Discriminative Feature Selection for Multiple Ocular Diseases Classification by Sparse Induced Graph Regularized Group Lasso

Xiangyu Chen[1], Yanwu Xu[1], Shuicheng Yan[2], Tat-Seng Chua[2], Damon Wing Kee Wong[1], Tien Yin Wong[2], and Jiang Liu[1]

[1] Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
[2] National University of Singapore, Singapore

Abstract. Glaucoma, Pathological Myopia (PM), and Age-related Macular Degeneration (AMD) are three leading ocular diseases worldwide. Visual features extracted from retinal fundus images have been increasingly used for detecting these three diseases. In this paper, we present a discriminative feature selection model based on multi-task learning, which imposes the exclusive group lasso regularization for competitive sparse feature selection and the graph Laplacian regularization to embed the correlations among multiple diseases. Moreover, this multi-task linear discriminative model is able to simultaneously select sparse features and detect multiple ocular diseases. Extensive experiments are conducted to validate the proposed framework on the *SiMES* dataset. From the Area Under Curve (AUC) results in multiple ocular diseases classification, our method is shown to outperform the state-of-the-art algorithms.

1 Introduction

Many of the leading causes of vision impairment and blindness worldwide are irreversible and cannot be cured [1]. Glaucoma, Pathological Myopia (PM), and Age-related Macular Degeneration (AMD) are three leading ocular diseases worldwide. Early detection of these ocular diseases utilizing effective visual features is highly needed [2][11].

With the advancement of retinal fundus imaging, several computer-aided diagnosis (CAD) methods and systems have been developed to automatically detect these three leading ocular diseases from retinal fundus images [6][4]. However, current work mainly focus on detecting Glaucoma, PM, and AMD individually. Classifying these three leading diseases simultaneously is still an open research direction. There are some correlations among these three leading ocular diseases. In recent decades, the problem of low vision and blindness in elderly people became a major and socially significant issue. The number of patients having age-related macular degeneration (AMD) in association with glaucoma is growing all over the world [8], which attaches great medical and social value to this multiple diseases diagnosis problem. Moreover, in a recent study, myopic eyes are less likely to have AMD and diabetic retinopathy (DR) but more likely to have nuclear cataracts and glaucoma [9].

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 11–19, 2015.
DOI: 10.1007/978-3-319-24571-3_2

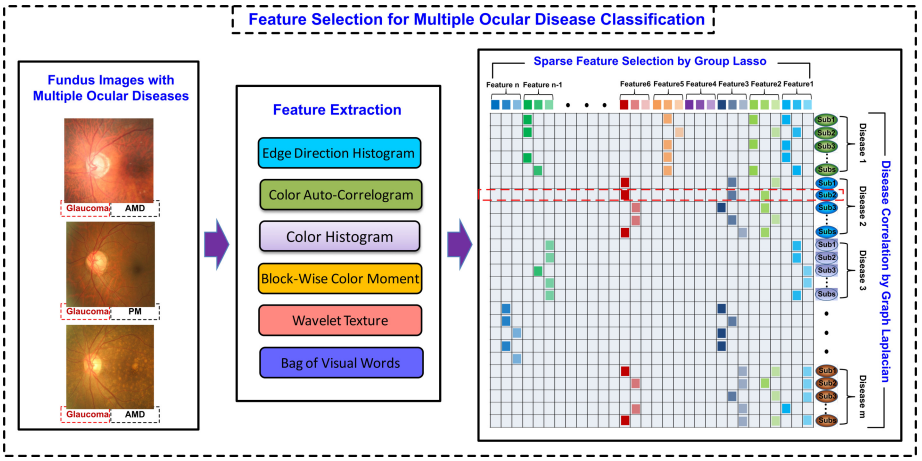

Fig. 1. System overview of our proposed discriminative feature selection scheme for multiple ocular diseases classification. Here "Sub1" stands for a sub-type disease of a leading ocular disease. As indicated within red dashed borders in the right of the figure, a sparse feature will be learned for each disease. For better viewing, please see the color pdf file.

In this paper, we adopt a Multi-task Learning (MTL) based method for discriminatively selecting sparse features, harmoniously integrating the correlation information of multiple diseases, and investigating the problem of learning to simultaneously diagnose them for a given fundus image. Different from previous algorithms that detect ocular disease independently, the proposed method formulates the correlation information of different diseases by a graph Laplacian regularizer and models the sparse feature selection by an exclusive group lasso regularizer. It then utilizes a multi-task linear model to learn a linear mapping from features to diseases. Since a patient may have two or three ocular diseases at the same time, multiple ocular diseases detection is well suited for real-world diagnosis scenarios.

