Incorporating Privileged Genetic Information for Fundus Image Based Glaucoma Detection

Lixin Duan[1], Yanwu Xu[1], Wen Li[2], Lin Chen[1], Damon Wing Kee Wong[1], Tien Yin Wong[3], and Jiang Liu[1]

[1] Institute for Infocomm Research, Singapore
[2] Nanyang Technological University, Singapore
[3] National University of Singapore, Singapore

Abstract. Visual features extracted from retinal fundus images have been increasingly used for glaucoma detection, as those images are generally easy to acquire. In recent years, genetic researchers have found that some single nucleic polymorphisms (SNPs) play important roles in the manifestation of glaucoma and also show superiority over fundus images for glaucoma detection. In this work, we propose to use the SNPs to form the so-called privileged information and deal with a practical problem where both fundus images and privileged genetic information exist for the training subjects, while the test objects only have fundus images. To solve this problem, we present an effective approach based on the learning using privileged information (LUPI) paradigm to train a predictive model for the image visual features. Extensive experiments demonstrate the usefulness of our approach in incorporating genetic information for fundus image based glaucoma detection.

1 Introduction

Glaucoma, one of the leading causes of blindness worldwide, is a chronic and irreversible neurodegenerative disease. It usually results in deterioration of vision, due to the progressive damage of the optic nerve in a patient. It is reported that, by estimation, up to 80 million people will contract glaucoma by the year of 2020 [7]. Researchers have proposed various methods to facilitate a fully automatical process in order to make fast and accurate glaucoma detection in the early stage. By this means, glaucoma patients will have chance to appoint an early time for clinical treatment and thus prevent it from going any worse.

Nowadays, the fast development of clinical hardware technology makes it generally easy and economical to take images of patients' retinal fundus. Because of such convenient acquisition, fundus image based glaucoma detection has attracted increasing attention from researchers. Till now, many works have been proposed by using fundus images [1–3, 13]. However, the performance is still not very satisfactory yet, because the important depth information, which is used to determine the optic nerves head structure (i.e., cupping) in anatomy, is missing

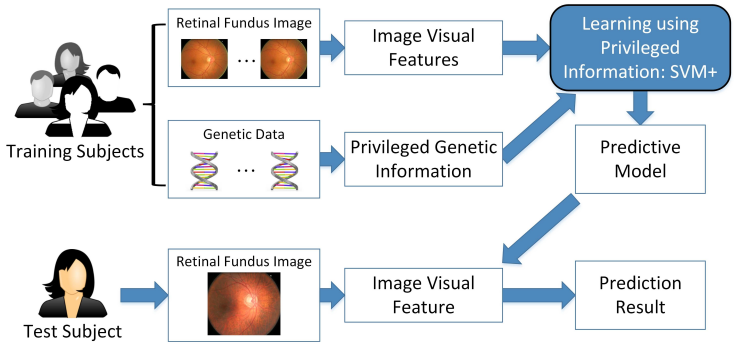

Fig. 1. Flowchart of our proposed approach for fundus image based glaucoma detection.

in the 2D fundus images. Moreover, the quality of fundus images is sometimes heavily affected by imaging device configurations, which prevents people from extracting discriminative visual features from the fundus images.

To help improve performance, other people are looking for other data modalities. Recently, the studies on certain diseases have found that some single nucleotide polymorphisms (SNPs) from genes have high associations with the development of glaucoma [12]. Based on this study, a more recent work [6] shows that a support vector machine (SVM) classifier trained by using SNPs as genetic features outperforms the SVM classifier trained by using visual features extracted from fundus images. This interesting result motivates us to additionally consider incorporating the genetic information (i.e., SNPs) for fundus image based glaucoma detection. However, as it is generally much easier to acquire fundus images than SNPs, especially in underdeveloped districts, we may face with the problem that only fundus images for newly coming subjects (i.e., test subjects) can be obtained. In spite of that, we can leverage SNPs of existing subjects (i.e., training subjects) to help train a better predictive model.

