33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# MRMR Optimized Classification for Automatic Glaucoma Diagnosis

Zhuo Zhang, Chee Keong Kwoh, Jiang Liu, Fengshou Yin, Adrianto Wirawan,
Carol Cheung, Mani Baskaran, Tin Aung, Tien Yin Wong

*Abstract*—**Min-Redundancy Max-Relevance (mRMR) is a feature selection methodology based on information theory. We explore the mRMR principle for automatic glaucoma diagnosis. Optimal candidate feature sets are acquired from a composition of clinical screening data and retinal fundus image data. An mRMR optimized classifier is further trained using the candidate feature sets to find the optimized classifier. We tested the proposed methodology on eye records of 650 subjects collected from Singapore Eye Research Institute. The experimental results demonstrate that the new classifier is much compact by using less than ¼ of the initial feature set. The ranked feature set also enables the clinicians to better access the diagnostic process of the algorithm. The work is a further step towards the advancement of the automatic glaucoma diagnosis.**

## I. Introduction

Glaucoma is a chronic and irreversible neurodegenerative eye condition in which the optic nerve fibers and astrocytes are progressively damaged [1-2]. It is often accompanied by an increased intraocular pressure. As the lost capability of the optic nerve cannot be recovered, early diagnosis and subsequent early treatment [3] are important to preserve the vision of the affected patients.

In clinical practice, glaucoma is diagnosed based on the analysis of patient's medical history, measurement of the intraocular pressure, testing of visual field loss, the manual assessment of the optic nerve head (ONH) via ophthalmoscopy or fundus imaging [4] and etc. However, due to the complexity and variety of the disease pathology, the diagnosis of glaucoma relies heavily on the experiences of the glaucoma specialists. Glaucoma's irreversibility, lacking of glaucoma specialist and patient unawareness demand for an economic, effective and automatic glaucoma screening system.

Traditionally, in an eye screening practice, patients' health records are recorded in a text database. With the popularity of high-resolution digital photography, retinal fundus imaging is widely used in eye screening programs nowadays. A large number of image characteristics can be collected or extracted. Among them, the retinal optic nerve head vertical cup-to-disc ratio (vCDR) has been recognized as an important risk factor for detecting the presence of glaucoma. Some researchers focus their efforts on automatic segmentation of the optic disc and cup [5-7] and estimation of vCDR values. Besides vCDR, retinal images also possess numerous pathological signs that are often referred by ophthalmologists for glaucoma diagnosis. For instances, the following signs usually suggest high possibility of glaucoma: presence of optic disc haemorrhages; thinning of neuroretinal rim (NRR); NRR thickness not following the 'ISNT Rule' [8]; nerve fiber layer defect; presence of peripapillary atrophy (PPA) and notch in NRR etc. Currently, these signs are often manually observed and annotated by trained graders in population-based eye studies.

In an population-based eye study, there are usually hundreds of features collected, which are of different levels of importance and sometimes redundant or even contradictory with each other. Computer Aided Diagnosis (CAD) systems built based on a few features or all features available often fail to give good performance. How to obtain an optimal subset of important and clinically relevant features to build a CAD system is a challenge to researchers and clinicians.

In machine learning, glaucoma diagnosis can be understood as a two class ("glaucoma" and "normal") classification problem, and the features used by the classifier can be selected in many different ways. One scheme is to select features that correlate strongest to the classification variable. However, this subset often contains materials that are relevant but redundant. To remove the redundant features, in this paper, we explore a feature selection scheme called Minimum Redundancy Maximum Relevance (mRMR)[9], a scheme mainly used in genes and phenotypes relevance analysis.

In our mRMR based glaucoma diagnosis approach, the features that have strong relevance on glaucoma and are minimally redundant are recursively selected to generate a ranking of the features. Ranking and selecting both clinical data and image features enable the clinicians to better access the diagnostic process of the algorithm instead of the black-box effect when using too many or too few features.

Experiments are conducted on eye records of 650 patients collected from Singapore Eye Research Institute. The experimental results clearly demonstrate that the new classifier contains only 1/4 of the features and in the mean time outperforms classifiers without mRMR optimization.

Manuscript received Mar 29, 2011. This work was supported in part by the Agency for Science, Technology and Research, Singapore, under SERC grand 092 148 0067.

Zhang Z., Liu J. and Yin F. are with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore (phone: +065-64082163; fax:+065-67761378; email:zzhang@ i2r.a-star.edu.sg).

