

Diagnosis of Diabetic Retinopathy using Machine Learning algorithms

M.Rajkumar¹, P.Charulatha², P.Hima Bindu³, A.V.Kiruthika⁴

¹Associate Professor, Department of Computer Science and Engineering, R.M.D Engineering College, Gummidipoondi

²Student, Department of Computer Science and Engineering, R.M.D Engineering College, Gummidipoondi

³Student, Department of Computer Science and Engineering, R.M.D Engineering College, Gummidipoondi

⁴Student, Department of Computer Science and Engineering, R.M.D Engineering College, Gummidipoondi

Abstract - Diabetic retinopathy is a prevalent eye disease in diabetic patients and is the most common cause of visual impairment/blindness among the diabetic patients. Diabetic retinopathy requires early detection in order to prevent the patients from losing their vision. Thus, this paper focuses on decision about the presence of disease by applying ensemble of machine learning classifying algorithms on features extracted from output of different retinal image processing algorithms, like diameter of optic disk, lesion specific (microaneurysms, exudates), image level (prescreening, AM/FM, quality assessment). Decision making for predicting the presence of diabetic retinopathy was performed using machine learning algorithms. The main motive is to automatically classify the grade of non-proliferative diabetic retinopathy in any retinal image. In order to do that, an initial image processing stage isolates the necessary features from the blood vessels (microaneurysms, hard-exudates) that can be used by a support vector machine to figure out the retinopathy grade of each retinal image and it is depicted in a graphical format. The next stage is to detect the level of accuracy using stochastic gradient descent algorithm and it was found that 91% of accuracy level was achieved in our proposed method.

Key Words: Diabetic retinopathy, Support Vector Machine, Stochastic Gradient Descent Algorithm, Optic disk, Lesion Specific Image.

1.INTRODUCTION

Diabetic retinopathy is a dreadful and a very predominant eye disease. It is the most common cause of vision loss in the working-age population of developed countries of an estimated 285 million people with diabetes mellitus worldwide, approximately one third have signs of DR and of these, a further one third of DR is vision-threatening DR. In

addition to vision loss, DR has also been shown to contribute to the development of other diabetes-related complications including nephropathy, peripheral neuropathy and cardiovascular events. Diabetic retinopathy is caused when diabetes damages the blood vessels inside the retina, leaking blood and fluids into the surrounding tissue. This fluid leakage produces microaneurysms, hemorrhages, hard exudates, and cotton wool spots (a.k.a., soft exudates). Diabetic retinopathy is a silent disease and may only be recognized by patients when changes in the retina have progressed to a level where treatment becomes difficult or even impossible. Automatic detection of diabetic retinopathy will lead to a large amount of savings of time and effort. Thus, Wu et al proposed a method for automatic detection of microaneurysms in retinal fundus images. Several image pre-processing techniques have also been proposed in order to detect diabetic retinopathy. However, despite all these previous works, automated detection of diabetic retinopathy still remains a field for improvement. Thus, this paper proposes a new computer assisted diagnosis based on the digital processing of retinal images in order to help people detecting diabetic retinopathy in advance. The main goal is to automatically classify the non-proliferative diabetic retinopathy grade and the level of advancement of the disease in any retinal image. In order to do that, an initial image processing stage isolates blood vessels, microaneurysms and hard exudates in order to extract features that can be used by a support vector machine (SVM) to figure out the retinopathy grade of each retinal image. A decision tree classifier is also implemented to contrast the results obtained with our SVM classifier. Our proposal has been tested on a database of hundreds of retinal images. The level of advancement of the disease (benign or malignant) has been depicted in a graphical format (Scatter plot, bar graph) which is easier to understand and interpret the level of advancement of the disease among the patients. The next step involves the accuracy detection of the classification using stochastic gradient descent and the accuracy was

found to be 84% and SVM was also used which also produced an accuracy level of 63%.

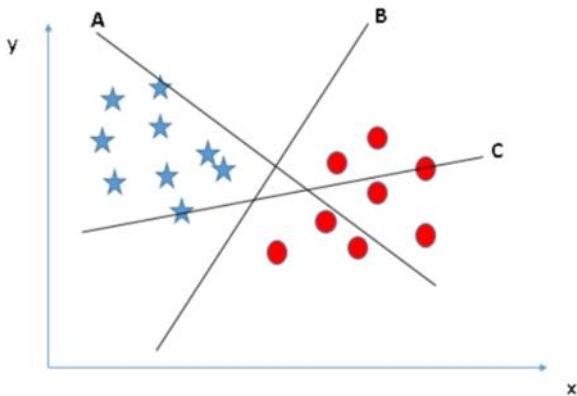


Fig -1: Advancement of disease among patients

1. EXISTING METHODS

In previous methods, they deployed supervised learning where, the segmented image is compared with a reference image. The main drawback here is that it cannot check all characteristics of the disease, instead compares only with the characteristics of the fed image. There are also other types of detection methods like kernel PCA, Gaussian DD, PCA, Rotation symmetric method, Lesion segmentation algorithm, Wavelet transforms, Probabilistic neural networks, Rational homography are used.

Algorithms like Kernel PCA, PCA DD and Gaussian DD lack accuracy and fail to correct background estimation errors. In Lesion transformation and Wavelet transformation the results are with high noise, these noises deviate the results.