In this work, we propose to solve a practical problem where a training subject has both genetic data and fundus image, while a test subject only has one fundus image. We note that our problem is different from multi-modality fusion problems [5, 6] in which both the training and test subjects must have the same data modalities. And the trained model cannot make prediction on any test subject, unless all of its data modalities are provided. To train one model for each modality, multi-view learning methods have been proposed. As a special case, existing two-view learning methods [4, 9] can be applied to our problem to train two models for genetic data and fundus image data, respectively, and only the model for image data will be used for testing. However, existing two-view learning methods usually assume there is a strong correspondence between the two views of data. Otherwise, the trained models will generally not perform well. Considering that the genetic view is totally different from the image view and the genetic view is much better than the image view [6], it is likely that the two-view learning methods may not achieve good performance in our problem.

Unlike the two-view learning methods [4, 9], in this work we propose a novel approach by first extracting the so-called *privileged genetic information* from the genetic data. And such privileged information is then used to model the slack variable in the hinge loss for the training subjects based on visual features in a max-margin formulation, under the learning using privileged information (LUPI) paradigm [11]. After training the predictive model for fundus images, we use it to make prediction on a given test subject by using its visual feature. Fig. 1 illustrates our overall approach. We extensively evaluate our approach on the Singapore Malay Eye Study (SiMES) database and demonstrate its effectiveness in incorporating the privileged genetic information.

2 Learning Using Privileged Genetic Information

In this section, we first introduce how we extract the privileged information from genetic data and then develop a max-margin method to minimize the averaged empirical error on the training subjects. Before we start, let us define some notations to be used in this paper. the operator $^\top$ represents the transpose of a vector or matrix. $\mathbf{0}$ (resp., $\mathbf{1}$) denotes a column vector of all zeros (resp., ones). And also, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^\top \geq 0$ means $\alpha_i \geq 0, \forall i = 0, \ldots, n$.

2.1 Privileged genetic information

In the literature of glaucoma detection, Liu *et al.* [6] has shown that the genetic data, consisting of SNPs, greatly outperforms the visual features extracted from retinal fundus images, which seems to be a good alternative to detect glaucoma, compared to using fundus images. However, in underdeveloped districts, it is impractical yet to perform genotyping for test subjects, while acquiring their fundus images is still feasible. Although the test subjects lack the genetic data, we can leverage already existing sources with additional genetic data. Specifically in this work, we propose to use the genetic data of the training subjects as important prior information to help train a good predictive model, together with the commonly used visual features extracted from retinal images. The intuition is two-fold: i) Genetic data are better than image data [6]; and ii) genetic data are totally different from image data, which can bring complementary discriminative knowledge for the model training.

As indicated in [11], using outputs of some pre-learned SVM classifier as privileged information show better performance than using original data features. Motivated by this, we propose to extract the privileged information from the genetic data through SVM. Specifically, we first use a set of genetic data $\{(\mathbf{z}_i, y_i)|_{i=1}^n\}$ with each from the i-th training subject. Also, we assume that there are n_+ positive and n_- negative training subjects. So we have $n = n_+ + n_-$. The genetic data are used to train an SVM classifier using the following max-margin formulation by minimizing the averaged empirical loss:

$$\min_{\mathbf{w},b,\xi_i} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\frac{1}{n_+}\sum_{i:\, y_i=1}\xi_i + \frac{1}{n_-}\sum_{i:\, y_i=-1}\xi_i\right), \tag{1}$$

$$\text{s.t.} \quad y_i(\mathbf{w}^\top \varphi(\mathbf{z}_i) + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, \ldots, n. \tag{2}$$

where $\mathbf{w}$ is the weight variable, b is the bias variable, and ξ_i is the slack variable; $\varphi(\cdot)$ is a pre-defined feature mapping function, and $C > 0$ is a tradeoff parameter. After solving (1) by using some existing toolbox such as LIBSVM, we can obtain the optimal solution of $\mathbf{w}$ and b, which are then used to compute the following residual value for each training subject i using the deviation function as below:

$$r_i = 1 - y_i(\mathbf{w}^\top \varphi(\mathbf{z}_i) + b). \tag{3}$$

Intuitively, the residual value r_i reflects the difficulty of classifying the i-th training subject. When the SVM classifier makes an error on z_i, r_i becomes larger than 1, and vice versa. Although original genetic features could also be used as privileged information, using the residual value r_i usually achieves better results as shown in [11] and is also much simpler, i.e., it only introduces two scalar variables, which will be shown in (4).

2.2 Privileged learning: A max-margin formulation

Vapnik and Vashist [11] recently proposed a new learning paradigm called learning using privileged information (LUPI) for SVM type of algorithms, which aims at improving predictive performance and reducing the number of required training samples. In the LUPI paradigm, the privileged information (i.e., prior knowledge) comes in the form of privileged data features which are available at the training time, but not at the testing time.

