International Journal of Engineering & Technology, 7 (2.21) (2018) 440-442

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Analysis of diabetic retinopathy using naive bayes classifier technique

V. Sudha[1*], C. Karthikeyan[2]

[1]Research Scholar, Anna University Chennai.
[2]Professor, K.S.R College of Engineering, Thiruchengode.
**Corresponding author E-mail: vsudhaece@gmail.com*

Abstract

The vital issues of diabetes are Diabetic retinopathy (DR) and Retinal Vascular Disease which leads to the blindness. The DR disease may be detected by the early regular screening. and the automatic detection of this disease is a great solution and which is more reliable to identify the normality level in Fundus images (FI). The FI contains the texture discrimination capacity to differentiate the healthy images. The Data mining technique are used for identifying the retinal features of DR disease. The Data mining technique contains two stages. In first stage the features of DR disease extract from the Retinal Images (RI). The highlights for DR disease determination incorporate blood vessels, optic nerve, neural tissue, neuroretinal edge, optic plate size, thickness and change and which are removed by applying Data mining strategy. The result of the different information mining arrangement systems was looked at utilizing quick excavator apparatus. Gullible bayes and Support Vector Machine classifiers are utilized to anticipate the early discovery of eye disease diabetic retinopathy and observed that Naive bayes technique to be enhance the exactness of 89% precise.

Keywords: *Diabetic Retinopathy (DR), Support Vector Machine, Fundus Images (FI), Naive bayes, Retinal Images(RI).*

1. Introduction

Diabetic retinopathy (DR) is a genuine eye disease that happens because of diabetes mellitus and it has developed as the most widely recognized reason for visual deficiency in the present world. In light of most recent reports by 2030 there is a pandemic ascent of 4.4% in the worldwide pervasiveness of diabetes. Patient's sight can be influenced by diabetes which causes waterfalls, glaucoma, and in particular, harm to blood vessels inside the eye, a condition known as "diabetic retinopathy". Powerful medications for DR are accessible however it requires early analysis and the constant checking of diabetic patients. Conclusion of DR is performed by the assessment of retinal (fundus) pictures. Manual reviewing of these pictures to decide the seriousness of DR is somewhat moderate and asset requesting. It happens when diabetes harms the little blood vessels inside the retina, the light touchy tissue at the back of the eye. This minor blood vessel will spill blood and liquid on the retina shapes highlights, for example, miniaturized scale aneurysms, hemorrhages, hard exudates, cotton fleece spots or venous loops.[1]

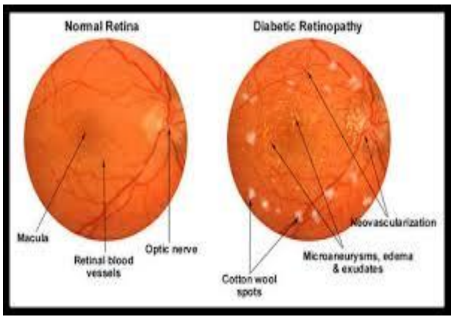

Fig. 1.1. Diabetic retinopathy

Diabetic retinopathy arises when excess glucose in the bloodstream resulting from diabetes mellitus causes damage to the blood vessels of the retina. Among US adults between the ages of 20 and 74 years, diabetic retinopathy is the leading cause of blindness.Diabetes influences an expected 29.1 million individuals in the United States. In a Centers for Disease Control and Prevention (CDC) assessment performed between 2005 and 2008, 4.2 million or 28.5% of people with diabetes aged 40 years or older in that time period had diabetic retinopathy

2. Related work

In Classification method Variables with 50% or more of their values missing were not included in the datasets used for machine learning. Missing data for the remaining variables (less than 50% of values missing) were handled by using imputation techniques. For the datasets analyzed, we performed feature subset selection. Since we used standard (single) classifiers in our previously published study, for this study, we use ensembles, which combine classifiers and may perform better than a single classifier approach. We learned ensemble classifiers based on decision tree learners designed to handle class imbalances such as RUSBoost,30 which utilizes majority class under sampling. For contrast, we also learned ensemble classifiers using AdaBoost.M1,31 which uses adaptive boosting to combine the weighted output of several weak learners to produce a boosted classification output. On its own, AdaBoost.M1 has no special accommodation for class imbalances. The ensemble classifiers were learned on the full feature set as well as the feature subsets obtained. We reserved 20% of each dataset for testing and then performed 10-fold cross validation on the remaining 80% of the dataset, selecting the best classifier from the cross-validation process for use on the reserved test set. For each classifier, we measured sensitivity or the true positive rate (the total number of cases delegated having diabetic retinopathy partitioned by the aggregate number of cases really including retinopathy), specificity or the genuine negative rate, the AUC, which speaks to the exchange off between the genuine positive rate/affectability and the false positive rate or specificity, and precision (the aggregate number of effectively characterized cases separated by the aggregate number of cases). [2]

3. Algorithm for diabetic retinopathy

Portrayal of dataset

The dataset was gotten at the Eye Clinic of the Sakarya University Educational and Research Hospital. The dataset comprises of 385 records. In dataset each record comprises of 9 highlights.