Kwoh C.K is with Nanyang Technological University, Singapore

Wirawan A. is with Yong Loo Lin School of Medicine, National University of Singapore.

Cheong C., Baskaran M. Aung T. and Wong T.Y. are with Singapore Eye Research Institute and National University of Singapore.

The content of this paper is organized as follows. Section II describes our methodologies for classification optimization with mRMR feature selection. Section III reports the experimental result. Section IV discusses the issues and future work.

## II. METHODOLOGY

A framework for mRMR-based classification optimization is proposed, as illustrated in figure 1. We compose the feature space using information obtained from image segmentation and features extracted from preprocessed medical screening data. MRMR is used as a filter approach to generate a ranked feature pool; we then implement a forward selection wrapper [9] to search for the optimal classifier.

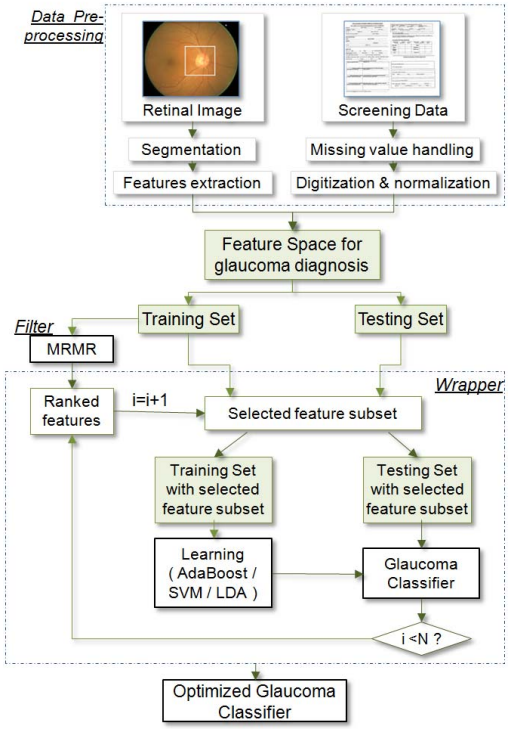



Figure 1. Framework for mRMR based Classification Optimization

### A. Minimum Redundancy–Maximum Relevancy (mRMR)

The mRMR method aims at selecting maximally relevant and minimally redundant set of features for discriminating glaucoma from normal subjects. We use mutual information quotient (MIQ) based mRMR criterion to find a maximally relevant and minimally redundant set of features obtained from retinal image segmentation and medical screening.

Let $D = \{x_{i,j}\}_{N\times M}$ denotes the feature matrix, where $x_{i,j}$ is the value of feature $i$ in subject $j$, $N$ denotes the number of features, and $M$ denotes the number of samples. Let $x_j = (x_{1,j}, x_{2,j}, \ldots x_{N,j})$ denote the $j^{th}$ subject and $x_i^T = (x_{i,1}, x_{i,2}, \ldots . x_{i,M})$ denote the values of $i^{th}$ feature across subjects. Let $F = \{1, 2, \ldots, N\}$ be the indexed set representing the features. In this paper, we address 2-class classification of feature sets into glaucoma or normal. Let the target class label of sample $j$ be $C_j = g \in \{1, -1\}$, taking values 1 for glaucoma and $-1$ for normal, respectively. The mutual information between class labels and feature $i$ will quantify the relevancy of feature $i$ for the classification. The relevancy $Rs$ of features in a subset $S \subset F$ is given by

$$R_S = \frac{1}{|S|}\sum_g \sum_{i\in S} Mu(g,i) \quad (1)$$

where $Mu(g,i) = \sum_{x_i^T} p(g, x_i^T) \log\left(\frac{p(g,x_i^T)}{p(g)p(x_i^T)}\right)$ is the mutual information between class labels $g$ and gene $i$, $p(g)$ and $p(x_i^T)$ represents the probability density function, $p(g, x_i^T)$ is the joint probability density function, and the summation is taken over the space of feature values. The redundancy of a feature subset is determined by the mutual information among the features. The redundancy of feature $i$ with the other features in the subset $n$ is given by

$$Q_{S,i} = \frac{1}{|S|^2}\sum_{i'\in S, i'\neq i} Mu(i,i') \quad (2)$$

In mRMR method, feature ranking is performed by optimizing the ratio of the relevancy of a feature to the redundancy of the features in the set. The maximally relevant and minimally redundant feature $i^*$ in the set $n$ is given by