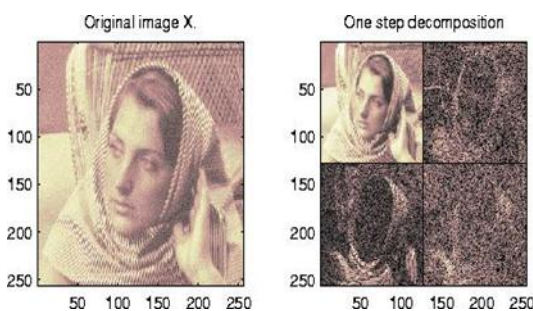


Fig -2: Image Processing in existing methods

In all existing methods neural networks are used, which requires high technical knowledge and involves more complex algorithms.

2. PROPOSED METHODS

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the coordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes.

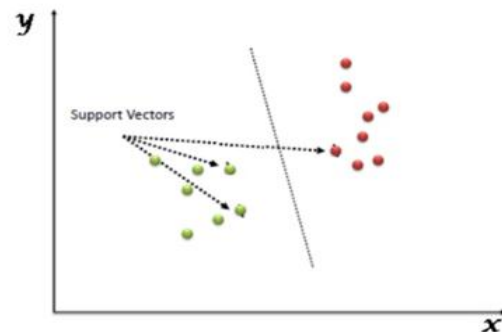


Fig -3: Support Vectors

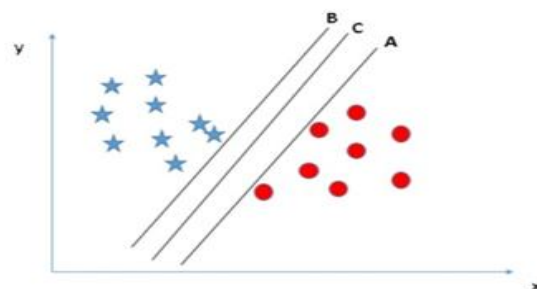


Fig -4: Image Processing

Identify the right hyper-plane (Scenario-1): Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane to classify star and circle.

You need to remember a thumb rule to identify the right hyper-plane: “Select the hyper-plane which segregates the two classes better”. In this scenario, hyper-plane “B” has excellently performed this job.

Identify the right hyper-plane (Scenario-2): Here, we have three hyper-planes (A, B and C) and all are segregating the classes well. Margin for hyper-plane C is high as compared to both A and B. Hence, we name the right hyper-plane as C. Another lightning reason for selecting the hyper-plane with higher margin is robustness. If we select a hyper-plane having low margin then there is high chance of misclassification. Identify the right hyper-plane (Scenario-3): Hint: Use the rules as discussed in previous section to identify the right hyper-plane

Implementing SVM in Python and R?

In Python, scikit-learn is a widely used library for implementing machine learning algorithms, SVM is also available in scikit-learn library and follow the same structure (Import library, object creation, fitting model and prediction). The e1071 package in R is used to create Support Vector Machines with ease. It has helper functions as well as code for the Naive Bayes Classifier. The creation of a support vector machine in R and Python follow similar approaches.

How to tune Parameters of SVM?

Tuning parameters value for machine learning algorithms effectively improves the model performance. Let's look at the list of parameters available with SVM.

3. CONCLUSIONS

There are many difficulties we faced while working. First of all, there's a lot more to do before an algorithm like this can be used widely. For example, as a classifier algorithm KNN and NNET was remarkable in percentage but we had to work more on binary classification. Secondly, if we could manage more of our training data we could train our algorithm more to achieve more accuracy. Furthermore, we also faced some problems while choosing algorithms. It was quite difficult for us to choose some specific machine learning algorithms that would give accurate classification of the disease. In addition we used simple techniques for feature selection and scaling and possibly we could arrive at better results by introducing more complex techniques for selecting and generating features. We looked at small subspaces for model parameters. Possibility there be other parameter spaces that would yield better performing models. For these few amount of data in our dataset we faced difficulties in implementing the classification.

REFERENCES

- [1] Gandhi M. and Dhanasekaran R. (2013). Diagnosis of Diabetic Retinopathy Using Morphological Process and SVM Classifier, IEEE International conference on Communication and Signal Processing, India pp: 873-877
- [2] Li T, Meindert N, Reinhardt JM, Garvin MK, Abramoff MD (2013) Splat Feature Classification with Application to Retinal Hemorrhage Detection in Fundus Images, IEEE Transactions on Medical Imaging, 32: 364-375
- [3] Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*. 2012;35:556-64
- [4] Boser B, Guyon I.G, Vapnik V., "A Training Algorithm for Optimal Margin Classifiers", Proc. Fifth Ann. Workshop Computational Learning Theory, pp. 144-152, 1992.
- [5] Mitchell, T. (1997). *Machine Learning*, McGraw Hill. ISBN 0-07-042807-7., McGrawHill, Inc. New York, NY, USA. Published on March 1, 1997
- [6] Alex C, Boston A. (2016). *Artificial Intelligence, Deep Learning, and Neural Networks, Explained* (16:n37)
- [7] Saimadhu P. *How the Random Forest Algorithm Works in Machine Learning*. Published on May 22, 2017
- [8] Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6), 493-507.
- [9] Jason B, Boinee P. "Machine Learning Algorithms" 2(3), 138-147. Published on 15, 2016.

[10] Boser B. E, Guyon I. M., Vapnik V. N. (1992). "A training algorithm for optimal margin classifiers". Proceedings of the 5th Annual Workshop on Computational Learning Theory COLT'92, 152 Pittsburgh, PA, USA. ACM Press, July 1992. On Page(s): 144-152

[11] Wikipedia. http://en.wikipedia.org/wiki/Support_vector_machine.

[12] Ben-Hur, A, Weston, J (2009). "A User's Guide to Support Vector Machines". Data Mining Techniques for the Life Science. Humana Press. On Page(s): 223-239.