CrossMark

PATIENT FACING SYSTEMS

Learning ECOC Code Matrix for Multiclass Classification with Application to Glaucoma Diagnosis

Xiaolong Bai[1,2] · Swamidoss Issac Niwas[3] · Weisi Lin[4] · Bing-Feng Ju[1] ·
Chee Keong Kwoh[4] · Lipo Wang[2] · Chelvin C. Sng[6] · Maria C. Aquino[5] ·
Paul T. K. Chew[6]

Abstract Classification of different mechanisms of angle closure glaucoma (ACG) is important for medical diagnosis. Error-correcting output code (ECOC) is an effective approach for multiclass classification. In this study, we propose a new ensemble learning method based on ECOC with application to classification of four ACG mechanisms. The dichotomizers in ECOC are first optimized individually to increase their accuracy and diversity (or interdependence) which is beneficial to the ECOC framework. Specifically, the best feature set is determined for each possible dichotomizer and a wrapper approach is applied to evaluate the classification accuracy of each dichotomizer on the training dataset using cross-valida-tion. The separability of the ECOC codes is maximized by selecting a set of competitive dichotomizers according to a new criterion, in which a regularization term is introduced in consideration of the binary classification performance of each selected dichotomizer. The proposed method is experimentally applied for classifying four ACG mechanisms. The eye images of 152 glaucoma patients are collected by using anterior segment optical coherence tomography (AS-OCT) and then segmented, from which 84 features are extracted. The weighted average classification accuracy of the proposed method is 87.65 % based on the results of leave-one-out cross-validation (LOOCV), which is much better than that of

---

✉ Swamidoss Issac Niwas
   issacniwas@ntu.edu.sg

   Xiaolong Bai
   bxl@zju.edu.cn

   Weisi Lin
   wslin@ntu.edu.sg

   Bing-Feng Ju
   mbfju@zju.edu.cn

   Chee Keong Kwoh
   asckkwoh@ntu.edu.sg

   Lipo Wang
   elpwang@ntu.edu.sg

   Chelvin C. Sng
   chelvin@gmail.com

   Maria C. Aquino
   mcdaquino@gmail.com

   Paul T. K. Chew
   ophchewp@nus.edu.sg

[1] State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310027, People's Republic of China

[2] School of Electrical and Electronics Engineering, Nanyang Technological University (NTU), Singapore 639798, Singapore

[3] School of Electrical and Electronics Engineering & School of Computer Engineering, Nanyang Technological University (NTU), Singapore 639798, Singapore

[4] School of Computer Engineering, Nanyang Technological University (NTU), Singapore 639798, Singapore

[5] Eye Surgery Centre, National University Health System (NUHS), Singapore 119228, Singapore

[6] Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore 119228, Singapore

Springer

the other existing ECOC methods. The proposed method achieves accurate classification of four ACG mechanisms which is promising to be applied in diagnosis of glaucoma.

Introduction

Angle closure glaucoma (ACG) is a prevalent eye disease in Asia, which is a major cause of blindness [1]. There are mainly four mechanisms underlying ACG, namely, pupil block (PB), plateau iris configuration (PL), thick peripheral iris roll (PIR) and exaggerated lens vault (LV). Specific treatments are required for different ACG mechanisms. So classification of these four mechanisms is clinically important in order to provide better treatment for ACG patients [2]. Anterior segment optical coherence tomography (AS-OCT), is capable of providing high resolution images of anterior chamber of eyes and it has been used extensively for glaucoma detection [1–3]. Anterior chamber (AC) and AC angle parameters provided by AS-OCT were used to evaluate different ACG mechanisms [2, 4]. Wirawan et al. selected ten discriminative features from 84 parameters which were measured from segmented AS-OCT images of the patients, and used AdaBoost classifier for classification of these four different mechanisms [5].

While there are many existing methods for glaucoma detection, differentiating glaucomatous eyes from normal ones using features extracted from the fundus images [6], optic nerve head stereo photographs [7], or OCT [8] in combination with various classifiers, few studies have been done for classification of different mechanisms of ACG in spite of its clinical importance. So, our study is motivated by developing a new multiclass classification method with high accuracy in recognition of different ACG mechanisms. Traditionally, there are two categories of approaches to address the multiclass classification problems. One is to develop a single model for all the classes. Some classifiers such as k-Nearest Neighborhood (kNN), decision tree, Naive Bayes, and linear discriminative analysis (LDA) are naturally designed to deal with multiclass classification.

In contrast, some more advanced classifiers such as support vector machine (SVM), and AdaBoost are binary classifiers. Hence, it is required to design a set of different binary classifiers (dichotomizers) and combine them to deal with multiclass classification tasks. On the other hand, ensemble learning by combining multiple dichotomizers is more advantageous than traditional single classifier in that it has superior generalization ability and is suitable for learning nonlinear classification boundary. Intuitively, to differentiate c classes, c dichotomizers are trained, and each of which discriminates

one specific class from all the remaining classes. For a test sample, the classifier with largest probability output will be taken as the predicted class label. This is the well-known one-versus-all (OVA) strategy [9, 10]. An alternative to this is one-versus-one (OVO), where $c.(c\text{-}1)/2$ dichotomizers are trained to separate each pair of classes.

