Multiple Ocular Diseases Classification with Graph Regularized Probabilistic Multi-label Learning

Xiangyu Chen[1], Yanwu Xu[1], Lixin Duan[1], Shuicheng Yan[2], Zhuo Zhang[1], Damon Wing Kee Wong[1], and Jiang Liu[1]

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

Abstract. Glaucoma, Pathological Myopia (PM), and Age-related Macular Degeneration (AMD) are three leading ocular diseases in the world. In this paper, we proposed a multiple ocular diseases diagnosis approach for above three diseases, with Entropic Graph regularized Probabilistic Multi-label learning (EGPM). The proposed EGPM exploits the correlations among these three diseases, and simultaneously classifying them for a given fundus image. The EGPM scheme contains two concatenating parts: 1) efficient graph construction based on k-Nearest-Neighbor (k-NN) search; 2) entropic multi-label learning based on Kullback-Leibler divergence. In addition, to capture the characteristics of these three leading ocular diseases, we explore the extractions of various effective low-level features, including Global Features, Grid-based Features, and Bag of Visual Words. Extensive experiments are conducted to validate the proposed EGPM framework on *SiMES* dataset. The results show area under curve (AUC) of the receiver operating characteristic curve in multiple ocular diseases detection are much better than the state-of-the-art algorithms.

1 Introduction

Vision is one of the most important senses which greatly influences an individuals quality of life. Studies have shown that 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. Diagnosing Glaucoma, PM and AMD is one of the most challenging problems in medical imaging.

Glaucoma is a chronic eye disease that leads to vision loss, in which the optic nerve is progressively damaged. It is one of the common causes of blindness, and is predicted to affect around 80 million people by 2020 [1]. Glaucoma diagnosis is typically based on the medical history, intraocular pressure, and visual field loss tests together with a manual assessment of the Optic Disc (OD) through ophthalmoscopy. OD or optic nerve head is the location where ganglion cell

axons exit the eye to form the optic nerve, through which visual information of the photo-receptors is transmitted to the brain [2].

As one of the leading causes of blindness worldwide, Pathological Myopia (P-M) is a type of severe and progressive nearsightedness characterized by changes in the fundus of the eye, due to posterior staphyloma and deficient corrected acuity. PM is primarily a genetic condition [3] [4]. It is accompanied by degenerative changes in the retina, which if left untreated can lead to irrecoverable vision loss. The accurate detection of PM will enable timely intervention and facilitate better disease management to slow down the progression of the disease. PM has been detected mostly from fundus image where retinal degeneration is observed in the form of Peripapillary Atrophy (PPA). PPA is the thinning of retinal layers around the optic nerve and is characterized by a pigmented ring like structure around the optic disc.

Age-related Macular Degeneration (AMD) causes vision loss at the central region and blur and distortion at the peripheral region. Depending on the presence of exudates, AMD is classified into dry AMD (non-exudative AMD) and wet AMD (exudative AMD). Dry AMD results from atrophy of the retinal pigment epithelial layer below the retina. It causes vision loss through loss of photoreceptors (rods and cones) in the central part of the retina. The major symptom and also the first clinical indicator of dry AMD is drusen, sub-retinal deposits formed by retinal waste. Wet AMD causes vision loss due to abnormal blood vessel growth (choroidal neovascularization) in the choriocapillaris, through Bruchs membrane, ultimately leading to blood and protein leakage below the macula. Bleeding, leaking, and scarring from these blood vessels eventually cause irreversible damage to the photoreceptors and rapid vision loss if left untreated. The major symptom of wet AMD is exudation [5][6].

For these three leading ocular diseases, there are some correlations among them. In recent decades, the problem of low vision and blindness in elderly people became major and socially significant issue. The number of patients having age-related macular degeneration (AMD) in association with glaucoma grows all over the world [8], which attaches medical and social value to this multiple diseases diagnosis problem. Moreover, in recent study, myopic eyes are less likely to have AMD and diabetic retinopathy (DR) but more likely to have nuclear cataract and glaucoma [9]. Inspired by the correlations among Glaucoma, P-M, and AMD, we propose a entropic graph regularized probabilistic multi-label learning framework for harmoniously integrating the above correlation information, and investigating the problem of learning to simultaneously diagnose these three leading ocular diseases for a given fundus image. Unlike previous approaches that detect individual ocular disease independently, our proposed EGPM scheme encodes the correlation information of different diseases of an image as a unit label confidence vector, which naturally imposes inter-label constraints and manipulates labels interactively. It then utilizes the probabilistic Kullback-Leibler divergence and Shannon Entropy for problem formulation on multi-label learning. This kind of disease screening is more oriented to real world

diagnosis scenario because a patient may have two or three diseases at the same time.

