Cross-Examination for Angle-Closure Glaucoma Feature Detection

Swamidoss Issac Niwas, Weisi Lin, *Senior Member, IEEE*, Chee Keong Kwoh, C.-C. Jay Kuo, *Fellow, IEEE*, Chelvin C. Sng, Maria Cecilia Aquino, and Paul T. K. Chew

***Abstract*—Effective feature selection plays a vital role in anterior segment imaging for determining the mechanism involved in angle-closure glaucoma (ACG) diagnosis. This research focuses on the use of redundant features for complex disease diagnosis such as ACG using anterior segment optical coherence tomography images. Both supervised [minimum redundancy maximum relevance (MRMR)] and unsupervised [Laplacian score (L-score)] feature selection algorithms have been cross-examined with different ACG mechanisms. An AdaBoost machine learning classifier is then used for classifying the five various classes of ACG mechanism such as iris roll, lens, pupil block, plateau iris, and no mechanism using both feature selection methods. The overall accuracy has shown that the usefulness of redundant features by L-score method in improved ACG diagnosis compared to minimum redundant features by MRMR method.**

***Index Terms*—Angle-closure glaucoma (ACG), anterior segment optical coherence tomography (AS-OCT), unsupervised feature selection, redundant features, Laplacian score (L-score), minimum redundancy maximum relevance (MRMR), machine learning classifier.**

I. Introduction

GLAUCOMA is a chronic eye disease, where a loss of vision occurs as a result of progressive optic nerve and astrocytes damage caused by high intraocular pressure (IOP) [1]. It is the second major cause of visual impairment and blindness worldwide with estimated 60.5 million glaucoma cases by 2010 and this number may increase to almost 80 million by 2020 [2]. Early diagnosis of this disease slows down the disease progression toward the complete vision loss. Due to the complex and diverse nature of disease pathology of glaucoma, its diagnosis heavily relies on the experience of glaucoma expert ophthalmologist. It is important to detect glaucoma in its early stages so that a patient's vision can be preserved. Detection of glaucoma is time consuming and need special skills and devices. Recent advances in medical image processing have enabled the development of different image modality-based computer aided detection systems for glaucoma using various features and classifier techniques [3]–[8].

Glaucoma can be classified into three major divisions such as angle-closure glaucoma (ACG), open-angle glaucoma, and developmental glaucoma. Each of the classes above mentioned can be further subdivided into primary and secondary categories. Among these, ACG are chronic and asymptomatic, which is more prevalent [9]. ACG causes more visual morbidity than open-angle glaucoma. Optical coherence tomography (OCT) has proven to be useful in the diagnosis of glaucoma; it uses near-infrared light to measure the distance of the anatomical structures within the eye, and is hence convenient due to its noncontact nature [10].

Anterior chamber angle (ACA) assessment is mostly used for the detection of ACG. It can be visualized and measured by using anterior segment OCT (AS-OCT) imaging techniques [11]. AS-OCT provides excellent repeatability and reproducibility for the measured corneal thickness, ACA, and depth. It has been observed that ACG could be the result of one or more mechanisms in the anterior segment of the eye and the patients with different ACG mechanisms differ in anterior segment measurements [12]. Analysis of the dimensions of the features obtained by AS-OCT and their classification into the right mechanisms, using feature selection and machine learning techniques, would be useful in the clinical diagnosis of ACG. The supervised feature selection technique by minimum redundancy maximum relevance (MRMR) method from the anterior segment measurements can determine the predominant angle closure mechanism with high accuracy [13].

Despite the fact that redundant information has the disadvantage such as being more computationally expensive and requiring larger memory space, the redundant representations may be more pliable than the nonredundant ones [14]. The removal of redundant information during the feature selection affects the precision result in text classification [15]. The advantage and significant role of redundancy in extracting useful information in signal and image analysis [16], medical image fusion [17], biological data [18], and complex medical diagnostic applications [19] has been explored. The redundant multiscale transforms such as undecimated wavelet transform, ridgelet transform, and curvelet transform which produce redundant information are widely used in many applications such as signal/image denoising, enhancement, and contour detection [14]. Since noise is usually unavoidable and spread over a small number of neighboring samples/pixels, the abovementioned redundant transforms are good for denoising signals/images [20].

Manuscript received August 27, 2014; revised November 26, 2014; accepted December 23, 2014. Date of publication January 1, 2015; date of current version December 31, 2015. This work was supported by the Ministry of Education AcRF Tire 1 Funding, Singapore, under Grant M4010981.020 RG36/11.

S. I. Niwas, W. Lin, and C. K. Kwoh are with the School of Computer Engineering, Nanyang Technological University, 639798 Singapore (e-mail: issacniwas@ntu.edu.sg; wslin@ntu.edu.sg; asckkwoh@ntu.edu.sg).

C.-C. Jay Kuo is with the Ming Hsieh Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA (e-mail: cckuo@sipi.usc.edu).

M. C. Aquino is with Eye Surgery Centre, National University Health System, 119228 Singapore (e-mail: mcdaquino@gmail.com).

C. C. Sng and P. T. K. Chew are with the Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, 119228 Singapore (e-mail: chelvin@gmail.com; ophchewp@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2387207

Reliable feature detection from AS-OCT images is important for improving the classification accuracy since the detected features are the basis of glaucoma detection and this has been less investigated. The motivation behind this study is to explore two different methods for selection of features from the anterior segment of the eye and to provide a better understanding of the use of selected redundant features for machine classification in improved glaucoma diagnosis. This paper, therefore, gives useful relevant insights based upon cross examination on the selected features and allows a detailed analysis to compare the results in terms of their accuracy and F-measure for possible interpretation. The proposed study is important for understanding of glaucoma detection, classification, and analysis using redundant features. The subsequent machine learning classification demonstrates the effectiveness toward different mechanisms of ACG.

In the rest of this paper, Section II reviews the more specific literature on ACG and feature selection methods, while the proposed methodology is presented in Section III. Section IV discusses the experimental results. A comparative study is carefully made with features by the two feature selection algorithms adopted, and common and similar/redundant features are identified to confirm the important features toward ACG mechanisms. The last section draws the conclusion.

II. Background

A. ACG and Its Mechanism

In ACG, aqueous outflow is obstructed due to iridocorneal apposition, which in turn causes a rise in IOP and optic nerve damage. The blocked drainage canals can be identified by a closed or narrow angle between the iris and the cornea (see Fig. 1), upon examination of the anterior segment of the eye [12]. AS-OCT is an imaging system, which is able to image the anterior segment of the eye using near-infrared light to measure distance of anatomical structures. From the AS-OCT scans, several features can be extracted, such as the angle-opening distance (AOD), trabecular-iris space area (TISA), and angle recess area (ARA) [21], as illustrated in the schematic diagram of anterior segment of an eye in Fig. 1.

AOD is defined as the measurement of the distance between a point of the cornea, which is "x" μm away from the scleral spur and the opposite point of the iris [22]. Values of "x" used in the dataset are 500 (AOD_500) and 750 (AOD_750). TISA is defined as the measurement of the area covering "x" μm located in the area bounded by the cornea and the iris. Values of "x" used in the dataset are 500 (TISA_500) and 750 (TISA_750). ARA is defined as the triangular area bordered by the anterior iris surface, corneal endothelium, and a line perpendicular to the corneal endothelium from a point "x" μm anterior to the scleral spur to the iris surface. Values of "x" used in the dataset are 500 (ARA_500), 750 (ARA_750), 1000 (ARA_1000), 1500 (ARA_1500), and 2000 (ARA_2000).