These are, in particular, Glycated Hemoglobin (HbA1C), Hemoglobin (HGB), URE, High-Density Lipoprotein (HDL), Low-Density Lipoprotein (LDL), Diabetes Duration, Triglyceride, Creatine and Glucose. Since Naïve Bayes calculation does not allow consistent information compose, every one of the qualities in the dataset are dealt with as straight out. In Table 1, the finding segments demonstrate the all-out qualities for the relating highlights. The conclusion segment is recognized as unsurprising element with esteem "1" for patients with diabetic retinopathy and esteem "0" for patients with non-diabetic retinopathy. All the unmitigated highlights in the dataset were chosen by specialists and assessment was made in view of these highlights.

Table 1: Demonstrates the Clinical Feature of the Patients in the Dataset

Feature Number	Description of Feature	Diagnosis (1)	Diagnosis(0)
1	Glycated Hemoglobin (HbA1C)	< 6.5	>= 6.5
2	Hemoglobin (HGB)	> 12	<= 12
3	High-Density Lipoprotein (HDL)	> 40	<= 40
4	Low-Density Lipoprotein (LDL)	< 130	>= 130
5	Diabetes Duration	< 5	>= 5
6	Triglyceride	> 150	<= 150
7	Creatine	> 1,2	<= 1,2
8	Glucose	> 140	<= 140
9	URE	> 45	<= 45

Gullible bayes

The Bayesian Classification speaks to a directed learning and additionally a factual technique for characterization. Accept a basic probabilistic model and it enables us to catch vulnerability about the model principledly by deciding probabilities of the results. It can take care of symptomatic and prescient issues. Innocent Bayes calculation depends on Bayesian Theorem.

1. Each information test is spoken to by a n dimensional component vector, X = (X1, X2,… , Xn), portraying estimations made on the example from n properties, individually A1, A2, An.

2. Suppose that there are m classes, C1, C2,… , Cm. Given an obscure information test, X (i.e., having no class name), the classifier will foresee that X has a place with the class having the most noteworthy back likelihood, adapted if and just if:

$$P(Ci|X) > P(Cj|X) \text{ for all } i \le j \le m \text{ and } j \ne I$$

In this way we amplify P(Ci|X). The class Ci for which P(Ci|X) is amplified is known as the most extreme posteriori theory. By Bayes hypothesis,

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

3. As P(X) is consistent for all classes, just P(X|Ci)P(Ci) should be amplified. In the event that the class earlier probabilities are not known, at that point it is usually expected that the classes are similarly likely, i.e. P(C1) = P(C2) = … = P(Cm), and we would in this way boost P(X|Ci). Else, we augment P(X|Ci)P(Ci). Note that the class earlier probabilities might be evaluated by P(Ci) = si/s, where si is the quantity of preparing tests of class Ci, and s is the aggregate number of preparing tests.

Cross approval

Cross-Validation (CV) is the standard information digging strategy for assessing execution of order method. Chiefly it's utilized to assess the blunder rate of a learning procedure. In CV a dataset is parceled in n folds, where each is utilized for testing and the rest of utilized for preparing. The method of testing and preparing is rehashed n times so each segment of crease is utilized once to test.

In a stratified 10-overlap Cross-Validation the information is isolated haphazardly into 10 sections in which the class is spoken to in roughly an indistinguishable extents from in full dataset. Each part is held out thus and the learning plan prepared on the staying nine-tenths; at that point its mistake rate is ascertained on the holdout set. The learning methodology is executed an aggregate of 10 times on various preparing sets, lastly the 10 blunder rates are found the middle value of to yield a general mistake evaluate.

Perplexity grid

Perplexity grid is a representation device which is generally used to introduce the precision of the classifiers in arrangement (Han and Kamber, 2006). It is utilized to demonstrate the connections amongst results and anticipated classes.

The passages in disarray grid have the accompanying implications with regards to our investigation:

- a is the quantity of right expectations that an example is negative,
- b is the quantity of mistaken forecasts that an example is certain,
- c is the quantity of mistaken forecasts that an example is negative,
- d is the quantity of right expectations that an example is certain

Table 2: Confusion Matrix

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

For Example

4. Performance and evaluation

To acquire and assess the test consequences of Naïve Bayes classifiers, 10-crease cross approval strategy was utilized. Consequently the dataset is haphazardly isolated into preparing set and testing set 10 times. Table 3, demonstrates the detail consequences of 10-overlap cross approval. Number of lines section speaks to testing set of each overlay though Sum of Diagonal segment speaks to the aggregate number of forecasts that were right. As we said before the exactness is figured utilizing condition (1).

