GLAUDIA: A predicative system for glaucoma diagnosis in mass scanning

Yasser Omar
Arab Academy for Science, Technology, and Maritime Transport, Egypt

Mohamed Abd-ElSalam ElSheikh
Cairo University, Egypt

Rania Hodhod
Columbus State University, USA

Abstract
Glaucoma is a serious eye disease characterized by dysfunction and loss of retinal ganglion cells (RGCs) which can eventually lead to loss of vision. Robust mass screening may help to extend the symptom-free life for the affected patients. The retinal optic nerve fiber layer can be assessed using optical coherence tomography, scanning laser polarimetry (SLP), and Heidelberg Retina Tomography (HRT) scanning methods which, unfortunately, are expensive methods and hence, a novel automated glaucoma diagnosis system is needed. This paper proposes a new model for mass screening that aims to decrease the false negative rate (FNR). The model is based on applying nine different machine learning techniques in a majority voting model. The top five techniques that provide the highest accuracy will be used to build a consensus ensemble to make the final decision. The results from applying both models on a dataset with 499 records show a decrease in the accuracy rate from 90\% to 83\% and a decrease in false negative rate (FNR) from 8\% to 0\% for majority voting and consensus model, respectively. These results indicate that the proposed model can reduce FNR dramatically while maintaining a reasonable overall accuracy which makes it suitable for mass screening.

Keywords
glaucoma disease, mass screening, machine learning, ensemble technique

Corresponding author:
Mohamed Abd-ElSalam ElSheikh, Department of Basic Science, Cairo University, 6th of October City 12566, Egypt.
Email: mohamedelsheikh@cu.edu.eg

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Introduction

Computer-based diagnosis from image data is important for medicine. Eye images provide an insight into important parts of the visual system and can also indicate the health of the entire human body. Glaucoma is an eye disease that can lead blindness. It is a disease in which structural changes occur to the optic nerve head, retinal nerve fiber layer (RNFL) thickness, and ganglion cell and inner plexiform layers as well as loss of the visual field. The retinal optic nerve fiber layer can be assessed using optical coherence tomography, scanning laser polarimetry (SLP), and Heidelberg Retina Tomography (HRT) scanning methods. However, these methods are expensive and hence a novel automated glaucoma diagnosis model that uses extracted features from digital fundus images is needed. Robust mass screening may help to extend the symptom-free life for the affected patients; the ocular fundus image can be easily obtained and can be used to automatically identify whether an eye is glaucomatous or not. However, the image has a two-dimensional distribution, and it is difficult to feature the whole image through some real-valued parameters in general.

Generally, raw datasets usually include useful information that cannot be extracted by traditional data classification; although traditional solutions can reveal some latent information, they often require longer time and usually include some human mistakes. Finding hidden relationships in datasets from different sources (e.g. medical science, transportation, news, social media, weather, etc.) require computer-based solutions, such as machine learning and data mining to provide reliable worthy models with high accuracy. The exploration of medical data is a significant issue as it has a close direct relationship to our life. Thus, the proposed models in this field should have the lowest error rates in terms of treatment and diagnosis. These popular models and techniques include support vector machine (SVM), Naïve Bayes model, K-Nearest neighbor (KNN), and random forests (RFs).

SVM is a supervised learning algorithm that depends on the statistics theory. It is basically used for dimensional pattern recognition and nonlinear regression. In both cases, sample data are split into two different feature groups: training data that is applied for training the SVM model network and testing data that is used for the model validation. SVM treats all data equally as it requires all the training feature data to be multiplied with the same weighted coefficient, which neglects the importance of the special feature data. Accordingly, fuzzy memberships are used to describe the corresponding input feature data, which is regarded as the affiliation of corresponding feature data to one of the classes.

Naïve Bayes algorithm was introduced to the text retrieval community in the 1960s. It is a machine learning tool that depends on applying Bayes’ theorem with strong independence assumptions between the features. Naïve Bayes classifier is a popular method for text categorization that rules if documents do belong to one category or another based on word frequencies as the features. This classifier can solve the problem of multi-class density-based classification. In other words, it can calculate explicit probabilities for each hypothesis based on the Bayes theorem. Although Naïve Bayes algorithm is less competitive than SVM, it was successfully applied in automatic medical diagnosis.

KNN is a non-parametric supervised machine learning algorithm that can solve both regression and classification problems. It relies on labeled input data to learn a function that produces an appropriate output when given new unlabeled data. The input consists of the k-closest training examples in the feature space. In KNN classification, the output is a class membership; an object is classified by a plurality vote of its neighbors, with the object being assigned to the most common class among its k nearest neighbors.