To solve the multiclass problems using binary classifiers, Dietterich et al. provided a unified framework based on a set of error-correcting output codes (ECOC), which was further improved by [11, 12]. This framework is well known for its nonlinear classification capability. A code matrix with n columns was designed to decompose the multiclass problems into n binary ones. Then the outputs of these N dichotomizers were combined to determine the class label of the test sample [13]. OVA and OVO can be seen as special cases of the ECOC framework. Many research works have been explored to improve the classification performance of ECOC [14–18], especially using data-driven approaches [19–21]. By observing the data distribution in the original feature space, dichotomizers are trained for easily separated pairs of class clusters. However, this observation of data distribution relies on a common feature space.

In this study, we propose a wrapper approach to learn the ECOC code matrix, in which the best feature set for each dichotomizer and best combination of dichotomizers for ECOC are both selected via cross-validation on the training dataset. This method is better than the other ECOC based methods in two aspects: 1) each dichotomizer has its own optimal feature set; 2) a new criterion is proposed that the best combination of dichotomizers is selected in consideration of not only of the separability of the codewords in ECOC framework but also the classification ability of the selected dichotomizers.

This paper is organized as follows. The ECOC framework is reviewed in "ECOC framework" section, followed by the proposed method in "Method for multiclass classification" section. Experimental results of multiclass classification on glaucoma dataset are shown in "Experimental results on classification of different glaucoma mechanisms" section. Conclusion is made in the final section.

ECOC framework

There are two major processes in ECOC framework: coding and decoding. The key of coding process lies in the design of a code matrix $M \in \{-1, 0, 1\}^{c \times n}$, with c rows and n columns, where c and n denote the numbers of classes and dichotomizers, respectively. The ith row of M provides the codeword C_i for the ith class ($i = 1, 2,…,c$). Meanwhile, each column of M represents the partition of classes by each dichotomizer. Classes coded by 1 and −1 are treated as positives and negatives, respectively, while those coded by 0 are

omitted in training the dichotomizers. For the four class classification problem, the OVA and OVO strategy are represented by two code matrixes as shown in Table 1a and b, respectively.

In the decoding process, the outputs of these n trained dichotomizers for the test sample are given as a vector $V = \{v_1, v_2, \ldots, v_n\}$, and compared with the codeword of each class to find the nearest one to determine the class label of the test sample. There are many decoding strategies to evaluate the distance between the vector V and each codeword C_i [13]. For example, in Loss-based decoding, the distance is formulated as

$$d_H(V, C_i) = \frac{1}{2} \sum_{j=1}^{n} L(V(j) \cdot C_i(j)) \qquad (1)$$

where $L(.)$ denotes the loss function which is dependent on the type of dichotomizer.

Most research is focused on the coding process to design an optimal code matrix. Allwein et al. proposed a dense random code matrix by maximizing the distances between the codewords of different classes, which is further extended as sparse random code matrix [21]. Pujol et al. proposed discriminative ECOC (DECOC) to learn the code matrix based on a hierarchical partition of the multiple classes, and used (c-1) dichotomizers in the ECOC code matrix [17]. In most of the existing works, the feature space is usually fixed to facilitate learning the data distribution and designing a problem-dependent code matrix. However, the common feature space is optimal for all the dichotomizers as a whole. So the dichotomizers may not be optimized specifically. The classification performance may not be the best, although the code matrix is designed well in terms of class separability. In this study, all the dichotomizers are first optimized with their own specific feature set, and then the ECOC code matrix is learned by selecting the best combination of dichotomizers. The details are shown in the next section.

Method for multiclass classification

In the proposed method, feature selection for each dichotomizer and dichotomizer selection in ECOC code matrix are performed in tandem to learn the code matrix as shown in Fig. 1. The proposed method is detailed as follows.

Step 1: Feature set optimization for each dichotomizer

Based on combinational analysis, we can calculate that there are totally N different dichotomizers using Eq. (2),

$$N = \frac{1}{2} \left(3^c - 2^{c+1} + 1\right) \qquad (2)$$

where c is the number of classes. For example, there are 25 dichotomizers for a four-class classification problem such as ACG diagnosis. By using the state-of-the-art feature selection method, such as minimum redundancy and maximum relevancy (mRMR) [22–24], the best feature set, which is closely dependent on the target class with minimum inter-redundancy, is identified for each dichotomizer. Wirawan et al. has shown that mRMR is suitable and effective in selecting informative and discriminative features for ACG classification [5]. In order to find the optimal feature set, a filter-wrapper approach is used [22, 23]. Features are first ranked according to the criteria of mRMR, and the highly ranked features are retained, then sequential forward selection (SFS) (or sequential backward selection (SBS), floating search methods) is performed to select the best feature set for each dichotomizer using cross-

Table 1 The ECOC code matrix in (a) OVA strategy and (b) OVO strategy

(a)

Class1	1	−1	−1	−1
Class2	−1	1	−1	−1
Class3	−1	−1	1	−1
Class4	−1	−1	−1	1

(b)