2 Discriminative Feature Selection by Graph Regularized Group Lasso

2.1 Graph Regularized Group Lasso Multi-task Learning

For a multiple ocular disease dataset $\{x_i, l_i\}_{i=1}^N$, $x_i \in R^d$ is the feature vector of the i-th fundus image and $l_i = \{l_i^k\}_{k=1}^K$ is the associated disease label. For simplicity and clarity, we set $K = 2$ and denote the Cartesian product of $\mathcal{L}^1$ and $\mathcal{L}^2$ as $\mathcal{L} = \mathcal{L}^1 \times \mathcal{L}^2$, as done in [10]. In addition, $y_i \in R^{|\mathcal{L}|}$ is the zero-one disease label vector ($|\mathcal{L}| = |\mathcal{L}^1| \times |\mathcal{L}^2|$) indicating whether x_i is jointly labeled as $l^1 \in \mathcal{L}^1$ and $l^2 \in \mathcal{L}^2$. For the case of multiple ocular diseases, $|\mathcal{L}^1|$ is equal to the number of types of ocular diseases and $|\mathcal{L}^2|$ stands for the number of sub-types of each disease. For example, since AMD is classified into dry AMD (non-exudative

AMD) and wet AMD (exudative AMD), $|\mathcal{L}^2|$ is equal to 2. Given the training feature-label set $\{x_i, y_i\}_{i=1}^N$, the disease label of a test image can be predicted via learning a linear model $y = Mx$. Our employed learning model is based on the following basic MTL formulation [5]:

$$\Theta(M) = \frac{1}{2}\sum_{j=1}^{|\mathcal{L}|} \|Y_j - M_j X\|^2, \tag{1}$$

where $Y_j \in \mathbb{R}^n$ and $M_j \in \mathbb{R}^d$ are the j-th row of Y and M, respectively. $X = [x_1, ..., x_n] \in \mathbb{R}^{d\times n}$ is the feature matrix with each column representing a training image feature, $Y = [y_1, ..., y_n] \in \mathbb{R}^{|\mathcal{L}|\times n}$ is the label matrix with each column as a training image label vector, and $M \in \mathbb{R}^{|\mathcal{L}|\times d}$ is the parameter to be estimated. We are to learn $|\mathcal{L}|$ different linear regression models (tasks) $Y_j = M_j X$, $j = 1, ..., |\mathcal{L}|$. In this naive formulation, the tasks are learned independently of each other.

In order to learn the discriminative features for multiple diseases detection, we consider the relationships across tasks by imposing two regularizers (*Exclusive Group Lasso Regularizer and Graph Laplacian Regularizer*) to the objective $\Theta(M)$ in (1) as in [10]. Before we start to utilize the improved objective $\Theta(M)$, let us introduce some notation for group lasso and graph Laplacian. Let $\mathcal{G}^1$ of size $|\mathcal{L}^1|$ be a group of label index sets in $\mathcal{L}$ constructed as follows: each element $g \in \mathcal{G}^1$ is an index set of combinational labels $(l^1, l^2) \in \mathcal{L}$ which share a common $l^1 \in \mathcal{L}^1$. And $\mathcal{G}^2$ of size $|\mathcal{L}^2|$ is constructed with label set $\mathcal{L}^2$. Given a similarity matrix $S \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{L}|}$ that stores the pairwise similarity scores between concepts, the larger S_{jk} is, the more similar two concepts j and k are, and vice versa. Hence, the objective $\Theta(M)$ in (1) can be written as:

$$\begin{aligned}\Theta(M) = \frac{1}{2}\sum_{j=1}^{|\mathcal{L}|} \|Y_j - M_j X\|^2 + \mu \sum_{i=1}^{d} \left(\|M^i_{\mathcal{G}^1}\|^2_{2,1} + \|M^i_{\mathcal{G}^2}\|^2_{2,1}\right) \\ + \eta \sum_{j,k=1}^{|\mathcal{L}|} S_{jk} \parallel M_j - M_k \parallel^2,\end{aligned} \tag{2}$$