2 Related Work

Multi-label learning is a hot and promising research direction in computer vision. In the past, there are several approaches proposed to exploit the multiple labels learning problem. For example, the work in [14] introduced a unified Correlative Multi-Label (CML) framework for classifying labels and modeling correlations between them. Chen et al. [18] solved the multi-label learning problem by utilizing a sylvester equation. However, in the medical imaging analysis, multiple ocular diseases detection is still an open problem. In this paper, the proposed scheme exploits the medical problem of simultaneously diagnosing the leading multiple ocular diseases based on entropic graph regularized probabilistic multi-label learning.

In the previous work [10], a graph-based semi-supervised learning (SSL) method was proposed for phone classification task. Unlike previous approaches, this method modeled the multi-class label confidence vector as a probabilistic distribution, and utilized the Kullback-Leibler (KL) divergence to gauge the pairwise discrepancy. The underlying philosophy is that such soft regularization term will be less vulnerable to noisy annotation or outliers. Here we adopt the same distance measure, yet in a different scenario (i.e. multiple ocular disease detection in medical imaging analysis), thus demanding new solution. In the setting of multi-label annotation in multimedia, the work in [19] proposed the Kullback-Leibler divergence based multi-label propagation, which encoded the label information of an image as a label vector and imposes inter-label constraints and manipulates labels interactively. In this paper, based on Kullback-Leibler Divergence and Shannon entropy, we propose a graph regularized probabilistic multi-label learning framework for harmoniously integrating the correlation information of different diseases, and investigating the problem of learning to simultaneously diagnose these three leading ocular diseases for a given fundus image.

3 Feature Extraction

Detecting Glaucoma, PM and AMD is one of the most challenging problems in medical image analysis. In order to effectively capture the characteristics of these three leading ocular diseases, we explore the extractions of various popular features adopted in medical imaging and computer vision in this section. We extract three types of low-level features: Global Features, Grid-based Features, and Bag of Visual Words.

3.1 Global Features

Color Histogram: The color histogram serves as an effective representation of the color content of an image. It is defined as the distribution of the number

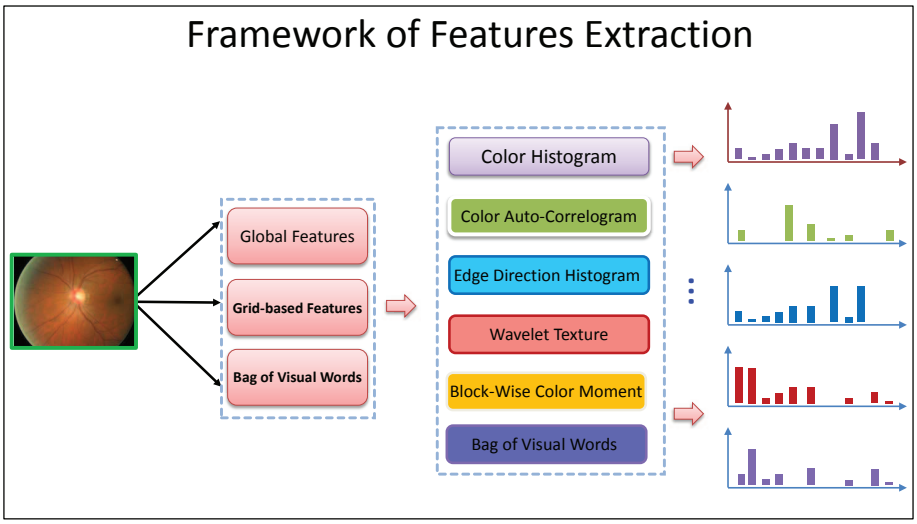

Fig. 1. Framework of features extraction. A set of effective and popularly used global and local features for each fundus image are extracted. Global Features: color histogram, color auto-correlogram, edge direction histogram, and wavelet texture; Grid-based Features: block-wise color moments are extracted; Local Features: bags of visual words.

of pixels for each quantized bin. We adopt the LAB color space [13] to model the color image, where L is lightness and A, B are color opponents. As LAB is a linear color space, we therefore quantize each component of LAB color space uniformly into four bins. Then the color histogram is defined for each component as follows:

$$L(i) = \frac{Z_i}{N}, i = 1, 2, ..., k, \tag{1}$$

where Z_i is the number of pixels with value i, N is the total number of pixels in the image, and k is the size of the quantized bins (with $k = 4$). The resulting color histogram has a dimension of $64 = 4 \times 4 \times 4$.

Color Auto-Correlogram: The color auto-correlogram was proposed to characterize the color distributions and the spatial correlation of pairs of colors together. The first two dimensions of the three-dimensional histogram are the colors of any pixel pair and the third dimension is their spatial distance. Let I represent the entire set of image pixels and $I_c(i)$ represent the subset of pixels with color $c(i)$, then the color auto-correlogram is defined as [15]:

$$r_{i,j}^{(t)} = P_{r_{p_1 \in I_{c_i}, p_2 \in I}}[p_2 \in I_{c(j)} \parallel |p_1 - p_2| = d], \tag{2}$$

where $i, j \in \{1, 2, ..., k\}$, $d \in 1, 2, ..., l$ and $|p_1 - p_2|$ is the distance between pixels p_1 and p_2. Color auto-correlogram only captures the spatial correlation between

identical colors and thus reduces the dimension from $O(N^2d)$ to $O(Nd)$. We quantize the HSV color components into 36 bins and set the distance metric to four odd intervals of $d = \{1, 3, 5, 7\}$. Thus the color auto-correlogram has a dimension of $144 = 36 \times 4$.

