

Glaucoma Detection by Learning from Multiple Informatics Domains

Yanwu Xu¹, Lixin Duan¹, Damon Wing Kee Wong¹, Tien Yin Wong^{2,3}, and Jiang Liu¹

¹ Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore

² Singapore Eye Research Institute, Singapore

³ Department of Ophthalmology, National University of Singapore, Singapore

Abstract. We present a comprehensive and fully automatic glaucoma detection approach that uses machine learning techniques over multiple informatics domains, consisting of personal profile data, genetic data, and retinal image data. This approach, referred to as MKL^{clm}, enriches the feature set of the multiple kernel learning (MKL) framework through the incorporation of *classemes*, which represent the outputs of multiple class-specific classifiers trained from the data of each informatics domain. We validate our MKL^{clm} framework on a population-based dataset consisting of 2258 subjects, achieving an AUC of 94.9% ± 1.7% and a specificity of 88.5% ± 2.7% at 85% sensitivity, which is significantly better than the current clinical standard of care which uses intraocular pressure (IOP) for glaucoma detection. The experiments also demonstrate that MKL^{clm} outperforms the standard SVM method using data from individual domains, as well as the traditional MKL method, showing that this deeper integration of data from different informatics domains can lead to significant gains in holistic glaucoma diagnosis and screening.

1 Introduction

Glaucoma is a disease of the optic nerve which is one of the leading causes of blindness worldwide. It is estimated to affect up to 80 million people by 2020 [1]. Based on a recent survey by the World Health Organization, glaucoma is the second leading cause of blindness, and it is considered the top cause of irreversible and permanent visual impairment. In glaucoma, the degeneration of the retinal ganglion nerve cells results in a gradual loss of sight inwards from the periphery of the visual field of view, leading to the characteristic ‘tunnel-vision’ experienced in cases of advanced glaucoma. Due to the lack of obvious visual symptoms in glaucoma, recent studies have reported that more than 90% of glaucoma cases remain undetected in the population [2].

With the permanence of visual loss in glaucoma, there is a strong need to develop models of disease prediction to detect the disease early to minimize vision loss. Broadly, there are three types of data: personal profile data, genetic data, and imaging data. Profile data, which includes both ocular measurements as well as non-ocular data such as age, has been previously used to develop predictive models for glaucoma [3, 4]. Specifically in [3], age, intraocular pressure (IOP), central corneal thickness, vertical cup to disk ratio and visual field pattern parameters were used to obtain a five-year glaucoma risk. Similarly, perimetric measurements were used in the Kaplan-Meier model to

estimate the probability of developing glaucoma in [4]. Besides personal profile data, the human genome also contains a wealth of predictive information. With the large increase in processing power for large scale data analysis, there has been renewed interest to mine the expanse of data in genome-wide association studies to identify genetic risk factors for disease prediction. Some of the recent studies for certain diseases have found several genes [5] being associated with the development of glaucoma. Besides personal profile data and genetic information, researchers also use imaging devices such as fundus cameras to examine retina for glaucomatous changes. As structural glaucomatous degeneration occurs before functional visual loss, a direct observation of retina provide an opportunity to detect early stages of glaucoma. A number of works have been reported for automatic glaucoma assessment in retinal image analysis, such as [6] which presented a glaucoma risk index based on visual features extracted from retinal images.

Many methods have been developed for automatic glaucoma detection, but they can only use data from the same source to train predictive models. How to effectively integrate data from different informatics domains is not yet well studied. Recently, researchers employed multiple kernel learning (MKL) techniques as a means to fuse data from different domains for medical image classification. For example, Gál *et al.* proposed to combine both visual features and meta-textual information from medical images to determine their modality [15]. More recently, Liu *et al.* applied the traditional MKL method to integrate data from the three informatics domains (i.e., personal data, genome data, and retinal image data) for glaucoma detection [14].

