Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images

YU WANG,1,2 YAONAN ZHANG,1,3,9 ZHAOMIN YAO,1,2 RUIXUE ZHAO,4,5 AND FENGFENG ZHOU4,5,6,7,8

1Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, Liaoning 110169, China
2Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China
3College of Electronics and Information Engineering, Xi’an Siyuan University, Xi’an 710038, China
4College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China
5Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China
6FengfengZhou@gmail.com
7ffzhou@jlu.edu.cn
8http://www.healthinformaticslab.org/ffzhou/
9zhangyn@bmie.neu.edu.cn

Abstract: Non-lethal macular diseases greatly impact patients’ life quality, and will cause vision loss at the late stages. Visual inspection of the optical coherence tomography (OCT) images by the experienced clinicians is the main diagnosis technique. We proposed a computer-aided diagnosis (CAD) model to discriminate age-related macular degeneration (AMD), diabetic macular edema (DME) and healthy macula. The linear configuration pattern (LCP) based features of the OCT images were screened by the Correlation-based Feature Subset (CFS) selection algorithm. And the best model based on the sequential minimal optimization (SMO) algorithm achieved 99.3% in the overall accuracy for the three classes of samples.

© 2016 Optical Society of America

OCIS codes: (100.0100) Image processing; (100.2960) Image analysis; (170.4470) Ophthalmology; (170.4500) Optical coherence tomography; (100.5010) Pattern recognition.

References and links

1. J. S. Schuman, T. Pedut-Kloizman, E. Hertzmark, M. R. Hee, J. R. Wilkins, J. G. Coker, C. A. Puliafito, J. G. Fujimoto, and E. A. Swanson, “Reproducibility of nerve fiber layer thickness measurements using optical coherence tomography,” Ophthalmology 103(11), 1889–1898 (1996).
2. J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, Optical Coherence Tomography of Ocular Diseases (SLACK Incorporated, 2004).
3. G. Virgili, F. Menchini, G. Casazza, R. Hogg, R. R. Das, X. Wang, and M. Michelessi, “Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy,” Cochrane Database Syst. Rev. 1, CD008081 (2015).
4. P. A. Keane, P. J. Patel, S. Liakopoulos, F. M. Heussen, S. R. Sadda, and A. Tufail, “Evaluation of age-related macular degeneration with optical coherence tomography,” Surv. Ophthalmol. 57(5), 389–414 (2012).
5. K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff, “Segmentation of the optic disc in 3-D OCT scans of the optic nerve head,” IEEE Trans. Med. Imaging 29(1), 159–168 (2010).
6. M. K. Garvin, M. D. Abramoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, “Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search,” IEEE Trans. Med. Imaging 27(10), 1495–1505 (2008).
7. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, “Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation,” Opt. Express 18(18), 19413–19428 (2010).
8. M. Zhang, J. Wang, A. D. Pechauer, T. S. Hwang, S. S. Gao, L. Liu, L. Liu, S. T. Bailey, D. J. Wilson, D. Huang, and Y. Jia, “Advanced image processing for optical coherence tomographic angiography of macular diseases,” Biomed. Opt. Express 6(12), 4661–4675 (2015).

9. S. Farsiu, S. J. Chiu, R. V. O’Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth; Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, “Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography,” Ophthalmology 121(1), 162–172 (2014).

10. G. Gregori, F. Wang, P. J. Rosenfeld, Z. Yehoshua, N. Z. Gregori, B. J. Lujan, C. A. Puliafito, and W. J. Feuer, “Spectral domain optical coherence tomography imaging of drusen in nonexudative age-related macular degeneration,” Ophthalmology 118(7), 1373–1379 (2011).

11. K. Horie-Inoue and S. Inoue, “Genomic aspects of age-related macular degeneration,” Biochem. Biophys. Res. Commun. 452(2), 263–275 (2014).

12. J. Merl-Pham, F. Gruhn, and S. M. Hauck, “Proteomic Profiling of Cigarette Smoke Induced Changes in Retinal Pigment Epithelium Cells,” Adv. Exp. Med. Biol. 854, 785–791 (2016).

