IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data Mining Techniques for Automatic Diagnosis of Glaucoma Detection

B. Aruna[1]., K.Manickaraj[2].,

[1]M.Phil Scholar in Computer Science, Bharathiar University., Coimbatore

[2]Assistant Professor in Computer Science Bharathiar University., Coimbatore

ABSTRACT

In the modern period, people are afflicted with numerous ailments. Finding treatments for these illnesses or ways to identify them in their earliest stages has become essential for their prevention or treatment. One of the main causes of blindness and one of the eye illnesses is glaucoma. It is a disease that causes the patient's ocular vessels to slowly deteriorate, impairing eyesight. The data mining methods used to diagnose glaucoma in retinal images, including Decision Tree, Linear Regression, and Support Vector Machine, are discussed in this work. To evaluate each technique's effectiveness in terms of accuracy, sensitivity, and specificity, parameters from perimetry and the Stratus Optic Coherence Test (OCT) were supplied into it. The accuracy of the decision tree and linear regression models for the diagnosis of glaucoma is 99.17%, 92.56%, and 70.25 percent, respectively. The researcher compared the outcomes obtained from the decision tree, linear regression, and support vector machine (SVM), and discovered that they perform much better than SVM. SVM and linear regression have specificities of 97.56% and 96.34%, respectively.

Keywords: Detection of Glaucoma, Cup to Disc Ratio(CDR), Data Mining, Decision Tree , SVM and Linear Regression.

1. INTRODUCTION

An eye condition known as glaucoma results in an surge in intraocular pressure (IOP). The optic nerve that sends images to the brain is damaged by an increase in IOP. Glaucoma can cause a patient to lose their vision if it is not caught in the early stages. Within a few years, it can potentially result in permanent eyesight loss. Glaucoma is second only to cataracts in terms of the likelihood that a patient would experience unexpected vision loss, therefore early detection and diagnosis are crucial to halting further vision loss. Conferring to the Glaucoma Community, 12.8% of persons in India have glaucoma, which causes blindness. In our nation, glaucoma is frequently misdiagnosed Glaucoma is thought to afflict 12 million individuals, and by 2020, that figure is predicted to rise to 16 million. According to statistical analysis, one in eight adults over the oldness of 40 both has glaucoma or are at risk for developing it. A person may develop glaucoma due to high IOP, degeneration of astrocytes and optic nerve fibres, and other factors. Due to damage to the optic nerve fibre, the Retinal Nerve Fibre Layer's (RNFL) thickness continues to steadily decrease. Additionally, this deterioration alters the Optic Nerve Head (ONH), which reduces the retina's functional capacity. The thinning of the neuroretinal rim is caused by the expansion of the cup caused by the degeneration of astrocytes and axons. [27]. IOP (if IOP is greater than 21 mm Hg [32], it is suspected that the patient has hypertension), optical nerve cupping, and loss of vision field are some of the findings that are used to make the glaucoma diagnosis. Glaucoma is detected and diagnosed by a variety of procedures. This study is organized as reviews: Section II provides information on data collection, Section III provides background information on the research, and Section IV discusses data mining techniques used to make a diagnosis of glaucoma. Section VI gives our commentary after Section V outlines our result.