Edge Direction Histogram: Edge direction histogram encodes the distribution of the directions of edges. It comprises a total of 73 bins, in which the first 72 bins are the count of edges with directions quantized at five degrees interval, and the last bin is the count of number of pixels that do not contribute to an edge. To compensate for different image sizes, we normalize the entries in histogram as follows [16]:

$$E_i = \begin{cases} \frac{E(i)}{M_s}, & \text{if } i \in [0, ..., 71] \\ \frac{E(i)}{M}, & \text{if } i = 72 \end{cases} \tag{3}$$

where $E(i)$ is the count of bin i in the edge direction histogram; M_s is the total number of edge points detected in the sub-block of an image; and M is the total number of pixels in the sub-block. We use Canny filter to detect edge points and Sobel operator to calculate the direction by the gradient of each edge point.

Wavelet Texture: The wavelet transform provides a multi-resolution approach for texture analysis. Essentially wavelet transform decomposes a signal with a family of basis functions $\psi_{mn}(x)$ obtained through translation and dilation of a mother wavelet $\psi(x)$ [21], i.e.,

$$\psi_{mn}(x) = 2^{\frac{m}{2}} \psi(2^{-m}x - n), \tag{4}$$

where m and n are the dilation and translation parameters. A signal $f(x)$ can be represented as:

$$f(x) = \sum_{m,n} c_{mn} \psi_{mn}(x). \tag{5}$$

Wavelet transform performed on image involves recursive filtering and subsampling. At each level, the image is decomposed into four frequency sub-bands, LL, LH, HL, and HH, where L denotes the low frequency and H denotes the high frequency. Two major types of wavelet transform often used for texture analysis are the pyramid-structured wavelet transform (PWT) and the tree-structured wavelet transform (TWT). The PWT recursively decomposes the LL band. On the other hand, the TWT decomposes other bands such as LH, HL or HH for preserving the most important information appears in the middle frequency channels. After the decomposition, feature vectors can be constructed using the mean and standard deviation of the energy distribution of each sub-band at each level. For the three-level de- composition, PWT results in a feature vector of $24 = 3 \times 4 \times 2$ components. For TWT, the feature will depend on how the sub-bands at each level are decomposed. A fixed decomposition tree can be obtained by sequentially decomposing the LL, LH, and HL bands, thus resulting in a feature vector of $104 = 52 \times 2$ components.

3.2 Grid-based Features

Block-wise Color Moments: The first (mean), the second (variance) and the third order (skew- ness) color moments have been found to be efficient and effective in representing the color distributions of images. Mathematically, the first three moments are defined as [7]:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij}. \tag{6}$$

$$\sigma_i = (\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - \mu_i)^2)^{1/2}. \tag{7}$$

$$s_i = (\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - \mu_i)^3)^{1/3}. \tag{8}$$

where f_{ij} is the value of the i-th color component of the image pixel j, and N is the total number of pixels in the image.

Color moments offer a very compact representation of image content as compared to other color features. For the use of three color moments as described above, only nine components (three color moments, each with three color components) will be used. Due to this compactness, it may not have good discrimination power. Thus for our dataset, we extract the block-wise color moments over 55 fixed grid partitions, giving rise to a block-wise color moments with a dimension of 225.

3.3 Bag of Visual Words

The bag of visual words model approach in computer vision, also known as bag-of-words model [22], is a simplifying representation used in natural language processing and information retrieval by treating local image features as words. In natural language processing, a bag-of-words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. In computer vision, a bag-of-words is a sparse vector of occurrence counts of a vocabulary of local image features (codebook), which is a location-indepedent global feature; however, the properties of local features, such as intensity, rotation, scale and affine invariants can also be preserved. In this paper, the generation of bag of words comprises three steps:

- First, apply the difference of Gaussian filter on the gray scale images to detect a set of key-points and scales respectively.
- Then, compute the Scale Invariant Feature Transform (SIFT) over the local region defined by the key-point and scale.

– Finally, perform the vector quantization on SIFT region descriptors to construct the visual vocabulary by exploiting the k-means clustering. Here we generated 500 clusters, and thus the dimension of the bag of visual words is 500.

4 Graph Regularized Probabilistic Multi-label Learning

Our proposed multi-label learning framework includes two concatenating parts: 1) k-Nearest-Neighbor (k-NN) Search based Graph Construction; 2) Entropic Multi-label Learning based on Kullback-Leibler Divergence.

