A comparative study of pattern recognition methods for predicting glaucoma diagnosis

Louis Williams[1], Salman Waqar[2], Tom Sherman[2], and Giovanni Masala[3]

[1] University of Plymouth, Plymouth, United Kingdom,
kossy@kossy.co.uk
[2] Royal Eye Infirmary, Derriford Hospital, Plymouth, United Kingdom,
salman.waqar@nhs.net thomas.sherman@nhs.net
[3] Manchester Metropolitan University, Manchester, United Kingdom,
G.Masala@mmu.ac.uk

Abstract. Glaucoma is the second leading cause of blindness globally, it is characterized by degeneration of the optic nerve with particular patterns of corresponding defects in the visual field. Aiding doctors in early diagnosis and detection of progression is crucial, as glaucoma is asymptomatic in nature. Furthermore there is good therapeutic results in early cases before irreversible visual loss occurs. Thus it is of great importance to find automated methods to discriminate glaucomatous diseases giving insight to doctors. In order to develop a Computer Aided Diagnosis system (CAD), we realised an extensive competitive study of pattern recognition methods should be undertaken. A range of methods have been evaluated including the use of Deep Neural Networks (DNN), Support Vector Machines (SVM), Decision Trees (DT) and K-Nearest Neighbours (KNN) for diagnosing glaucoma. Using a range of classification techniques, this paper aims to diagnose glaucomatous diseases. Results have been produced with data comprising of Visual Field and OCT Disc readings from anonymous patients with and without glaucoma. Multiple systems are proposed that can predict diagnosis for ocular hypertension, primary open angle glaucoma, normal tension glaucoma and healthy patients with a reasonable confidence. Best performance has been obtained from voting classier comprised of SVM and KNN at 0.87 (AUC) and DNN at 0.87 (AUC) which possibly could be used as an automatic diagnosis aid in order to streamline the diagnosis of glaucoma for complex cases or flagging of urgent cases.

Keywords: glaucoma; Visual Field, pattern recognition, Computer Aided Diagnosis, Deep Neural Network, Support Vector Machine, Decision Trees.

1 Introduction

Glaucoma is a group of diseases that are characterized by degeneration of the optic nerve with particular patterns of corresponding defects in the visual field. This is usually, but not exclusively, associated with a raised intraocular pressure. Visual field defects are permanent and if the disease progresses enough, can result

in permanent blindness. It is the second leading cause of blindness globally. However early detection and treatment, can often protect against serious vision lose [13]. Repeated measurements of the visual field over time are required to detect early changes.

Typically, eyes affected by glaucoma have a raised intraocular pressure. This can be related to anatomical factors affecting the depth of the front chamber of the eye. Where the front chamber is too narrow, the condition is termed chronic narrow angle glaucoma (CNAG). Where glaucoma occurs despite no narrowing it is termed primary open angle glaucoma (POAG). POAG accounts for the vast majority of glaucoma cases.

In some cases of glaucoma there are clear signs of optic nerve degeneration when the nerve is examined clinically or imaged. However, there may be no evidence of visual field defect, in which case preperimetric glaucoma is applied. The biological mechanism that allows the visual field to remain intact despite optic nerve damage is not fully understood, as one would expect a strong anatomical:physiological correlation. Imaging of the optic nerve is performed by a technique termed optical coherence tomography (OCT). The most common abnormality of the optic nerves in glaucoma is termed 'cupping'. Some optic nerves display signs of cupping but never develop field defects. In this situation the changes are termed physiologic disc cupping (PDC).

Ocular hypertension (OHT) exists where there is a raised intraocular pressure but no signs of optic nerve damage. A small proportion of people with this will eventually develop glaucoma, so regular monitoring is necessary to detect these cases.

There have been many previous works on the classification of glaucoma by machine classifiers. [7] used a support vector machine using eight parameters achieving a AROC of 0.981. Classifying Healthy, early and advanced glaucoma. Linear discriminant analysis (LDA), generalised linear model (GDM) and generalised additive model (GAM) were also tested. [4] focused on predicting preperimetric glaucoma using a DNN with 52 parameters, two hidden layers and connected with stacked de-noising autoencoders thus. Producing a AUC of 92.6% and distinguishing preperimetric glaucoma visual fields from healthy visual fields.