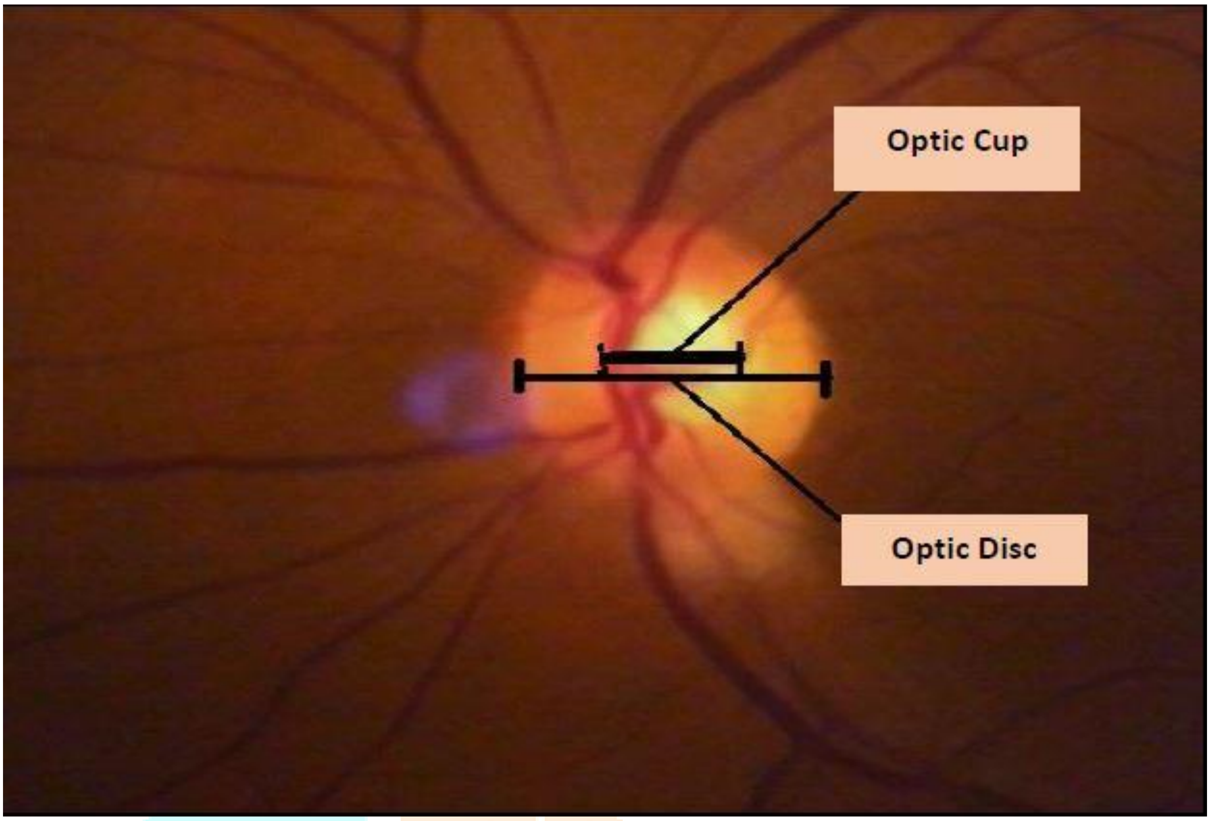
2. DATA ACQUISITION

Our analysis compares the most widely used techniques using the openly accessible databases DRIONS-DB, RIM-ONE v.3, and DRISHTI-GS. While obtaining quality that is comparable to current state-of-the-art techniques for optic disc and cup segmentation, our method exceeds them in terms of prediction time.

3. RESEARCH METHODOLOGIES

Intraocular pressure and Cup Disk Cup are the variables that are taken into account in cases of suspected glaucoma. IOP is not seen to be a reliable indicator for the analysis of glaucoma because it might rise for a variety of reasons, including increased making of aqueous fluid in the eye, certain drugs, eye damage, etc. Currently, CDR is used to diagnose glaucoma. The CDR is calculated as follows:

$$CDR = \frac{Area\ of\ optic\ Cup}{Area\ of the\ optic\ disc} \qquad (1)$$



The optic cup proceeds as it grows in size relative to the optic cup in Figure 1A, and the optic disc continues to gradually thin as the optic cup grows, as illustrated in Figure 1B above. In general, if the CDR is larger than 0.3, glaucoma is assumed to be present [1]. An algorithm (CDR ALGO) has been created by Kinjan Chauhan et al. to measure CDR. They separated the disc and cup from the retinal pictures using the OTSU histogram and morphological techniques. Using CDR, they were able to diagnose glaucoma with a 94% accuracy rate. S. Chandrika et al. employed Cup Disk Ratio for glaucoma identification utilising fuzzy c-means clustering and k-means clustering. According to their study, if the Cup Disk Ratio value is greater than 0.6, glaucoma is diagnosed and the patient is referred to an ophthalmologist for additional testing [3]. A convex hull-based ellipse optimization approach that effectively detects neuro-retinal optical cups was proposed by Zhuo Zhang and colleagues. Since their approach produced superior CDR calculation results than the ARGALI system, they were able to diagnose glaucoma with accuracy. A large-scale evaluation was conducted including 15,000 patients from Australia and Singapore [12]. CDR has been employed by Jyotika Pruthi et al. to separate the cup and disc from retinal pictures. Glaucoma diagnosis has been made with the help of Back propagation neural networks, SVM, and ANFIS. They found that the accuracy for SVM, ANFIS, and neural networks was, respectively, 98.12%, 97.77%, and 97.35% [9]. For the diagnosis of glaucoma, It will use Radial Basis Function (RBF) kernels, SVM linear, and polynomials of orders 1, 2, and 3. Higher Order Spectra (HOS) and Discrete Wavelet Transform (DWT) characteristics have been used to segment the disc and cup. To diagnose glaucoma and healthy eyes, they created the Glaucoma Risk Index (GRI). They have attained a 95% accuracy rate, a 93.33% sensitivity rate, and a 96.67% specificity rate, respectively [22]. For the purpose of categorizing retinal pictures normal, glaucoma- or diabetic retinopathy-affected, Ramani Geetha et al. have created methods for picture analysis and data mining. [11]. Syed SR. Abidi has put up a system for diagnosing glaucoma that combines data mining and image processing techniques. Their methodology examines optic nerve pictures for glaucoma diagnosis and surveillance [24]. Bowd et al. trained SVM, linear discriminant functions, and multi-layer perceptron (MLP) for ocular diagnosis using retinal tomography pictures and forward and backward feature selection approaches [7]. In order to extract features from the optic disc utilizing correlation analysis and the wrapper model, Park et al. [10] used SVM classifiers used a forward wrapper model for feature selection when training SVM classifiers [8].