RFs is a nonlinear regression or classification model that consists of ensembles of regression/classification trees such that each tree depends on a random vector sampled independently from the
data. Random forests algorithm was introduced by Breiman\textsuperscript{11} to overcome some of the shortcomings of the decision trees algorithm particularly its instability with small perturbations in a learning sample. RFs uses many randomly built decision trees to combine their predictions and reduces the possible correlation between decision trees by selecting different subsamples of the feature space. It turns out that the random forests algorithm has become a very powerful, efficient, and popular tool for the survival analysis.\textsuperscript{12}

This paper is organized as follows: The following section provides literature review on the use of the above machine learning models and few others to screen glaucoma patients. Section 3 presents the proposed system. Section 4 lays out the experimental results, and finally conclusions are provided in Section 6.

**Related works**

Several works attempted to help with glaucoma diagnosis. For example, an inductive logic programming (ILP) system called GKS was developed, not only to deal with low-level measurement data such as images, but also to produce diagnostic rules that are readable and comprehensive for interactions with medical experts.\textsuperscript{13} Another work provided automated identification of normal and glaucoma classes using Higher Order Spectra (HOS) and Discrete Wavelet Transform (DWT) features.\textsuperscript{14} The extracted features are fed into the SVM classifier with linear, polynomial order 1, 2, 3, and Radial Basis Function (RBF) to select the best kernel function for automated decision making. In this work, SVM classifier with kernel function of polynomial order 2 was able to identify the glaucoma and normal images automatically with an accuracy of 95\%, and sensitivity and specificity of 93.33\% and 96.67\%, respectively. Mookiah et al.\textsuperscript{14} present a computer-based glaucoma screening system in which optic nerve defects detection, visual field examination, and expert system fuzzy rules are combined to increase the sensitivity and specificity.\textsuperscript{14} The system is cost effective and is suitable for detecting early-stage glaucoma, especially for large-scale screening.

Several computational algorithms have been used for image-based glaucoma diagnosis. Most of these methods follow the two-stage pipeline structure: feature extraction and classification. This includes Discrete Wavelet Transform (DWT), Higher Order Spectra (HOS), Gabor transform for feature extraction and SVM, Artificial Neural Network (ANN), and KNN for prediction. The design of such hand-crafted features is a tedious job and time consuming. These features are strongly related to expert knowledge with a restricted representation power. Thus, for a huge dataset it cannot show the discriminative power. Deep features are essential to overcome this problem and enhance the classification performance. Using deep neural network techniques like CNN were applied on colored retinal images and achieved promising results as shown in Raghavendra et al.\textsuperscript{15,16}

The detailed literature reviews of the existing work for automated glaucoma diagnosis are summarized in Table 1. Moreover, the table includes a comparative summary with different performance parameters such as: accuracy (acc), sensitivity (sen), and specificity (spc).

**GLAUDIA: A predicative system for glaucoma diagnosis in mass scanning**

In the following sub-sections, we describe the dataset and the proposed model.

**Dataset**

The dataset is retrieved from Kim et al.\textsuperscript{34} It consists of 499 numerical records with seven input features extracted from digital fundus images and one label feature in a CSV file format. The
dataset is distributed as follows: 202 records of non-glaucoma cases and 297 of glaucoma cases 
((primary open angle glaucoma (POAG) or normal tension glaucoma (NTG)). The dataset contains 
a set of features that are used to describe the glaucoma cases including age, ocular pressure, cornea 