Motivated by [14], we propose an improved MKL framework that effectively incorporates pre-learned classifiers (e.g., binary SVM classifiers) which can be leveraged as useful priors for further boosting classification performance. Specifically, we consider the outputs of pre-learned SVM classifiers on the data as additional features. The prior knowledge represented by these classifiers is used to more deeply integrate the information from different domains and also train a more effective prediction model for glaucoma detection. Another contribution of our work is the design of a new visual feature for retinal image representation, which will be shown to outperform the visual feature used in [14]. This new visual feature in conjunction with the classifier-based priors is shown to bring appreciable performance improvements to the MKL framework, and has the potential to be applied not only to glaucoma detection, but also for diagnosing other diseases that have indicators over different domains.

2 Comprehensive MKL-based Glaucoma Detection

In this work, we learn a predictive model for glaucoma detection by using multiple informatics domains under the proposed MKL^{clm} framework. As illustrated in Fig. 2, our framework first trains several base SVM classifiers using different kernel functions and parameters for each individual feature. Together with the pre-computed base kernels using the original features, the decision values from these classifiers then serve as additional features in training the final MKL^{clm} classifier.

2.1 Feature representation of data from different domains

Following [14], we conducted a univariate analysis of all demographic and clinical variables that compose the personal profile data. A total of 46 variables are then used as a 46-dimensional vector for profile features. According to a recent genetics study [5],

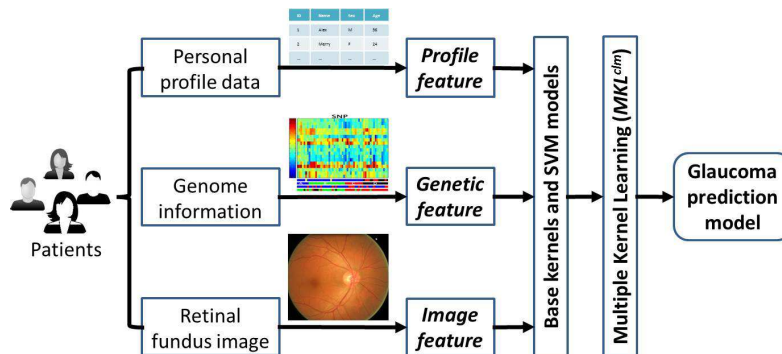


Fig. 1. Overview of the proposed MKL^{clm} framework for glaucoma detection using multiple informatics domains.

three single nucleotide polymorphism (SNP) markers (rs11024102 in PLEKHA7, rs3753841 in COL11A1 and rs1015213) were found to have significant associations with glaucoma. In total, the study identified 178 relevant SNPs, which are used as genetic features for glaucoma assessment in our work.

To represent image data, Liu *et al.* [14] modeled each retinal image as a 569-dimensional vector consisting of the standard deviations of color and texture descriptors within image blocks. This simple representation of local variances, however, lacks clinical support as a discriminant feature for glaucoma detection. Unlike that, in this work we extract a new image feature that is closely correlated to optic cup-to-disc ratio (CDR), which is a major visual cue for glaucoma diagnosis in clinical practice [7]. Since CDR is costly to obtain by manual labeling, we instead incorporate features that are indicative of its value. At present, disc segmentation can be computed reliably, but cup segmentation remains a challenge because current automatic methods either lack sufficient accuracy [9] or are time-consuming [8]. Inspired by the sliding window and ranking-based automatic cup localization method in [8], histogram-based features in sliding windows (which represent virtual cups) can be trained to estimate how close the virtual cup is to the real one, and thus are correlated with the actual CDR. For efficiency, we extract simplified histogram-based features from four virtual cups in the optic disc automatically segmented using the level-set method in [9], with following steps:

1. Each optic disc is normalized to a uniform circle with a diameter of 256 pixels, with the cup and disc centers aligned. Virtual cups with CDR values $\rho \in \{0.2, 0.4, 0.6, 0.8\}$ are then generated.
2. As in [8], histograms are obtained from the green channel with different bin numbers $\beta \in \{8, 16, 32, 64\}$ such that each bin has an equal (or as equal as possible due to quantization) number of pixels, giving equalized channels.
3. For each number of bins β and each virtual cup with CDR ρ , we form two types of features: 1) an L1-normalized histogram of the virtual cup region; 2) an L1-normalized histogram of the virtual rim (*i.e.*, non-cup) region within the disc;
4. The final image feature \mathbf{I}_i is obtained by concatenating the two types of features over the four CDR values and four bin numbers, which leads to a feature dimension of $|\mathbf{I}_i| = 2 \times 4 \times (8 + 16 + 32 + 64) = 960$.

Feature normalization In our framework, feature normalization is individually performed for profile features, genetic features, and image features. Specifically, for each feature dimension, the corresponding value is normalized to the range of $[0, 1]$ in order to avoid bias caused by magnitude differences among the feature dimensions.

2.2 Multiple kernel learning using classemes

Having enhanced interpretability over a single kernel, multiple kernel learning (MKL) [10] assumes that the final kernel is a linear combination of a set of base kernels. Promising results of MKL have been shown in the literature [14, 15]. Motivated by [14], we propose an improved MKL framework to better integrate the aforementioned three types of features from multiple informatics domains. Specifically, our framework proposes to utilize the outputs of pre-learned class-specific classifiers, referred to as *classemes* in [13], to further improve performance for glaucoma detection. The intuition for using classemes is that the glaucoma prediction of a patient \mathbf{x} from a pre-learned classifier f_p basically describes how likely \mathbf{x} has glaucoma. Thus, such classemes provide a rich encoding of the informatics data and thus can be considered as valuable prior information. Because of this use of classemes in the MKL framework, we denote our method as MKL^{clm} .

The goal of our work is to learn a binary MKL^{clm} classifier by using multiple features from different informatics domains for glaucoma detection. Suppose for each patient \mathbf{x}_i , we have M types of features $\mathbf{x}_i = [\mathbf{x}_i^{\{1\}\top}, \dots, \mathbf{x}_i^{\{M\}\top}]^\top$, where $\mathbf{x}_i^{\{m\}}$ represents the m -th feature. Let us denote the training samples as $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $y_i \in \{1, -1\}$ is the class label for which $y_i = 1$ means a patient \mathbf{x}_i has glaucoma and $y_i = -1$ otherwise. We organize the training samples so that the first n_+ samples are positive and the remaining n_- are negative, where $n = n_+ + n_-$. We also denote the set of pre-learned classifiers by $\{f_p\}_{p=1}^P$, which are learned from standard SVMs using the RBF kernels of different types of features as well as different kernel parameters (see the experimental setup in Section 3.2 for more details). Moreover, for each patient, we augment the set of original features \mathbf{x}_i by including the classemes from the P pre-learned classifiers, which leads to the augmented feature representation $\mathbf{z}_i = [\mathbf{x}_i^\top, \mathbf{f}_i^\top]^\top$, where $\mathbf{f}_i = [f_1(\mathbf{x}_i), \dots, f_P(\mathbf{x}_i)]^\top$. By introducing nonlinear feature mapping functions $\{\phi_m(\cdot)\}_{m=1}^M$ for $\mathbf{x}_i^{\{m\}$ ¹ and $\varphi(\cdot)$ for \mathbf{f}_i , we map \mathbf{z}_i into $\psi(\mathbf{z}_i) = [\phi_1(\mathbf{x}_i^{\{1\}\top}), \dots, \phi_M(\mathbf{x}_i^{\{M\}\top}), \varphi(\mathbf{f}_i)^\top]^\top$. With the aforementioned notations, we formulate the classifier for our classification problem as

$$f(\mathbf{z}_i) = \mathbf{w}^\top \psi(\mathbf{z}_i) + b = \sum_{m=1}^M \mathbf{w}_m^\top \phi_m(\mathbf{x}_i^{\{m\}}) + \mathbf{w}_f^\top \varphi(\mathbf{f}_i) + b, \quad (1)$$

where $\mathbf{w} = [\mathbf{w}_1^\top, \dots, \mathbf{w}_M^\top, \mathbf{w}_f^\top]^\top$, \mathbf{w}_m and \mathbf{w}_f are the feature weights for the m -th feature and the classemes, respectively, and b is the bias term. Then the formulation of MKL^{clm} is as follows:

¹ Without loss of generality and for simplicity, each feature $\mathbf{x}_i^{\{m\}}$ here is associated with one feature mapping function $\phi_m(\cdot)$. However, in the experiments of Section 3, we will use seven multiple feature mappings (leading to seven multiple base kernels) for each $\mathbf{x}_i^{\{m\}}$.

$$\begin{aligned}
 \min_{\mathbf{w}_m, \mathbf{w}_f, \mathbf{d}, b, \xi_i} & \frac{1}{2} \sum_{m=1}^M \frac{\|\mathbf{w}_m\|^2}{d_m} + \frac{\mu}{2} \|\mathbf{w}_f\|^2 + \frac{\theta}{2} \|\mathbf{d}\|^2 + C \left(\frac{1}{n_+} \sum_{i=1}^{n_+} \xi_i + \frac{1}{n_-} \sum_{i=n_++1}^n \xi_i \right), \quad (2) \\
 \text{s.t.} & y_i f(\mathbf{z}_i) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n, \quad (3) \\
 & \sum_{m=1}^M d_m = 1, \quad d_m \geq 0, \quad m = 1, \dots, M,
 \end{aligned}$$

where $\mathbf{d} = [d_1, \dots, d_M]^\top$, d_m controls the squared norm of \mathbf{w}_m [10], $\mu, \theta > 0$ are pre-defined parameters which respectively regularize the complexity of \mathbf{w}_f and \mathbf{d} , $C > 0$ is also a pre-defined parameter that balances the objective and empirical error, and ξ_i is the slack variable for each training sample. Note that we normalize the empirical errors of the positive and negative data in order to avoid the imbalance problem. We also note that μ controls the influence of prior information in MKL^{clm} . A larger μ will lead to a smaller \mathbf{w}_f , which reduces the impact of prior information in MKL^{clm} .

To solve (2), we first introduce a dual variable α_i for each constraint in (3). Then the dual form of (2) with respect to $\mathbf{w}_m, \mathbf{w}_f, b$ and ξ_i can be expressed as

$$\begin{aligned}
 \min_{\mathbf{d}} \max_{\boldsymbol{\alpha}} & \frac{\theta}{2} \|\mathbf{d}\|^2 + \mathbf{1}^\top \boldsymbol{\alpha} - \frac{1}{2} (\boldsymbol{\alpha} \circ \mathbf{y})^\top \tilde{\mathbf{K}} (\boldsymbol{\alpha} \circ \mathbf{y}), \quad (4) \\
 \text{s.t.} & \boldsymbol{\alpha}^\top \mathbf{y} = 0, \quad 0 \leq \alpha_i \leq C/n_+, \quad i = 1, \dots, n_+; \\
 & 0 \leq \alpha_i \leq C/n_-, \quad i = n_+ + 1, \dots, n,
 \end{aligned}$$

where the notation \circ represents the element-wise product operator, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_n]^\top$, $\mathbf{y} = [y_1, \dots, y_n]^\top$, and $\tilde{\mathbf{K}} \in \mathbb{R}^{n \times n}$ is the kernel matrix with elements $k(\mathbf{z}_i, \mathbf{z}_j) = \sum_{m=1}^M d_m k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}_j^{\{m\}}) + \frac{1}{\mu} \cdot k_f(\mathbf{f}_i, \mathbf{f}_j)$. Here $k_m(\cdot, \cdot)$ is the m -th base kernel function, which induces the nonlinear feature mapping $\phi_m(\cdot)$ (i.e., $k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}_j^{\{m\}}) = \phi_m(\mathbf{x}_i^{\{m\}})^\top \phi_m(\mathbf{x}_j^{\{m\}})$). And $k_f(\cdot, \cdot)$ is also a kernel function, which induces $\varphi(\cdot)$ (i.e., $k_f(\mathbf{f}_i, \mathbf{f}_j) = \varphi(\mathbf{f}_i)^\top \varphi(\mathbf{f}_j)$) for classes.