4. DATA MINING TECHNIQUE

Applications of classifiers to aid in medical diagnosis have greatly expanded in recent years. The process of data mining is participatory and iterative. Three types of mining issues can be identified: recognising classifications, detecting sequential patterns, and identifying relationships. Data mining is a potent method for discovering new information, but it has limitations, including the need for specific data types and formats as well as application-specific requirements for different mining methods. Glaucoma diagnosis has made extensive use of classifiers for machine learning. The diagnosis of new cases can be predicted using trained classifiers. Using a predictive model, decision tree learning creates inferences about an item's target value from observations about it. It is a predictive modelling technique that is employed in statistics, machine learning, and data mining. A model or tree-like graph is called a decision tree. Given that its root is at the top and it develops downward, it is more akin to an inverted tree. Based on a variety of input variables from the dataset, a classification model known as a decision tree predicts the value of a target attribute, also known as a label value in fast miner. Attributes including information gain, gain ratio, gini index, and accuracy are used to segment the decision tree. While gain ratio modifies the information for each attribute to achieve the uniformity of the attribute values, information gain is biassed towards selecting large values. Gain ratio has therefore been utilised for splitting, with a value of 0.1Minimal gain must be raised in order to lower the size of the tree. Depending on the quantity and make-up of the data collection, a tree's depth changes. This parameter is used to limit the Decision Tree's size. As a result, the tree's depth is set to 20, which is rapid miner's default value. A new family of learning algorithms called Support Vector Machines (SVM) may tackle a wide range of classification and regression (model fitting) issues. SVM is a potent technique for both regression and classification. In order to reduce the generalisation error when training a classifier, it makes use of statistical learning theory. SVM comes with linear, polynomial, and radial basis function kernels (rbf). SVM have generalised well in breast cancer diagnosis and prognosis, handwritten digit recognition, facial recognition, and 26 text categorizations. In most cases, linear separable data are classified using the linear kernel. In medical diagnostics, where there are many attributes and a small amount of data, linear kernel outperforms rbf. Regression is a mathematical prediction method. It is a statistical measurement that seeks to ascertain the degree of the correlation between a single dependent variable (i.e., the label characteristic) and a number of different fluctuating variables known as independent variables (regular attributes). Regression is used to predict a continuous value, just like classification is used to predict categorical labels.

A DECISION TREE, LINEAR REGRESSION, AND SVM MODEL FOR THE DETECTION OF GLAUCOMA

Support Vector Machine (SVM), Decision Trees, and competent classifiers can all be used to predict the analysis of novel instances of glaucoma. In the medical field, classifiers are being used more frequently to aid in disease diagnosis. The classification technique can aid in lowering judgement errors and provides immediate results for medical data, which cuts down on the time needed to identify an illness. To further aid in the diagnosis of glaucoma, to classify the retinal pictures into glaucoma and normal.