4.1 Graph Construction

The first step of the proposed framework is the construction of an directed weighted graph $\mathcal{G} =< V,\ E >$, where the cardinality of the node set V is $m = l + u$ (denote the labeled and unlabeled data respectively), and the edge set $E \subseteq V \times V$ describes the graph topology. Let V_l and V_u be the sets of labeled and unlabeled vertices respectively. $\mathcal{G}$ can be equivalently represented by a weight matrix $\mathbf{W} = \{w_{ij}\} \in \mathbb{R}^{m \times m}$. To efficiently handle the large-scale data, we enforce the constructed graph to be sparse. The weight between two nodes w_{ij} is nonzero only when $j \in \mathcal{N}_i$, where $\mathcal{N}_i$ denotes the local neighborhood of the i-th image. The graph construction can thus be decomposed into two sub-problems: 1) Neighborhood Selection; and 2) Edge Weight Computation.

Neighborhood Selection For the issue of neighborhood selection, there are two conventional strategies in previous work: ϵ-ball neighborhood and k-nearest-neighbor based neighborhood.

For ϵ-ball neighborhood selection, given a pre-specified distance measure between two nodes $d_{\mathcal{G}}(x_i,\ x_j)$ and a threshold ϵ, any vertex x_j that satisfies $d_{\mathcal{G}}(x_i,\ x_j) \leq \epsilon$ will be incorporated in the neighborhood of the vertex x_i, resulting in nonzero w_{ij}. It is easy to observe that the weight matrix of the constructed graph is symmetric. However, for some vertices beyond a distance from the others, there is probably no edge connecting to other vertices.

For k-nearest-neighbor based neighborhood selection, w_{ij} is nonzero only if x_j is among the k-nearest neighbors to the i-th datum. Obviously, graphs constructed in this way may ensure a constant vertex degree, avoiding over-dense sub-graphs and isolated vertices. In this paper, we employ k-nearest-neighbor based neighborhood for graph construction.

Edge Weight Computation A proper inter-sample similarity definition is the core for graph-based label learning. The message transmitted from the neighboring vertices with higher weights will be much stronger than the others. Generally, the more similar a sample is to another sample, the stronger the interaction (thus

larger weight) exists between them. There are two traditional ways to computing the edge weight: unweighted k-NN similarity and exponentially weighted similarity.

For unweighted k-NN similarity, the similarity w_{ij} between x_i and x_j is 1 if x_j is among the k-NN of x_i; otherwise 0. For undirected graph, the weight matrix is symmetric and therefore $w_{ij} = w_{ji}$ is enforced.

For exponentially weighted similarity, given all chosen k-NN neighbors, their weights are determined as below:

$$w_{ij} = \exp\left(-\frac{d_{\mathcal{G}}(x_i, x_j)}{\sigma^2}\right), \tag{9}$$

where $d_{\mathcal{G}}(x_i, x_j)$ is the ground truth distance and σ is a free parameter to control the decay rate.

In this paper, we utilized an efficient weight computation method–*weighted linear neighborhood similarity* [17]. In this scheme sample x_i is assumed to be linearly reconstructed from its k-NN. The weights are obtained via solving the following optimization problem:

$$\min_{w_{ij}} \parallel x_i - \sum_{j \in \mathcal{N}_i} w_{ij} x_j \parallel^2 . \tag{10}$$

Typically additional constraints are given to w_{ij}. For example, in [17], the constraints $w_{ij} \geq 0$ and $\sum_j w_{ij} = 1$ are imposed. The kind of constraints could help exploit the correlations of the three ocular diseases.

4.2 Entropic Multi-label Learning based on Kullback-Leibler Divergence

Let $M_l = \{x_i, r_i\}_{i=1}^{l}$ be the set of labeled images, where x_i is the feature vector of the i-th image and r_i is a multi-label vector (its entry is set to be 1 if it is assigned with the corresponding label, otherwise 0). Let $M_u = \{x_i\}_{i=l+1}^{l+u}$ be the set of unlabeled images, and $M = \{M_l, M_u\}$ is the entire data set. The graph-based multi-label learning is intrinsically a transductive learning process, which propagates the labels of M_l to M_u.

For each x_i, we define the probability measure p_i over the measurable space $(Y, \mathcal{Y})$. Here $\mathcal{Y}$ is the σ-field of measurable subsets of Y and $Y \subset \mathbb{N}$ (the set of natural numbers) is the space of classifier outputs. $|Y| = 2$ yields binary classification while $|Y| > 2$ implies multi-label. In this paper, we focus on the multi-label case. Hereafter, we use p_i and r_i for the i-th image, both of which are subject to the multinomial distributions, and $p_i(y)$ is the probability that x_i belongs to class y. As mentioned above, $\{r_j, j \in V_l\}$ encodes the supervision information of the labeled data. If it is assigned a unique label by the annotator, r_j becomes the so-called "one-hot" vector (only the corresponding entry is 1, the rest is 0). In case being associated with multiple labels, r_j is represented to be a probabilistic distribution with multiple non-zero entries.