The aftereffect of the precision that is acquired is great utilizing Naïve Bayes calculation in the genuine dataset. The precision rate is 89%. [8]

Table 3: Detail Results of 10 Fold Cross Validation

Fold	Sum of Diagonal	Number of rows	Accuracy
1	34	39	87.18%
2	31	39	89.49%
3	36	39	93.31%
4	35	39	90.74%
5	33	39	84.62%
6	33	39	84.62%
7	38	39	97.44%
8	38	39	98.44%
9	35	39	89.74%
10	31	35	88.57%
Accuracy: 89%			

5. Conclusion

This investigation obviously demonstrates that the outcomes are promising for the use of the information mining methods into forecasts of issue in therapeutic databases. In this paper, a choice emotionally supportive network was intended for diabetic retinopathy. The framework can be filled in as preparing instrument for medicinal understudies. Likewise, it will help hand for specialists. The framework can be additionally improved and extended; it can consolidate other restorative highlights other than in the Table 1, likewise it can join other information mining strategies. Persistent information can be utilized rather than simply clear cut information.

Future work

Several research fields are remained open in the field of diabetic retinopathy management and temporal diseases. Some future research works directly visualized by the advances of this thesis and from the experience in working in this field are listed below.

References

[1] Gupta S & Karandikar, A.M. Diagnosis of diabetic retinopathy using machine learning. *Journal of Research and Development*, Vol.3, No.3, (2015), pp.1-6.

[2] Sopharak A, Dailey M.N, Uyyanonvara B, Barman S, Williamson T, New K.T & Moe Y.A. Machine learning approach to automatic exudate detection in retinal images from diabetic patients. *Journal of Modern optics*, Vol.57, No.2, (2010), pp.124-135.

[3] Gandhi M & Dhanasekaran R, "Diagnosis of Diabetic Retinopathy Using Morphological Process and SVM Classifier", *IEEE International conference on Communication and Signal Processing, India*, (2013), pp.873-877.

[4] Li T, Meindert N, Reinhardt JM, Garvin MK & Abramoff MD, "Splat Feature Classification with Application to Retinal Hemorrhage Detection in Fundus Images", *IEEE Transactions on Medical Imaging*, Vol.32, (2013), pp.364-375.

[5] Ramani RG, Geetha R, Lakshmi B & Jacob SG, "Data Mining Method of Evaluating Classifier Pre- diction Accuracy in Retinal Data", *IEEE International Conference on Computational Intelligence and Computing Research*, (2012), pp.1-4.

[6] Rocha A, Carvalho T, Jelinek HF, Goldenstein S & Wainer J, "Points of Interest and Visual Dictionaries for Automatic Retinal Lesion Detection", *IEEE Transactions on Biomedical Engineering*, Vol.59, (2012), pp.2244 – 2253.

[7] Giancardo L, Meriaudeau F, Karnowski T, Tobin K & Grisan E, "Textureless macula swelling detection with multiple retinal fundus images", IEEE Trans Biomed Eng., Vol.58, (2011), pp.795-799

[8] Hayrettin E & Menduh Ç, "Prediction and Diagnosis of Diabetic Retinopathy using Data Mining Technique", *Journal of science and technology* Vol. 4, No.3.

[9] Agurto C, Murray V, Barriga E, Murillo S & Pattichis M, "Multiscale Am-Fm methods for diabetic retinopathy lesion detection", *IEEE Transactions on Medical Imaging*, Vol.29, (2010), pp.502-512.

[10] Kumar A, Gaur AK & Srivastava M, "A Segment based Technique for detecting Exudate from Retinal Fundus image", *Procedia Technology*, Vol.6, (2012), pp.1-9.

[11] Sopharak A, Dailey MN, Uyyanonvara B, Barman S & Williamson T, "Machine learning approach to automatic Exudates detection in retinal images from diabetics patients", *Journal of Modern Optics.*, Vol.57, (2008), pp.124-135.

[12] Yazid H, Arof H & Isa HM, "Exudates segmentation using inverse surface adaptive thresholding", *Measurement*, Vol.45, (2012), pp.1599-1608.

[13] Osare A, Shadgar B & Markham R, "A Computational Intelligence-Based Approach for De- tection of Exudates in Diabetic Retinopathy Images", *IEEE Transactions on Information Technology in Biomedicine,* Vol.13, (2009), pp.535-545.

[14] Kumari CJ & Maruthi R, "Detection of Hard Exudates in Color Fundus Images of the Human Retina", *Procedia Engineering*, Vol.30, (2012), pp.297-302.

[15] Saleh MD & Eswaran C, "An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and Has Detection", *Computer Methods and Programs in Biomedicine*, Vol.108, (2012), pp.186-196.

[16] Lazar I & Hajdu A, "Retinal Microaneurysm Detection Through Local Rotating Cross-Section Profile Analysis", *IEEE Transactions On Medical Imaging*, Vol. 32, (2013), pp.400-407.

[17] Antal B & Hajdu A, "An Ensemble Based System for Microaneurysm Detection and Diabetic Retinopathy Grading", *IEEE Transactions On Biomedical Engineering,* Vol.59, (2012), pp.1720-1726.

[18] Deepak KS & Sivaswamy J, "Automatic Assessment of Macular Edema from Colour Retinal Images", *IEEE Transactions Medical Imaging*, Vol. 31, (2012), pp.766-776.