$$i^* = argmax_{i\in n}\left(\frac{R_n}{Q_{n,i}}\right) \quad (3)$$

MRMR starts by seeking a single feature $x_i$ that is maximally relevant to glaucoma, and includes $x_i$ in the set $S_1$. From the remaining features $X - S_1$, it then seeks next feature that is maximally relevant to target class but minimally redundant with features that have been selected in $S_1$. The selected feature with feature set $S_1$ forms feature set $S_2$. This process iterates until the feature number has reached N. As result, $S_1 \subset S_2 \subset \cdots \subset S_N$, are supplied as the candidate feature sets for supervised learning process.

### B. Build classifiers via supervised learning

Supervised learning technologies infer classifier from supervised training data. Good learning algorithms should be able to generalize from the training data to unseen situations in a "reasonable" way, e.g. to learn how clinicians make the decision on glaucoma diagnosis.

We select several supervised classifiers for our classification task to discriminate glaucoma subjects from normal ones. All classifiers were implemented in R statistical software [10].

#### 1) Adaptive Boosting for glaucoma classification

Adaptive boosting (AdaBoost) was introduced in 1995 by Freund and Schapire [11]. It is a serial ensemble approach that builds an additive model. It begins by training a weak learner on a data set to generate a hypothesis $H_1$. The distribution of weights of the training samples is updated by a function of the classification error. This ensures that misclassified subjects have larger weights so that the next hypothesis $H_2$ is generated by training a weak classier on the same set of samples again but with the updated weight distribution. This process continues iteratively until a target error bound or maximum number of rounds has been reached. The final hypothesis $H$ is formed by linearly combining the set of hypotheses $(H_1, H_2, \ldots H_t)$ generated at each round with their weighted votes.

*2) Support Vector Machines for glaucoma classification*

Support vector machines (SVM) was introduced by Vapnik [12] in 1995 and since then became a popular method in pattern recognition. In this study, we use SVM to map the multidimensional parameters into a feature space and to create a hyperplane to separate the glaucoma and normal subjects with maximal distance between all subjects and the hyperplane.

*3) Linear discriminant analysis*

Linear discriminant analysis (LDA) [13] is used in statistics and machine learning to find a linear combination of features which characterize or separate two or more classes of objects or events. It assumes Gaussian distribution of the data, and separates the data with linear discrimination boundaries that maximize the variance between the two classes while minimizing the variance within classes. Each new data point is classified based on the likelihood generated by each of the categories (glaucoma or normal).

### C. MRMR optimized classifiers

According to Occam's Razor, one should include model complexity when evaluating a model. In this study, we search for the optimal classifier using the following two criteria: 1) Feature set compactness – evaluated by number of features; 2) Classification performance – judged by receiver operating characteristic (ROC) via calculating the area under curve (AUC). The classification optimization is achieved via feature selection. In machine learning, feature selection algorithms fall into three categories: Embedded, Wrappers and Filters. Embedded methods perform feature selection in the process of training, while wrappers use learning machines of interest as a black box to score subsets of features according to their predictive power. mRMR is a filter approach, independently of the chosen learning machines.

According to [9], wrappers can be implemented for a 2-stage optimization to find the compact feature set. As the greedy wrapper approach may be trapped by local minimum, we performe exhaustive search in the whole search space $S_1 \subset S_2 \subset \cdots \subset S_N$ to identify the most compact feature set.

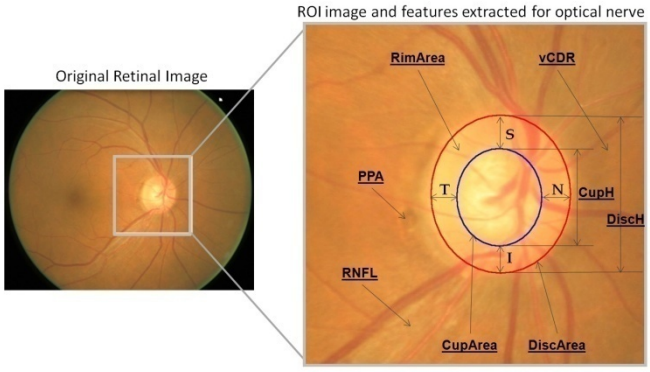
## III. Experiment and Result

### A. Data Set

The presented work is based on Retinal image data obtained from ORIGA [14] database and medical screening data collected in a population based study, Singapore Malay Eye Study (SiMES) [15]. Named as image data and screening data respectively.