Similar to [10], we develop an algorithm to solve the optimization problem in (4) by iteratively updating \mathbf{d} and $\boldsymbol{\alpha}$. Due to space constraints, we only briefly describe the proposed algorithm. First, by fixing \mathbf{d} , we obtain a standard SVM problem and solve it by using existing tools such as LIBSVM. Subsequently, we fix $\boldsymbol{\alpha}$ and solve a quadratic programming problem for \mathbf{d} . We iterate the above procedures until the value of the objective function in (4) converges. After obtaining the optimal \mathbf{d} and $\boldsymbol{\alpha}$, we can rewrite the decision function in (1) as follows:

$$f(\mathbf{z}) = \sum_{i=1}^n \alpha_i y_i \left(\sum_{m=1}^M d_m k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}^{\{m\}}) + \frac{1}{\mu} \sum_{p=1}^P k_f(\mathbf{f}_i, \mathbf{f}) \right) + b,$$

where \mathbf{z} is the augmented feature of a test patient, $\mathbf{x}^{\{m\}}$ is its m -th feature, and $\mathbf{f} = [f_1(\mathbf{x}), \dots, f_P(\mathbf{x})]^\top$ is a vector of its classes.

3 Experiments

To assess the performance of the proposed framework using multiple informatics domains, we compare results of seven settings that use different features and combina-

tions: 1) profile features (P); 2) genetic features (G); 3) retinal image features (I); 4) $P+G$; 5) $P+I$; 6) $G+I$; and 7) $P+G+I$. The notation ‘+’ means that data features from multiple informatics domains are used for both the training and testing phases.

3.1 Dataset description

We use the Singapore Malay Eye Study (SiMES) dataset in the experiments. SiMES is a population-based study conducted from 2004 to 2007 to assess the causes and risk factors of blindness and visual impairment in the Singapore Malay community [2]. The database contains 2258 subjects each with complete personal profile data, genome data and retinal fundus image data. In addition, diagnostic information of glaucoma is available, and is used as the class label (i.e., +1 for glaucoma and -1 for normal) in our experiments. Since IOP is clinically used for glaucoma detection, it is employed as a baseline method in our experiments and removed from the personal profile data.

3.2 Experimental setup

Besides IOP, we also compare our proposed MKL^{clm} with the standard SVM and the traditional MKL presented in [14]. For all the methods, we use a nonlinear radial basis function (RBF) kernel, i.e., $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, where we set the kernel parameter as $\gamma = 2^\sigma\gamma_0$, with γ_0 set as the reciprocal of the average squared distance between all pairs of training samples. In the experiments, σ is set as $\{-3, -2, -1, 0, 1, 2, 3\}$, so that for each type of feature we have seven RBF kernels. For the standard SVM, we train one SVM classifier by using one RBF kernel constructed using one type of feature. For each of the aforementioned settings 1)–3), we obtain the final SVM classifier as an average of the seven trained SVM classifiers; and for settings 4)–7), we average fourteen (or twenty-one) trained SVM classifiers obtained from two (or three) types of corresponding features to get the final SVM classifier. For MKL and our MKL^{clm} , we construct a number of base kernels by using the RBF kernel with each type of feature and seven kernel parameters. So we have seven base kernels for settings 1)–3), fourteen for settings 4)–6) and twenty-one for setting 7). Note that those base kernels are exactly the same ones used to train each SVM classifier.