The characteristic features in the anterior segment of the eye that may result in ACG can be categorized into several mechanisms: thick peripheral iris roll, exaggerated lens vault, pupil block, and plateau iris [10]. Each mechanism has several

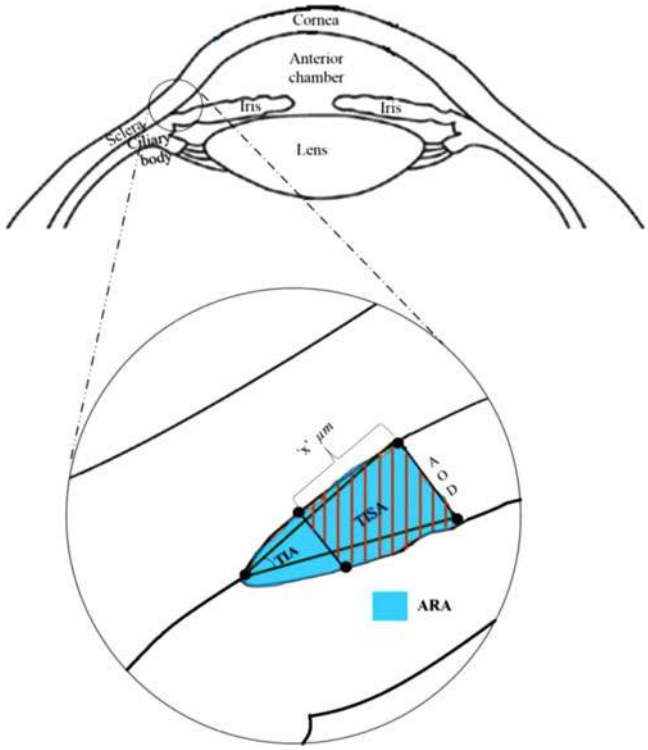

Fig. 1. Anterior segment sketch of an eye with narrow angle between iris and cornea. Inner figure representing the parameters that are used for quantifying the ACA; AOD; TISA; TIA; ARA.

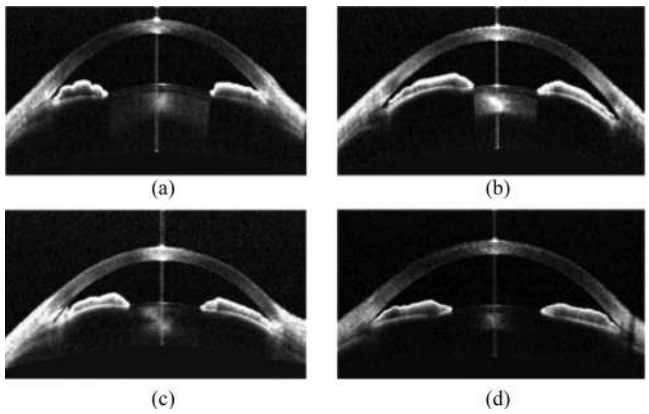

Fig. 2. AS-OCT image of an eye with the (a) iris roll mechanism, (b) exaggerated lens vault mechanism, (c) pupil block mechanism, (d) plateau iris mechanism.

characteristics that can be identified by visual observation of the AS-OCT images, and will be also detectable by feature selection algorithms and machine learning classifiers.

The iris roll mechanism [see Fig. 2(a)] can be identified by a thick iris, which narrows the angle between the iris and cornea due to the circumferential folds along the periphery of the iris. The exaggerated lens vault mechanism [see Fig. 2(b)] can be identified by the lens pushing the iris forward (upward in the image), hence reducing the angle between the iris and the cornea. The pupil block mechanism [see Fig. 2(c)] can be identified

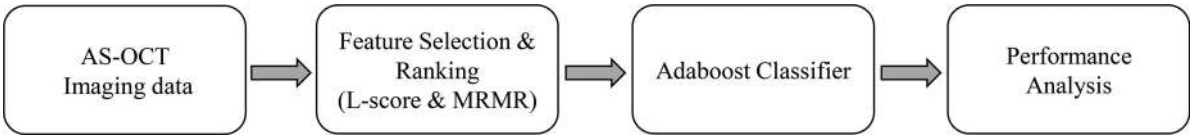

Fig. 3. Block diagram of the proposed method.

by a convex forward iris profile (arched upward in the image) causing a shallow peripheral anterior chamber. The plateau iris mechanism [see Fig. 2(d)] can be identified by a sharp rise of the iris at the periphery, close to the angle wall, before sharply turning away from the angle wall toward the visual axis.

B. Related Work in Feature Selection

This section focuses on the feature selection methods closely related to this study, as well as existing research related to ACG detection from AS-OCT images. There are various feature selection algorithms in the literature, including the maximum dependency-maximum relevance-minimum redundancy [23], Laplacian score (L-score) [24], Variance and Fisher score methods [25].

The MRMR feature selection algorithm, proposed by Peng *et al.* [23], selects features that are optimal for classification and its fast computation. It aims to reduce the correlation between the selected features themselves. The idea is that, if two features are highly correlated to each other, it would be undesirable to include them in the selected feature set, despite their relevance to the target class and, therefore, only the more relevant feature out of the two correlated features would be selected.

Hence, this requires the calculation of mutual information of a feature with not only the target class, but with other features as well. It is explored that the MRMR feature selection produces classification error rates that are significantly lower when compared to other feature selection methods such as max-dependency and max-relevance, which do not take into account the correlation between selected features. In their study, experiments were performed from 1 to 50 features. However, considering the fact that the raw datasets tested had 278 to 9703 features, it might have been useful to know the performance of the MRMR algorithm beyond 50 features selected.

The L-score feature selection algorithm, proposed by He *et al.* [24] selects features by calculating its L-score, which is an indication of its locality preserving power. The algorithm is able to select features in both supervised and unsupervised settings. L-score do not handle feature redundancy and require more computational time. The study reported significantly higher accuracy of the unsupervised L-score algorithm when compared to the unsupervised variance method [25]. In sorting four features from the Iris dataset from the UCI ML repository [26], it was able to achieve the same result as the supervised Fisher score [24], proving the capability of the L-score algorithm to detect discriminative features even in the absence of class labels.

The study on the feature selection of anterior segment features from AS-OCT images performed by Wirawan *et al.* [13] concluded that, using the MRMR feature selection method and the AdaBoost machine classifier, an accuracy of 84.39% is achieved using only ten out of the 84 features provided in the AS-OCT dataset (11.90% of the available features). This appears to be consistent with the study of the MRMR and its good performance result at smaller feature subsets [23]. Comparisons were also made against other machine learning methods, namely classification tree, support vector machine, Random forest, and Naïve Bayes classifiers.

It was noted that, while making comparisons between the different machine learning methods, the AdaBoost-MRMR method was the only method that did not include the entire available feature set. All the other machine learning methods, including AdaBoost without MRMR, were tested using all 84 features. Among the methods that were tested using the entire set of features, the AdaBoost algorithm was shown to provide the highest accuracy at 83.03%. A combination of AdaBoost with MRMR feature selection algorithm was then shown to boost the accuracy to 84.39%. It could hence be deduced that having the machine learning algorithm classify the entire set of provided features may not necessarily yield the best accuracy, and that classification of selected features could instead provide better results. Since the previous study [13] focused only on the MRMR feature selection method which is supervised, another feature selection method could be studied with the AdaBoost machine learning algorithm, which was shown to be superior to the other machine algorithms compared against, in an attempt to further boost the accuracy of the classification of the mechanisms of ACG. A comparative study of the features selected by the various feature selection algorithms could also be performed for reliability in feature selection and insight derivation toward glaucoma diagnosis. To the best of our knowledge, a cross comparison of the features selected by various feature selection algorithms for the identification of meaningful features and the usefulness of redundant features for improving ACG detection has not been performed.