In this work, the residual value r_i in (3) is used as the privileged feature and the image visual feature $\mathbf{x}_i$ as the original feature for each training subject. Based on the LUPI paradigm, we propose to minimize the averaged empirical error on the positive and negative training subjects to avoid imbalanced problems (consider that we always have much more normal people than glaucoma patients). And unlike the standard SVM in (1) which directly optimizes for the slack variables ξ_i, the LUPI paradigm models each slack variable as a function of the privileged information. Specifically, we define each slack variable as the following linear function with respect to r_i in this work:

$$\xi_i(u, e) = u \cdot r_i + e, \tag{4}$$

where u is a linear weight and e is a bias term. With the above definitions, we replace the slack variable in SVM with (4) and propose the following optimization problem under the max-margin framework to learn a binary classifier for glaucoma detection, with the decision function as $f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b$:

$$\begin{aligned} \min_{\mathbf{w},b,u,e} \quad & \frac{1}{2}\left(\|\mathbf{w}\|^2 + \lambda \cdot u^2\right) + C\left(\frac{1}{n_+}\sum_{i:\, y_i=1} \xi_i(u,e) + \frac{1}{n_-}\sum_{i:\, y_i=-1} \xi_i(u,e)\right), \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i\,(u,e)\,,\ \xi_i\,(u,e) \geq 0,\ i = 1, \ldots, n. \end{aligned} \tag{5}$$

where $\phi(\cdot)$ is a feature mapping function to map each $\mathbf{x}_i$ into a higher dimensional space, $\lambda > 0$ is a pre-defined parameter to control the magnitude of u, $C > 0$ is also pre-defined to balance the averaged empirical loss and the regularizations of the weight variables. The optimization problem in (5) is often solved in its dual alternative. So here, we first introduce dual variables α_i and β_i for all the constraints in (5). Then its Lagrangian can be obtained as:

$$L(\mathbf{w}, b, u, e) = \frac{1}{2}\big(\|\mathbf{w}\|^2+\lambda\|u\|^2\big)+C\left(\frac{1}{n_+}\sum_{i:\, y_i=1}\xi_i(u,e)+\frac{1}{n_-}\sum_{i:\, y_i=-1}\xi_i(u,e)\right) - \sum_{i=1}^{n}\alpha_i\big[y_i(\mathbf{w}^\top\phi(\mathbf{x}_i)+b)-1+\xi_i\,(u,e)\big] - \sum_{i=1}^{n}\beta_i\cdot\xi_i\,(u,e)\,.$$

By setting the derivatives of the above Lagrangian to zeros with respect to $\mathbf{w}, b, u, e$ and then substituting the subsequent derivative results back into (5), we arrive at the dual problem of (5) as follows:

$$\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\beta}}\quad & -\mathbf{1}^\top\boldsymbol{\alpha}+\frac{1}{2}(\boldsymbol{\alpha}\circ\mathbf{y})^\top\mathbf{K}(\boldsymbol{\alpha}\circ\mathbf{y})+\frac{1}{2\lambda}(\boldsymbol{\alpha}+\boldsymbol{\beta}-C\cdot\mathbf{d})^\top\tilde{\mathbf{K}}(\boldsymbol{\alpha}+\boldsymbol{\beta}-C\cdot\mathbf{d}), \quad (6)\\ \text{s.t.}\quad & \mathbf{y}^\top\boldsymbol{\alpha}=0,\ \mathbf{1}^\top(\boldsymbol{\alpha}+\boldsymbol{\beta}-C\cdot\mathbf{d})=0,\ \boldsymbol{\alpha}\geq 0,\ \boldsymbol{\beta}\geq 0,\end{aligned}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^\top$ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_n]^\top$ are vectors of the dual variables, $y = [y_1, \ldots, y_n]^\top$ is the label vector, and $\mathbf{d} = [d_1, \ldots, d_n]^\top$ is a vector with $d_i = \frac{1}{n_+}$ (if $y_i = 1$) or $\frac{1}{n_-}$ (if $y_i = -1$); $\mathbf{K}$ and $\tilde{\mathbf{K}}$ are both kernel matrices with each element as $K_{ij} = \phi(\mathbf{x}_i)^\top\phi(\mathbf{x}_j)$ and $\tilde{K}_{ij} = r_i \cdot r_j$, respectively. Note that (6) is a standard quadratic programming problem, and it can thus be efficiently solved by existing software such as MOSEK[1] for MATLAB. After obtaining the optimal $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (6), we can get the optimal $\mathbf{w} = \sum_{i=1}^{n}\alpha_i y_i\phi(\mathbf{x}_i)$. Therefore, the decision function of SVM+ can be rewritten as $f(\mathbf{x}) = \sum_{i=1}^{n}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x})+b$, which will be used to make prediction for any test subject $\mathbf{x}$.