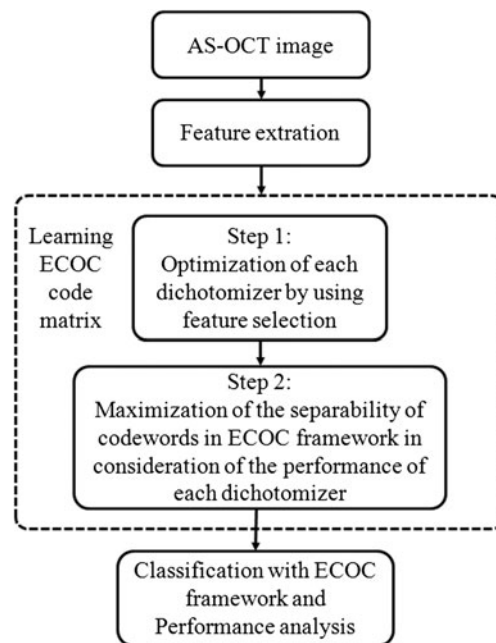|						
Class1	1	1	1	0	0	0
Class2	−1	0	0	1	1	0
Class3	0	−1	0	−1	0	1
Class4	0	0	−1	0	−1	−1

Fig. 1 Block diagram of the proposed method

Table 2 The algorithm of the proposed method

Algorithm-Learning the ECOC code matrix

Input: Training Data: $X=\{x_1, x_2,\ldots, x_i,\ldots, x_m\}$, $x_i \in R^D$ with the labels $Y=\{y_1, y_2,\ldots, y_i,\ldots, y_m\}$, $i=1,2,\ldots,m$. (D is the original feature dimension, m is the number of training samples)

Output: The ECOC code matrix M

1: **Generate** all the possible dichotomizers $B=\{b_1, b_2,\ldots, b_j,\ldots, b_N\}$, $j=1,2,\ldots,N$, by searching each possible binary partition of the training data;

2: **for each** dichotomizer $b_j \in B$ **do**

3: Best feature set $F_j \leftarrow$ FEATURE SELECTION(X, Y, c_f, s_f); (c_f is the criterion used to sort the features, s_f is the selection method, i.e., SFS)

4: The accuracy $a_j \leftarrow$ CROSS-VALIDATION(X, Y, F_j);

5: **end for**

6: **for each** weighting coefficient $\lambda_p \in [\lambda_{min}, \lambda_{max}]$ and code length $n_q \in [n_{min}, n_{max}]$ **do**

7: Select a set of dichotomizers B_{pq} which maximize d_s'' in Eq.(5);

8: **end for**

9: **Find** the best set of dichotomizers B^* which performs best via cross-validation on the training dataset;

10: ECOC code matrix $M \leftarrow B^*$.

validation on the training dataset.

However, since the number of dichotomizers N grows exponentially with the number of classes c in Eq. (2), it may be time-consuming to perform a wrapper approach to select the best feature set for each dichotomizer. For simplicity, filter approach is preferred for fast feature selection when N is large. Feature selection not only improves the classification performance (i.e., accuracy) of each dichotomizer, but also decreases the dependency among the dichotomizers by selecting different optimal feature sets for different dichotomizers, which is beneficial to the error-correcting ability of ECOC.

Step 2: Maximization of the separability of codewords in ECOC framework in consideration of the performance of each dichotomizer

To improve the classification performance of ECOC framework, the separability of ECOC codes is maximized, which is defined as

$$d_s = \min_{\substack{1 \leq i \leq c \\ i < k \leq c}} \left(\frac{1}{2} \sum_{j=1}^{n} L(C_i(j) \cdot C_k(j)) \right) \quad (3)$$

Fig. 2 Illustrative samples of AS-OCT images of glaucoma patients with different ACG mechanisms **a** PIR; **b** LV; **c** PB; **d** PL.
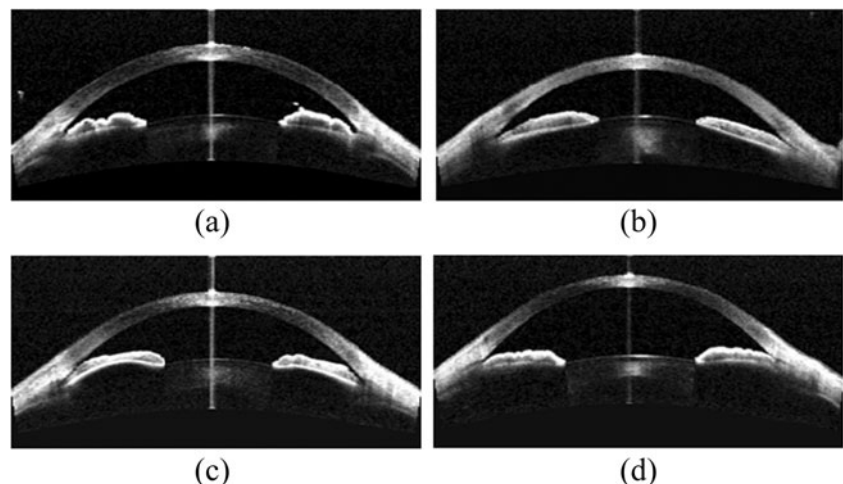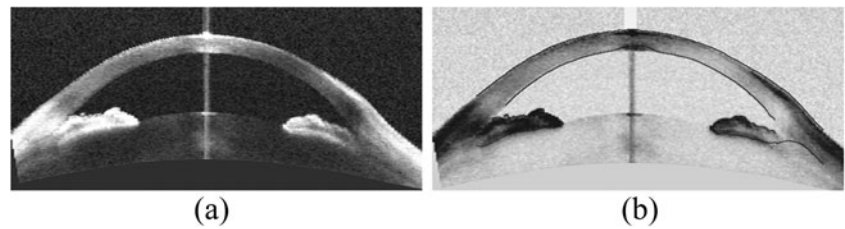
Fig. 3 An example of AS-OCT image **a** and its corresponding segmentation result in **b** by using ASAP software