SiMES is a population-based study conducted from 2004 to 2007, aiming to assess the causes and risk factors of blindness and visual impairment in Singapore Malay community. It falls into 4 categories: 1) demographical data such as age, gender, height; 2) medical histories data acquired via interview; 3) ocular examination data, e.g. intro-ocular pressure (IOP) and cornea thickness etc; 4) Retinal images taken by retinal fundus camera. Moreover, diagnostic information such as glaucoma and cataract were identified by clinicians based on the data from screening. Categories 1-3 are used as our screen dataset and diagnostic information is used as class labels in our study, e.g. 'normal' or 'glaucoma'.

ORIGA contains 650 retinal images randomly selected from SiMES study (the 4th category mentioned above). The images were segmented semi-automatically and verified by a group of professionals from Singapore Eye Research Institute. Figure 2 illustrates some features obtained from image segmentation, which possess valuable information for glaucoma diagnosis. For example, one can use I-S-N-T values to check the compliance of ISNT rule, which means a normal eye's neuro-retinal rim is thickest Inferiorly and thinnest Temporally. We combined the image data and screening data by matching their subject IDs. The data is preprocessed by removing features with missing values, digitizing and normalizing the values before further steps. The fused feature space contains 104 features in total, from which 19 features are from retinal image and 85 features are from screening data.




a b

Figure 2. Features extracted from retinal image segmentation. a. Original Retinal Image. b. Region of Interest (ROI) image for optic nerve head analysis. RimArea - area of neuroretinal rim; vCDR - vertical cup-disc ratio; PPA - peripapillary atrophy; I – rim inferior thickness; S – rim superior thickness; N – rim nasal thickness; T – rim temporal thickness; RNFL - retinal nerve fiber layer defect; CupArea – area of optic cup; DiscArea – area of optic disc; cupH – height of optic cup; DiscH – height of optic disc

### B. Classifiers build on full feature set

LDA, SVM and AdaBoost classifiers are build using image features, screening data features and fused full feature set respectively. Table I illustrates the performance of different classifiers evaluated via 10-fold cross validation. The result shows that, using only screening data, the classifier can barely

TABLE I
CLASSIFIERS PERFORMANCE IN 10-FOLD CROSS VALIDATION

| classifier | Dataset | AUC | Accuracy |
|---|---|---|---|
| **AdaBoost** | Image Feature | 0.886 | 79.4% |
| | Screening data | 0.61 | 64.1% |
| | **Full Feature Set** | **0.916** | **85.3%** |
| SVM | Image Feature | 0.889 | 79.9% |
| | Screening data | 0.58 | 59.3% |
| | Full Feature Set | 0.90 | 82.4% |
| LDA | Image Feature | 0.883 | 80.1% |
| | Screening data | 0.57 | 61.3% |
| | Full Feature Set | 0.88 | 81.2% |
| vCDR | | 0.868 | 78.5% |

discriminate glaucoma from normal. However, add screening data to image data, the classifier build upon full dataset outperforms classifier build on image features only, which outperform vCDR-only based prediction.

### C. Comparison of mRMR-classifiers

We use mRMR to generate ranked feature set, the feature sets are then used for building classifiers in a forward selection manner. Figure 3 shows the classification performance evaluation based on AUC against size of feature sets. At feature set size of 23, i.e., both SVM and LDA classifiers achieve an AUC of 0.92, which, is equivalent to the AUC achieved by AdaBoost learning from a feature set with 104 features.

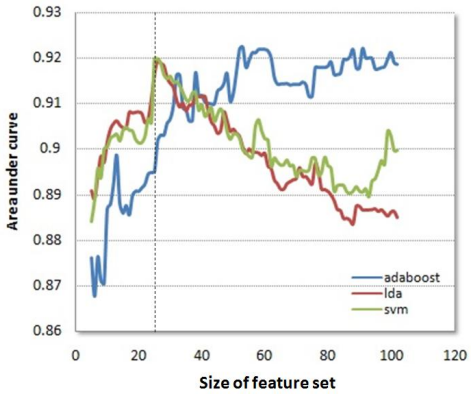



Figure 3. Classifier performance evaluation on feature sets. Dotted line shows at a feature set of size 23, SVM and LAD classifier achieve a optimal module with the AUC equivalent to AdaBoost classifier build on 104 features.