Parameter Description

Parameters	Description
CDR_CALC	Calculated CDR using an algorithm
CDR_OCT	retrieved from OCT CDR
PER_OS_S	Superior value for Perimetry Left Eye
PER_OD_I	Value for Left Eye's inferior perimetry
PER_OD_I	Right Eye Lower Perimetry Value
PER_OD_S	Superior value for Perimetry Right Eye
OCT_OS_S	Excellent value for OCT Left Eye
OCT_OS_I	Left eye inferior value for OCT
OCT_OD_I	Right Eye OCT Lower Value
OCT_OD_S	Superior OCT Right Eye Value

The decision tree models that were created, which uses 10 fold cross validation and stratified sampling, has been used to validate linear regression and SVM. The data set will be split into 10 equal-sized subsets for the 10-fold cross validation, leaving one example set as test data. Up until all of the subsets have been utilised as test data sets once, this process is repetitive. The use of stratified sampling ensures that subsets are generated randomly and that class distributions match the original data set. Since the X-validation operator is a nested validation operator, the training data set is then passed on to the testing data set, where the model is applied using the Apply Model operator to the test data. The Performance operator can be used to determine the model's performance. The confusion matrix is produced by the performance operator, which also gauges a model's correctness. Medical data have more properties than other types of data, which causes the classifier to overfit. Researchers have used scaling by employing the quick miner tool's normalise operator to address this issue.

RESULTS

Performance of classifiers has been assessed in terms of Accuracy, Precision, Specificity, and Sensitivity for each classification model using the confusion matrix produced by the classifiers. A confusion matrix displays the comparison of the classification model's properly and erroneously predicted data to the actual data. Confusion matrix is used to gauge how well the classifier-built model is performing. The confusion matrix's explanation is provided in Table 5.10.

Confusion Matrix		Glaucoma condition	
		Disease Present	Disease Absent
Model	Predictive Present	True Positive	False Positive
	Predictive Absent	False Negative	True Negative

The calculation of the confusion matrix takes into account True Positive, True Negative, False Positive, and False Negative values. False Positive (FP) means that glaucoma in the patient is detected even when the patient is not suffering from it, and False Negative (FN) means that it does not detect glaucoma even when the patient is suffering from glaucoma. True Positive (TP) means that glaucoma is detected when the condition is present in the patient; True Negative (TN) means that it does not detect glaucoma when the patient is not suffering from it.

The following definitions apply to the terms Accuracy, Precision, Sensitivity, and Specificity:

Accuracy is the total number of accurately identified predictions.

Precision: the total number of positively detected positive cases.

Specificity: the percentage of accurately identified true negative cases.

Sensitivity: the number of validly identified genuine positive cases.

$$Accuracy = \frac{(True\ Positive\ +\ True\ Negative)}{True\ Positive\ +\ True\ Negative + False\ Positive + False\ Negative}$$

$$Presicion = \frac{True\ Positive}{(True\ \ positive + false\ positive)}$$

$$Specificity = \frac{True\ Negative}{(True\ \ Negative + false\ positive)}$$

$$Sensitivity = \frac{True\ Positive}{(True\ \ positive + false\ Negative)}$$

Confusion matrix for Decision Tree

	True Positive	True Negative
Pred. Positive	78	1
Pred. Negative	2	40

The confusion matrix for the decision tree, which has 99.17% accuracy, is shown in Table 5.11. The value of the glaucomatic eye's True Positive classification is 81, and the True Negative classification value is 39.

Confusion matrix for Linear Regression

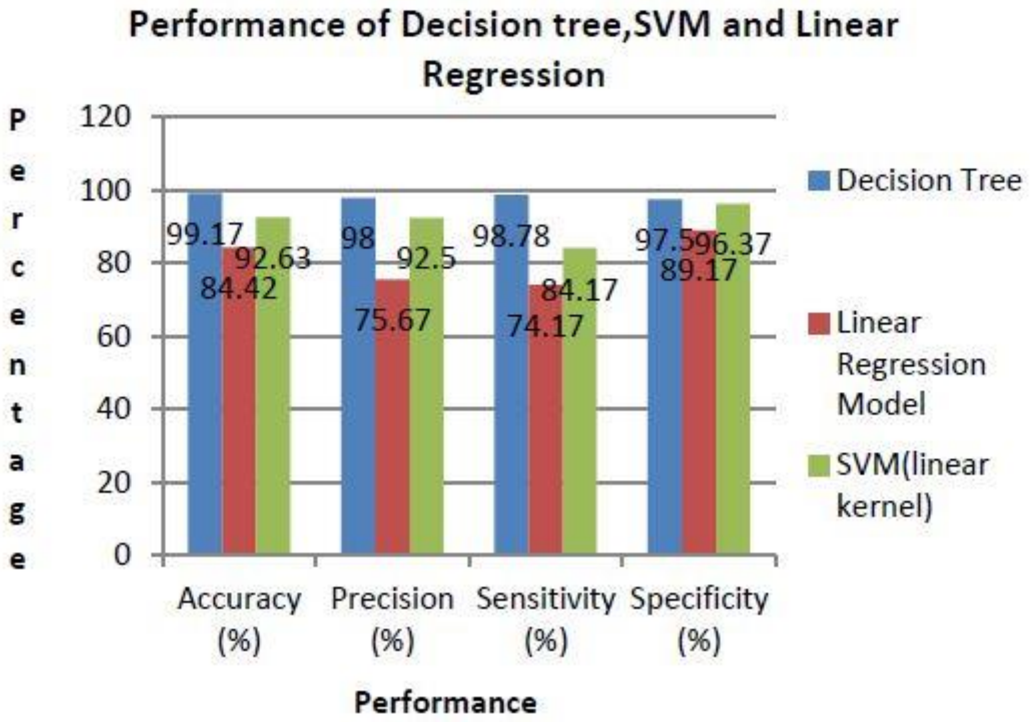
	True Positive	True Negative
Pred. Positive	78	32
Pred. Negative	4	6