We adopt the following criterion to guide the propagation of the supervision information, which is based on the concepts of Kullback-Leibler divergence [10] defined on two distributions:

$$Z_1(p) = \sum_{l=1}^{l} Z_{KL}\big(r_i \parallel p_i\big) + \mu \sum_{i=1}^{m} Z_{KL}\big(p_i \parallel \sum_{j \in N(i)} w_{ij} p_j\big), \tag{11}$$

and the optimal solution $p^* = \arg_p \min Z_1(p)$.

Here $Z_{KL}(r_i \parallel p_i)$ denotes the KL divergence between r_i and p_i, whose formal definition for the discrete case is expressed as $Z_{KL}(r_i \parallel p_i) = \sum_y r_i(y) \log \frac{r_i(y)}{p_i(y)}$. The first term in $Z_1(p)$ trigger a heavy penalty if the estimated value p_i deviates from the pre-specified r_i. The second term of Z_1 stems from the assumption that p_i can be linearly reconstructed from the estimations of its neighbors, thus penalizing the inconsistency between the p_i and its neighborhood estimation. μ is a free parameter to balance these two terms.

Note that Z_1 in Equation (11) is not amenable to alternating optimization. We further propose a modified version by introducing a new group of variables $\{q_i\}$ and Shannon entropy $H(q_i)$, which is shown as below:

$$\begin{aligned} Z_2(p,q) = &\sum_{l=1}^{l} Z_{KL}(r_i \parallel q_i) + \mu \sum_{i=1}^{m} Z_{KL}(p_i \parallel \sum_{j \in \mathcal{N}(i)} w_{ij} q_j) \\ &+\eta \sum_{i=1}^{m} Z_{KL}(p_i \parallel q_i) + \xi \sum_{i=1}^{m} H(q_i). \end{aligned} \tag{12}$$

In the above, a third measure q_i is introduced to decouple the original term $\mu \sum_{i=1}^{m} Z_{KL}\big(p_i \parallel \sum_{j \in N(i)} w_{ij} p_j\big)$. q_i can actually be regarded as a relaxed version of p_i. To enforce consistency between them, the third term $\sum_{i=1}^{m} Z_{KL}(p_i \parallel q_i)$ is incorporated. Here $H(q) = \sum_y q(y) log q(y)$. The Equation (12) could be solved by utilizing the similar method in [10].

5 Experiments

To evaluate the multiple diseases diagnosis performance of our proposed Graph Regularized Probabilistic Multi-label Learning (EGPM), we conduct extensive experiments on the Singapore Malay Eye Study (SiMES) database [12] for detecting the three leading ocular diseases: Glaucoma, Pathological Myopia (PM), and age-related macular degeneration (AMD). We consider using three different types of features as well as their combination in the experiments, which gives us 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 provide quantitative study on SiMES, with an emphasis on the comparison with six state-of-the-art related methods.

Table 1. The Baseline Algorithms.

Name	Methods
KNN	k-Nearest Neighbors [24]
SVM	Support Vector Machine [23]
LNP	Linear Neighborhood Propagation [17]
SPM	State-of-the-art algorithm for PM Detection [11]
SAMD	State-of-the-art algorithm for AMD Detection [20]
SGL	State-of-the-art algorithm for Glaucoma Detection [2]

5.1 Dataset

SiMES is a population-based study conducted from 2004 to 2007. It examined a cross-sectional and age stratified sample of 3,280 randomly selected Malays aged from 40 to 80 years old living in Singapore. For each subject in this database, personal demographic/clinical data, a retinal fundus image, and a blood sample (used for genotyping) were collected during the clinic visit, which thus gives us three informatics domains containing completely different types of data.

Moreover, the detection of three leading ocular diseases (i.e., Glaucoma, AMD, and PM) have been made by clinicians at the same time. The detection of different diseases made by clinicians during the visit are used as the gold standard to evaluate the classification performance of all the methods in this work. In this work, we select a subset of SiMES for experiments, which contains 2,258 subjects. Among the 2,258 subjects, 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.

5.2 Low-level Features

As detailed in Section 2, to facilitate experimentation and comparison of results, we extract a set of effective and popularly used global and local features for each image. For global features, four types of features are extracted: 64-dimensional color histogram [13], 144-dimensional color auto-correlogram [15], 73-dimensional edge direction histogram [16], and 128-dimensional wavelet texture [21]. For grid-based features, 225-dimensional block-wise color moments are extracted [7]. For local features, 500-dimensional bags of visual words [22] are generated.

