IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

Prediction of Chances Diabetic Retinopathy using Data Mining Classification Techniques

A. Ramalaniya[1]

M. Phil, Research Scholar, H.H. The Rajah's College (Autonomous), Pudukkottai, India[1]

Abstract: Diabetic retinopathy the most common diabetic eye disease, is caused by complications that occurs when blood vessels in the retina weakens or distracted. The correct diagnosis of proliferative diabetic retinopathy is essential; because it is a treatable disease and missing the diagnosis can lead to the patient becoming blind. We examined the ability of internists and ophthalmologists to diagnose proliferative retinopathy under optimal conditions. Several data mining technique serves different purposes depending on the modeling objective. I have used various data mining techniques to predict the early detection of eye disease diabetic retinopathy. Image enhancement, mass screening and monitoring of disease are the main methodologies of this work.

Keyword: Naïve bayes, Support Vector Machine, Daibetic Retinopathy.

I.INTRODUCTION

Diabetic retinopathy (DR) occurs when diabetes damages the tiny blood vessels inside the retina, the light-sensitive tissue at the back of the eye. This tiny blood vessel will leak blood and fluid on the retina, forming features such as microaneurysms, haemorrhages, hard exudates, cotton wool spots, or venous loops. DR can be broadly classified as non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Depending on the presence of features on the retina, the stages of DR can be identified. . Diabetes is the fifth deadliest disease in the USA, and still there is no cure. The total annual economic cost of diabetes in 2002 was estimated to be US $132 billion, or one out of every 10 health care dollars spent in the USA. Diabetes is the chronic state caused by an abnormal increase in the glucose level in the blood and which causes the damage to the blood vessels. The tiny blood vessels that nourish the retina are damaged by the increased glucose level

II. REVIEW OF LITERATURE

Fleming et al.(2010) have shown the role of microaneurysm and hemorrhage in automatic grading of diabetic retinopathy. One of the most important steps in the automated detection of DR is the detection of microaneurysms. Microaneurysms are amongst the earliest observable signs of the presence of diabetic retinopathy. Due to a large number of patients, the available ophthalmologists are not sufficient in handling all the patients, especially in rural areas. Therefore, automated early detection of microaneurysms could ease the burden of ophthalmologists. Automated microaneurysms detection can also help the ophthalmologistsininvestigating and treating the disease more efficiently.

Akarasopharak et al(2011)proposed a series of experiments on feature selection and exudates classification using naive bayes and Support Vector Machine (SVM) Classifiers. At first, they used naive bayes model to a training set consisting of 15 features extracted from positive and negative examples of exudates pixels. Next, to obtain the best SVM, they used the best feature set from the naive bayes classifier and continually appended the removed features to the classifier. For each combination of features, they carried out a grid search to find the best combination of hyper parameters like tolerance for training error and radial basis function width. They compared the best naive bayes and SVM classifier to a Nearest Neighbour classifier. They proved that the naive bayes and SVM classifiers executed better than the NN classifier.

AlirezaOsareh et al **(2011)** We address the development of a method to quantitatively diagnose these random yellow patches in colour retinal images automatically. After a colour normalisation and contrast enhancement preprocessing step, the colour retinal image is segmented using Fuzzy C-Means clustering. We then classify the segmented regions into two disjoint classes, exudates and non-exudates, comparing the performance of various classifiers. We also locate the optic disk both to remove it as a candidate region and to measure its boundaries accurately since it is a significant landmark feature for ophthalmologists. Three different approaches are reported for optic disk localisation based on template matching, least squares arc estimation and snakes. The system could achieve an overall diagnostic accuracy of 90.1% for identification of the exudate pathologies and 90.7% for optic disk localisation. Another useful by-product of the proposed FCM approach is that it also segments the blood vessels. We hope to turn our attention to these in our future work. The study presented here indicates that automated diagnosis of exudative retinopathy based on colour

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

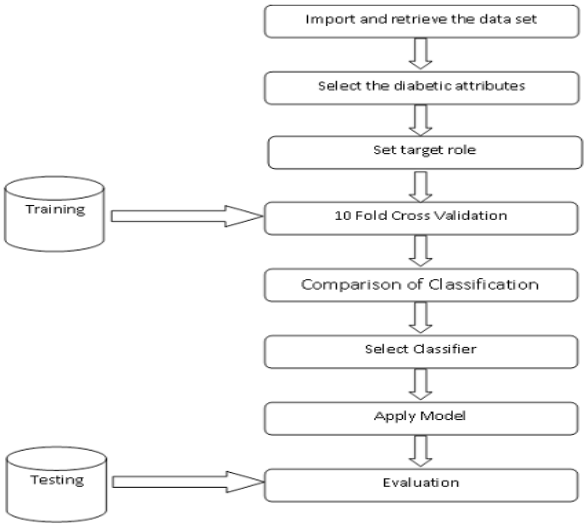
retinal image analysis is very successful in detecting EXs. Even though not all exudates may sometimes be found in a retinal image, it is important that some are found.