[19] evaluates the performance of deep convolutional neural networks (DCNNs) for glaucoma discrimination using images producing AUC of 90% or more.

We aim to produce results with less parameters, as found in actual practice due to incomplete data and pre-guideline recording. Diagnosing a wider range of glaucomatous diseases than works, which just distinguish between healthy and glaucomatous visual fields. Whilst pathognomic changes in the intraocular pressure, disc appearance, disc/macula OCT and visual field tests are easily interpreted as glaucomatous, we aim to show that a system of analysing just the visual field and optic disc OCT can have similar sensitivity and specificity to clinical observer. This opens up the possibility of glaucoma diagnosis and monitoring with limited tests (whilst also opening up a discussion around prognostic indicators) and without the need for image analysis of the optic disc appearance (which to date remains a challenge to replicate for automated systems).

This comparative study of Deep Neural Networks, Decision Trees, Support Vector Machine, and k-Nearest Neighbours for the classification of glaucoma will determine which method performs best for this task. The proposed system aims to predict diagnosis for ocular hypertension, chronic glaucoma and healthy patients.

The proceeding sections are organized as follows. Section 2 describes concepts of classification algorithms, including deep neural networks. Section 3 discusses configuration and implementation of these algorithms. Section 4 shows the dataset used for training and testing, detailing demographics and composition. Section 5 discusses the results. Finally, in Section 6 conclusions regarding optimal methods and possible future works are presented.

2 Background on classifiers

2.1 Deep Neural Network

Deep Neural Networks (DNN) also known as deep feedforward neural networks, are a popular method for classification and regression [11]. This bio-inspired method, is modelled on how our brains are structured. Built with layers of neurons (nodes) which react to a given stimulus (input), producing a response. These neurons are modelled by an activation function such as a sigmoid or Gaussian function. This method is trained by adjusting weights which connect the nodes to gain a desired decision surface, so when given input x a prediction of y is given. DNNs are Artificial Neural Networks which have multiple layers of neurons. [10] Doing this increases the networks ability to accurately represent a decision boundary.

For multi-class classification, the Softmax function is used on the output layer. Softmax normalises the output, a discrete variable with n possible values. Equating to a categorical probability distribution, producing the probability of the classification being correct.

2.2 k-Nearest Neighbours

k-Nearest Neighbours (KNN) is a simple algorithm which stores given data and produces predictions based on a similarity measure, such as euclidean distance. A prediction is given using a majority vote based on the k closest data-points. [9]

For this type of deterministic classifier, it is necessary to have a training set which is not too small, and a good discriminating distance. KNN performs well in multi-class simultaneous problem solving. There exists an optimal choice for the value of the parameter k, which results in better performance of the classifier.

2.3 Support Vector Machine

The Support vector machine (SVM) is a kernel based method. Given training examples labelled either "yes" or "no", a maximum-margin hyperplane is identified

which splits the "yes" from the "no" training examples, such that the distance between the hyperplane and the closest examples (the margin) is maximized [12] Since generally a larger margin equates to a lower generalization error.

There is a way to create non-linear classifiers by applying the kernel trick to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a non-linear kernel function.

The margin determines the offset of the vector to the data points γ controls the range in which points are included. These points on the edge of the optimal hyperplane are known as support vectors, giving the method its name.

2.4 Decision Trees

A Decision Tree consists of a set of ordered rules for classifying data. Each node in the tree addresses an input variable. Leaves assign labels or values to the data. Decision trees are quite different to the methods already described as they are not kernel based. They are intuitive when compared to ANNs, due to the fact the reasoning behind a classification can be easily followed.

However if greedy learning algorithms are used then DTs can be prone to getting stuck in local optima. Furthermore algorithms can produce overly complex or large trees that do not generalise well (cause overfitting). Decision trees are constructed by splitting the dataset recursively into sub-sets based on a input variable. This is done until all subsets have the same class, the sets become too small or splitting no longer improves classification [17]. This is measured by one of the following commonly used metrics, Gini impurity or Shannon entropy. They both measure diversity in a set of discrete data, also know as purity [18].