To evaluate different methods, balanced accuracy \bar{P} and the area under the ROC (receiver operating characteristic) curve (AUC) are used as evaluation metrics, where the ROC is plotted as a curve which shows the tradeoff between sensitivity P_+ and specificity P_- . As the screening set point, we maintain a baseline specificity P_- of 85% to limit the rate of false negatives, and determine the corresponding balanced accuracy \bar{P} of the various methods.

In the experiments, we randomly partition the dataset into training and testing subsets which contain 80% and 20% of the 2258 subjects, respectively. For each machine learning method, five-fold cross-validation is performed to automatically select the parameters C from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ during the training phase. And the final classifier of each method is obtained by training this method again using all the training data together with the selected parameters found through cross-validation. For fair and thorough evaluations of the different methods, we conduct ten tests with random partitioning of the training and testing data. Finally, we report the means and standard deviations of both AUC and balanced accuracy \bar{P} over the ten rounds of experiments for each method.

Table 1. Performance comparisons of different methods under different settings on the SiMES dataset. The best results over all the settings are highlighted in boldface. Note that intraocular pressure (IOP) is the glaucoma assessment method currently used in clinics. IOP only achieves an AUC of 0.572 ± 0.062 and a \bar{P} of 0.565 ± 0.04 .

Setting	SVM		MKL		MKL ^{clm}	
	AUC	\bar{P}	AUC	\bar{P}	AUC	\bar{P}
1) <i>P</i>	0.634±0.054	0.547±0.036	0.615±0.05	0.522±0.02	0.621±0.05	0.532±0.03
2) <i>G</i>	0.922±0.022	0.835±0.041	0.923±0.02	0.842±0.05	0.922±0.02	0.832±0.05
3) <i>I</i>	0.815±0.031	0.727±0.057	0.809±0.03	0.730±0.03	0.814±0.03	0.740±0.04
4) <i>P+G</i>	0.852±0.048	0.757±0.065	0.916±0.025	0.830±0.050	0.926±0.024	0.840±0.047
5) <i>P+I</i>	0.816±0.030	0.715±0.049	0.816±0.026	0.730±0.044	0.822±0.029	0.727±0.045
6) <i>G+I</i>	0.925±0.021	0.822±0.048	0.935±0.020	0.850±0.037	0.945±0.017	0.872±0.028
7) <i>P+G+I</i>	0.913±0.025	0.832±0.031	0.933±0.019	0.847±0.034	0.949±0.017	0.885±0.027

3.3 Performance comparisons

Table 1 lists the means and standard deviations of both AUC and \bar{P} for each method under the seven settings. From the results, the following observations can be made:

1. SVM, MKL [14] and our MKL^{clm} generally perform much better than the current clinical assessment method based on intraocular pressure (IOP) in terms of both AUC and \bar{P} . This clearly demonstrates the potential for clinical usage of machine learning methods using additional fundus image data and genome information.
2. When using individual features under settings 1)–3), SVM performs comparably or even better than the two MKL-based methods. An explanation is that even though the base kernels are different (using different kernel parameters), they are all constructed by using the same feature, which in return makes them intrinsically similar. As a result, the MKL-based methods do not achieve better performance than SVM. However, when using two or three types of features under settings 4)–7), the MKL-based methods outperform SVM in almost all cases. It is because that the three types of features are quite different in nature and are used in a complementary manner in the MKL framework, which helps learn a better predictive model.
3. By utilizing classes as prior information, our proposed MKL^{clm} framework achieves better performance than the MKL method [14] under almost all the settings, which clearly demonstrates the effectiveness of our MKL^{clm} in better integrating the heterogeneous data from different domains through the seamless incorporation of classes into the MKL-based framework. Using all the domains in setting 7), MKL^{clm} achieves the best result overall.

We also conduct one more experiment to compare the performance of our newly developed image feature to that used in [14]. In this comparison, we train standard SVMs with each image feature. Using our visual feature (see setting 3 in Table 1), SVM performs much better than when using the visual feature in [14] which only achieves a 0.752 ± 0.04 AUC and 0.642 ± 0.04 \bar{P} .