5.3 Evaluation Criteria

In this work, we utilize the area under the curve (AUC) of receiver operation characteristic curve (ROC) to evaluate the performance of glaucoma diagnosis. The ROC is plotted as a curve which shows the tradeoff between sensitivity TPR (true positive rate) and specificity TNR (true negative rate), defined as

$$TPR = \frac{TP}{TP + FN},\ TNR = \frac{TN}{TN + FP}, \tag{13}$$

where TP and TN are the number of true positives and true negatives, respectively, and FP and FN are the number of false positives and false negatives, respectively.

Table 2. The AUCs of different algorithms for simultaneously detecting the three leading ocular diseases (i.e., Glaucoma, PM and AMD) on SiMES dataset. The combined visual features (global features + grid-based features + bag of words) are utilized in the experiment. The results of AUC marked in boldface are significantly better than others.

Methods	**Glaucoma**	**Pathological Myopia**	**AMD**
KNN	74.2 %	86.5 %	72.9%
SVM	76.7 %	89.1%	75.0%
LNP	78.8 %	90.1%	76.6%
SGL	81.0 %	-	-
SPM	-	91.0%	-
$SAMD$	-	-	77.8%
$OurProposed$	**82.5** %	**92.3** %	**79.3**%

5.4 Baselines and Experimental Setup

In the experiments, we compare our proposed Graph Regularized Multi-label Learning (EGPM) with six baseline methods (as shown in Table 1): Support Vector Machine (SVM) [23], k-Nearest Neighbors (KNN) [24], Linear Neighborhood Propagation (LNP) [17], SPM [11], SAMD [20], and SGL [2].Amongst them, SVM is originally developed to solve binary-class or multi-class classification problem. Here we use its multi-class version by adopting the one-vs-one method. LNP is the state-of-the-art algorithms for semi-supervised learning. It bases on a linear construction criterion to calculate the edge weights of the graph, and disseminates the supervision information by a local propagation and updating process. SPM is the state-of-the-art algorithm for PM detection, which is a sparse learning based framework to recognize PM in retinal fundus images. SAMD is the state-of-the-art algorithm for AMD detection, which is an automatic framework for the detection of drusen images for AMD assessment. SGL is the state-of-the-art algorithm for Glaucoma detection, which is a reconstruction-based learning technique for glaucoma screening. Since SPM, SAMD and SGL are the individual ocular disease detection algorithms, we only give the AUCs of PM, AMD and Glaucoma for SPM, SAMD and SGL in Table 2, respectively.

For KNN, SVM, and LNP, we implement them under the aforementioned three settings using different feature types as well as their combinations. For

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

Methods	KNN	SVM	LNP	Our Proposed
Global Features	71.2 %	73.5 %	75.2%	78.7%
Grid based Features	69.1 %	71.2 %	73.0%	76.5%
Bag of Words	68.4 %	70.9%	72.6%	75.0%
Combined Features	**74.2** %	**76.7**%	**78.8** %	**82.5**%

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

Methods	KNN	SVM	LNP	Our Proposed
Global Features	70.2 %	72.5%	73.9 %	76.4%
Grid based Features	69.3 %	71.8 %	72.5%	76.5%
Bag of Words	68.1 %	70.3%	71.6%	73.5%
Combined Features	**72.9** %	**75.0**%	**76.6** %	**79.3**%

each setting, all the methods for the automatic detections of the three leading ocular diseases (i.e., glaucoma, AMD and PM) are evaluated on SiMES dataset. All the experiments are implemented with Matlab and tested on a four core 3.4GHz PC with 12GB RAM.

5.5 Experiment Results Analysis

In the experiments, we systematically compare our proposed EGPM with six baselines (SVM, KNN, LNP, SPM, SAMD, and SGL) on SiMES. Below are the parameters and the adopted values for each method:

- For SVM algorithm, we adopt the RBF kernel. For its two parameters γ and C, we set $\gamma = 0.5$ and $C = 1$ in experiments after fine tuning.
- For KNN, there is only one parameter k for tuning, which stands for the number of nearest neighbors and is trivially set as 500.
- For EGPM, we set the two parameters as $\mu = 9$, $\eta = 4$, and $\xi = 0.01$.
- For SGL, SPM, and SAMD, we use the similar setting as in their papers.

The AUCs of the seven methods for detecting the three leading ocular diseases (i.e., Glaucoma, PM, and AMD) on SiMES dataset are illustrated in Table 2. The combined visual features (global features + grid-based features +bag of words) are utilized in this experiment. For SGL, SPM and SAMD, we adopt the similar setting in their papers. Our proposed algorithm EGPM outperforms the other baseline algorithms significantly. For example, EGPM has an improvement 7.6% over SVM, 11.2% over KNN, 4.7% over LNP for detecting Glaucoma. For PM, EGPM has an improvement 3.6% over SVM, 6.7% over KNN, and 2.5%

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

Methods	KNN	SVM	LNP	Our Proposed
Global Features	81.5 %	84.1%	85.6 %	87.3%
Grid based Features	79.3 %	82.3 %	83.7%	85.1%
Bag of Words	83.8 %	86.5%	87.9%	89.5%
Combined Features	**86.5** %	**89.1**%	**90.1** %	**92.3**%

over LNP. For AMD, EGPM has an improvement 5.7% over SVM, 8.8% over KNN, and 3.5% over LNP. Comparing with the state-of-the-art algorithms of individual disease detection, the proposed EGPM outperforms SGL, SPM, and SAMD by achieving the AUC 82.5%, 92.3%, 79.3%, respectively. The improvement is supposed to stem from the fact that our proposed algorithm encodes the disease label information of each image as a unit confidence vector, which imposes extra inter-label constraints. In contrast, other methods consider each disease label independently.