3 Methods

3.1 DNN

The Deep Neural Network architecture used consisted of a 4 layer network with 6 input nodes, two hidden layers with 30 and 20 nodes respectively, and four output nodes (one for each class) using softmax. Input features are detailed in Section 4.1. The hidden layers used a rectified linear unit (ReLU) [15] activation function, as this produced the best results in testing. The AdaGrad algorithm was used, a modified stochastic gradient descent was used for gradient descent optimisation [8]. AdaGrad adapts learning rates by scaling them, inversely proportional to the square root of the sum of all the historic values of the gradient. The effect of this is greater progress in more shallow sloped directions of the parameter space [11]. The addition of a dropout layer was tested, though did not provide an improvement in results. The DNN was simulated using Tensor-Flow [2].

3.2 SVM

The SVM was created with the scikit-learn library [16], as where all algorithms except the DNN. A range of C values and kernels where tested. Kernels tested where sigmoid, radial basis function (RBF), polynomial and linear.

A configuration with a $C = 50$, $\gamma = 1/\text{n\_features}$ and a non-linear RBF activation function, proved to be the most optimal configuration for the model. Out performing over tested configurations.

C is the penalty parameter for the error term. It balances the size of the margin with the amount of data-points correctly classified. It defines how much you want to avoid misclassifying a data-point. The larger the C the smaller the hyperplane of the margin though this hyperplane will give a better decision boundary.

3.3 KNN

The KNN uses $k = 3$ neighbours, using the Minkowski metric with a power of 2 ($p = 2$) which is equivalent to the euclidean distance. Using uniform weights, where all in a neighbourhood are equal.

3.4 DT

The implemented decision tree uses the Gini impurity metric as classification criteria. On creating a split, the best split is taken. A maximum depth was set to avoid all leaves becoming pure. A range of random forests where tested for performance with different number of estimators.

4 Dataset

The initial dataset consisted of 119 patients OCT nerve fibre layer thicknesses and Visual Field indices of sensitivity over three visits starting in 2014/2015. To total 236 eyes. All data was completely anonymized and only included test readings (shared prior to GDPR regulations introduction).

The dataset comprises of multiple diagnoses OHT, PDC, Normal Tension Glaucoma (NTG), POAG, CNAG and Healthy. CNAG and PDC cases where excluded due to limited examples in the dataset, which could lead to inaccurate prediction. After cleaning the dataset size reduced to 196 eyes, Figure 1 shows the exact breakdown of classes.
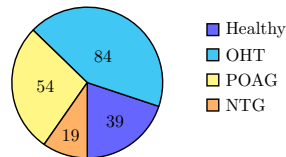
Fig. 1. A breakdown of classes in dataset. A uneven class distribution within the dataset can be seen.

Only the most recent visit readings were used in the dataset for the methods tested, as data was incomplete in previous visits. This was due to the incomplete nature of the earlier sets as not all of the tests and measurements take place at every visit. Any patterns which had missing data where removed. The following Table 1 shows the demographic of the dataset by class.

Parameter	Healthy $(n = 39)$	Glaucoma $(n = 115)$	OHT $(n = 84)$
Gender (male/female)	20/19	57/58	43/41
Age, years (mean±SD)	69 ± 14.10	72 ± 9.56	62 ± 10.88
Visual Field (MD)	-1.78 ± 2.71	-5.51 ± 6.88	-2.16 ± 2.32
Visual Field (PSD)	2.18 ± 1.53	4.26 ± 3.37	2.51 ± 1.40

Table 1. Dataset Demographics

4.1 Feature Selection

Features where selected based on there relevancy to diagnosis, MD (Mean Deviation) and PSD (Pattern Standard Deviation) on the Visual Field, and Superior (Sup), Inferior (Inf), Nasal, Temporal (Temp) average RNFL segment thickness on the disc OCTs. Analysis of principle components (PCA) was carried out though produced the same results as the manually selected features. OCT Disc retinal nerve fibre layer thickness readings were only provided in four segment averages. Due to having limited parameters form the OCT, PCA did not produce better feature selector.

Various methods where tested to normalise the dataset, centering to the mean and doing a component wise scale to the unit variance produced the best results.

A small amount of noise was added to the dataset, this was done to increase the number of examples. This was implemented using a Gaussian noise with a standard deviation of available $\sigma = 0.2$. Wherewe use an original sample X and a virtual sample $X + noise$ The generated noise was added to the whole set doubling the amount of examples. This resulted in a slight improvement in classifier accuracy of a couple percent.