| Method | Classifier | Author |
|--------|------------|--------|
| Fundus disk parameters Cup to disk ratio (CDR) | (ANN) | Mvoulana et al.17 |
| Retinal nerve fiber layer thickness | (ANN) | Wang et al.18 |
| Histogram of oriented gradients (HOG) | https://en.wikipedia.org/wiki/Least-squares_support-vector_machine (SVM) | Balasubramanian et al.19 |
| Higher order statistics (HOS) and discrete wavelet transform (DWT) based features | Sequential minimal optimization (SMO) | Sharma et al.20 |
| Higher order spectra (HOS), trace transform (TT), and discrete wavelet transform (DWT) | Naïve Bayes | Kirar et al.21 |
| Gabor transformation and texture and entropy features | https://en.wikipedia.org/wiki/Least-squares_support-vector_machine (SVM) | Acharya et al.23 |
| Empirical wavelet transform (EWT) and correntropy | Least-Squares Support-Vector Machine (LS-SVM) | Maheshwari et al.24 |
| Variational mode decomposition (VMD), entropy, and fractal dimension | Least-Squares Support-Vector Machine (LS-SVM) | Maheshwari et al.25 |
| Nonparametric spatial envelop energy spectrum | (SVM) | Raghavendraet al.16 |
| Bit plane slicing (BPS) and local binary patterns (LBP) | https://en.wikipedia.org/wiki/Least-squares_support-vector_machine (SVM) | Maheshwari et al.26 |
| Self-organizing maps | Probabilistic Neural Network (PNN) | Abidi et al.27 |
| Wavelet energy features | Probabilistic Neural Network (PNN) | Annu and Justin28 |
| Haralick texture features | (KNN) | Simonthomas et al.29 |
| Wavelet energy features | (ANN) | Gayathri et al.30 |
| Wavelet and geometric moment features | (SVM), (KNN) | Gajbhiye and Kamthane31 |
| Morphological operations, Hough transform, and an anchored active contour model | Linear discriminant analysis | Fatima Bokhari et al.32 |
| Independent component analysis | Classification trees | Wang et al.33 |
| Deep convolutional neural network (CNN) | K-nearest neighbor (KNN) | Raghavendrae et al.15 |
| Eighteen-layer convolutional neural network (CNN) | Convolutional neural network (CNN) | | 
| | Latent Dirichlet allocation (LDA) | Raghavendra et al.16 |
thickness, retinal nerve fiber layer (RNFL) thickness, glaucoma hemifield test (GHT), history of macular degeneration (MD), and pattern standard deviation (PSD). In addition to one dependent variable (diagnosis) that takes either a value of 1 to indicate a Glaucoma case or a value of 0 to indicate a normal case. All features are used as an input to the classifiers.

For more information on feature extraction from the dataset, see Kim et al.34

**Proposed model**

Figure 1 represents the proposed model, which consists of six steps. In the first step, the data is preprocessed by normalizing the values of each feature to be between 0 and 1 as shown in equation (1). The data is then divided into two sets: 80% for training and 20% for testing as indicated in step 2. Different types of classifiers are applied to the training parts with 10-fold validation to obtain generalized models as shown in step 3. In step 4, we use accuracy from the average of 10-fold divisions to rank the models, where accuracy is defined as the percentage of true predicated cases either positive or negative to the total number of all cases as shown in equation (2).
In step 5, the top five models with the highest accuracy are ensembled together; all the models have to be in consensus about a case to be classified as a negative case (no glaucoma disease), while if one model is positive about a case, then the case will be classified as a positive case (glaucoma disease). The final decision of the stacked model is defined as shown in equation (4). Finally, the proposed model is evaluated with the other staging models that are based on the majority voting, not on the consensus, from two different metrics: accuracy and FNR. Equation (3) shows the FNR that represents the percentage of positive cases (glaucoma disease) that were falsely predicted as negative cases (no glaucoma disease).

\[
\begin{align*}
    f_{\text{normalized}} &= \frac{f_{\text{original}} - f_{\text{min_val}}}{f_{\text{max_val}} - f_{\text{min_val}}} \\
    \text{Accuracy} &= \frac{TP + TN}{TP + FP + TN + FN} \\
    \text{FNR} &= \frac{FN}{TP + FN} \\
    \text{Ensemble Decision} &= \text{No Glaucoma, if all models decide no glaucoma} \\
    &\quad= \text{Glaucoma, Otherwise}
\end{align*}
\]

Where \(f_{\text{original}}, f_{\text{normalized}}\) feature value before and after normalization, respectively.

\(f_{\text{min_val}}, f_{\text{max_val}}\) minimum, maximum value in each feature respectively.

**Experimental results**

In this section we describe the results obtained and give a detailed discussion about it. We used the Google Colab (a tool for machine learning) for the implementation of the machine learning techniques, Python and Keras (a machine learning library) were used. We started our experiment by normalizing the data to ensure that it is between 0 and 1. After that, we divided the data into 80% for training and 20% for testing. The division is done using a built-in library in Skit Learn that grants random and balanced distributions of the output labels. It is worth mentioning that the data distribution is 40.4% non-glaucoma cases and 59.6% glaucoma cases; the data is not fully balanced. However, the bias is on the side of positive cases which can be more suitable for mass screening as the goal is to reduce FNR even if this leads to reducing the overall accuracy for the sake of early diagnosis.