where $\sum_{i=1}^{d} \left(\|M^i_{\mathcal{G}^1}\|^2_{2,1} + \|M^i_{\mathcal{G}^2}\|^2_{2,1}\right)$ is the *Exclusive Group Lasso Regularizer* (denoted as $\Gamma(M)$) and $\sum_{j,k=1}^{|\mathcal{L}|} S_{jk} \parallel M_j - M_k \parallel^2$ is the *Graph Laplacian Regularizer*. $\|M^i_{\mathcal{G}^k}\|^2_{2,1} = \left(\sum_{g\in\mathcal{G}^k} \|M^i_g\|_2\right)^2$, $k = 1, 2$, and $M^i \in \mathbb{R}^{|\mathcal{L}|}$ is the i-th column of M, and $M^i_g \in \mathbb{R}^{|\mathcal{L}|}$ is the restriction of vector M^i on the subset g by setting $M^i_j = 0$ for $j \neq g$.

According to [13], $\|M^i_{\mathcal{G}^k}\|^2_{2,1}$ is sparseness inducing and it encourages exclusive selection of features at the level of group $g \in \mathcal{G}^k$. Hence, for each feature i, it tends to assign larger weights to some important groups while assigning small or even zero weights to the other groups. Different from the *Exclusive Group Lasso Regularizer* that describes the negative correlation among tasks, the *Graph Laplacian*

Regularizer models the positive correlation among tasks by transferring the weight information among multiple ocular diseases.

2.2 Solution

The objective $\Theta(M)$ in (2) is convex but non-smooth since all the three components are convex whereas the *Exclusive Group Lasso Regularizer* $\Gamma(M)$ is non-smooth. The non-smooth structure of $\Gamma(M)$ makes the optimization of problem $\min_M \{\Theta(M)\}$ a non-trivial task. The subgradient method as used in [13] is applicable but it typically ignores the structure of the problem and suffers from a slow rate of convergence. As indicated in [14], the optimization can be achieved by approximating the original non-smooth objective by a smooth function, and then solving the latter by utilizing some off-the-shelf fast algorithms. Here, we introduce a Nesterov-like smoothing optimization method [3] to achieve this purpose. Once the optimal parameter M^* is obtained, the multiple diseases labels of a test fundus image with feature x are learned by $y = M^*x$. The whole algorithm of the proposed feature selection and multiple diseases detection is described in Algorithm 1.

For any vector $z \in \mathbb{R}^n$, its ℓ_2-norm $\|z\|_2$ has a max-structure representation $\|z\|_2 = \max_{\|v\|_2 \leq 1} \langle z, v \rangle$. Based on this simple property and the smoothing approximation techniques originally from [3], function $\Gamma(M)$ can be approximated by the following smooth function:

$$\Gamma_\delta(M) = \frac{1}{2} \sum_{i=1}^{d} \left(z^2_{\mathcal{G}^1,\delta}(M^i) + z^2_{\mathcal{G}^2,\delta}(M^i) \right), \tag{3}$$

where $z_{\mathcal{G}^k,\delta}(M^i) := \max_{\|V^{i,k}_{\mathcal{G}^k}\|_{2,\infty} \leq 1} \langle M^i, V^{i,k} \rangle - \frac{\delta}{2} \|V^{i,k}\|_2^2$, and δ is a parameter to control the approximation accuracy. We use the following Theorem [10] to show that $\Gamma_\delta(M)$ is differentiable and its gradient can be easily calculated.

Algorithm 1. Discriminative Feature Selection for Multiple Ocular Diseases Detection

Input: Feature matrix of training data $X \in \mathbb{R}^{d \times n}$, diseases label matrix $Y \in \mathbb{R}^{|\mathcal{L}| \times d}$, groups $\mathcal{G}^1$ and $\mathcal{G}^2$, parameters μ, η, δ.
Output: Learned feature selection matrix $M^t \in \mathbb{R}^{|\mathcal{L}| \times d}$
Initialization: Initialize M_0, V_0 and let $\alpha_0 \leftarrow 1$, $t \leftarrow 0$.
repeat
  $U_t = (1 - \alpha_t) M_t + \alpha_t V_t$,
  Calculate $\nabla \Gamma_\delta(U_t)$ based on (4), (5), and obtain L_δ according to (6).
  $V_{t+1} = V_t - \frac{1}{\alpha_t L_\delta} \left(-(Y - MX)X^T + \mu \nabla \Gamma_\delta(U_t) + \eta L_\delta M \right)$,
  $M_{t+1} = (1 - \alpha_t) M_t + \alpha_t V_{t+1}$,
  $\alpha_{t+1} = \frac{2}{t+1}$, $t \leftarrow t + 1$.
until Convergence

Table 1. The Baseline Algorithms.