Tomi Kauppi1 et al **(2010)** proposes an evaluation methodology is proposed and an image database with ground truth is described. The database is publicly available for benchmarking diagnosis algorithms. With the proposed database and protocol, it is possible to compare different algorithms, and correspondingly, analyse their maturity for technology transfer from the research laboratories to the medical practice.The work will continue and the research group's main objective is to publish an ultimate tool for the evaluation of diabetic retinopathy detection methods. The tool will provide accurate and reliable information of method performance to estimate their maturity before starting the technology transfer from the research laboratories to practice and industry.

Ms.Nilamchandgude et al It occurs when pancreas does not produce sufficient insulin, or body can not sufficiently use insulin it produces. Diabetes person has increase blood glucose in the body. People with diabetes may develop serious problems such as heart disease, stroke, kidney failure, blindness, and premature death. WHO reported, in 2013 it was found that over 382 million people throughout the world had diabetes and mostly occurred in women than men due to improper food habit or low quality of food. Early diagnosis of diabetes is an important challenge. This survey present various classification are used for diagnosis of diabetes such as artificial neural network, support vector machine, naïve bayes, decision tree. PIMA Indian dataset are chosen for diagnosis of diabetes. The research hopes to propose a quicker and more efficient technique of diagnosing the disease, leading to timely treatment of the patients. Diagnosis of diabetes is real world important problem in medical field. This paper shows how to classify techniques such as Neural Network, Naive Bayes, SVM, C4.5, CART, ID3which are used for diagnosis of diabetes. Then accuracy of classification techniques are compared and plot the graph shows in (fig.[2])according to accuracy prediction. This type of diabetes called juvenile-onset because it's occurring a very young age of below 20 years. It also called insulin dependent because of human body does not produce sufficient insulin. Near about 10% of all diabetes cases are found in types1.Injection of insulin along with frequent blood test and dietary restriction has to be followed by patient suffering from type 1 diabetes.This type of diabetes called adult onset. It also called non-insulin dependent. The human body does not produce sufficient insulin for proper function in the body. Near about type 2 have 90% of all cases of diabetes in worldwide. Obesity, Being overweight, being physically inactive can lead to type2 diabetes. [3]

III. METHODOLOGIES

Three main methodologies were applied in this paper
1. Image enhancement
2. Mass screening
3. Monitoring of diseases



Proposed architecture

Standard contrast stretching techniques that allow enhancement of certain features (e.g., only microaneurysms) have been proposed image restoration techniques for images of very poor quality (e.g., due to cataracts).

A **ROI** performs two major functions: it defines a specific shape which can be examined for morphological reasons, and it specifies voxels within an image volume which are of interest. In either case, defining and storing the ROI is usually the same process.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

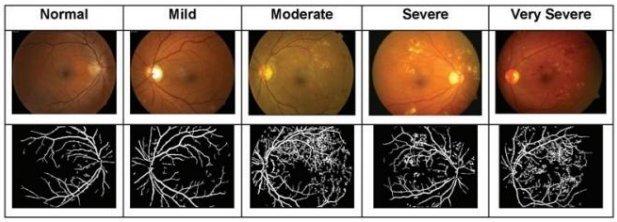
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

Hough transform is a technique which can be used to isolate features of a particular shape within an image. because it requires that the desired features be specified in some parametric form, the classical hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc.

Suport Vector Machine have attracted a great deal of attention in the last decade and actively applied to various domains applications. SVMs are typically used for learning classification, regression or ranking function. SVM are based on statistical learning theory and structural risk minimization principal and have the aim of determining the location of decision boundaries also known as hyperplane that produce the optimal separation of classes .Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalization error . Efficiency of SVM based classification is not directly depend on the dimension of classified entities. Though SVM is the most robust and accurate classification technique, there are several problems.
The data analysis in SVM is based on convex quadratic programming, and it is computationally expensive, as solving quadratic programming methods require large matrix operations as well as time consuming numerical computations. Training time for SVM scales quadratically in the number of examples, so researches strive all the time for more efficient training algorithm, resulting in several variant based algorithm.
candidateSV = { closest pair from opposite classes }
While there are violating points do
Find a violator
candidateSV = candidateSV S violator
If any αp< 0 due to addition of c to S then
candidateSV = candidateSV \ p
Repeat till all such points are pruned
End if
End while

Naive Bayes classifiers assume that the effect of a variable value on a given class is independent of the values of other variable. This assumption is called class conditional independence. An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the variable values necessary for classification.