III. Proposed Methodology

The proposed methodology for the feature selection and classification is shown in Fig. 3. As the range of values of input raw data varies widely, it was normalized before performing any feature selection. The L-score and MRMR algorithms were then performed on the normalized data, and six feature lists (supervised L-score technique with (1) Binary and (2) Heat kernels, unsupervised L-score technique with (3) Binary and (4) Heat kernels, supervised MRMR technique with (5) mutual information difference (MID), and (6) mutual information quotient (MIQ) methods) were generated. The feature rank lists consisted of the 84 features in the data set, arranged in order of the

importance of the features as determined by L-score and MRMR algorithms.

The sorted lists of features were then trained by the AdaBoost machine learning classifier using incremental selection subsets of $S_i(S_1 \subset S_2 \subset \ldots\ldots \subset S_n)$, i.e., for each list of 84 features, the AdaBoost training was performed 84 times with the *n*th training run being performed on the top *n* features in the list, so as to find the optimal number of top features generated per list. The AdaBoost algorithm was also run using 100 iterations and with leave-one-out cross validation to prevent overfitting of training data.

The performance of the AdaBoost classification based on the generated feature selected lists were then analyzed by generating a confusion matrix from the results of the classification, and the accuracy, F-measure, specificity, and sensitivity are calculated. Features selected by the algorithms which yielded favorable accuracy were analyzed and compared to explore the usefulness of redundant features.

A. L-score Method

As introduced in the previous section, L-score is a feature selection method that determines the importance of a feature by the construction of a graph using the sample data points and identifying which features best represent the structure of the graph [24]. The L-score algorithm is able to operate in both unsupervised and supervised contexts. In the unsupervised approach, a nearest-neighbor graph is constructed using data points, with an edge placed between two data points that are deemed to be close to each other using *k*-nearest neighbors (*k*-nn). For example, for a data point x_i with $k = 5$, the five nearest data points to x_i will have an edge placed between them. In supervised L-score, an edge is placed between two data points that share the same classification. The edges are then assigned weights using either of the following methods: Binary and Heat kernel. The Binary method is a straightforward method where all edges are assigned a weight of 1, and 0 indicates that there is no edge between the two data points. Formally, for two data points x_i and x_j, $S_{ij} = 1$ if there is an edge between them. Otherwise, $S_{ij} = 0$. In the Heat kernel method, if there is an edge between two data points x_i and x_j, and given a suitable constant *t* (which was set to a value of 1 in this study), the edge weight will be calculated as

$$S_{ij} = e^{\frac{\|x_i - x_j\|^2}{t}}. \tag{1}$$

Otherwise, if there is no weight, then $S_{ij} = 0$.

After the edge weights have been assigned, the weight matrix *S* is then formed which represents the local structure of the data space. A matrix *L*, called the graph Laplacian, may also be defined. For the *y*th feature using *n* data points, we can then define

$$\boldsymbol{f}_y = [f_{y1}, f_{y2}, \ldots, f_{yn}]^T, D = diag\,(S\mathbf{1}),\ \mathbf{1} = [1, \ldots, 1]^T. \quad L = D - S \tag{2}$$

Let

$$\hat{\boldsymbol{f}}_y = \boldsymbol{f}_y - \frac{\boldsymbol{f}_y^T D\mathbf{1}}{\mathbf{1}^T D\mathbf{1}}\mathbf{1}. \tag{3}$$

The L-score for the *y*th feature may be calculated as

$$L_y = \frac{\hat{\boldsymbol{f}}_y^T L \hat{\boldsymbol{f}}_y}{\hat{\boldsymbol{f}}_y^T D \hat{\boldsymbol{f}}_y}. \tag{4}$$

The features will be then sorted by their L-score and provided to the machine learning classifier.

B. MRMR Method

MRMR [23] is a supervised feature selection algorithm that, as the name suggests, aims to find features that are most relevant to the target classifications, while reducing the redundancy between selected features. To find features that are relevant, the mutual information between a feature and the target classification should be maximized. The mutual information between two variables *x* and *y* is defined as

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p\,(x)\,p(y)} dxdy \tag{5}$$

where *D(F, y)* represents the mutual information between a feature in set *F* and class *y*, with set *F* containing *n* features $\{x_1, x_2, \ldots, x_n\}$, the mutual information is defined as

$$\max D(F, y) = \frac{1}{|F|} \sum_{x_i \in F} I(x_i, y). \tag{6}$$

While finding relevant features is important, the MRMR algorithm is based on the idea that similar or correlated features should not be included in the feature set, regardless of the mutual information between the features and the target classification, resulting in a feature set that is compact yet accurate, where *R(F)* represents the mutual information between two features x_i and $x_j (i, j = 1, \ldots, n)$ in set *F*, the mutual information is defined as

$$\min R(F) = \frac{1}{|F|^2} \sum_{x_i, x_j \in F} I\,(x_i, x_j)\,. \tag{7}$$

The two methods in which MRMR was used in this study are the combined criteria known as MID and MIQ.

MID is defined as

$$\max \Phi\,(D, R)\,, \Phi = D\,(F, y) - R(F). \tag{8}$$

MIQ is defined as

$$\max \Phi\,(D, R)\,, \Phi = \frac{D(F, y)}{R(F)}. \tag{9}$$

In practice, candidate feature sets may be created by using incremental search methods to find the near optimal features defined by Φ. Hence, in this study, the following implementations were used

$$\text{MID}: \max_{x_j \in X - F_{m-1}} \left[I\,(x_j, y) - \frac{1}{m-1} \sum_{x_i} I(x_j, x_i) \right]. \tag{10}$$

Algorithm for AdaBoost Classifier

Input:

Sequence of N samples with labels
Distribution D over N samples
Weak learning algorithm W
Number of iterations I

Start;

Initialize weight vector $w_i^1 = \frac{1}{N}$ for $i = 1, \ldots, N$;
Do for $t = 1$ to I
Set $p' = \frac{w'}{\sum_{i=1}^{N} w_i^t}$;
Run weak algorithm W on distribution p';
Set hypothesis $h_t: X = \pm 1$;
Find weight error of $h_t: e_t = \sum_{i=1}^{N} p_i^t \left| \frac{h_t(x_i) - y_i}{2} \right|$;
Set $\alpha_t = \log\left(\frac{e_t}{1-e_t}\right)$;
Set updated weight vector $w_i^{t-1} = w_i^t e^{\left(\frac{a_i |h_t(x_i) - y_i|}{2}\right)}$;
End for
Output hypothesis $F(x) = \begin{cases} +1, \text{if } \sum_{i=1}^{T} a_i h_t(x) \geq 0 \\ -1, \text{otherwise} \end{cases}$;

End;

Fig. 4. Pseudocode of the AdaBoost algorithm.

$$\text{MIQ}: \max_{x_j \in X - F_{m-1}} \left[\frac{I(x_j, y)}{\frac{1}{m-1} \sum_{x_i} I(x_j, x_i)} \right]. \tag{11}$$

The above incremental algorithms suppose that when we have F_{m-1}, consisting of m–1 features, we will then select the mth feature from the set $\{X - F_{m-1}\}$, which is done by selecting the feature that maximizes Φ.