Confusion matrix for SVM

	True Positive	True Negative
Pred. Positive	73	10
Pred. Negative	9	29

Table 6 Results of Comparative Experiments Using Data Mining Methodologies

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Decision Tree	99.17	98.78	97.5
SVM	92.63	84.17	96.37
Linear Model	84.42	74.17	89.17



6. CONCLUSION

Glaucoma can be caught early enough to preserve vision. The enactment analysis of several data mining approaches used to diagnose glaucoma using perimetric and OCT data is reported by the researchers in this paper. Using Decision Tree, Linear Regression, and SVM classifier, glaucoma has been diagnosed. All data mining algorithms had a 99.17%, 84.42%, and 92.63% accuracy in detecting the glaucoma and normal classifications, respectively. A comparison of the three methodologies reveals that Decision Tree and SVM model have provided results that are more accurate and precise, as well as sensitive and specific. By utilising better features and more reliable data mining approaches, the models can be further enhanced. With the aid of this diagnostic instrument, ophthalmologists will be able to diagnose glaucoma more accurately and quickly.

7. REFERENCES

[1] Chauhan K., Chauhan P., Gulati R., "Diagnosis system using data mining approach for Glaucoma (a social threat),": Technology and Society in Asia (T&SA), 2012 IEEE Conference on, 27-29 Oct. 2012. doi: 10.1109/TSAsia.2012.6397985 UR. pp.1,4.

[2] Juan Xu et al, J. 3D optical coherence tomography super pixel with machine classifier analysis for glaucoma detection Boston, MA: IEEE, 2011. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 3395-3398. ISSN 1094-687X, E-ISBN: 978-1-4244-4122-8 Print ISBN:978- 1-4244-4121-1.

[3] K. Nirmala, S. Chandrika, Comparative Analyis of CDR Detection for Glaucoma Diagnosis, The International Journal of Computer Science &Applications (TIJCSA), ISSN–2278-1080, Vol 2, No. 04, pp. 1-9, June 2013.

[4] Kavita Choudhary et al, Glaucoma Detection using Cross Validation Algorithm, Fourth International Conference on Advanced Computing & Communication Technologies. pp. 478–482. 2014.

[5] Saksham Sood, Aparajita Shukla, Arif Arman, Poonkodi, Automated Glaucoma Detection using Structural Optical Coherence Tomography with Data Mining, International Journal of Electrical, Computing Engineering and Communication (IJECC), Vol. 1 Issue 3, pp. 12-17. 2015.

[6] Kinjan Chauhan and Dr. Ravi M., Diagnosis of Glaucoma Using Cup to Disc Ratio in Stratus OCT Retinal Images. Ahmedabad: Smart Innovation, Systems and Technologies, Springer International Publishing, Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Vol 1, pp. 507-516. 10.1007/978-3- 319-30933-0_51. Series ISSN 2190-3018. 2016.

[7] Bowd C, Cban K, Zangwill LM, Goldbaum MH, Lee T,Sejnowski TJ. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser opthalmoscopy of the optic disc. Investigative Opthalmology & Visual Science, Vol. 43 (11), 2002.

[8] Swindale NV, Stjepanovic G, Cbin A, Mikelberg F. Automated analysis of normal and glaucomatous optic nerve head topography images. Investigative Opthalmology & Visual Science, Vol. 41 (7), 2000.