13. D. Iejima, M. Nakayama, and T. Iwata, “HTRA1 Overexpression induces the Exudative Form of Age-related Macular Degeneration,” J. Stem Cells 10(3), 193–203 (2015).

14. J. R. Evans and J. G. Lawrence, “Antioxidant vitamin and mineral supplements for slowing the progression of age-related macular degeneration,” Cochrane Database Syst. Rev. 11, CD000254 (2012).

15. M. M. Engelgau, L. S. Geiss, J. B. Saaddine, J. P. Boyle, S. M. Benjamin, E. W. Gregg, E. F. Tierney, N. Rios-Burrows, A. H. Mokdad, E. S. Ford, G. Imperatore, and K. M. Narayan, “The evolving diabetes burden in the United States,” Ann. Intern. Med. 140(11), 945–950 (2004).

16. P. J. Kertes and T. M. Johnson, Evidence-Based Eye Care (Lippincott Williams & Wilkins, 2007).

17. R. J. Tapp, J. E. Shaw, C. A. Harper, M. P. de Courten, B. Balkau, D. J. McCarty, H. R. Taylor, T. A. Welborn, and P. Z. Zimmer, “The prevalence of and factors associated with diabetic retinopathy in the Australian population,” Diabetes Care 26(6), 1731–1737 (2003).

18. M. R. Mookiah, U. R. Acharya, J. E. Koh, V. Chandran, C. K. Chua, J. H. Tan, C. M. Lim, E. Y. Ng, K. Noronha, L. Tong, and A. Laude, “Automated diagnosis of Age-related Macular Degeneration using greyscale features from digital fundus images,” Comput. Biol. Med. 53, 55–64 (2014).

19. B. Hassan, G. Raja, T. Hassan, and M. Usman Akram, “Structure tensor based automated detection of macular edema and central serous retinopathy using optical coherence tomography images,” J. Opt. Soc. Am. A 33(4), 455–463 (2016).

20. A. Albarrak, F. Coenen, and Y. Zheng, “Age-related Macular Degeneration Identification In Volumetric Optical Coherence Tomography Using Decomposition and Local Feature Extraction,” in The 17th Annual Conference in Medical Image Understanding and Analysis (MIUA)(2013), pp. 59–64.

21. Y. Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg, “Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding,” Med. Image Anal. 15(5), 748–759 (2011).

22. P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farsiu, “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images,” Biomed. Opt. Express 5(10), 3568–3577 (2014).

23. S. Mehta, “Age-Related Macular Degeneration,” Prim. Care 42(3), 377–391 (2015).

24. M. R. Mookiah, U. R. Acharya, H. Fujita, J. E. Koh, J. H. Tan, K. Noronha, S. V. Bhandary, C. K. Chua, C. M. Lim, A. Laude, and L. Tong, “Local configuration pattern features for age-related macular degeneration characterization and classification,” Comput. Biol. Med. 63, 208–218 (2015).

25. I. I. Esener, S. Ergin, and T. Yuksel, “A new ensemble of features for breast cancer diagnosis,” in International Convention Information Technology. Electronics and Microelectronics(2015).

26. E. N. G. T. K. Senthil Kumar and R. Umamaheswar, “Automatic lung nodule segmentation using autocoded region growing with morphological masking (ARGMM),” EuroMediterranean Biomedical Journal 10, 99–119 (2015).

27. V. K. Sudarshan, U. R. Acharya, E. Y. Ng, R. S. Tan, S. M. Chou, and D. N. Ghista, “Data mining framework for identification of myocardial infarction stages in ultrasound: A hybrid feature extraction paradigm (PART 2),” Comput. Biol. Med. 71, 241–251 (2016).

28. G. Z. Ying Guo, Mifti Pietikäinen, “Texture Classification using a Linear Configuration Model based Descriptor,” in The 22nd British Machine Vision Conference(BMVA Press, University of Dundee, 2011), pp. 119.111–119.110.