The comparison results of the detecting performance under four feature setting are listed in Table 3, Table 4, and Table 5. Since the state-of-the-art algorithms (SGL, SPM, SAMD) of the individual ocular disease detection are based on their own special visual features and retinal structures, the AUC results are not given in Table 3, Table 4, and Table 5. From Table 3, we are able to observe that, for glaucoma detection, our proposed algorithm EGPM outperforms the three baseline algorithms based on the combined features. The AUC of the receiver operating characteristic curve in glaucoma detection is 82.5%. The similar results are shown in Table 4 and Table 5 for AMD and PM detection respectively. For AMD detection, our proposed EGPM algorithm achieves 79.3%. For PM detection, the AUC of EGPM is 92.3%.

Recall that the proposed algorithm is a graph based probabilistic multi-label learning algorithm, wherein $p_i(y)$ expresses the probability for the i-th image to be associated with the y-th label, as detailed in Section 3.2. Figure 3 gives eight sample results by our proposed EGPM algorithm. For each fundus image, we attach the ground truth diagnosed by clinicians and the predicted labels with probabilities by EGPM. In the real world, the number of patients usually have AMD in association with glaucoma [8]. PM eyes are less likely to have AMD, but more likely to have glaucoma [9]. Hence, in our experimental results, Glaucoma and AMD are usually detected at the same time, as well as Glaucoma and PM (as shown in the fourth sample in Figure 3). As shown in the fourth row of Figure 3, even the quality of the fundus images is not good, our proposed EGPM still detects the glaucoma and AMD diseases. This validates the robustness and stability of the proposed method.

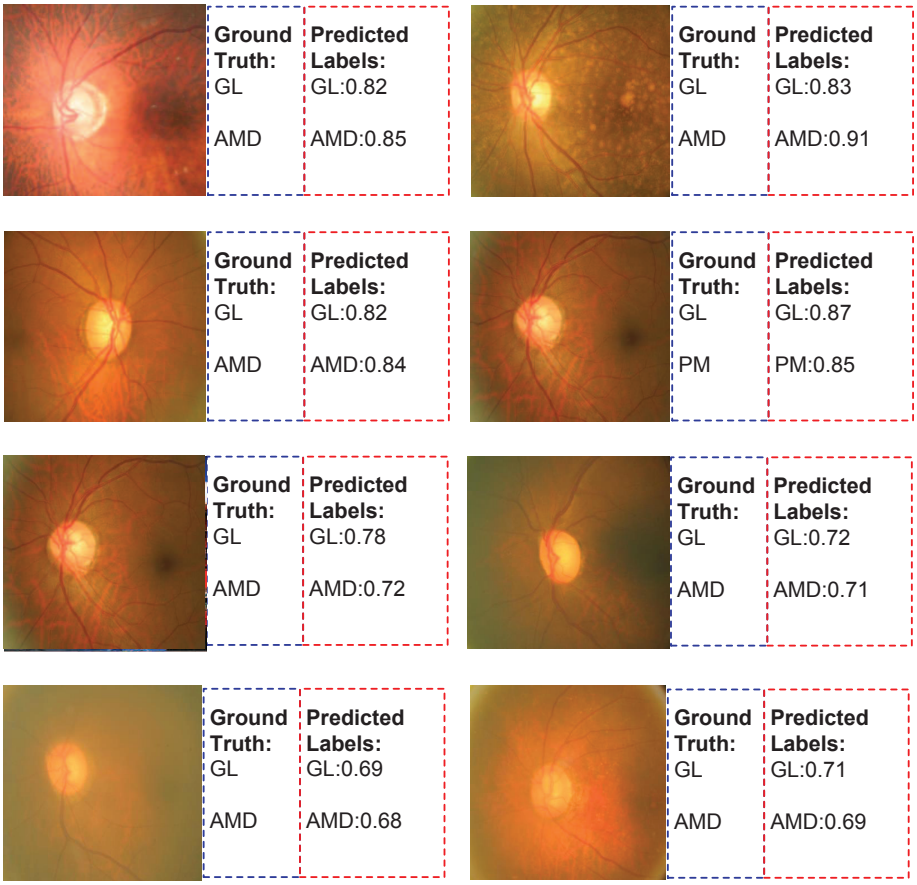

Fig. 2. Sample results diagnosed by our proposed EGPM algorithm. The order is from left top to right bottom. Each fundus image is attached the ground truth diagnosed by clinicians and the predicted labels with probabilities by EGPM. GL, AMD, and PM stand for Glaucoma, Age-related Macular Degeneration, and Pathological Myopia, respectively.