C. Adaboost Classifier

AdaBoost, used in this study, is the primary machine learning algorithm [27], due to the algorithm being shown to be superior in classifying ACG mechanisms in [13]. The algorithm works by boosting a weak learner for a predetermined number of iterations until a hypothesis is generated. At each iteration, classifications that have been wrongly labeled by the weak learner are more heavily weighted, and the weak learner is reapplied. The detailed procedure of the AdaBoost algorithm is shown in Fig. 4.

D. Performance Analysis

The confusion matrix [28] is used to measure the performance of a machine learning classifier upon training on a dataset. It contains information about the classifications predicted by the machine learning classifier, as well as the actual classification of the data. Table I shows a confusion matrix for a two-class classifier. *A* and *D* indicate the number of samples that have been correctly classified into positive and negative samples, respectively. *B* indicates the number of positive classes that have been erroneously classified as negative, and *C* indicates the number of negative classes that have been erroneously classified as positive (disease). Since the data in this study consists of five different classes, 5 × 5 confusion matrices will be predominantly used in this study.

TABLE I
CONFUSION MATRIX FOR A TWO-CLASS CLASSIFIER

Class		Predicted class	
		+	-
Actual class	+	A	B
	-	C	D

Using the confusion matrix, the performance of the machine learning classifier can be measured by accuracy, F-measure, sensitivity, and specificity, which in turn can be derived from true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). A TP occurs when a classifier correctly classifies a sample into its correct classification. Example: a "Lens" classification is classified as "Lens." A TN occurs when a classifier correctly does not classify a sample into a classification it should not belong to. Example: a sample that is not "Lens" is not classified as "Lens." A FP occurs when a classifier wrongly classifies a sample. Example: a "Lens" sample is not classified as "Lens." A FN occurs when a classifier wrongly classifies a sample into a classification it should not belong to. Example: a sample that is not "Lens" is classified as "Lens". Sensitivity is the measure of the classifier's ability to identify positive results and specificity is the measure of the classifier's ability to identify negative results.

Accuracy is used to measure the overall discrimination power of the classifier. It is a proportion of the total number of predictions made by the classifier that were correct. It can be defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{12}$$

Concerning the statistical significance, the F-measure is also measured to calculate the test's accuracy. It is used to measure the identification of positive class (disease) and it can be interpreted as a weighted average of the precision and recall

$$F\text{-measure} = 2. \left(\frac{p * r}{p + r} \right) \tag{13}$$

where "*p*" is the precision and "*r*" is the recall of the test to compute the score. Precision is the number of correct results divided by the number of all returned results and it can be defined as

$$\text{Precision}(p) = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{14}$$

Recall is the number of correct results divided by the number of results that should have been returned

$$\text{Recall}(r) = \text{TP}/(\text{TP} + \text{FN}). \tag{15}$$

F-measure is a composite measure which benefits algorithms with higher sensitivity and challenges algorithms with higher specificity. High values of accuracy, F-measure, sensitivity, and specificity indicate good performance of a machine learning classifier on the trained data.

TABLE II
CLASSES OF ACG IN PROVIDED DATASET

Class Number	Mechanism	Number of Samples
1	Iris Roll (IR)	21
2	Lens (L)	51
3	No Mechanism	4
4	Pupil Block (PB)	44
5	Plateau Iris (PL)	36

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The dataset used in this study consists of data samples provided by the Department of Ophthalmology in the National University Hospital (NUH), Singapore. It consists of 156 samples split into five classes which include the four mechanisms, as well as a "No mechanism" class, which indicates that the sample does not contain any of the characteristic features of any of the mechanisms. Ethics approval was obtained from the review board of NUH and the written consent was obtained from all subjects prior to AS-OCT imaging.

One eye from each patient (only nasal and temporal quadrants) was captured with the images centered on the pupil using ZEISS Visante Anterior Segment OCT Model 1000 device (Carl Zeiss Meditec, Inc., Dublin, CA, USA) under standardized dim illumination conditions (0 lux) in a room that has no windows and no lights. The scans were obtained with the standard anterior segment single-scan protocol, which produces 256 scans in 0.125 s. Each eye image was captured several times with undilated state of the pupil and only images with clearly visible scleral spurs were analyzed qualitatively by three glaucoma specialists (P. T. K. Chew, M. C. Aquino, and C. C. Sng). They were categorized into four groups of images based on ACG mechanism.

When images revealed more than one mechanisms of ACG, the dominant mechanism of angle closure was established otherwise a consensus on the dominant mechanism of angle closure was established after consideration among the three specialists. The customized software (Anterior Segment Analysis Program-ASAP, NUH, Singapore) [10] was used to quantify the ac parameters (features). Each sample consists of 84 features and a classification label. The mechanism classes, as well as their labels and number of samples per class, are listed in Table II.

Our experiment was conducted using MATLAB Toolbox from the original authors for feature selections with L-score method [24] and MRMR method [29]. The classification of the ACG with the both supervised and unsupervised selected features was conducted and evaluated using AdaBoost in MATLAB 8.0 R2012b (The Mathworks Inc., Natick, MA, USA).

The data were first normalized prior to feature selection methods, to have zero mean and a standard deviation of one. It is useful for ensuring that all features will contribute evenly during the feature selection and machine learning processes, instead of having skewed results due to some features having greater variance than others and being erroneously identified as a significant feature. The top accuracy and F-measure of the AdaBoost classification in each of the six sorted lists of features, as well as the number of features which yielded the top accuracy, are listed in Table III. It is noted that the feature lists generated by the unsupervised L-score algorithm using the Heat kernel criterion and the MRMR algorithm using the MIQ criterion yielded peak accuracy and high F-measure results (see Table III).

TABLE III
TOP ACCURACY OF ADABOOST CLASSIFICATION USING LAPLACIAN AND MRMR ALGORITHM

Feature Selection Criterion			Features	Accuracy (%)	F-measure (%)
L-score	Unsupervised	Binary	84	84.23	63.91
		Heat kernel	40	86.66	70.10
	Supervised	Binary	84	84.29	64.01
		Heat kernel	80	84.20	64.01
MRMR	Supervised	MID	84	84.23	63.90
		MIQ	10	84.39	65.60

TABLE IV
ADABOOST ACCURACY COMPARISON BETWEEN THE FEATURE SELECTION METHODS AT 10 AND 40 FEATURES

Ranked features	Unsupervised L-score	Supervised MRMR
10 features	65.23%	84.39%
40 features	86.66%	79.32%

The L-score algorithm (unsupervised) was able to produce a higher than average accuracy of 86.66% and F-measure of 70.00% using top 40 features, while the MRMR algorithm (supervised) was able to produce an accuracy of 84.39% and F-measure of 65.60% using a small set of top 10 features. The results from the MRMR algorithm were consistent with the findings made in [13].

Table IV shows the comparison of accuracies of AdaBoost training on the top 10 and 40 ranked features of both the unsupervised L-score using heat kernel (unsupervised L-score) and MRMR with MIQ (supervised-MRMR) methods. Using top 10 ranked features, unsupervised L-score giving less accuracy (65.23%), which is not a comparable performance with supervised-MRMR (84.39%). Using top 40 ranked features, supervised-MRMR giving less accuracy (79.32%), which is a comparable performance to supervised-MRMR (86.66%). The detailed performance using top 10 and 40 ranked features of both unsupervised L-score and supervised-MRMR are given in Tables V–VIII. The following sections will discuss the cross comparison between unsupervised L-score (i.e., more redundant features) using top 40 ranked features and Supervised-MRMR using top 10 (i.e., less redundant features) and 40 ranked features which are giving top accuracy and F-measure.