29. T. Ojala, M. Pietikäinen, and D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” in Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision &amp; Image Processing., Proceedings of the 12th IAPR International Conference on(1994), pp. 582–585 vol.581.

30. A. Oliva and A. Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope,” Int. J. Comput. Vis. 42(3), 145–175 (2001).

31. C. Schmid, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” CVPR 2006, 2169–2178 (2006).

32. R. Ge, M. Zhou, Y. Luo, Q. Meng, G. Mai, D. Ma, G. Wang, and F. Zhou, “McTwo: a two-step feature selection algorithm based on maximal information coefficient,” BMC Bioinformatics 17(1), 142 (2016).
1. Introduction

The established three-dimensional imaging technology, *i.e.* Optical coherence tomography (OCT), has been extensively used to capture the subtle changes in the retina, and serves as one of the standard procedures in the clinical ophthalmology examination [1, 2]. The retinal structures may be clearly captured by the OCT’s micrometer-level resolution. Some ophthalmological diseases like age-related macular degeneration (AMD) and diabetic macular edema (DME) are diagnosed based on the OCT images [3, 4]. The existing guideline diagnoses these retinal diseases based on the visual inspection of the OCT images by the well-trained ophthalmologists. The recent development of the OCT imaging technology has enabled its much wider clinical deployments and the OCT image data has accumulated rapidly. Considering that an ophthalmologist needs to handle dozens or even more patients in Asia, this also introduced a major challenge for the clinical diagnosis, with a heavier visual inspection workload for smaller lesions and larger data volume. Effective computer-aided diagnosis models provide quantitative and objective measurements to facilitate the clinical decisions. A sensitive machine learning model may facilitate the initial screening of the OCT images, and clinicians may pay special attentions to the automatically detected suspect lesions that were missed by human screening.

A number of computer algorithms were proposed to facilitate the diagnosis process of retinal disorders. Lee *et al.* extracted 15 imaging features to discriminate the retinal surface segments from rim, optic disc cup or background, which provide important information to evaluate the development stages of glaucoma [5]. Retina consists of multiple layers, of which each may respond differently by retinal disorders. The automatic segmentation of the OCT-based retina image into these intra-retinal layers has been solved by the Minimum-cost closed set algorithm [6] and Dynamic Programming algorithm [7] and Directional Graph Search algorithm [8]. A few studies tried to investigate the binary classification of age-related macular degeneration (AMD) [9] or drusen [10] based on the OCT images.

Human eye has a pigmented area in the retinal center, the macula, of which lesion is one of the major causes of vision weakening, eye shadow and blindness [11]. Two major macular
lesions receive intensive research interests, i.e. age-related macular degeneration (AMD) and diabetic macular edema (DME). AMD is an irreversible macular lesion, with the symptoms of blurred or completely no vision in the visual field center [12]. This degeneration gets worsen with age, and has two sub-types, i.e. dry and wet AMDs [13]. Dry type AMD accounts for up to 90% of the AMD patients in the US, but a significantly different ratio was observed for the Japanese population [13]. No cure is known for the lost sight, but proper supplements and living habits may slow the degeneration process if being diagnosed in the early stages [14].

Diabetic macular edema (DME) is another common macular lesion, and it is one of the leading causes of complete blindness [10, 15]. Most of the diabetes patients develop DME after 20 years of diabetes [16] but the symptoms may be slowed down with proper monitoring and treatments [17]. Although the early-stage symptoms of DME are difficult to be noticed by the patient, experienced clinicians may see the narrowed or completely blocked retinal blood vessels using the fundus photography. The disordered blood vessels will start to burst, bleed and blur the vision. Eventually the sight may be completely gone. These macular diseases may develop at all ages, and their early diagnosis will greatly improve the treatment effect and life quality.