6 Conclusion

The proposed EGPM harmoniously integrates the correlation information of Glaucoma, PM and AMD, and exploits the problem of learning to simultaneously detect these three ocular diseases. Two concatenating parts are included in EGPM: 1) k-Nearest-Neighbor (k-NN) search based graph construction ; 2) Kullback-Leibler divergence based entropic multi-label learning. In addition, in order to capture the characteristics of Glaucoma, PM and AMD, the extractions of various effective low-level features are explored, including Global Features, Grid-based Features, and Bag of Visual Words.

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. Xu, Y., Lin, S., Wong, T.Y., Liu, J., Xu, D.: Efficient Reconstruction-Based Optic Cup Localization for Glaucoma Screening. In: MICCAI 2013
3. Young, T.L., Ronan, S.M.,Alvear, A.B., Wildenberg, S.C., Oetting, W.S., Atwood, L.D., Wilkin, D.J., King, R.A.: A second locus for familial high myopia maps to chromosome 12q. : Am J Hum Genet., 63(5), 1419–24 (1998)
4. Xu, Yanwu, Liu, Jiang, Zhang, Zhuo, Tan, Ngan Meng, Wong, Damon Wing Kee, Saw, Seang Mei, Wong, Tien Yin: Learn to recognize pathological myopia in fundus images using bag-of-feature and sparse learning approach. In: International Symposium on Biomedical Imaging, (2013)
5. Bressler, N.M., Bressler, S.B., Fine, S.L.: Age-related macular degeneration. Survey of ophthalmology. : Survey of ophthalmology, 32(6), 375–413 (1988)
6. De Jong, P.T.: Age-related macular degeneration. : New England Journal of Medicine, 355(14), 1474–1484 (2006)
7. Stricker, M., Orengo, M. : Similarity of color images. : SPIE Storage and Retrieval for Image and Video Databases III, (1995)
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), 284–91, (2013)
10. Subramanya, A., Bilmes, J.: Entropic graph regularization in non-parametric semi-supervised classification. In: NIPS, (2009)
11. Xu, Yanwu, Liu, Jiang, Zhang, Zhuo, Tan, Ngan Meng, Wong, Damon Wing Kee, Saw, Seang Mei, Wong, Tien Yin: Learn to recognize pathological myopia in fundus images using bag-of-feature and sparse learning approach. In: ISBI, 888–891, (2013)
12. 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. : Investigative Ophthalmology and Visual Science , 49(9), 3846–3851, (2008)
13. Shapiro, L. G., Stockman, G. C.: Computer Vision. : Prentice Hall, (2003)
14. Qi, Guojun, Hua, Xiansheng, Rui, Yong, Tang, Jinhui, Mei, Tao, Zhang, Hongjiang: Correlative multi-label video annotation. In: ACM Multimedia, (2007)
15. Huang, J., Kumar, S., Mitra, M., Zhu, W.-J., Zabih, R.: Image indexing using color correlogram. In: IEEE Conf. on Computer Vision and Pattern Recognition, (1997)
16. Park, D. K., Jeon, Y. S., Won, C. S.: Efficient use of local edge histogram descriptor. In: ACM Multimedia, (2000)
17. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. In: ICML, (2006)
18. Chen, G., Song, Y., Wang, F., Zhang, C.: Semi-supervised multi-label learning by solving a sylvester equation. In: SIAM International Conference on Data Mining, (2008)
19. Chen, Xiangyu, Mu, Yadong, Yan, Shuicheng, Chua, Tat-Seng: Efficient large-scale image annotation by probabilistic collaborative multi-label propagation. In: ACM Multimedia, (2010)

20. Wong, Damon Wing Kee, Liu, Jiang, Cheng, Xiangang, Zhang, Jielin, Yin, Fengshou, Bhargava, Mayuri, Cheung, Chui Ming Gemmy, Wong, Tien Yin: THALIA - An automatic hierarchical analysis system to detect drusen lesion images for amd assessment. In: ISBI, 884–887, (2013)
21. Manjunath, B. S., Ma, W.-Y.: Texture features for browsing and retrieval of image data. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), 837–842, (1996)
22. Lowe., D.: Distinctive image features from scale-invariant keypoints. : Intl J. Computer Vision, 2(60), 91–110, (2004)
23. Collobert, R., Sinz, F. H., Weston, J., Bottou, L.: Large scale transductive svms. : Journal of Machine Learning Research, 7, 1687–1712, (2006)
24. Duda, R., Stork, D., Hart, P.: Pattern Classification. : JOHN WILEY, (2000)