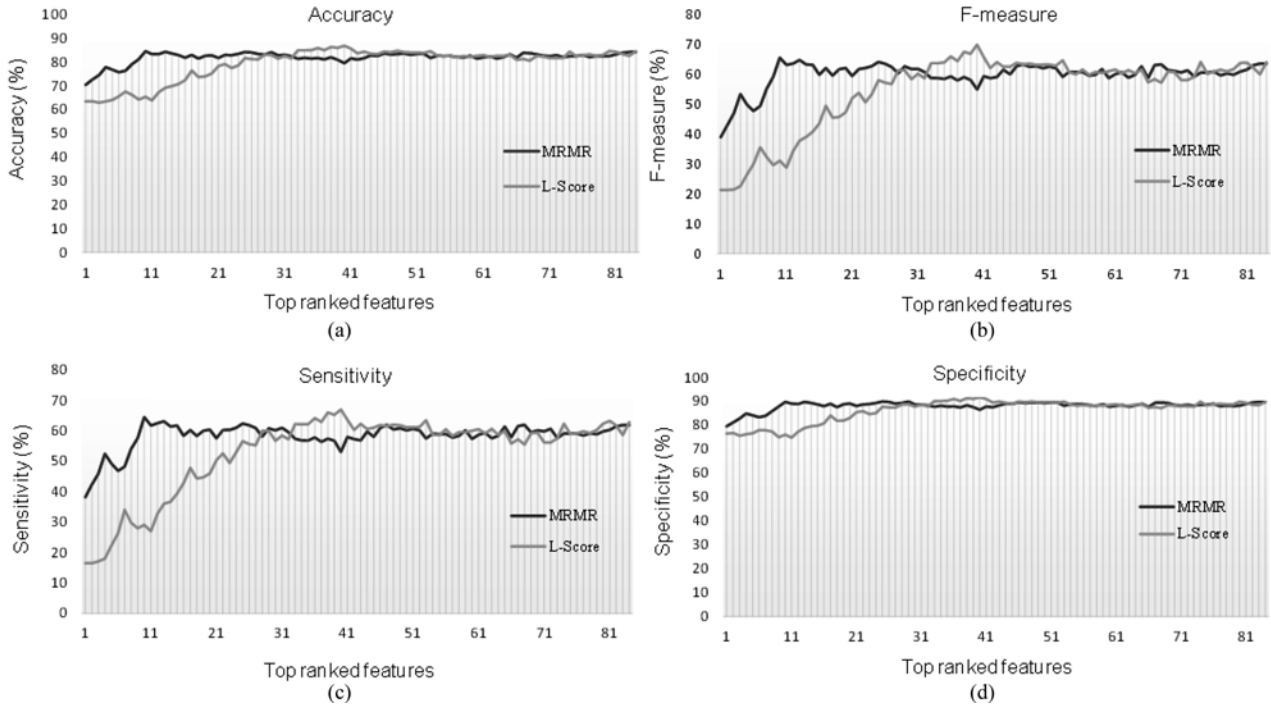

Fig. 5. Comparison graph between the AdaBoost-L-score and MRMR algorithms (a) Accuracy, (b) F-measure, (c) Sensitivity, (d) Specificity.

A. Results of AdaBoost Algorithm on Feature Selections of Unsupervised L-score and Supervised MRMR

Fig. 5 shows the comparison of unsupervised L-score and supervised MRMR. From the graph, it is found that the accuracy, F-measure, sensitivity, and specificity of the unsupervised L-score algorithm grow gradually to a peak of 86.66%, 70.10%, 67.13%, and 91.97%, respectively, at top 40 features [see Fig. 5(a)–(d)], while the accuracy, F-measure, sensitivity, and specificity of the supervised-MRMR algorithm were found to grow quickly to a peak of 84.39%, 65.60%, 64.66%, and 90.00%, respectively at its top ten features [see Fig. 5(a)–(d)].

Beyond these peaks, the accuracy and F-measure of both algorithms were found to dip slightly and eventually stabilize. The MRMR feature selection algorithm was able to perform well on a small feature set of 11.90% set (ten out of 84 features) of the entire feature, due to the selection of features that had high relevance to the target class, while reducing features that may have been correlated with the features already selected. The ten selected features would hence be very significant in the detection of the ACG mechanism. The peak accuracy for the features selected by the unsupervised L-score algorithm required 40 features, or 47.62% of the feature set, but was able to produce an accuracy of 2.27% higher than the features selected by the supervised-MRMR algorithm.

The L-score algorithm provided higher accuracy and better F-measure than MRMR algorithm since the algorithm does not remove features that are correlated with features that had already been selected, i.e., the nth and $(n+1)$th features could be very correlated with each other. Despite their correlation, small differences could exist that may influence the machine learning classification. For example, ARA_500 and ARA_750 would be much correlated due to being measured in the exact same manner only 250 μm apart from each other, but the difference in measurements 250 μm away could be significant enough to distinguish between the ACG mechanisms. This is particularly impressive, considering the fact that the L-score algorithm was unsupervised method, while the MRMR algorithm required class labels, i.e., a supervised method. This could indicate that the different ACG mechanisms had some inherently discriminating features that could be detected by the unsupervised L-score algorithm. The accuracy, sensitivity, and specificity of each class using unsupervised L-score algorithm using top 40 features is shown in Table V.

TABLE V
ACCURACIES, SENSITIVITIES, AND SPECIFICITIES OF EACH CLASS FROM UNSUPERVISED L-SCORE RESULTS WITH TOP 10 RANKED FEATURES

Class No.	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	82.69	00.00	86.00
2	55.13	35.82	69.66
3	96.79	00.00	97.42
4	60.90	31.11	72.97
5	71.15	37.84	81.51
Weighted	65.23	29.22	76.24

It is observed that the weighted accuracy, sensitivity, and specificity of the unsupervised L-score algorithm results (see Table VII) are higher than those in the supervised-MRMR results (see Table VI). Upon closer observation at the accuracies of the individual classes, it is also observed that classes 2, 4,

TABLE VI
ACCURACIES, SENSITIVITIES, AND SPECIFICITIES OF EACH CLASS FROM SUPERVISED MRMR RESULTS WITH TOP 10 RANKED FEATURES

Class No.	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	83.97	38.89	89.86
2	83.33	71.19	90.72
3	96.79	00.00	97.42
4	83.97	72.09	88.49
5	85.25	68.57	90.08
Weighted	84.39	64.66	90.00

TABLE VII
ACCURACIES, SENSITIVITIES, AND SPECIFICITIES OF EACH CLASS FROM UNSUPERVISED L-SCORE RESULTS WITH TOP 40 RANKED FEATURES

Class No.	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	81.41	25.00	87.86
2	85.26	73.33	92.71
3	97.44	00.00	97.44
4	87.82	76.60	92.66
5	89.10	78.79	91.87
Weighted	86.66	67.13	91.97

TABLE VIII
ACCURACIES, SENSITIVITIES AND SPECIFICITIES OF EACH CLASS FROM SUPERVISED MRMR RESULTS WITH TOP 40 RANKED FEATURES

Class No.	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	81.41	16.67	86.81
2	76.92	62.30	86.32
3	96.79	00.00	97.42
4	78.85	61.70	86.24
5	80.13	57.14	86.78
Weighted	79.32	53.20	86.75

and 5 (corresponding to the Lens, Pupil Block, and Plateau Iris mechanisms) had improved accuracy using the unsupervised L-score algorithm (see Table VII). This suggests that the unsupervised L-score algorithm was able to better identify the features inherent to these three classes. Class 1 (iris roll) was observed to have a higher accuracy with the supervised-MRMR algorithm (see Table VII). However, it should be noted that the sensitivities of Class 1 in both algorithms were very low (38.89% using supervised-MRMR and 25.00% using unsupervised L-score algorithm) compared to other classes. It is investigated that the iris roll mechanism could have a high probability of occurring (dominant) with other mechanisms, as it has been recognized that ACG can be a result of one or a combination of a number of mechanisms (mixed mechanisms) [15].