4 Conclusion

To perform fully automatic and comprehensive glaucoma detection, we proposed an improved framework called MKL^{clm} to learn an effective predictive model by using

multiple informatics domains containing data of different nature (i.e., personal profile data, genetic data and retinal image data). Different from traditional MKL, we incorporate classemes (i.e., outputs of pre-learned SVM classifiers) into MKL^{clm} and use them as additional features to boost the classification performance of the predictive model. Moreover, we developed a new visual feature for retinal image representation, and it is demonstrated to perform better than the one used in [14]. MKL^{clm} is validated by extensive experiments conducted on the large population-based SiMES dataset, and its strong performance demonstrates its ability to effectively integrate heterogeneous data from different informatics domains.

References

1. Quigley H.A., Broman A.T.: The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 90(3), 262–7 (2006)
2. Shen S.Y., Wong T.Y., Foster P.J., Loo J.L., Rosman M., Loon S.C., Wong W.L., Saw S.M., Aung T.: The prevalence and types of glaucoma in malay people: the Singapore Malay Eye Study. *Invest Ophthalmol Vis Sci* 49(9), 3846–51 (2008)
3. Gordon M.O., Torri V., Miglior S., *et al.*: Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology* 114(1), 10–9 (2007)
4. Hattenhauer M.G., Johnson D.H., Ing H.H., Herman D.C., Hodge D.O., Yawn B.P., Butterfield L.C., Gray D.T.: The probability of blindness from open-angle glaucoma. *Ophthalmology* 105(11), 2099–104 (1998)
5. Vithana E.N., Khor C.C., *et al.*: Genome-wide association analyses identify three new susceptibility loci for primary angle closure glaucoma. *Nature Genet* 44(10), 1142–6 (2012)
6. Bock R., Meier J., Nyúl L.G., Hornegger J., Michelson G.: Glaucoma risk index: automated glaucoma detection from color fundus images. *Med Image Anal* 14(3), 471–81 (2010)
7. Jonas J., Budde W., Panda-Jonas S.: Ophthalmoscopic evaluation of the optic nerve head. *Survey of Ophthalmology* 43, 293–320 (1999)
8. Xu Y., Xu D., Lin S., Liu J., Cheng J., Cheung C.Y., Aung T., Wong T.Y.: Efficient optic cup detection from intra-image learning with retinal structure priors. In: *MICCAI* 1–8 (2011)
9. Wong D.W.K., Liu J., Lim J.H., Tan N.M., Zhang Z., Lu S., Li H., Teo M.H., Chan K.L., Wong T.Y.: Intelligent fusion of cup-to-disc ratio determination methods for glaucoma detection in ARGALI. In: *IEEE Int. Conf. Engin. in Med. and Biol. Soc. 2009*, 577780 (2009)
10. Rakotomamonjy A., Bach F.R., Canu S., Grandvalet Y.: SimpleMKL. *Journal of Machine Learning Research* 9, 2491–521 (2008)
11. Vedaldi A., Gulshan V., Varma M., Zisserman A.: Multiple kernels for object detection. In: *ICCV* 606–13 (2009)
12. Duan L., Xu D., Tsang I.W., Luo J.: Visual event recognition in videos by learning from web data. In: *CVPR* 1959–66 (2010)
13. Torresani L., Szummer M., Fitzgibbon A.: Efficient object category recognition using classemes. In: *ECCV* 776–89 (2010)
14. Liu J., Zhang Z., Wong D.W.K., Xu Y., Yin F., Cheng J., Tan N.M., Kwok C.K., Xu D., Tham Y.C., Aung T., Wong T.Y.: Automatic glaucoma diagnosis through medical imaging informatics. *Journal of American Medical Informatics Association* 20(6), 1021–27 (2013)
15. V. Gál E. K., Nachtgaeel M.: Multiple kernel learning based modality classification for medical images. In: *CVPR* 76–83 (2012)