Computer aided diagnosis technologies based on image texture and spatial features have been attempted for the early detection of the macular diseases. Farsiu et al. assessed the thickness of human retinal layers to discriminate AMD patients from the normal controls [9]. The distributions of grayscale features in the digital fundus images may also be integrated with the entropies and higher order spectra to detect the AMD samples [18]. By using five features analyzed from thickness profiles and cyst fluids, Hassan et al. built a model to classify macular edema and central serous retinopathy (CSR). The Support Vector Machine (SVM) classify trained on 90 images and realized a 97.77% accuracy [19]. T-test was utilized to rank all the features and the top 54 ranked features were chosen to train a Support Vector Machine (SVM) model. The model achieved a recognition rate of 95.07%. Albarrak et al. demonstrated that 3D-OCT based images may be used to discriminate AMD from the normal controls using the local description factors, e.g. local binary pattern from three orthogonal panels (LBP-TOP) and histograms of gradient features [20]. The Bayesian network trained with these features achieved an accuracy of 91.4%. Liu et al. employed the multi-scale spatial pyramid features and reduced the dimensions of the local binary patterns (LBPs). The optimized model used multiple binary SVM classifiers, and achieved at least 93% in the area under the receiver operator characteristic curve (AUC) for the four classes of samples, i.e. normal, macular edema, macular hole and age-related macular degeneration [21].

This work proposed a multi-class model of detecting age-related macular degeneration (AMD), diabetic macular edema (DME) and normal controls using the linear configuration patterns (LCPs). Only 23 were optimized from the 5,000 + LCP pyramid features, and an overall accuracy of 98.0% was achieved based on the individual OCT images. The patient level accuracy of 100% was achieved for both the DME and normal samples. The AMD samples received a slightly smaller accuracy of 93.33%. This work is organized in the following sections. Section 2 introduced the rationale of the proposed algorithm. Section 3 described the experimental results. The last section concludes the whole manuscript.

### 2. Material and methods

This work investigated the classification problem of three retinal OCT images, i.e. age-related macular degeneration (AMD), diabetic macular edema (DME) and the normal macula. The experiment procedure has four steps, i.e. OCT image preprocessing, feature extraction and selection, building the classification model, and predicting each pathology group, as shown in Fig. 1.
2.1. Data set

This study evaluated the proposed procedure using the publically available OCT data set provided by the joint efforts from Duke University, Harvard University and University of Michigan [22]. This data set is denoted as SD-OCT, and consists of over 3,000 OCT images from 45 participants. There are 15 patients with dry AMD, 15 patients with DME, and 15 healthy human controls.

The 2D horizontal cross-sectional retinal slice from the 3D macular imaging data was used in the proposed procedure, as shown in Fig. 2. The ophthalmological diagnoses of both AMD and DME are based on the patterns in such 2D horizontal cross-sectional OCT slices according to the clinical guidelines [23]. AMD retina begins with accumulating drusen between the retinal pigment epithelium (RPE) and the underlying choroid. The usually dome-shaped retina atrophy and scar will be developed and cause severe damage to the sight gradually, as shown in Fig. 2(a). DME-induced micro vascular changes in retina will cause the thickening of the basement membrane. The mal-functioning and liquid accumulating vascular walls will show black blobs in the OCT images, as shown in Fig. 2(b). Normal macula shows clear boundaries between layers and regularly shaped vessels at the center, as in Fig. 2(c).

2.2. Linear configuration patterns (LCP)

This study utilized the linear configuration pattern (LCP) features to represent a retina image. The topological LCP features of image texture are widely utilized by biomedical researchers to classify disease patterns from the control tissues. Mookiah et al. utilized the linear
configuration pattern method to automatically detect age-related macular degeneration in the fundus images [24]. They ranked the features individually and achieved 97.8% in accuracy for the binary classifier, which focused on disease and control samples. The LCP-based features were also utilized on image-based biomedical problems, such as mammogram-based breast cancer diagnosis [25], CT-based lung nodule segmentation and identification [26], and ultrasound-based myocardial infarction staging [27], etc. LCP was demonstrated to work well on describing both the microscopic configurations and the local structural features [28]. The local structural information is calculated by the local binary patterns (LBPs) [29]. The rotation invariant uniform patterns (riu) are extended from the original version of LBP operator.