B. Cross Comparison of Selected Features

In this section, the comparison of features between the unsupervised L-score and the supervised-MRMR algorithm is discussed and identification of redundant features (common and similar features) between the two, in hopes of finding features that are most significant in the detection of ACG mechanisms are studied. Table IX shows the top 40 features selected by both algorithms, with the features that contributed to the peak accuracy shaded in gray.

1) Common Features Between Unsupervised L-score and Supervised-MRMR Methods: The top 40 features of unsupervised L-score and top 10 features of supervised-MRMR (see Table IX) had two features in common: "TISA_L500" and "ARA_R750." These two features also happened to be in the top 10 features of the L-score set. They would hence be deemed to be very significant in detection of ACG mechanisms. Comparing the top 40 features of both the L-score and MRMR set, the following 15 features were observed to be common: ARA_R500, TISA_L500, ARA_R750, Iris_end_concavity_R, PCA_L 500, ARA_L2000, ACD_iris_R_ML, AC_Area, ARA_R2000, mean_iris_thickness_R, ACD_LC, Iris_area_PR, LC_C, Iris_area_ML, and Iris_thickness_R_2000.

2) Similar Features Between Unsupervised L-score and Supervised-MRMR Methods: Similar features in the selected feature sets were identified by obtaining the difference of all the samples between two features and obtaining the variance of the differences, from which a similarity matrix can be generated involving all features. A low variance score between the two features would indicate that the features are highly correlated. Fig. 6 shows a graph of two features that are highly correlated: "TISA_R750" and "ARA_R750," with a variance score of 0.0017. As these two features are measurements of similar areas measured 750 μm anterior from the right scleral spur to iris surface, they would be expected to be correlated, albeit with minor differences. We have investigated through some experiments that below the variance score value 0.15, the maximum number data points tend to be very close to each other; if not, they spread out around. So, this study considers any two features to be similar if they have a variance score of under 0.15.

It was observed that, in the set of 40 features from the L-score method, there were some other features more similar to those in the set of ten from MRMR, as displayed in Table X. From Table X, it is observed that nine of the features in the unsupervised L-score set were similar to three in the supervised-MRMR set, indicating that there is a redundancy in selected features, particularly for the five features in the L-score set that were similar to ARA_R750.

This is to be expected from the L-score algorithm, as it does not consider feature-to-feature correlation. None of the top ten selected features of the MRMR set were similar to each other. A further study on the similar features in the top 40 of the L-score set is made in the next section. When comparing the top 40 features of both sets, 14 features in the L-score set were observed similar to six features in the MRMR set, as shown in Table XI, indicating that 26 out of 40 features in the L-score set were common with or similar to 18 out of the top 40 features in the MRMR set.

TABLE IX
TOP FEATURES WHICH YIELDED PEAK ACCURACIES FROM UNSUPERVISED L-SCORE AND SUPERVISED-MRMR ALGORITHM

	Unsupervised L-score		Supervised-MRMR
1	'TISA_R500'	1	'ACD_C'
2	'ARA_R500'	2	'TISA_L500'
3	'ARA_L500'	3	'Iris_thickness_IR'
4	'TISA_L500'	4	'Iris_end_concavity_L'
5	'AOD_R500'	5	'ARA_R750'
6	'AOD_L500'	6	'Iris_tilt_R'
7	'ARA_R750'	7	'Iris_thickness_L_1000'
8	'TISA_R750'	8	'Iris_area_L'
9	'CD_R'	9	'Iris_area_R500'
10	'TISA_L750'	10	'ACD_iris_L_MR'
11	'ARA_L750'	11	'LC_C'
12	'ARA_L1000'	12	'ACD_RC'
13	'ARA_R1000'	13	'Iris_thickness_R_2000'
14	'AOD_L750'	14	'Iris_end_concavity_R'
15	'Iris_area_PL'	15	'ACD_iris_L_R'
16	'Iris_end_concavity_R'	16	'ARA_R2000'
17	'Iris_length_R'	17	'ACD_R2000'
18	'Iris_thickness_R_750'	18	'PCA_R250'
19	'Iris_area_R'	19	'Iris_thickness_L_DMR'
20	'PCA_L500'	20	'PCA_L500'
21	'ARA_L2000'	21	'ACD_iris_R_L'
22	'ACD_iris_R_ML'	22	'Iris_thickness_IL'
23	'AC_Area'	23	'AOD_R750'
24	'ARA_R2000'	24	'ARA_L2000'
25	'mean_Iris_thickness_R'	25	'Iris_thickness_R_SMR'
26	'ARA_R1500'	26	'Iris_tilt_L'
27	'Iris_area_MR'	27	'Iris_area_PR'
28	'ACD_LC'	28	'Iris_thickness_L_750'
29	'ACD_L500'	29	'ACD_iris_R_ML'
30	'Iris_area_PR'	30	'AC_Area'
31	'Iris_thickness_MR'	31	'Iris_thickness_R_1000'
32	'ACD_R1000'	32	'PCA_R500'
33	'LC_C'	33	'ACD_L1500'
34	'Pupil_distance'	34	'Iris_thickness_R_DMR'
35	'Iris_length_L'	35	'ARA_R500'
36	'Iris_area_ML'	36	'ACD_LC'
37	'Iris_thickness_L_500'	37	'ACD_R1500'
38	'Iris_thickness_R_2000'	38	'CD_L'
39	'ACD_R500'	39	'mean_Iris_thickness_R'
40	'ARA_L1500'	40	'Iris_area_ML'

Abbreviations: AC-Anterior Chamber; ACD-Anterior Chamber Depth; AOD-Angle-Opening Distance; ARA-Angle Recess Area; C-Center; CD-Corneal Diameter; DMR-Dilator Muscle Region; IL-Iridolenticular contact on the Left side; ILC-Iridolenticular contact in the Center; IR-Iridolenticular contact on the Right side; L-Left side of AS-OCT image; LC-Left half of anterior Chamber; ML-Mid of iris and Lens; MR-Muscle Region; PCA-Posterior Ciliary Artery; PL-Peripheral in the Left side; PR-Peripheral on the Right side; R-Right side of AS-OCT image; RC-Right half of anterior Chamber; SMR-Sphincter Muscle Region; TISA- Trabecular Iris Space Area.

* The detailed description of each feature can be found in [10], [13].

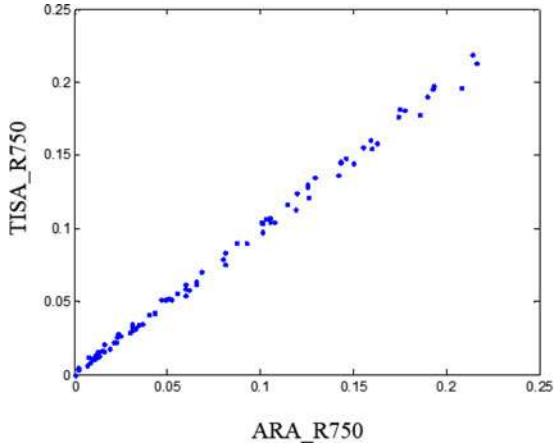

Fig. 6. Example of correlation between two features: TISA_R750 Vs ARA_R750.