In the machine learning step, nine different techniques were applied: Logistic Regression, K-Neighbors, Random Forest, Decision Tree, Naive Bayes with both Gaussian and Bernoulli distributions, and SVM with different kernels (Linear, Poly, and RBF). For each model, we used 10-fold validation for the training data to grant generalization for the obtained model and to avoid any overfitting by averaging the resulting error from the validation results. In addition, a random search approach for tuning each model parameters were used to achieve best accurate result. Figure 2 shows a bar chart for the mean average accuracy from applying 10-fold validation for the nine models in descending order.
From Figure 2, the highest five accuracies were achieved by Logistic Regression, Random Forests, SVM with linear kernel, K-Nearest Neighbors, and decision trees. We have then built an ensemble model that constitutes these top five models only. We evaluated the model for both majority voting and consensus using 20% of the dataset, that is, to decide if a case is negative, majority voting must have a highest count for a negative case across all the techniques while the consensus model needs just one technique to determine that the case is negative. The consensus among the five models does also help with avoiding overfitting as its target is to reduce false negative error based on reducing the overall accuracy to be suitable for mass scanning.

For the first phase, we chose 10-fold validation to overcome the sample size limitation and avoid any overfitting. While in the second phase we divided the dataset into 80% for training and 20% for testing which is a rule of thumb in dividing small size datasets reasonably.

A comparison between the FNR is shown in Figure 3. The results show that FNR has decreased to 0% using the ensemble model. In Table 2, we compare between the accuracy of the ensemble while considering the majority voting approach and consensus approach, it is noticed that accuracy has decreased from 90% for the majority voting to 83% for the consensus approach.

**Conclusions and future work**

This paper presents a new model to screen for the glaucoma disease based on different features. The proposed model compares between different machine learning techniques and ensemble the top five techniques that provide the highest accuracies based on 10-fold validation. The ensemble of the best top five models is done via two different approaches: majority voting and consensus. The results obtained a decrease in the accuracy from 90% using majority voting to 83% using consensus for testing data that was never seen before by the trained models. The results also show a decrease in false negative rate (FNR) from 8% using majority voting to 0% using consensus. When it comes to mass screening, FNR is more critical than accuracy as the main goal is to reduce the number of cases that needs to be screened by the physicians to positive cases only. This makes
the ensemble approach based on consensus an attractive one. The main contribution of this work is to identify five top best classifiers based on the resulting accuracy rate, in addition to providing a consensus model that aims to reduce the false negative rate to make the proposed model more suitable for mass scanning. Future work includes combining different modalities to diagnose glaucoma, such as retinal images associated with risk factors to increase the accuracy of the model without losing the advantage of having low FNR, in addition to differentiating between Glaucoma types. In addition, other features that do not exist in the current dataset would be considered, such as gender and race as they might have a role in Glaucoma diagnosis.

**Funding**
The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**
Mohamed Abd-ElSalam ElSheikh [ID](https://orcid.org/0000-0003-1494-2130)

**References**
1. Kim SJ, Cho KJ and Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One* 2017; 12(5): e0177726.
2. Ohwada H, Daidoji M, Shirato S, et al. Learning first-order rules from image applied to glaucoma diagnosis. In: *Pacific Rim international conference on artificial intelligence*, November 1998, pp.494–505. Berlin, Heidelberg: Springer.
3. Alkeshuosh AH, Moghadam MZ, Al Mansoori I, et al. Using PSO algorithm for producing best rules in diagnosis of heart disease. In: ICCA international conference on computer and applications, Dubai, United Arab Emirates, 6–7 September 2017, pp.306–311. IEEE.

4. Abdar M and Makarenkov V. CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer. Measurement 2019; 146: 557–570.

5. Fan M, Wei L, He Z, et al. Defect inspection of solder bumps using the scanning acoustic microscopy and fuzzy SVM algorithm. Microelectron Reliab 2016; 65: 192–197.

6. Maron ME. Automatic indexing: an experimental inquiry. JACM 1961; 8(3): 404–417.

7. Niazi KAK, Akhtar W, Khan HA, et al. Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier. Solar Energy 2019; 190: 34–43.

8. Rish I. An empirical study of the naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence, August 2001, vol. 3, no. 22, pp.41–46. New York: IBM.

9. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 1992; 46(3): 175–185.

10. Pan X, Luo Y and Xu Y. K-nearest neighbor based structural twin support vector machine. Knowl Based Syst 2015; 88: 34–44.

11. Breiman L. Random forests. Mach Learn 2001; 45(1): 5–32.

12. Utkin LV, Konstantinov AV, Chukanov VS, et al. A weighted random survival forest. Knowl Based Syst 2019; 177: 136–144.