(a)   (b)

where C_i and C_k denote the codewords of the ith and k^{th} classes, respectively [12]. In [21], the separability is modified as

$$d_s^{'} = \min_{\substack{1 \leq i \leq c \\ i < k \leq c}} \left(\frac{1}{2} |C_i(j) \cdot C_k(j)| \sum_{j=1}^{n} L(C_i(j) \cdot C_k(j)) \right) \tag{4}$$

to ignore the contributions from codes with value of 0. This definition is more reasonable. However, in all the existing ECOC methods, only the code information of the dichotomizers is considered. So some dichotomizers with unsatisfactory ability of binary classification may also be selected into the ECOC framework, which may deteriorate the final classification performance. In the proposed method, the separability is reformulated as

$$d_s^{''} = d_s^{'} + \lambda \bar{a} = \min_{\substack{1 \leq i \leq c \\ i < k \leq c}} \left(\frac{1}{2} |C_i(j) \cdot C_k(j)| \sum_{j=1}^{n} L(C_i(j) \cdot C_k(j)) \right) + \lambda \bar{a} \tag{5}$$

where $\bar{a}$ is the average binary classification accuracy of the selected dichotomizers, and λ is the coefficient weighting the relative importance of $\bar{a}$ compared with $d_s^{'}$. There are two key parameters to determine: the weighting coefficient λ and the code length n. Cross-validation is applied on the training dataset to find the optimal parameters λ^* and n^*, and also the optimal set of dichotomizers. Finally, the trained and selected dichotomizers are used as base learners in ECOC framework. The algorithm of the proposed method is shown in Table 2.

Experimental results on classification of different glaucoma mechanisms

Data preparation and experiment results

A dataset of 152 ACG samples provided by National University Hospital Singapore (NUHS) collected over 2 years is used for classification of four different ACG mechanisms [5, 24, 25]. The dataset is small due to the limited number of ACG patients are recruited using Ministry of Education (MoE) AcRF Tire 1 Funding, Singapore. AS-OCT images of glaucoma patients with different ACG mechanisms are shown in

Fig. 2. PIR is characterized by the thick and folded iris, while PB is characterized by the convex forward iris profile. Eyes with PL and LV mechanisms have the largest and smallest AC volumes, respectively [1]. Customized software (Anterior Segment Analysis Program-ASAP, National University Hospital, Singapore) was used to measure anterior chamber (AC) characteristics. ASAP software used the level set method for the segmentation of AC area [1, 5] of AS-OCT image as shown in Fig. 3. The quantifiable parameters of the anterior chamber (AC) measured by ASAP software includes anterior chamber depth (ACD), anterior chamber volume (ACV), anterior chamber width (ACW), angle recess area (ARA), angle opening distance (AOD), post closure area (PCA), angle opening distance (AOD), trabecular-iris space area (TISA), lens vault (LV) distance, iris area (IA), iris thickness (IT), iris concavity, etc., as illustrated in Fig. 4. The samples are labeled by medical experts in NUHS (C. C. Sng, M. C. Aquino and P. T. K. Chew) and the basic information of the glaucoma dataset used in this study is shown in Table 3.

Since, each mechanism has several characteristics, from medical point of view, 84 features are extracted, which are all clinically important parameters measured from the segmented AS-OCT image. Some of the important features are identified based on our previous studies using the same dataset [5, 24, 25] are given as follows: AC_Area (Anterior chamber area); AC_Volume (Anterior chamber volume); ACD (Anterior chamber depth); ACW (Anterior chamber width);
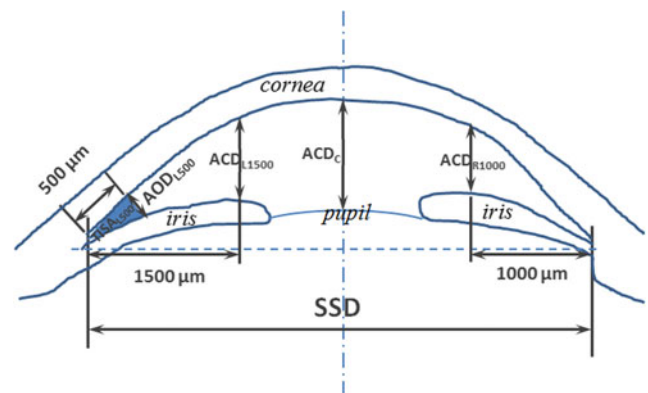

Fig. 4 The parameters measured from the AS-OCT image of the AC segment of the eyes (not all of the 84 parameters are shown here). For example, ACD_{L1500} means anterior chamber depth of the left hand side measured at 1500 μm from the scleral spur