Name	Methods
KNN	k-Nearest Neighbors
SVM	Support Vector Machine
LNP	Linear Neighborhood Propagation [10]
MTL	Naive Multi-task Learning in (1)
SPM	State-of-the-art algorithm for PM Detection [4]
SAMD	State-of-the-art algorithm for AMD Detection [15]
SGL	State-of-the-art algorithm for Glaucoma Detection [2]
GRML	State-of-the-art algorithm for Multiple Ocular Diseases Detection [7]

Theorem 1. *Function $\Gamma_\delta(M)$ is well defined, convex and continuously differentiable with gradient*

$$\nabla\Gamma_\delta(M) = \left[\nabla\Gamma_\delta(M^1), ..., \nabla\Gamma_\delta(M^d)\right], \tag{4}$$

where for $i = 1, ..., d$,

$$\nabla\Gamma_\delta(M^i) = q_{\mathcal{G}^1,\delta}(M^i)V^{i,1}(M^i) + q_{\mathcal{G}^2,\delta}(M^i)V^{i,2}(M^i). \tag{5}$$

Moreover, $\nabla\Gamma_\delta(M)$ is Lipschitz continuous with the constant

$$L_\delta = \left(\frac{2\sqrt{2}R}{\delta} + |\mathcal{L}^1|^2 + |\mathcal{L}^2|^2\right) d. \tag{6}$$

3 Experiments

3.1 Dataset and Evaluation Criteria

We conduct the experiments on the SiMES dataset [12], in which the detection of Glaucoma, AMD, and PM have been made by clinicians. We choose a subset of SiMES for experiments, which contains 2,258 subjects. In this subset dataset, there are 100 with glaucoma, 122 with AMD, and 58 with PM. For each disease, the distribution of the subjects who contracted the disease in the selected dataset is representative of the disease prevalence in the population. We extract three different types of visual features (popular global and local features as adopted in [7]) for conducting experiments, and study the performance of the proposed approach with a total of four settings: 1) global features; 2) grid-based features; 3) bag of words; 4) global features + grid-based features + bag of words. The notation + indicates a combination of four types of features in the corresponding setting. We utilize the area under the curve (AUC) of the receiver operation characteristic curve (ROC) to evaluate the performance of classification.

3.2 Experiment Analysis

In order to evaluate the performance of our proposed algorithm, we provide a quantitative study on SiMES, with an emphasis on the comparison with eight

state-of-the-art related methods as listed in Table 1. Below are the parameters and the adopted values for each method:

- For the SVM algorithm, we adopt the RBF kernel. For its two parameters γ and C, we set $\gamma = 0.6$ and $C = 1$ in the experiments after fine tuning.
- For KNN, there is only one parameter k for tuning, which stands for the number of nearest neighbors and is set to 150.
- For SGL, SPM, SAMD and GRML, we use the same settings as in their papers.
- For our proposed approach, we set $|\mathcal{L}^1| = 3$ and $|\mathcal{L}^2| = 1$ because the selected dataset lacks ground truth labels of the sub-types of each ocular disease for each fundus image. For the control parameters, we set $\mu = 0.5$, $\eta = 0.2$,

The AUCs of the baseline methods for detecting the three leading ocular diseases on the SiMES dataset are illustrated in Table 2. The combined visual features (global features + grid-based features + bag of words) are utilized for

Table 2. The AUCs of different algorithms for simultaneously detecting Glaucoma, PM and AMD on the SiMES dataset. The combined visual features (global features + grid-based features + bag of words) are utilized by the eight baseline methods. The AUC results marked in boldface are significantly better than the others.

Methods	KNN	SVM	LNP	MTL	SGL	SPM	SAMD	GRML	Our Proposed
Glaucoma	74.2 %	76.7 %	78.8%	80.0%	81.0%	-	-	82.5%	**85.3**%
PM	86.5 %	89.1 %	90.1%	90.6%	-	91.0%	-	92.3%	**94.4**%
AMD	72.9 %	75.0%	76.6%	77.0%	-	-	77.8%	79.3%	**81.8**%

Table 3. The AUCs of different algorithms under four settings of features on the SiMES dataset for Glaucoma diagnosis. The AUC results marked in boldface are significantly better than the others.