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$



Severity of the Daibetic Retinopathy

Probability that a training pattern with attribute array A belongs to class CK, also known as the posterior probability is given Bayes theorem of probability,

$$(C_k/A) = P(C_k) * P(A/C_k)/P(A)$$

Where,
A is an array of M >= 1 attributes A1 ,A2 ,….,AM for the patterns of a training set.
P (CK) is the probability that a training pattern belongs to class A, also called prior probability.
P (CK|A) is also called the posterior probability because it is probability of training pattern with attribute array B belongs to class Ck
P (A|CK) is the conditional probability of B given A. It is also called the likelihood shows probability of class A has attribute array B.
P (A) is the probability that a training pattern has attribute array B, regardless of the class to which the pattern belongs.
Steps in algorithm are as follows:
1. Each data sample is represented by an n dimensional feature vector, X = (X1, X2,…,Xn), depicting measurements made on the sample from n attributes, respectively A1, A2, An.
2. Suppose that there are m classes, C1, C2,…, Cm. Given an unknown data sample, X (i.e., having no class label), the

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

classifier will predict that X belongs to the class having the highest posterior probability, conditioned if and only if:
P(Ci|X) > P(Cj|X) for all i<= j <= m and j != i
Thus we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes theorem,

$$P(Ci|X) = (P(X|Ci)P(Ci))/P(X)$$

As P(X) is constant for all classes, only P(X|Ci)P(Ci) needs to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e. P(C1) = P(C2) = … = P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). Note that the class priorprobabilities may be estimated by P(Ci) = si/s, where si is the number of training samples of class Ci, and s is the total number of training samples.
A complete formulation of Support Vector Machines can be found at a number of publications. Here, the basic principles will be presented and then their implementation and application to Object Based Image Analysis will be evaluated.
Let us consider a supervised binary classification problem. If the training data are represented by $\{x_i, y_i\}$, i = 1, 2, …, N, and $y_i \in \{-1, +1\}$, where N is the number of training samples, y_i=+1 for class ω1 and y_i=-1 for class ω2. Suppose the two classes are linearly separable. This means that it is possible to find at least one hyper plane defined by a vector w with a bias w0, which can separate the classes without error:

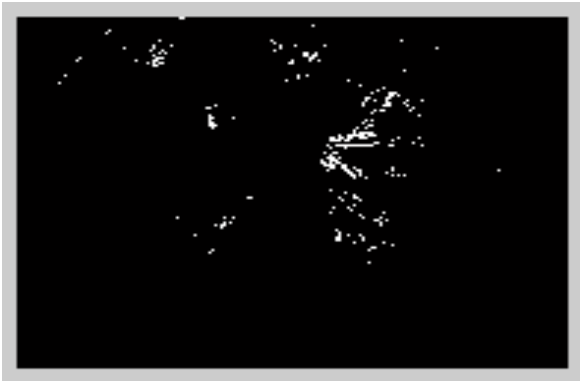
$$f(x) = w \cdot x + w0 = 0 \qquad (1)$$

To find such a hyperplane, w and w0 should be estimated in a way that $y_i(w \cdot x_i + w_0) \geq +1$ for $y_i = +1$ (class ω_1) and $y_i(w \cdot x_i + w_0) \leq -1$ for $y_i = -1$ (class ω_2). These two, can be combined to provide equation 2:

$$y_i(w \cdot x_i + w_0) - 1 \geq 0 \qquad (2)$$

Many hyper planes could be fitted to separate the two classes but there is only one optimal hyper plane that is expected to generalize better than other hyper planes.
The goal is to search for the hyper plane that leaves the maximum margin between classes. To be able to find the optimal hyper plane, the support vectors must be defined. The support vectors lie on two hyper planes which are parallel to the optimal and are given by:

$$w \cdot x_i + w_0 = \pm 1 \qquad (3)$$



SVM Classified Eye Retina Image

IV. CONCLUSION

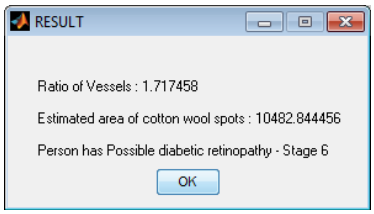
Automatic detection of micro aneurysm presents many of the challenges. The size and colorof micro aneurysm is very similar to the blood vessels. Its size is variable and often very small so it can be easily confused with noise present in the image. In human retina, there is a pigmentation variation, texture, size and location of human features from person to person. The more false positives occur when the blood vessels are overlapping or adjacent with micro aneurysms. So there is a need of an effective automated micro aneurysm detection method so that diabetic retinopathy can be treated at an early stage and the blindness due to diabetic retinopathy can be prevented.
The SVM algorithm could detect Micro aneurysms on very poor quality images. Although further development of this algorithm is still required, the results are satisfying. The outcome is quite successful with sensitivity and specificity of 81.61% and 99.99%, respectively. The system also provided ophthalmologists with the number of Micro aneurysms for grading the Diabetic retinopathy stage. In order to apply to a clinical application, the proposed method will be combined with an exudates detection system.

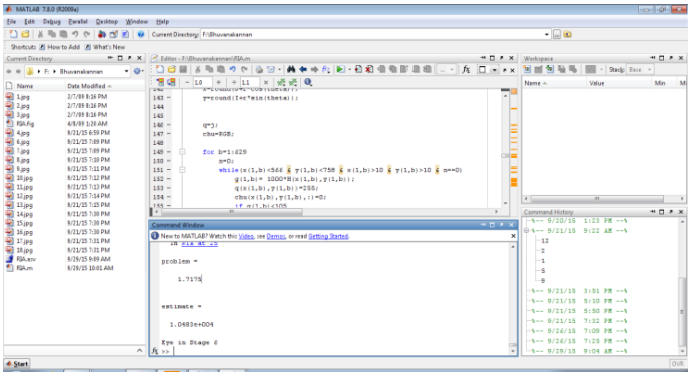
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

Performance Evaluation





REFERENCES

[1] A. D. Fleming, K. A. Goatman, and J. A. Olson, "The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy," British Journal of Ophthalmology, vol 94, no. 6, pp. 706- 711, 2010.

[2] Hipwell JH, Strachan F, Olson JA, McHardy KC, Sharp PF, Forrester JV. Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. Diabet Med 2012; 17: 588–594.

[3] AkaraSopharak, Mathew N. Dailey, BunyaritUyyanonvara, Sarah Barman, Tom Williamson, Yin Aye Moe, "Machine Learning approach to automatic Exudates detection in retinal images from diabetic patients", Journal of Modern optics, Vol. 57, No. 2, pp. 124-135, Nov 2015.

[4] Matei, Daniela, and R. Matei., "Detection of diabetic symptoms in retina images using analog algorithms," International Journal of Biological and Life Sciences, pp. 224-227, 2016.

[5] W. Hsu, P.M.D.S Pallawala, Mong Li Lee, Kah-Guan Au Eong, The Role of Domain Knowledge in the Detection of Retinal Hard Exudates, Proc. 2013 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2, 2001, II-246 - II-251. Classification and Localisation of Diabetic-Related Eye Disease by AlirezaOsareh ,MajidMirmehdi, Barry Thomas , and Richard Markham.

[6] DIARETDB1 diabetic retinopathy database and evaluation protocol by TomiKauppi, ValentinaKalesnykiene, Joni-KristianKamarainen, LasseLensu, IirisSorri, AstaRaninen, RaijaVoutilainen, HannuUusitalo, HeikkiK¨alvi¨ainen and JuhaniPietil¨a

[7] The electroretinogram in diabetic retinopathy by Tzekov R, Arden GB.

[8] Detection of Diabetic Retinopathy in Fundus Photographs by PavlePrentasi˘ c´.

[9] Spencer T, Olson JA, Mchrdy CK, Sharp FP, Forrester J. An image processing strategy for the segmentation and quantification of microaneurysems in fluorescein angiograms of the ocular fundus. Comput Biomed Res. 1999; 29:284–302.

[10] Niemeijer M, Abramoff MD, van Ginneken B. Segmentation of the Optic Disc, Macula and Vascular Arch in Fundus Photographs. Medical Imaging IEEE Transactions. 2007; 26:116–127.

[11] Ege BM, Hejlesen OK, Larsen OV, Møller K, Jennings B, Kerr D. et al. Screening for diabetic retinopathy using computer based image analysis and statistical classification. Comput Method Prog Biomed.2000; 62:165–175.