3 Experiments

We compare SVM+ with the standard SVM and two existing two-view learning methods, i.e., kernel canonical correlation analysis (KCCA) [9] and SVM-2K [4]. For the standard SVM, we train its model by using the genetic features only for both the training and test subjects. For the two-view learning emthods KCCA and SVM-2K, because the training subjects in our work have two views (i.e., image and genetic features), each method can be applied to train two models respectively for both views[2]. Note that the test subjects do not have genetic

[1] `http://www.mosek.com/`

[2] Note that KCCA first finds a common space where the correlation between the two views is maximized. After obtaining the projected features in the common space for both views, we apply SVM to train one model for each view.

features available in this work, so we can only use the model learned for image features to make prediction on any test subject. Besides those machine learning methods, we also compare with intraocular pressure (IOP) which is a glaucoma assessment method currently being used in clinics.

3.1 Database setup and feature description

The Singapore Malay Eye Study (SiMES) database is used for experimental evaluations of different methods. As a population-based study conducted by Singapore Malay community [10] from 2004 to 2007, SiMES records the assessments on the causes and risk factors of blindness and visual impairment for randomly selected Malays aged from 40 to 80 years old living in Singapore. It contains 2258 subjects with complete genetic data and retinal fundus image data for each subject. In the experiments, the diagnostic results of glaucoma are available for all the subjects and are thus used as the class label (i.e., +1 for glaucoma and -1 for normal). And in SiMES, 100 subjects are found to have glaucoma. Since IOP is clinically used for glaucoma detection, it is then removed from the personal profile data. In the experiments, we equally partition the database into training and testing sets. More specifically, we randomly select half of the glaucoma (i.e., positive) and normal (i.e., negative) subjects to form the training data, and the remaining subjects are used for testing. We assume the training subjects have both genetic and image data, while the test subjects only have image data. We believe this is a practical setting as in the real world. Also note that we form the training and test sets by conducting the random partitioning for ten times, therefore we have ten independent rounds of experiments.

We follow [6] to represent each retinal image as a 569-dimensional feature vector which is obtained from the standard deviations of color and texture descriptors extracted from image grids (please refer to [6] for more details). And for the genetic data, the recent study [12] has identified 178 single nucleotide polymorphisms (SNPs) which have high associations with glaucoma. Therefore, we use these SNPs as genetic features in the experiments.

3.2 Experimental configuration

For all the machine learning methods, we use the nonlinear 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 $\gamma = \frac{1}{A}$ and A is the mean value of all the square pairwise distances between any two training subjects. For SVM, KCCA and SVM-2K, five-fold cross-validation is performed to automatically select the tradeoff parameter C from the set $\mathcal{S} = \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$ during the training phase. And for our SVM+, as we have an additional parameter λ in (5), we perform the grid search to automatically select C and λ, both from $\mathcal{S}$. After cross-validation, the final classifier of each method is obtained by training this method again by using all the training data together with the selected parameters found through cross-validation. And this final classifier is used to make prediction on test subjects.

Table 1. Means and standard deviations of both AUC and balanced accuracy $\bar{P}$ of different methods. The best results are highlighted in boldface.

	IOP	SVM	KCCA	SVM-2K	SVM+
AUC	0.577±0.034	0.731±0.030	0.745±0.035	0.748±0.019	**0.772±0.012**
$\bar{P}$	0.555±0.021	0.625±0.017	0.647±0.029	0.633±0.013	**0.669±0.018**

Following [6], we use balanced accuracy $\bar{P}$ and the area under the receiver operating characteristic (ROC) curve, referred to as AUC, as evaluation metrics in this work, where the ROC is plotted as a curve which shows the tradeoff between sensitivity P_+ and specificity P_-. At the screening setpoint, 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.