Methods	KNN	SVM	LNP	MTL	GRML	Our Proposed
Global Features	71.2 %	73.5 %	75.2%	77.5%	78.7%	79.6%
Grid based Features	69.1 %	71.2 %	73.0%	75.4%	76.7%	78.0%
Bag of Words	68.4 %	70.9%	72.6%	73.9%	75.0%	76.6%
Combined Features	**74.2** %	**76.7**%	**78.8** %	**80.0**%	**82.5**%	**85.3**%

Table 4. The AUCs of different algorithms under three settings of features on the SiMES dataset for AMD diagnosis. The AUC results marked in boldface are significantly better than the others.

Methods	KNN	SVM	LNP	MTL	GRML	Our Proposed
Global Features	70.2 %	72.5%	73.9 %	75.0%	76.4%	76.9%
Grid based Features	69.3 %	71.8 %	72.5%	73.8%	75.1%	75.8%
Bag of Words	68.1 %	70.3%	71.6%	72.1%	73.5%	74.9%
Combined Features	**72.9** %	**75.0**%	**76.6** %	**77.0**%	**79.3**%	**81.8**%

Table 5. The AUCs of different algorithms under three settings of features on the SiMES dataset for PM diagnosis. The AUC results marked in boldface are significantly better than the others.

Methods	KNN	SVM	LNP	MTL	GRML	Our Proposed
Global Features	81.5 %	84.1%	85.6 %	86.4%	87.3%	89.1%
Grid based Features	79.3 %	82.3 %	83.7%	84.2%	85.2%	87.3%
Bag of Words	83.8 %	86.5%	87.9%	88.4%	89.5%	91.6%
Combined Features	**86.5** %	**89.1**%	**90.1** %	**90.6**%	**92.3**%	**94.4**%

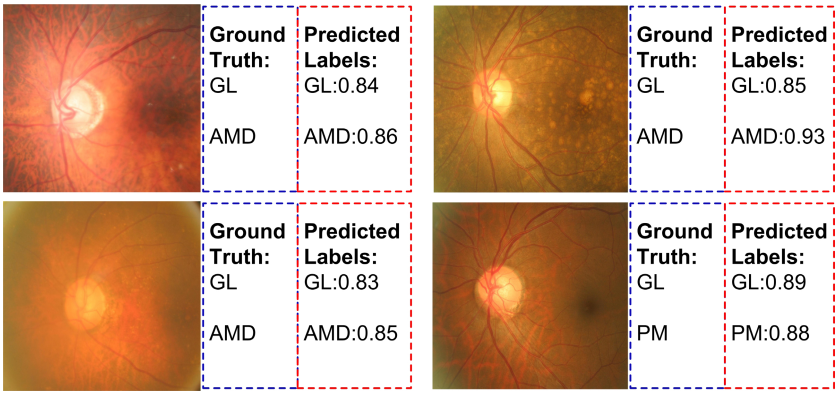

Fig. 2. Sample diagnosis results from our proposed algorithm.

all methods in this experiment. Our proposed algorithm performs the feature selection on the combined visual features, which outperforms the other baseline algorithms significantly. For example, our proposed method has an improvement of 11.2% over SVM, 15.0% over KNN, and 8.2% over LNP for detecting Glaucoma. For PM, our proposed method has an improvement of 5.9% over SVM, 9.1% over KNN, and 4.2% over LNP. For AMD, Ours has an improvement of 9.1% over SVM, 12.2% over KNN, and 6.8% over LNP. Compared with the state-of-the-art algorithms of individual disease detection, the proposed method outperforms SGL, SPM, and SAMD by achieving an AUC of 85.3%, 94.4%, 81.8%, respectively. The improvement stems from the fact that our proposed algorithm encodes the exclusive group lasso regularization for competitive sparse feature selection, and the graph Laplacian regularization to impose the correlations among multiple ocular diseases. GRML considers inter-label constraints but does not model the exact correlations of diseases in the objective formulation. Moreover, GRML is based on genetic features and is not able to select discriminative features for each disease.

The comparison results for detection performance under the four feature settings are listed in Table 3, 4, 5. Since the state-of-the-art detection algorithms (SGL, SPM, SAMD) of individual ocular diseases are based on their own special visual features and retinal structures, the AUC results are not given in

these tables. From Table 3, we are able to observe that, for glaucoma detection, our proposed algorithm outperforms the five baseline algorithms based on the combined features. The AUC of the receiver operating characteristic curve in glaucoma detection is 85.3%. Similar results are shown in Table 4 and 5 for AMD and PM detection respectively. Figure 2 gives four sample results by our proposed algorithm. Each fundus image is presented with the ground truth diagnosed by clinicians and the predicted labels with probabilities by our algorithm. GL stands for Glaucoma. Though the second row has low image quality, our method still detects the glaucoma, PM, and AMD, indicating its robustness and stability.