13. Mizoguchi F, Ohwada H, Daidoji M, et al. Using inductive logic programming to learn rules that identify glaucomatous eyes. In: Lavrač N, Keravnou ET and Zupan B (eds.) Intelligent data analysis in medicine and pharmacology. Boston, MA: Springer; 1997, pp.227–242.

14. Mookiah MRK, Acharya UR, Lim CM, et al. Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. Knowl Based Syst 2012; 33: 73–82.

15. Raghavendra U, Fujita H, Bhandary SV, et al. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf Sci 2018; 441: 41–49.

16. Raghavendra U, Bhandary SV, Gadigar A, et al. Novel expert system for glaucoma identification using non-parametric spatial envelope energy spectrum with fundus images. Biocybern Biomed Eng 2018; 38(1): 170–180.

17. Mvoulana A, Kachouri R and Akil M. Fully automated method for glaucoma screening using robust optic nerve head detection and unsupervised segmentation based cup-to-disc ratio computation in retinal fundus images. Comput Med Imaging Graphics 2019; 77: 101643.

18. Wang Y, Lu W, Xie Y, et al. Functional alterations in resting-state visual networks in high-tension glaucoma: an independent component analysis. Front Hum Neurosci 2020; 14: 330.

19. Balasubramanian T, Krishnan S, Mohanakrishnan M, et al. HOG feature based SVM classification of glaucomatous fundus image with extraction of blood vessels. In: IEEE annual india conference (INDICON), Bangalore, India, 16–18 December 2016. IEEE.

20. Sharma R, Sircar P, Pachori RB, et al. Automated glaucoma detection using center slice of higher order statistics. J Mech Med Biol 2019; 19(01): 1940011.

21. Kirar BS, Agrawal DK and Kirar S. Glaucoma detection using image channels and discrete wavelet transform. IETE J Res. Epub ahead of print 27 July 2020. DOI: 10.1080/03772063.2020.1795934.

22. Noronha KP, Acharya UR, Nayak KP, et al. Automated classification of glaucoma stages using higher order cumulant features. Biomed Signal Process Control 2014; 10: 174–183.

23. Acharya UR, Ng EYK, Eugene LWJ, et al. Decision support system for the glaucoma using Gabor transformation. Biomed Signal Process Control 2015; 15: 18–26.

24. Maheshwari S, Pachori RB and Acharya UR. Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. IEEE J Biomed Health Inf 2016; 21(3): 803–813.

25. Maheshwari S, Pachori RB, Kanhangad V, et al. Iterative variational mode decomposition based automated detection of glaucoma using fundus images. Comput Biol Med 2017; 88: 142–149.

26. Maheshwari S, Kanhangad V, Pachori RB, et al. Automated glaucoma diagnosis using bit-plane slicing and local binary pattern techniques. Comput Biol Med 2019; 105: 72–80.
27. Abidi SS, Roy PC, Shah MS, et al. A data mining framework for glaucoma decision support based on optic nerve image analysis using machine learning methods. *J Healthcare Inf Res* 2018; 2(4): 370–401.
28. Annu N and Justin J. Automated classification of glaucoma images by wavelet energy features. *Int J Eng Technol* 2013; 5(2): 1716–1721.
29. Simonthomas S, Thulasi N and Asharaf P. Automated diagnosis of glaucoma using Haralick texture features. In: *International conference on information communication and embedded systems (ICICES2014)*, Chennai, India, 27–28 February 2014, pp.1–6. IEEE.
30. Gayathri R, Rao PV and Aruna S. Automated diagnosis of glaucoma using wavelet energy features and ANN. In: *International conference on advances in computing, Communications and Informatics (ICACCI)*, Delhi, India, 24–27 September 2014, pp.2808–2812. IEEE.
31. Gajbhiye GO and Kamthane AN. Automatic classification of glaucomatous images using wavelet and moment feature. In: *Annual IEEE India conference (INDICON)*, New Delhi, India, 17–20 December 2015. IEEE.
32. Fatima Bokhari ST, Sharif M, Yasmin M, et al. Fundus image segmentation and feature extraction for the detection of glaucoma: a new approach. *Curr Med Imaging* 2018; 14(1): 77–87.
33. Wang M, Shen LQ, Pasquale LR, et al. An artificial intelligence approach to assess spatial patterns of retinal nerve fiber layer thickness maps in glaucoma. *Transl Vision Sci Technol* 2020; 9(9): 41–41.
34. Kim S, Cho KJ and Oh S. *Data from: development of machine learning models for diagnosis of glaucoma*. Dryad, Dataset, 2018.