TABLE X
SIMILAR FEATURES IN L-SCORE SET CORRESPONDING TO TOP TEN FEATURES IN MRMR SET

Features in Unsupervised L-score	Similar Feature in Supervised-MRMR
'AC_Area'	'ACD_C'
'TISA_R500' 'ARA_R500' 'AOD_R500' 'TISA_R750' 'ARA_R1000'	'ARA_R750'
'ARA_L500' 'TISA_L750' 'ARA_L750'	'TISA_L500'

3) Redundant/Similar Features on the Unsupervised L-score Set: As discussed in the previous section, several features in the top 40 of the unsupervised L-score set were identified to be similar to each other (see Table IX). For example, five features "TISA_R500," "ARA_R500," "AOD_R500," "TISA_R750," and "ARA_R1000" in the L-score set were found to be similar to a single feature "ARA_R750" in the supervised-MRMR set. The feature "ARA_R750" itself was also found in the L-score set and it is indicated that the L-score set had six features that were correlated with each other. This section also details the additional study of progressive removal of these similar features, and the observation of the resultant accuracies.

The features that were progressively removed, from least important to most important, were "ARA_R1000," "TISA_R750," "ARA_R750," "AOD_R500," and "ARA_R500." The feature "TISA_R500," which was the top feature in the L-score set, was not removed. As the similar features were progressively removed, it was found that the accuracy dipped beneath the peak accuracy where no features were removed, as shown in Table XII. Hence, it could be concluded that despite the features being mostly correlated with each other, as long as they are not perfectly correlated, small differences in the features could still influence the accuracy and lead to a better classification result.

TABLE XI
SIMILAR FEATURES IN L-SCORE SET CORRESPONDING TO TOP 40 FEATURES IN MRMR SET

Features in Unsupervised L-score	Similar Feature in Supervised-MRMR
'AC_Area	'ACD_C'
'ARA_L1500'	'ARA_L2000'
'ARA_R1500'	'ARA_R2000'
'TISA_R500' 'ARA_R750' 'TISA_R750'	'ARA_R500'
'TISA_R500' 'ARA_R500' 'AOD_R500' 'TISA_R750' 'ARA_R1000'	'ARA_R750'
'ARA_L500' 'TISA_L750' 'ARA_L750'	'TISA_L500'

TABLE XII
ACCURACY FROM REMOVAL OF SIMILAR FEATURES FROM L-SCORE

Features Removed	Accuracy (%)
None	86.66
'ARA_R1000'	86.30
'ARA_R1000' 'TISA_R750'	85.91
'ARA_R1000' 'TISA_R750' 'ARA_R750'	85.98
'ARA_R1000' 'TISA_R750' 'ARA_R750' 'AOD_R500'	85.45
'ARA_R1000' 'TISA_R750' 'ARA_R750' 'AOD_R500' 'ARA_R500'	85.71

C. Result Analysis

This study focused on the comparison of two feature selection algorithms, unsupervised L-score, and supervised-MRMR for understanding the importance of redundant features for ACG mechanism. It is observed that the results of classification using the AdaBoost machine learning algorithm on a dataset of 84 features and 156 samples splits into five classes. The top features selected by the unsupervised L-score method and supervised-MRMR method were compared, due to these algorithms producing high accuracies or performing on low feature sets. An analysis of the usefulness of redundant features was performed. The unsupervised L-score was able to perform classification at a relatively high accuracy of 86.66% using 40 features (47.62% of the entire feature set which are redundant), while the supervised-MRMR method was able to perform classification at an accuracy of 79.32% using a large set of 40 features and reasonable accuracy of 84.39% using a small set of 10 features (11.90% of the entire feature dataset).

Also, when comparing the F-measure, sensitivity, and specificity, unsupervised L-score made a significant result than the supervised-MRMR methods. While observing the performance of the algorithms based on each of the five classes, high accuracies, sensitivities, and specificities were observed in the classification of the lens, pupil block, and plateau iris mechanisms, with the unsupervised L-score feature selection algorithm showing improved results over the supervised-MRMR feature selection algorithm due to its redundant features. However, specificities were observed to be poor for the iris roll mechanism and samples with no mechanism type. This was attributed to a low sample count for both classes, and in the case of the iris roll class, a probable combination with other mechanisms may affect the classification.

A cross comparison between the top 40 features of the unsupervised L-score algorithm and the top 10 features of the supervised-MRMR algorithm was performed. Two features were common to both these feature sets, and nine other features in the L-score feature set were observed to be similar or correlated with three features in the MRMR feature set. It was also observed that the unsupervised L-score feature set contained features that were similar to each other, while there was no similar redundancy in the supervised-MRMR feature set. This conforms to the theories behind the unsupervised L-score and supervised-MRMR algorithms. An experiment performed on the unsupervised L-score feature set, which involved progressively removing similar features from the feature set, showed that the accuracies dipped from the peak accuracy upon removal of the similar features. This suggests that the similar features which are redundant actually contributed to the peak accuracy, and could have contained significant differences in influencing the accuracy, despite being correlated with each other.

V. CONCLUSION

It is resulted that inclusion of redundant features by the L-score method provides better performance in ACG detection than less-redundant features selected by MRMR method for glaucoma detection. From this study, it is explored that the unsupervised L-score feature selection algorithm has the capability to provide improved accuracy and F-measure with a larger feature set which consists of redundant features. On the other hand, the supervised-MRMR feature selection algorithm can be useful in conjunction with the AdaBoost machine learning classifier in the detection of ACG mechanisms if a small feature set is desirable while producing a reasonable accuracy. Employing redundant information can provide more substantive support for the complex medical diagnostic conditions, toward improved detection and classification of ACG mechanism. Hence, the selected 40 features with redundancy by unsupervised L-score method are more significant and reliable for ACG detection than the ten less-redundant features selected by the supervised MRMR method. Practically the unsupervised feature selection will be more beneficial in medical diagnosis, since the manual labeling of the huge number of samples is a more tedious task for

clinicians. Future study could involve a larger sample size for training, particularly for the iris roll mechanism and samples with no mechanism. It could also focus on the correlation between the mechanism of angle closure identified using feature selection, and response to treatment which targets each specific mechanism.

REFERENCES

[1] Glaucoma Research Foundation, Glaucoma Research Foundation. (2013). [Online]. Available: http://www.glaucoma.org/glaucoma/types-of-glaucoma.php

[2] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *Brit. J. Ophthalmol.*, vol. 90, no. 3, pp. 262–267, 2006.

[3] U. R. Acharya, S. Dua, X. Du, S. V. Sree, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 449–455, May 2011.

[4] S. S. García and E. H. Galilea, "Using artificial neural networks to identify glaucoma stages," in *The Mystery of Glaucoma*, Rijeka, Croatia: Intech, 2011, pp. 331–352.

[5] A. Pachiyappan, U. N. Das, T. V. S. P. Murthy, and R. Tatavarti, "Automated diagnosis of diabetic retinopathy and glaucoma using fundus and OCT images," *Lipids Health Disease*, vol. 11, no. 73, pp. 1–10, 2012.

[6] M. R. K. Mookiah, U. R. Acharya, C. M. Lim, A. Petznick, and J. S. Suri, "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features," *Knowl. Based Syst.*, vol. 33, pp. 73–82, 2012.

[7] M. M. R. Krishnan and O. Faust, "Automated glaucoma detection using hybrid feature extraction in retinal fundus images," *J. Mech. Med. Biol.*, vol. 13, no. 1, pp. 1–21, 2013.