2.3. Multi-scale feature extraction

LCP features were calculated on multiple scales of the OCT images in this study. Depending on the confusing structure appearance between shadow and macular edema, the context which shows the overall appearance of retina area are required being correctly interpreted. Figure 3 shows the shadowing effects which caused by uneven lighting or light block from the opaque media. Although the LBP(riu) features are apply to rotation invariant patterns, the extracted feature cannot present the retina characteristics properly. We demonstrate that in the experiment of Section 3.3. Besides, previous studies demonstrated that micro- and macro-scale images contribute complementary discriminating characteristics for an image-based classification problem [30]. Features of different image scales may be formulated using a pyramid strategy, as shown in Fig. 4. A 3 level spatial pyramid strategy was utilized in this study [31]. For each level $m$, we divided the image length and width by $2^m$. The overlapped blocks are also taken into consideration. The LCP features on the each block of different level were calculated and combined as a global image descriptor. The spatial pyramid calculation of linear configuration patterns was abbreviated as SP-LCP.

Fig. 3. Shadowing effect.

Fig. 4. Spatial pyramid image feature extraction. This study utilized three-level LCP image features.
2.4. Feature optimization and classification

It is hypothesized that not all the features calculated from the OCT images may contribute to the classification performance of AMD, DME and healthy controls. A feature subset was chosen by the following feature selection algorithms, and evaluated for their classification performances by the 10-fold cross validation strategy. Attribute/feature evaluation algorithms are divided into two major classes, i.e. filters and wrappers. Filters evaluate the association of each feature with the class labels, with the assumption that this association is independent between features. The top-ranked features will be chosen as the feature subset for classification modeling. Wrappers screen for a subset of features with an optimal performance measurement, which is usually the classification accuracy or error rate. Wrappers are usually slower but more accurate than filters [32]. So this study chose two wrapper feature selection algorithms CFS (Correlation-based Feature Subset) and CSE (Classifier Subset Evaluator) to find the feature subset significantly associated with the phenotypes [33, 34]. Previous studies focused on the discrimination power of the Linear Configuration Pattern (LCP) features and did not try to find a better and smaller feature subset using the feature selection strategies [35, 36]. This study proposed a hypothesis that a subset of the LCP features may generate a better classification performance as well as a simpler classification model. CFS with the best first searching heuristic rule evaluates the worth of a subset of features by considering the individual predictive ability of each feature as well as the redundancies degree between them. Subsets of features that are highly correlated with the class while having low inter correlation are chosen which shrink the feature subset from 5000 + into 493. The experimental data supports this hypothesis by achieving an averaged 3.0% improvement in classification accuracy with only 9.4% of the total feature set. CSE with the best fist searching heuristic rule evaluates attribute subsets on training data or a separate hold out testing set. The classification algorithm SMO is embedded to estimate the 'merit' of a set of features. CSE reduced the numbers of features from 493 to only 23 features. Previous comparative studies demonstrated that the above feature selection algorithms usually performed best on selection a feature subset with good classification performances, while feature selection algorithms based on the method of filter like t-test optimize the phenotype association significances [32].

Classification algorithms are also essential to train a good model [32, 37]. We tested our data set using one representative from each classification algorithm group, i.e. the quadratic programming based algorithm sequential minimal optimization (SMO) [38], the neural network algorithm multi-layer perceptron based on back propagation (BP) [39], the kernel based algorithm Support Vector Machine with polynomial kernel (SVM) [40], the linear regression based classification algorithm Logistic Regression (LR) [41], the Bayesian algorithm Naïve Bayes (NBayes) [42], the tree based algorithm J48 decision tree (J48) [43], and the ensemble forest algorithm Random Forest (RF) [44]. The experimental data in the later section suggested that classification algorithm SMO tends to achieve better performances, compared with the other algorithms.