3.3 Performance analysis

Table 1 shows the means and standard deviations of both AUC and balanced accuracy $\bar{P}$ for all the methods. We have the following observations:
i) All the machine learning methods, including SVM, KCCA, SVM-2K and our SVM+, perform much better than the clinically-used IOP method, showing that machine learning methods are effective and have great potentials for clinical use.
ii) KCCA, SVM-2K and our SVM+ have better performance than SVM in terms of both AUC and $\bar{P}$. This observation clearly demonstrate that both the two-view learning methods and the privileged learning methods can train better predictive models by using features from two different views, when compared to learning from a single view.
iii) Two-view learning methods KCCA and SVM-2K perform considerably worse than SVM+. The explanation is that in order to achieve good performance, the two-view learning methods generally require that each single view of data should perform comparably good to each other. However, in our experiments, the genetic features are much better than the image visual features[3]. Therefore, due to the considerably large difference between the two features, they may not highly correlate with each other in KCCA, or their single classifiers cannot make close predictions on each training subject in SVM-2K. As a result, KCCA and SVM-2K cannot achieve good performance. In contrast, our SVM+ considers the genetic information as prior knowledge and does not have any strong assumption made on the correspondence between the genetic and visual features.
iv) Our SVM+ achieves the best performance over the other baselines, which demonstrates that SVM+ is effective in leveraging privileged genetic information to learn a better predictive model for fundus image based glaucoma detection. When compared with the second best mean results, SVM+ achieves relative improvements of 3.21% in AUC over SVM-2K and 3.40% in $\bar{P}$ over KCCA.

[3] As reported in Table 3 in [6], genetic features enjoy a large improvement of 0.088 AUC over image visual features on the SiMES dataset.

4 Conclusion

We have developed an effective approach based on the LUPI paradigm for fundus image based glaucoma detection, which leverages privileged genetic information during the model training phase. The privileged genetic information of training subjects is obtained by first training an SVM classifier using genetic features and then computing the residual values from the SVM classifier predictions. We have conducted extensive experiments on the SiMES database, where results show the superiority of our method over SVM and other two-view learning methods.

References

1. Abràmoff, M.D., Alward, W.L., Greenlee, E.C., Shuba, L., Kim, C.Y., Fingert, J.H., Kwon, Y.H.: Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. Invest Ophthalmol Vis Sci, vol. 48, no. 4, pp. 1665–1673 (2007)
2. Bock, R., Meier, J., Nyúl, L.G., Hornegger, J., Michelson, G.: Glaucoma risk index: automated glaucoma detection from color fundus images. Med Image Anal, vol. 14, no. 3, pp. 471-481 (2010)
3. Chrástek, R., Wolf, M., Donath, K., Niemann, H., Paulus, D., Hothorn, T., Lausen, B., Lämmer R., Mardin C.Y., Michelson G.: Automated segmentation of the optic nerve head for diagnosis of glaucoma. Med Image Anal, vol. 9, no. 4, pp. 297-314 (2005)
4. Farquhar, J.D.R., Hardoon, D., Meng, H., Shawe-Taylor, J., Szedmak, S.: Two view learning: SVM-2K, Theory and Practice. In: NIPS, pp. 355–362 (2006)
5. V. Gál, E.K., Nachtegael, M.: Multiple kernel learning based modality classification for medical images. In: CVPR, pp. 76–83 (2012)
6. Liu, J., Zhang, Z., Wong, D.W.K., Xu, Y., Yin, F., Cheng, J., Tan, N.M., Kwoh, 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, vol. 20, no. 6, pp. 1021–1027 (2013)
7. Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol, vol. 90, no. 3, pp. 262–267 (2006)
8. Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y.: SimpleMKL. Journal of Machine Learning Research, vol. 9, pp. 2491–2521 (2008)
9. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press (2004)
10. 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, vol. 49, no. 9, 3846–3851 (2008)
11. Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged information. Neural Networks, vol. 22, pp. 544–557 (2009)
12. Vithana, E.N., Khor, C.C., *et al.*: Genome-wide association analyses identify three new susceptibility loci for primary angle closure glaucoma. Nature Genet, vol. 44, no. 10, pp. 1142–1146 (2012)
13. Xu, Y., Xu, D., Lin, S., Liu, J., Cheng, J., Cheung, C.Y., Aung, T., Wong, T.Y.: Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 1–8. Springer, Heidelberg (2011)