Table 3 The basic information of the Glaucoma dataset used in this study

Number of features	Number of samples			
84	Class 1(PIR)	Class 2 (L)	Class 3 (PB)	Class 4 (PL)
	21	51	44	36
	Total: 152			

Anterior_lens_curvature (Curvature of the anterior lens surface); ILC_L (Iridolenticular contact in the left side); ILC_R (Iridolenticular contact in the right side); Iris_area_IL (Iris area contacts with Iridolenticular); Iris_area_L500 (The scleral spur is used as the centre of a circle with radius of 500 μm. The area of iris region inside this circle on left side is Iris_area_L500). Iris_Chord_Length_L (The distance from the tip of the iris to the periphery); Iris_Chord_Length_R (The distance from the tip of the iris to the periphery on right side); Iris_end_concavity_L (Concavity of Iris area at the end

in left side); Iris_thickness_L1000 (The intersection point on the anterior surface of the iris is identified when the scleral spur (SS) is used as the centre of a circle with radius of 1000 μm. Iris Thickness is the shortest distance from this point of intersection to the posterior surface of the iris on left side; Iris_thickness_L_DMR (Thickness of the iris region in the dilator muscle region (DMR) on left side; Iris_thickness_L_Max (Maximum thickness of Iris length); Iris_thickness_L_SMR (Iris thickness in the sphincter muscle region on left side); Iris_thickness_PL (Iris thickness contacts with plateau); Iris_thickness_R_DMR (Thickness of the iris region in the dilator muscle region on right side); Iris_thickness_R_SMR (Iris thickness in the sphincter muscle region on right side); Lens vault (The perpendicular distance between the horizontal line joining the two scleral spurs and the anterior pole of the crystalline lens, represents the anterior portion of the lens); Pupil_distance (Distance between the centers of the pupils).

The experimental investigation of our proposed method was conducted and implemented using Matlab 8.0 R2012b

Table 4 The Ranking of all the Dichotomizers according to their classification accuracy on the training dataset

Dichotomizer	Class 1	Class 2	Class 3	Class 4	No. of selected features	Classification Accuracy (%)	Selected by OVO	Selected by OVA	Selected by Sparse Random	Selected by the proposed method
D1	0	+1	−1	0	17	95.79	√			√
D2	0	+1	0	−1	15	93.10	√			
D3	−1	+1	0	−1	8	91.67				√
D4	−1	−1	−1	+1	6	91.45		√		√
D5	+1	0	−1	0	6	90.77	√			√
D6	+1	−1	0	0	7	90.27	√		√	√
D7	0	+1	−1	−1	9	90.08			√	√
D8	0	−1	−1	+1	5	90.08			√	
D9	−1	−1	0	+1	9	89.82			√	
D10	−1	+1	−1	0	12	89.66			√	
D11	0	−1	+1	−1	14	88.55			√	
D12	−1	+1	−1	−1	10	88.16		√	√	
D13	+1	0	−1	−1	10	88.12			√	
D14	+1	−1	−1	0	9	87.93			√	
D15	−1	−1	+1	0	9	87.93				
D16	+1	−1	−1	−1	5	86.84		√		
D17	−1	−1	+1	−1	16	86.20		√		
D18	0	0	+1	−1	6	85.00	√			
D19	−1	0	−1	+1	4	84.16				
D20	+1	+1	−1	−1	14	83.55				
D21	−1	0	+1	−1	5	83.17				
D22	+1	−1	+1	−1	16	80.92				
D23	+1	0	0	−1	4	80.70	√			
D24	+1	−1	0	−1	6	80.56				
D25	+1	−1	−1	+1	6	79.61				

(The Mathworks Inc., Natick, MA, USA) and Microsoft Visual Studio (C++). All the 84 features are normalized to have zero mean and unity variance. Wirawan et al. has shown that Adaboost performs better than SVM in application of ACG classification, and the performance of Adaboost combined with OVA strategy is also better than traditional multiclass classifiers in terms of classification accuracy, such as classification tree, and Naive Bayes [5]. WEKA (Waikato Environment for Knowledge Analysis) data mining tool [26] was used for comparing the proposed method with the traditional multiclass classifiers based on its default parameters as same as reported in [5] which used the same dataset for their study. Thus Adaboost is the choice of binary classifier in this experiment. Besides, for fair comparison with the results in [5], we also used mRMR for feature selection. In this four-class classification problem, there are totally 25 possible dichotomizers, which are easily obtained by exhaustive search. In the first step of the proposed method, all the 84 features are ranked according to the mRMR criteria for each dichotomizer. Each feature is incrementally included into a ranking list according to the following equations,

$$\max_{f_j \in F-F_{m-1}} \left[I(f_j, y) - \frac{1}{m-1} \sum_{f_j \in F_{m-1}} I(f_j, f_i) \right] \quad (6)$$

or

$$\max_{f_j \in F-F_{m-1}} \left[I(f_j, y) \bigg/ \left(\frac{1}{m-1} \sum_{f_j \in F_{m-1}} I(f_j, f_i) \right) \right] \quad (7)$$