This study employed the implementations of these feature selection and classification algorithms in the data mining program Weka version 3.7.12 [45] and libSVM version 3.21 [46].

2.5. Experimental settings

There were 15 participants for each of the three groups, AMD, DME and normal control. Each of the participants has multiple scans of OCT images, and the normal tissue slices of AMD or DME patients were excluded [22]. Due to the technical reasons, including irregular lighting and motion blurring, some images didn’t provide a clear view of retina and were excluded. The final OCT image data set consists of 453 AMD images, 511 DME images and 1,403 normal images.

10 fold cross validation strategy was utilized to evaluate the classification performance of a given feature subset. Ten repeated runs with different random seeds were conducted to
avoid the data set splitting bias. The accuracy, specificity, sensitivity and area under the receiver operator characteristic curve (AUC) were also employed to evaluate the parameter-independent classification performances. The mean value and standard deviation of each index are calculated. The ROC curve is an intuitive and parameter-independent way to illustrate a classification model, but it’s difficult to generate the ROC plots of each algorithm due to the multiple runs of 10-fold cross validations. So the following sections do not provide the ROC plots.

3. Experimental results and discussion

The proposed automatic macular disease detection algorithm was evaluated from the following aspects. Different pyramid frameworks were evaluated for their classification performances in the section 3.1. Different feature extraction and selection algorithms were compared to find a best feature subset in the section 3.2. Section 3.3 compared different classification algorithms, so that a best classification model may be trained. The comparison with the existing studies was also conducted in the section 3.4.

3.1. Evaluation of pyramid settings

A number of pyramid settings were investigated for their classification performances, as similar in [21]. The image pyramid (IP) setting extracted the LCP features from the three scaling levels of the original images, as shown in Fig. 5(a). The Spatial Pyramid (SP) setting calculated the LCP features from the three levels of images scaled to the same size $80 \times 248$ in pixels, as in Fig. 5(b). The Multi-scale spatial pyramid setting (MSSP) carried out both the scaling and sub-image splitting before calculating the LCP features, as shown in Fig. 5(c). The overlapping pyramid setting is to calculate the LCP features from the fixed-size sub-images, as illustrated as a rectangle with a dashed frame in Fig. 5.

The performance measurement sensitivity (TPR, true positive rate) is defined as (number of true positives)/(number of true positives + number of false negatives) = TP/(TP + FN). Specificity is usually defined as (number of true negatives)/(number of negatives) for a binary classification problem. In our case of three-class classification problem, specificity is defined in the same way for each class label, where the negative samples are the samples not in the current class label. So the overall sensitivity and specificity were averaged over the three class labels. The sensitivity of a class label is also the prediction accuracy. The overall accuracy is defined as the ratio (number of correctly predicted samples)/(total number of samples).
The first round of feature selection was conducted to reduce the features for training a classifier, as shown in Table 1. All the pyramid settings generate at least 500 features, and there is a possibility of classifier over-fitting. So the CFS with the best first search strategy was carried out to find a feature subset with reasonable classification performances. The setting IP extracted the minimum number of features (507) from the data set, whereas SP:O extracted many more features (5,239). After the screening of CFS on the feature sets, at least 84% decreasing in the feature number was achieved for all the five pyramid settings. Only 8.9% of the features were left for the setting MSSP:O.

Table 1. Feature summary of the five pyramid settings. The tag “:O” means overlapping for this pyramid setting. The row “Original” is the number of features generated by each pyramid setting, and the row “Selected” is the number of features selected by the CFS algorithm. The row “S/O” gives the ratio between “Selected” and “Original”.

|        | MSSP | SP | IP  | MSSP:O | SP:O |
|--------|------|----|-----|--------|------|
| Original | 3549 | 3549 | 507 | 5239   | 5239 |
| Selected | 397  | 407 | 81  | 464    | 493  |
| S/O     | 11.2%| 11.5%| 16.0%| 8.9%   | 9.4% |