4 Conclusion

In this paper, we develop a discriminative feature selection scheme for multiple ocular diseases classification. We formulate this challenging problem as a multi-task discriminative analysis model, where individual tasks are defined by learning the linear discriminative model for each disease. We considered all the tasks in a joint manner by utilizing two types of regularization on parameters, namely the graph Laplacian regularization and exclusive group lasso regularization. A Nesterov-type smoothing approximation method is adopted for model optimization. In the future, we will attach a few sub-categories to each category of the three leading ocular diseases to expand our classification range.

References

1. Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. Br. J. Ophthalmol. 90(3), 262–267 (2006)
2. Xu, Y., Lin, S., Wong, D.W.K., Liu, J., Xu, D.: Efficient reconstruction-based optic cup localization for glaucoma screening. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 445–452. Springer, Heidelberg (2013)
3. Nesterov, Y.: Smooth minimization of non-smooth functions. Mathematical Programming 103(1), 127–152 (2005)
4. Xu, Y., Liu, J., Zhang, Z., Tan, N.M., Wong, D.W.K., Saw, S.M., Wong, T.Y.: Learn to recognize pathological myopia in fundus images using bag-of-feature and sparse learning approach. In: International Symposium on Biomedical Imaging (2013)
5. Yuan, X., Yan, S.: Visual classification with multi-task joint sparse representation. In: Computer Vision and Pattern Recognition (CVPR) (2010)
6. Bressler, N.M., Bressler, S.B., Fine, S.L.: Age-related macular degeneration. Survey of Ophthalmology 32(6), 375–413 (1988)
7. Chen, X., Xu, Y., Duan, L., Zhang, Z., Wong, D.W.K., Liu, J.: Multiple ocular diseases detection by graph regularized multi-label learning. In: Proceedings of the Ophthalmic Medical Image Analysis First International Workshop (OMIA) Held in Conjunction with MICCAI (2014)

8. Avetisov, S.E., Erichev, V.P., Budzinskaia, M.V., Karpilova, M.A., Gurova, I.V., Shchegoleva, I.V., Chikun, E.A.: Age-related macular degeneration and glaucoma: intraocular pressure monitoring after intravitreal injections. Vestn. Oftalmol. 128(6), 3–5 (2012)
9. Pan, C.W., Cheung, C.Y., Aung, T., Cheung, C.M., Zheng, Y.F., Wu, R.Y., Mitchell, P., Lavanya, R., Baskaran, M., Wang, J.J., Wong, T.Y., Saw, S.M.: Differential associations of myopia with major age-related eye diseases: the Singapore Indian Eye Study. Ophthalmol. 20(2), 284–291 (2013)
10. Chen, X., Yuan, X., Yan, S., Tang, J., Rui, Y., Chua, T.-S.: Towards multi-semantic image annotation with graph regularized exclusive group lasso. In: ACM Multimedia (2011)
11. Xu, Y., Duan, L., Lin, S., Chen, X., Wong, D.W.K., Wong, T.Y., Liu, J.: Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 788–795. Springer, Heidelberg (2014)
12. Foong, A.W.P., Saw, S.M., Loo, J.L., Shen, S., Loon, S.C., Rosman, M., Aung, T., Tan, D.T.H., Tai, E.S., Wong, T.Y.: Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). Ophthalmic Epidemiology 14(1), 25–35 (2007)
13. Zhou, Y., Jin, R., Hoi, S.C.: Exclusive lasso for multi-task feature selection. In: International Conference on Artificial Intelligence and Statistics (AISTATS) (2010)
14. Chen, X., Yuan, X., Chen, Q., Yan, S., Chua, T.-S.: Multi-label visual classification with label exclusive context. In: International Conference on Computer Vision (ICCV) (2011)
15. Wong, D.W.K., Liu, J., Cheng, X., Zhang, J., Yin, F., Bhargava, M., Cheung, C.M.G., Wong, T.Y.: THALIA - An automatic hierarchical analysis system to detect drusen lesion images for amd assessment. In: ISBI, pp. 884–887 (2013)