where F is the whole feature set, and F_{m-1} is the set of $(m-1)$ features that are already selected, by maximizing the mutual information $I(f_j, y)$ between the jth feature and the class label y and minimizing the mutual information $I(f_j, f_i)$ between the jth feature and the feature f_i in the set F_{m-1} [23]. To further increase the classification ability of each dichotomizer, a wrapper approach is used to select the best feature set which leads to the least classification error. The dichotomizers are ranked according to their classification accuracy in descending order as shown in Table 4. In the Step 2, for $\lambda_{(p-1)} \in [0, 600]$ with a step of 30 and $n_q \in [1, 9]$ with a step of 1 ($p = 0, 1, 2, \ldots, 20$; $q = 1, 2, \ldots 9$), the best set of dichotomizers $B_{p,q}$ is determined by maximizing d_s'' in Eq.(5) where $L(z) = 1/e^z$. The separability d_s'' is dominantly determined by the average classification accuracy of the selected dichotomizer when the weighting coefficient $\lambda = 600$. According to the suggestion of [17, 20], the code length n should be $15\log(c) \approx 9$. Here let n vary from 1 to 9.

The classification performance of these different sets of dichotomizers $\{B_{p,q}\}$ is evaluated on the training dataset by leave-one-out cross-validation (LOOCV). Since we have done our study with the available limited data sources, leave-one-out cross-validation is used to prevent over-fitting of training data. The best set of dichotomizers is determined to

Table 5 The ECOC code matrix determined in the proposed method

Classes	D1	D3	D4	D5	D6	D7
Class 1 (PIR)	0	−1	−1	+1	+1	0
Class 2 (L)	+1	+1	−1	0	−1	+1
Class 3 (PB)	- 1	0	−1	−1	0	−1
Class 4 (PL)	0	−1	+1	0	0	−1

be $B_{13,6} = \{$D1, D3, D4, D5, D6, D7$\}$, and the optimal values of the parameters are $\lambda^* = 360$ and $n^* = 6$. The ECOC code matrix formed by the selected set of dichotomizers $B_{13,6}$ is shown in Table 5. Loss-based decoding strategy is used in the decoding process. The loss function of Adaboost classifier is $L(z) = 1/e^z$. The confusion matrix obtained using LOOCV on the glaucoma dataset is shown in Table 6. The weighted average classification accuracy is 87.65 % as shown in Table 7, which is better than the accuracy of 84 % obtained by [5] (The dataset in [5] is slightly different from ours in that four more patients with no mechanism of glaucoma are added; however, this effect is negligible).

Comparison with other ECOC methods, including OVO, OVA, sparse random ECOC

In most traditional methods, the dichotomizers are not optimized specifically. Only Wang et al. and Maghsoudi et al. used feature selection to optimize the dichotomizers in OVA scheme [14, 15]. In this paper, we apply feature selection to each dichotomizer for all the three ECOC methods based on OVO, OVA [9, 20] and sparse random [21] strategies. The classification accuracy for each class and the weighted average accuracy of these three existing popular ECOC methods are shown in Table 7. The highest weighted average accuracy of the three ECOC methods is 85.81 %, better than the traditional multiclass classifiers, such as classification tree (72.22 %), random forest (76.58 %), SVM combined with OVA strategy (78.22 %) and Naive Bayes (77.93 %).

We also randomly select 80 % of the dataset for training and the other 20 % for testing to compare the proposed method with the other ECOC methods. This process is repeated 2000 times. The classification accuracy (mean value ± standard deviation) of the proposed method and the three other ECOC methods mentioned above is 84.86 ± 3.56 %,

Table 6 Confusion matrix obtained by using the proposed method (Leave-one-out cross-validation)

|	Y_1	Y_2	Y_3	Y_4	$\sum Y$
Y_1'	**9**	1	1	8	19
Y_2'	5	**47**	4	3	59
Y_3'	3	2	**35**	3	43
Y_4'	4	1	4	**22**	31
$\sum Y'$	21	51	44	36	152

Table 7   Comparison of classification accuracy of the proposed method with other ECOC-based methods with dichotomizer-specific feature selection

Classes	Classification accuracy of different ECOC methods with dichotomizer-specific feature selection (%)			
	OVO	OVA	Sparse Random	The proposed method
Class 1 (PIR)	82.23	69.73	84.87	**85.53**
Class 2 (L)	86.84	84.87	82.89	**89.47**
Class 3 (PB)	88.16	82.24	83.55	**88.82**
Class 4 (PL)	83.55	82.89	77.63	**84.87**
Weighted average	85.81	81.55	82.11	**87.65**

83.69 ± 3.75 %, 79.76 ± 3.75 %, and 81.45 ± 3.70 %, respectively. And the histograms of the classification accuracy for the proposed method and the three ECOC methods are shown in Fig. 5a–d, respectively, from which we can see that the proposed method performs best. In the proposed method, all the dichotomizers are first optimized specifically to increase the interdependence and classification accuracy, and then the ECOC code matrix is learned by maximizing Eq.(5) and selecting a set of competitive dichotomizers. Not only the code information but also the classification ability of the dichotomizers is considered in maximizing the separability of the codewords in the ECOC matrix. However, in most traditional ECOC methods, the dichotomizers are not selected in consideration of their classification performance.