[9] Jyotika Pruthi, Dr.Saurabh Mukherjee , Computer Based Early Diagnosis of Glaucoma in Biomedical Data Using Image Processing and Automated Early Nerve Fiber Layer Defects Detection using Feature Extraction in RetinalColored Stereo Fundus Images. International Journal of Scientific & Engineering Research, Vol 4, Issue 4, April-2013 1822 ISSN 2229-5518 IJSER © 2013

[10] Park J, Reed J, Zhou Q. Active feature selection in optic disc nerve data using support vector machine. IEEE World Congress on Computational Intelligence, 2002.

[11] R. Geetha Ramani, et.al , Automatic Prediction of Diabetic Retinopathy and Glaucoma through Retinal Image Analysis and Data Mining Techniques, 2012 IEEE

[12] Zhuo Zhanget.al ,Convex Hull Based Neuro- Retinal Optic Cup Ellipse Optimization in Glaucoma Diagnosis 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009.

[13] Huiqi Li, Opas Chutatape,A Model- Based Approach for Automated Feature Extraction in Fundus Images Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Vol. Set 0-7695- 1950- 4/03 ,2003

[14] Michael H. Goldbaum; Pamela A. Sample; Kwokleung Chan; Julia Williams; Te-Won Lee; Eytan Blumenthal; Christopher A. Girkin; Linda M. Zangwill; Christopher Bowd; Terrence Sejnowski; Robert N. Weinreb [2002] Comparing Machine Learning Classifiers for Diagnosing Glaucoma from Standard Automated Perimetry, Investigative Ophthalmology & Visual Science, January 2002, Vol. 43, No. 1**162**

[15] Rudiger Bock et al, “Glaucoma Risk Index Automated glaucoma detection from color fundus images” Elsevier 30 December 2009

[16] Nagarajan et al, Neural network model for early detection of glaucoma using multi-focal visual evoked potential (MVep), Investigative Ophthalmology & Visual Science 42 (2002)

[17] Dimitrios Bizio et al, Integration and fusion of standard automated perimetry and optical coherence tomography data for improved automated glaucoma diagnostics BMC Ophthalmology, 11:20 http://www.biomedcentral.com/1471-2415/11/20, 2011

[18] JagadishNayak et al, Automated Diagnosis of Glaucoma Using Digital Fundus Images, Journal of Medical Systems, Vol. 33, Issue 5, pp-337-346, October 2009.

[19] Madhusudhanan Balasubramanian et al, Clinical Evaluation of the Proper Orthogonal Decomposition Framework for Detecting Glaucomatous Changes in Human Subjects, Investigative Ophthalmology & Visual Science, Vol. 51, No. 1, January2010.

[20] Mei-LingHuang,Hsin-YiChen,Jian-JunHuang,Glaucoma Detection using Adaptive Neuro-Fuzzy Inference System, Expert Systems with Applications Vol.32458–468,2007.

[21] U Rajendra AcharyaI, et al, Automated Diagnosis of Glaucoma Using Texture and Higher Order Spectra Features IEEE Transaction on Information Technology in Biomedicine 15(3):449 – 455 JUNE 2011.

[22] Muthu Rama Krishnan Mookiah a et.al ,Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features Knowledge-Based Systems 33 73–82, 2012.

[23] Nitha Rajandran Glaucoma Detection Using Dwt Based Energy Features and Ann Classifier IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p- ISSN: 2278-8727, Vol. 16, Issue 5, Ver. IV (Sep – Oct. 2014), PP 35-42 www.iosrjournals.org 2014.

[24] Syed SR. Abidia, Paul H. Artesb, Sanjan Yuna, ,Jin Yua Automated Interpretation of Optic Nerve Images: A Data Mining Framework for Glaucoma Diagnostic Support - retrieved from https://web.cs.dal.ca/~sraza/papers/MEDINFO07a.pdf

[25] Mrs. Kinjan Chauhan, Dr. Ravi Gulati [2013], A Survey on various image processing techinques for glaucoma diagnosis. Council for Innovative Research Peer Review Research Publishing System Journal, International Journal of Management & Information Technology, Vol. 5, No. 1 ISSN 2278-5612 pp 393-396 Impact Factor: 1.103.

[26] Mrs. Kinjan Chauhan, Dr. Ravi Gulati , Pre-processing of Retinal Image and Image Segmentation using OTSU Histogram International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682 Vol.29, No.29, September 2014.

[27] Kleyton Arlindo Barella et al, Glaucoma Diagnostic Accuracy of Machine Learning Classifiers Using Retinal Nerve Fiber Layer and Optic Nerve Data from SD-OCT, Publishing Corporation Journal of Ophthalmology Volume 2013, Article ID 789129, 7 pages http://dx.doi.org/10.1155/2013/789129, 2013.