5 Experimental results

All the results are described in terms of sensitivity and specificity in this case they are the weighted averages of each class's sensitivity and specificity. Results are also measured in AUC, also known as AUROC or the area under receiver operating characteristic.

5.1 General chronic glaucoma diagnosis

In this experiment, all type of glaucoma are defined under the same label e.g POAG, CNAG $\rightarrow$ Glaucoma. This was carried out to test the general ability to classify between glaucoma, OHT and healthy cases.

DNN With this dataset a sensitivity of 77.52% and a specificity of 86.63% was achieved, with a k-Fold $k = 5$ cross validation (CV) on the DNN [14]. Figure 2 shows the Receiver Operating Characteristics (ROC) curve for each fold.
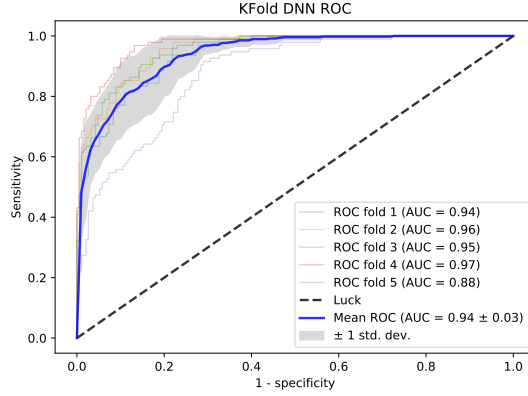
Fig. 2. Receiver operating characteristic for each class using DNN.
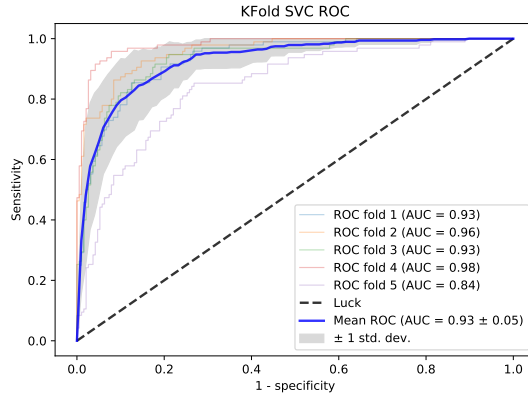
Fig. 3. Receiver operating characteristic for each class using SVM.

SVM The following Table 2 shows the results using LooCV.

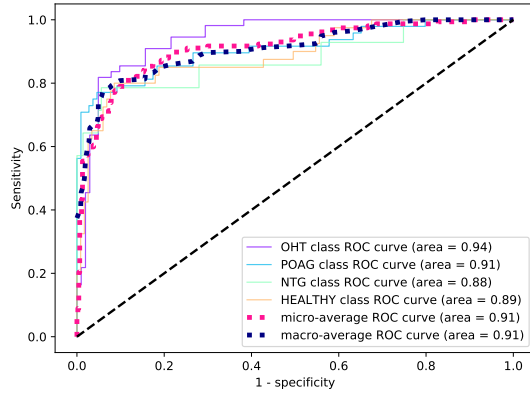| Method	Sensitivity	Specificity	AUC
Voting (SVM + KNN)	81.30 ± 0.13	87.92 ± 0.05	0.85 ± 0.04
SVM	80.67 ± 0.14	86.80 ± 0.06	0.84 ± 0.04
DNN	78.99 ± 0.08	86.87 ± 0.05	0.83 ± 0.02
k-Neighbors	76.68 ± 0.10	86.31 ± 0.04	0.81 ± 0.03
Random Forest	68.07 ± 0.24	77.28 ± 0.12	0.73 ± 0.06
Decision Tree	66.60 ± 0.19	76.71 ± 0.11	0.72 ± 0.04

Table 2. Comparison of classifier results, using LooCV $(k = n)$.

5.2   More detailed glaucoma diagnosis

This experiment aims to gain resolution into diagnosing different types of glaucoma. Labels are kept with the diagnosis as given by doctors, which consists of four classes OHT, POAG, NTG and healthy cases.