The 23-feature classifier is the optimal module we search for. Table 2 lists the features for the optimized classification module. It shows that vCDR is indeed the number one critical factor for glaucoma diagnosis, in the mean time, ocular hypertension, age etc are also important determinants.

## IV. Discussion and Conclusion

In clinical practice, glaucoma diagnosis is based on multi-factors taking into consideration of patient's medical history, ocular examination as well as the assessment of the optic nerve head via ophthalmoscopy. Combining multiple data sources for computer-aid diagnosis can better simulate the clinical decision making process in the real world.

One of the limitations of supervised learning is their black box manner. With mRMR, not only a simpler module can be trained, but more importantly, a clear list of features can be presented to clinicians so that it's much easier to explain what the classifier learn from the data and where the expert knowledge is come from. Furthermore, the features extracted can guide the mass screening process in glaucoma early detection, leading to reduced information to be collected. The proposed framework can be applied to other eye diseases such as cataract and retinopathy.

TABLE II
Top 23 Features from final feature set Of Glaucoma Classifier

| no | Feature symbol | Description | no | Feature symbol | Description |
|---|---|---|---|---|---|
| 1 | vCDR | vertical cup-disc ratio | 13 | ocular_htnr | right eye ocular hypertension |
| 2 | ocular_htnl | ocular hypertension | 14 | job_cat | job category |
| 3 | alphaPPA | alpha PPA at optic nerve | 15 | anisometropia | anisometropia indication |
| 4 | cupArea | area of optic cup | 16 | l_drlevel | diabetic retinopathy level |
| 5 | I | rim inferior width | 17 | N | rim nasal width |
| 6 | glyn | medical history of glaucoma | 18 | eyehist_re | eye history of retinopathy |
| 7 | T | rim temporal width | 19 | bestPPA | PPA info |
| 8 | ISNT | ISNT rule compliance | 20 | ocular_htn | overall ocular hypertension |
| 9 | cupHeight | optic cup height | 21 | discArea | optic disc area |
| 10 | S | rim superior width | 22 | mi | self reported heart attack |
| 11 | RNFL | retinal nerve fiber layer defect | 23 | race_cat | race category |
| 12 | age | age of screening | | | |

## References


[1] A. L. Coleman, "Glaucoma," *Lancet,* vol. 354, pp. 1803-10, Nov 20 1999.

[2] L. M. Weih, *et al.*, "Prevalence and predictors of open-angle glaucoma: results from the visual impairment project," *Ophthalmology,* vol. 108, pp. 1966-72, Nov 2001.

[3] G. Michelson, *et al.*, "The papilla as screening parameter for early diagnosis of glaucoma," *Dtsch Arztebl Int,* vol. 105, pp. 583-9, Aug 2008.

[4] S. C. Lin, *et al.*, "Optic nerve head and retinal nerve fiber layer analysis: a report by the American Academy of Ophthalmology," *Ophthalmology,* vol. 114, pp. 1937-49, Oct 2007.

[5] N. Inoue, *et al.*, "Development of a simple diagnostic method for the glaucoma using ocular Fundus pictures," *2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-7,* pp. 3355-3358, 2005.

[6] J. Liu, *et al.*, "ARGALI : An Automatic Cup-to-Disc Ratio Measurement System for Glaucoma Analysis Using Level-set Image Processing," in *13th International Conference on Biomedical Engineering, Vols 1-3*, 2009, pp. 559-562.

[7] D. W. Wong, *et al.*, "Intelligent fusion of cup-to-disc ratio determination methods for glaucoma detection in ARGALI," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2009, pp. 5777-80, 2009.

[8] C. f. E. R. Australia, *Assessment of Glaucomatous Optic Disk Signs*.

[9] H. C. Peng, *et al.*, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 1226-1238, Aug 2005.

[10] "R Project: http://cran.r-project.org."

[11] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences,* vol. 55, pp. 119-139, Aug 1997.

[12] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning,* vol. 20, pp. 273-297, Sep 1995.

[13] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics,* vol. 7, pp. 179-188, 1936.

[14] Z. Zhang, *et al.*, "ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2010, pp. 3065-8, 2010.

[15] T. Y. Wong, *et al.*, "Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES)," *Ophthalmic Epidemiology,* vol. 14, pp. 25-35, 2007.