were a great challenge for the authors because we would compete with professionals who have more experience in the field of science of data.

REFERENCES

Antal, B., & Hajdu, A. (2014). An ensemble-based system for automatic screening of diabetic retinopathy. *Knowledge-based systems*, *60*, 20–27.

Bensoussan, A., Li, Y., Nguyen, D. P. C., Tran, M. B., Yam, S. C. P., & Zhou, X. (2020). Machine learning and control theory. *arXiv preprint arXiv:2006.05604*.

Buizza, C., Casas, C. Q., Nadler, P., Mack, J., Marrone, S., Titus, Z., ... & Arcucci, R. (2022). Data Learning: Integrating Data Assimilation and Machine Learning. *Journal of Computational Science*, *58*, 101525.

Castillo-Otí, J. M., Cañal-Villanueva, J., García-Unzueta, M. T., Galván-Manso, A. I., Callejas-Herrero, M. R., & Muñoz-Cacho, P. (2020). Prevalencia y factores de riesgo asociados a la retinopatía diabética en Santander. Norte de España. *Atención Primaria*, *52*(1), 29-37.

Herrero Tudela, M. (2021). Clasificación automática de la severidad de la retinopatía diabética mediante técnicas de Deep Learning.

Lucero, M. Á. M., Castolo, E. M. C., Sánchez, D. N. L., González, J. F. U., Solís, O. C., Alayola-Sansores, A., & Dolci, G. F. (2020). Modelos de aprendizaje automático para clasificar patologías oculares en imágenes de fondo de ojo como prueba de tamizaje. *Res. Comput. Sci.*, *149*(8), 267–279.

Moreno, B. D. V., Ortiz, R. I. H., & Rivera, D. A. M. (2019, November). A new approach hybrid recommender system of item bundles for group of users. In *2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (pp. 1–7). IEEE.

Ocampo Arbeláez, H. F. (2018). Categorización de retinopatía diabética utilizando algoritmos de inteligencia artificial.

Peris-Martínez, C., Shaha, A., Clarida, W., Amelon, R., Hernáez-Ortega, M. C., Navea, A., ... & van der Heijden, A. A. (2021). Uso en la práctica clínica, de un método de cribado automatizado de retinopatía diabética derivable mediante un sistema de inteligencia artificial de diagnóstico. *Archivos de la Sociedad Española de Oftalmología*, *96*(3), 117–126.

Rojas, M. G., Carballido, J. A., Olivera, A. C., & Vidal, P. (2020). Optimización de Support Vector Machine mediante metaheurísticas para clasificación de retinopatia diabética. In *XXI Simposio Argentino de Inteligencia Artificial (ASAI 2020)-JAIIO 49 (Modalidad virtual)*.