In this experiment, three dichotomizer sets, namely {D1, D2, D5, D6, D18, D23}, {D4, D12, D16, D17}, {D6, D7, D8, D9, D10, D11, D12, D13, D14} are used in the three ECOC methods based on OVO, OVA and Sparse random strategies, respectively (the details of the dichotomizers are shown in Table 3). Due to the dichotomizers with low accuracy such as {D18, D23, D16, D17, D13, D14} are incorporated in the code matrix of these three methods, the final performance are deteriorated accordingly. The performance of OVO is relatively better than OVA and sparse random based ECOC methods, because only two relatively inaccurate dichotomizers {D18, D23} are included and the others have very high classification accuracy. In the proposed method, the dichotomizers, namely {D1, D3, D4, D5, D6, and D7}, all with high accuracy, are

Fig. 5   The histograms of the classification accuracy of the **a** proposed method **b** OVO; **c** sparse random; **d** OVA based ECOC methods based on 2000 rounds of experiments when 80 % of the dataset randomly selected for training and the other 20 % for testing
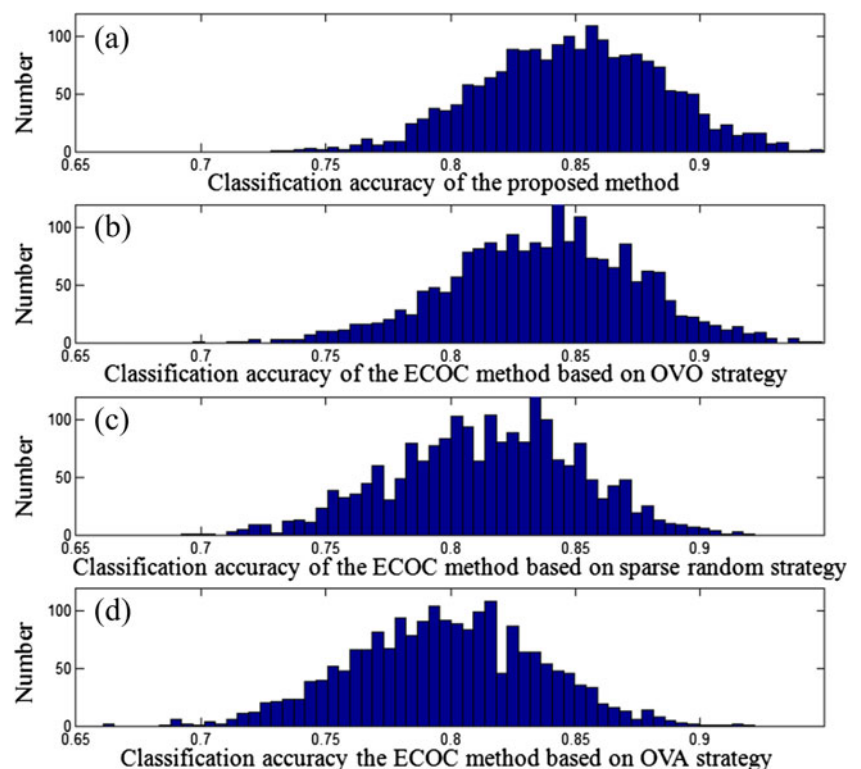
selected which ensures the performance is better than that of the other traditional ECOC methods.

Conclusions

Angle closure glaucoma is a prevalent eye disease worldwide, especially in Asia. There are four different mechanisms of ACG, which need different clinical treatments accordingly. Therefore, classification of these four mechanisms is important in automatic diagnosis of glaucoma. In this paper, a new ECOC based ensemble learning method is proposed for multiclass classification, with application to classification of four mechanisms of ACG. In the proposed method, for each possible dichotomizer, its best feature set is determined and classification accuracy is obtained by using cross-validation on the training glaucoma dataset. The dichotomizers are selected based on the maximization of both the separability of the codewords in the ECOC matrix and the classification ability of the dichotomizers. The selected dichotomizers are included into the ECOC framework. The proposed method has been experimentally applied on a glaucoma dataset including 152 patients with four different mechanisms. The classification accuracy is experimentally validated to be better than that of the other three existing ECOC methods.

There are two points make the proposed method perform better than the others: 1) the dichotomizers are optimized individually and their binary classification abilities are quantified prior to dichotomizer selection. The classification accuracy and diversity of the dichotomizers in the ECOC framework are improved by using different optimal feature set for each dichotomizer; 2) the ECOC code matrix in which the dichotomizers are all competitive with high binary classification performance and the codewords are all separated largely is determined to ensure the final classification performance. The proposed method is promising to be applied to automatic classification of different ACG mechanisms and helps doctors to make specific treatment for each mechanism.

Compliance with ethical standards

Conflict of interest There is no conflict of interest.

References

1. Shabana, N., Aquino, M. C., See, J., Tan, A. M., Nolan, W. P., Hitchings, R., Young, S. M., Loon, S. C., Sng, C. C., Wong, W., and Chew, P. T., Quantitative evaluation of anterior chamber parameters using anterior segment optical coherence tomography in primary angle closure mechanisms. *Clin. Exp. Ophthalmol.* 40: 792–801, 2012.

2. Ritch, R., Tham, C., and Lam, D., Long-term success of argon laser peripheral iridoplasty in the management of plateau iris syndrome. *Ophthalmology* 111:104–108, 2004.