[8] K. Chan, T. -W. Lee, P. A. Sample, M. H. Goldbaum, R. N. Weinreb, T. J. Sejnowski, "Comparison of machine learning and traditional classifiers in glaucoma diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 9, pp. 963–974, Sep. 2002.

[9] S. K. Seah, P. J. Foster, P. T. Chew, A. Jap, F. Oen, H. B. Fam, and A. S. Lim, "Incidence of acute primary angle-closure glaucoma in Singapore-An island-wide survey," *Arch. Ophthalmol.*, vol. 115, no. 11, pp. 1436–1440, 1997.

[10] N. Shabana, M. C. Aquino, J. See, A. M. Tan, W. P. Nolan, R. Hitchings, S. M. Young, S. C. Loon, C. C. Sng, W. Wong, and P. T. K. Chew, "Quantitative evaluation of anterior chamber parameters using anterior segment optical coherence tomography in primary angle closure mechanisms," *Clin. Exp. Ophthalmol.*, vol. 40, no. 8, pp. 792–801, 2012.

[11] A. Coyne and J. Shovlin. (2012). AS-OCT Technology: Analyzing the anterior segment *Rev. Optometry* [Online]. Available: http://www.revoptom.com/continuing_education/tabviewtest/lessonid/108148/

[12] J. Tian, P. Marziliano, M. Baskaran, H.-T. Wong, and T. Aung, "Automatic anterior chamber angle assessment for HD-OCT images," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 11, pp. 3242–3249, Nov. 2011.

[13] A. Wirawan, C. K. Kwoh, P. T. K. Chew, M. C. D. Aquino, C. L. Seng, J. See, C. Zheng, and W. Lin, "Feature selection for computer-aided angle closure glaucoma mechanism detection," *J. Med. Imag. Health Inform.*, vol. 2, no. 4, pp. 438–444, 2012.

[14] J. L Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component separation," *Adv. Imag. Electron. Phys.*, vol. 132, pp. 132–195, 2004.

[15] Y. Xu, Y. Qiu, and X. Zhao, "The effectiveness of redundant information in text classification," in *Proc. IEEE Int. Conf. Granular Comput.*, 2012, pp. 579–584.

[16] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[17] R. Singh, M. Vatsa, and A. Noore, "Multimodal medical image fusion using redundant discrete wavelet transform," in *Proc. Int. Conf. Adv. Pattern Recognit.*, Feb. 2009, pp. 232–235.

[18] M. Sadyś, A. Strzelczak, A. Grinn-Gofroń, and R. Kennedy, "Application of redundancy analysis for aerobiological data," *Int. J. Biometeorol.*, vol. 59, no. 1, pp. 25–36, 2015.

[19] R. F. Bloch, D. Hofer, S. Feller, and M. Hodel, "The role of strategy and redundancy in diagnostic reasoning," *BMC Med. Educ.*, vol. 3, no. 1, pp. 1–12, 2003. doi:10.1186/1472-6920-3-1

[20] S. Marusic, G. Deng, and D. B. H. Tay, "Image denoising using overcomplete wavelet representations," in *Proc. Eur. Signal Process. Conf*, 2005, pp. 1–4.

[21] H. Li, V. Jhanji, S. Dorairaj, A. Liu, D. S. C. Lam, and C. K. Leung, "Anterior segment optical coherence tomography and its clinical applications in glaucoma," *J. Current Glaucoma Practice*, vol. 6, no. 2, pp. 68–74, 2012.

[22] R. Koprowski, Z. Wróbel, S. Wilczyński, A. Nowińska, and E. Wylęgała. (2013). Methods of measuring the iridocorneal angle in tomographic images of the anterior segment of the eye *BioMed. Eng.* [Online]. Available: http://www.biomedical-engineering-online.com/content/12/1/40

[23] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[24] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, vol. 17, pp. 1–8.

[25] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing feature selection research: ASU feature selection repository,", School Comput., Informat. Decision Syst. Eng., Arizona State Univ., Tempe, AZ, USA, TR-10-007, 2007.

[26] R. A. Fisher, Machine Learning Repository, UCI. (1988). [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Iris

[27] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning," *J. Comput. Syst. Sci.*, vol. 55, pp. 119–139, 1997.

[28] Confusion Matrix, Dept. Comput. Sci., Univ. Regina, [Online]. Available: http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html

[29] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinform. Comput. Biol.*, vol. 3, pp. 185–205, 2005.

Swamidoss Issac Niwas received the B.E. degree in Electronics and Communication Engineering from Madurai Kamaraj University, Madurai, India, the M.E. degree in Communication Systems from Anna University, Chennai, India, and the Ph.D. degree in Medical Imaging from the National Institute of Technology, Tiruchirappalli, India.

He is currently a Postdoctoral Researcher at the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include medical imaging, biomedical signal processing, medical informatics, machine learning, and data mining.

Weisi Lin (M'92–SM'98) received the Ph.D. degree from King's College London, London, U.K.

He is currently an Associate Professor at the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, perceptual multimedia modeling and evaluation, and video compression. He has published more than 300 refereed papers in international journals and conferences.

Dr. Lin is on the editorial board of the *Journal of Visual Communication and Image Representation*, and IEEE SIGNAL PROCESSING LETTERS.

Chee Keong Kwoh received the Ph.D. degree from the Department of Computing, Imperial College of Science, Technology and Medicine, University of London, London, U.K. in 1995.

Since 1993, he has been with the Center for Computational Intelligence, School of Computer Engineering, Nanyang Technological University, Singapore. He has done significant research work and has published more than 135 international conference papers and more than 70 journal papers. His research interests include data mining, soft computing, and graph-based inference, with applications including biomedical engineering and bioinformatics.

C.-C. Jay Kuo (F'99) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, MA, USA.

He is currently the Director of the Media Communications Laboratory and a Professor of electrical engineering, computer science, and mathematics with the University of Southern California, Los Angeles, CA, USA, and the President of the Asia-Pacific Signal and Information Processing Association. He is the coauthor of about 230 journal papers, 870 conference papers and 12 books. His current research interests include digital image/video analysis and multimedia data compression.

Chelvin C. Sng received the Graduate degree from Gonville and Caius College, Cambridge University, Cambridge, U.K., with triple first class honours and distinctions.

She is a Consultant at National University Health System, Singapore, with a special interest in glaucoma, cataract surgery, and anterior segment imaging. She is actively involved in research on glaucoma surgery and imaging. She has more than 20 published papers in international journals, including *Ophthalmology and Investigative Ophthalmology* and *Visual Sciences*.

Maria Cecilia Aquino received the Doctor of Medicine degree from Faculty of Medicine and Surgery, University of Santo Tomas, Philippines and the Master of Medicine degree in Ophthalmology from National University of Singapore. She is a Resident Physician at National University Hospital, National University Health System, Singapore, with a special interest on imaging, laser, and surgical treatment of glaucoma. She received glaucoma subspecialty training under Professor Paul T.K. Chew at National University Hospital, Singapore. She has published as a main author and co-author in major Ophthalmic journals.

Paul T. K. Chew is currently an Associate Professor at Ophthalmology Department, Yong Loo Lin School of Medicine, National University of Singapore, a Senior Consultant at the Department of Ophthalmology, National University Health System (NUHS), the Head of the Glaucoma Division, NUHS, and the Strategic Director of the International Outreach, NUHS. He has spearheaded the Department of Ophthalmology as the Head for ten years from August 2001 to December 2010. His specialty is in glaucoma surgery and research, particularly angle-closure glaucoma. He is helping to develop various new laser therapies for glaucoma such as modified iridoplasty and iridotomy. He has published extensively in Ophthalmic Journals.