SVM With this dataset the best results where produced using a SVM. With a sensitivity of 82.90%, specificity of 92.18% and AUC of 0.87 however as can be seen in Figure 4 the characteristic of the NTG ROC curve with a AUC of 0.88 does not perform as well as other classes. This is due to the low amount of patterns which are contained in the dataset for this class. Which causes a weaker "understanding" in the model of said class. Table 3 shows results for all the classifiers tested.

Fig. 4. Class ROC for SVM. Effects of low NTG class count can be seen.

Method	Sensitivity	Specificity	AUC
Voting (KNN + SVM)	81.37 ± 0.12	91.93 ± 0.03	0.87 ± 0.05
DNN	80.61 ± 0.12	92.33 ± 0.02	0.86 ± 0.05
SVM	79.85 ± 0.11	91.14 ± 0.03	0.85 ± 0.04
k-Neighbors	75.51 ± 0.14	88.16 ± 0.06	0.82 ± 0.04
Decision Tree	64.29 ± 0.34	79.40 ± 0.14	0.72 ± 0.11
Random Forest	63.52 ± 0.28	78.42 ± 0.15	0.71 ± 0.08

Table 3. Comparison of classifier results, using LooCV ($k = n$).

5.3   Voting Classifier

A combination of the SVM and the KNN was used to produce a voting classifier using majority rule voting, in an attempt to improve classification accuracy. As can be seen in Table 3 this was successful.

6   Conclusion

In this paper we show that the proposed multiple system can predict diagnosis for ocular hypertension, chronic glaucoma and healthy patients with a reasonable confidence.

Health services around the globe face challenges dealing with the growing demand placed on them by an increasing prevalence of glaucoma. Glaucoma clinics have particular difficulties compared to other eye diseases, given that the condition is a lifelong disease requiring multiple visits to the hospital to diagnose and monitor. The use of artificial intelligence to streamline the diagnosis of glaucoma means there is improved clinician availability for complex cases.

Furthermore, a system such as the one we have detailed could potentially be used in triaging of community-based referrals to hospital eye services. It could 'flag' a referral that is more likely to be true glaucoma from an ocular hypertension or a healthy case. At present a referral is graded as urgent or routine by an optician, but there is no grading system for 'likelihood of glaucoma'. An alert that could be added to referrals for this would facilitate the task of triaging for the clinician, and improve timely access to glaucoma clinics for patients according to clinical need.

Due to the resolution of data collected, only having segment averages limits the results of this study. If we had the full set of parameters then a deeper analysis could have been carried out to select better features to comprise the input space.

Results are promising with DNN and SVM methods producing diagnosis with similar accuracy to healthcare professionals [3] with a AUC = 87%. This could be improved in further works, with larger sample and increased resolution of data collected.

Acknowledgements

References

1. National eye institute, facts about glaucoma, https://nei.nih.gov/health/glaucoma/glaucoma_facts, Sep 2015.
2. Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.
3. Sabina Andersson, Anders Heijl, Dimitrios Bizios, and Boel Bengtsson. Comparison of clinicians and an artificial neural network regarding accuracy and certainty in performance of visual field assessment for the diagnosis of glaucoma. *Acta ophthalmologica*, 91(5):413–417, 2013.

4. Ryo Asaoka, Hiroshi Murata, Aiko Iwase, and Makoto Araie. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*, 123(9):1974–1980, 2016.

5. Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

6. Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

7. Zvia Burgansky-Eliash, Gadi Wollstein, Tianjiao Chu, Joseph D Ramsey, Clark Glymour, Robert J Noecker, Hiroshi Ishikawa, and Joel S Schuman. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Investigative ophthalmology & visual science*, 46(11):4147–4152, 2005.

8. John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

9. Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

10. John Fulcher. Computational intelligence: an introduction. In *Computational intelligence: a compendium*, pages 3–78. Springer, 2008.

11. Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

12. Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

13. Sharon Kingman. Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization*, 82:887–888, 2004.

14. Frederick Mosteller and John Wilder Tukey. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977.

15. Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

16. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

17. J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

18. Brian D Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.

19. S. Phan, S. Satoh, Y. Yoda, K. Kashiwagi, T. Oshika, J. O. I. R. R. Group, et al. Evaluation of deep convolutional neural networks for glaucoma detection. *Japanese journal of ophthalmology*, 63(3):276–283, 2019.