3. Tian, J., Marziliano, P., Baskaran, M., Wong, H.-T., and Aung, T., Automatic anterior chamber angle assessment for HD-OCT images. *IEEE Trans. Biomed. Eng.* 58:3242–3249, 2011.

4. Nongpiur, M. E., He, M., Amerasinghe, N., Friedman, D. S., Tay, W. T., Baskaran, M., Smith, S. D., Wong, T. Y., and Aung, T., Lens vault, thickness, and position in Chinese subjects with angle closure. *Ophthalmology* 18:474–479, 2011.

5. Wirawan, A., Kwoh, C. K., Chew, P. T. K., Aquino, M. C. D., Seng, C. L., See, J., Zheng, C., and Lin, W., Feature selection for computer-aided angle closure glaucoma mechanism detection. *J. Med. Imag. Health Inform.* 2(4):438–444, 2012.

6. Nayak, J., Acharya, U. R., Bhat, P. S., Shetty, A., and Lim, T. C., Automated diagnosis of glaucoma using digital fundus images. *J. Med. Syst.* 33(5):337–346, 2009.

7. Balasubramanian, M., Zabic, S., Bowd, C., Thompson, H. W., Wolenski, P., Iyengar, S. S., Karki, B. B., and Zangwill, L. M., A framework for detecting glaucomatous progression in the optic nerve head of an eye using proper orthogonal decomposition. *IEEE Trans. Inf. Technol. Biomed.* 13(5):781–793, 2009.

8. Huang, M. L., Chen, H. Y., and Huan, J. J., Glaucoma detection using adaptive neuro-fuzzy inference system. *Expert Syst. Appl.* 32(2):458–468, 2007.

9. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F., An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn.* 44:1761–1776, 2011.

10. Garcıa-Pedrajas, N., and Ortiz-Boyer, D., Improving multiclass pattern recognition by the combination of two Strategies. *IEEE Trans. Pattern Anal. Mach. Intell.* 28:1001–1006, 2006.

11. Dietterich, T. G., and Bakiri, G., Solving multiclass learning problems via error-correcting output codes. *J. Artif. Intell. Res.* 2:263–286, 1995.

12. Allwein, E. L., Schapire, R. E., and Singer, Y., Reducing multiclass to binary: A unifying approach for margin classifiers. *J. Mach. Learn. Res.* 1:113–141, 2001.

13. Escalera, S., Pujol, O., and Radeva, P., On the decoding process in ternary error-correcting output codes. *IEEE Trans. Pattern Anal. Mach. Intell.* 32:120–134, 2010.

14. Wang, L., Zhou, N., and Chu, F., A general wrapper approach to selection of class-dependent features. *IEEE Trans. Neural Netw.* 19:1267–1278, 2008.

15. Maghsoudi, Y., Zoej, M. J. V., and Collins, M., Using class-based feature selection for the classification of hyperspectral data. *Int. J. Remote Sens.* 32:4311–4326, 2011.

16. Bagheri, M. A., Gao, Q., and Escalera, S., A genetic-based subspace analysis method for improving Error-Correcting Output Coding. *Pattern Recogn.* 46:2830–2839, 2013.

17. Pujol, O., Radeva, P., and Vitria, J., Discriminant ECOC: A heuristic method for application dependent design of error correcting output codes. *IEEE Trans. Pattern Anal. Mach. Intell.* 28:1007–1012, 2006.

18. Escalera, S., Pujol, O., and Radeva, P., Error-correcting ouput codes library. *J. Mach. Learn. Res.* 11:661–664, 2010.

19. Zhou, J., Peng, H., and Suen, C. Y., Data-driven decomposition for multi-class classification. *Pattern Recogn.* 41:67–76, 2008.

20. Rocha, A., and Goldenstein, S. K., Multiclass from binary: Expanding one-versus-all, one-versus-one and ECOC-based approaches. *IEEE Trans. Neural Netw. Learn. Syst.* 25:289–302, 2014.

21. Escalera, S., Pujol, O., and Radeva, P., Separability of ternary codes for sparse designs of error correcting output codes. *Pattern Recogn. Lett.* 30:285–297, 2009.

22. Saeys, Y., Inza, I., and Larranaga, P., A review of feature selection techniques in bioinformatics. *Bioinfomatics* 23:2507–2517, 2007.

23. Peng, H., Long, F., and Ding, C., Feature selection based on mutual information criteria of maxdependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27:1226–1238, 2005.

24. Issac Niwas, S., Lin, W., Bai, X., Kwoh, C. K., Sng, C. C., Aquino, M. C., and Chew, P. T. K., Reliable feature selection for automated angle closure glaucoma mechanism detection. *J. Med. Syst.* 39:21, 2015.

25. Issac Niwas, S., Lin, W., Kwoh, C. K., Kuo, C. C., Aquino, M. C., Sng, C. C., and Chew, P. T. K., Crossexamination for angle-closure glaucoma feature detection. *IEEE J. Biomed. Health Informatics* 20:343–354, 2016.

26. Waikato Environment for Knowledge Analysis: University of Waikato New Zealand. Available from: http://www.cs.waikato.ac.nz/ml/weka/index.html