Table 2. Classification performances of different feature subsets. All the feature subsets were selected by the CFS algorithm from the original features of the five pyramid settings. The tag “:O” means overlapping for this pyramid setting. Each of the first three rows gave the classification accuracy for the classes AMD, DME and Normal, respectively. The other rows gave the overall classification accuracies (Acc), sensitivities, specificities and Area Under the receiver operator characteristic Curve (AUC) of the five pyramid settings. The highest result in each row was highlighted in the bold font.

|        | MSSP | SP | IP  | MSSP:O | SP:O |
|--------|------|----|-----|--------|------|
| AMD    | 0.997 ± 0.003 | 0.992 ± 0.004 | 0.980 ± 0.002 | 0.998 ± 0.004 | 0.996 ± 0.002 |
| DME    | 0.979 ± 0.004 | 0.976 ± 0.009 | 0.892 ± 0.004 | 0.969 ± 0.003 | 0.986 ± 0.007 |
| Normal | 0.992 ± 0.001 | 0.995 ± 0.001 | 0.984 ± 0.001 | 0.994 ± 0.001 | 0.996 ± 0.001 |
| Acc    | 0.990 ± 0.001 | 0.990 ± 0.001 | 0.964 ± 0.001 | 0.990 ± 0.001 | 0.993 ± 0.001 |
| Sensitivity | 0.994 ± 0.001 | 0.995 ± 0.001 | 0.977 ± 0.001 | 0.994 ± 0.001 | 0.996 ± 0.001 |
| Specificity | 0.993 ± 0.001 | 0.993 ± 0.001 | 0.973 ± 0.009 | 0.992 ± 0.001 | 0.996 ± 0.001 |
| AUC    | 0.993 ± 0.001 | 0.993 ± 0.001 | 0.973 ± 0.009 | 0.992 ± 0.001 | 0.996 ± 0.001 |
The five pyramid settings were evaluated for their classification performances using the sequential minimal optimization (SMO) [38], which proved to be best classifier in the later section. Firstly, all the five pyramid settings demonstrated good classification performances, as shown in Table 2. The worst single-class accuracy 89.2% was observed for the DME samples based on the IP features. The IP features also achieved the worst overall accuracy 96.4%. All the other four pyramid settings achieved at least 99% in the overall accuracy, and the setting SP:O achieved 99.3% in the overall accuracy. The SP:O features performed slightly worse than the MSSP:O features, by 0.2% decreased accuracy for the sample class AMD. So the following sections utilized the pyramid setting SP:O as the default feature subset.

3.2. Evaluation of feature extraction and selection algorithms

The above section demonstrated the importance of having a good feature set for training a classifier. The clinical diagnosis of AMD and DME relies on the local macular structure abnormalities. So we compared the aforementioned features with the widely used local binary pattern (LBP) features [29] and the local configuration pattern (LCP) features [28].

Table 3. The overall accuracies of different feature sets. The columns LBP and LCP represented the features calculated by the LBP and LCP algorithms. The column SP:O was the best feature subset evaluated in the above section. The column SP:O + CSE represented the data based on the SP:O features further screened by the feature selection algorithm CSE. The first three rows gave the accuracies for the sample classes AMD, DME and Normal, respectively. And the row Acc gave the overall accuracies for different feature sets. Sensitivities, specificities and AUC are also listed below.

|          | LBP  | LCP  | SP:O | SP:O + CSE |
|----------|------|------|------|------------|
| AMD      | 0.742 ± 0.006 | 0.953 ± 0.005 | 0.996 ± 0.002 | 0.978 ± 0.008 |
| DME      | 0.434 ± 0.005 | 0.905 ± 0.005 | 0.986 ± 0.007 | 0.940 ± 0.004 |
| Normal   | 0.928 ± 0.001 | 0.980 ± 0.003 | 0.996 ± 0.001 | 0.996 ± 0.001 |
| Acc      | 0.786 ± 0.001 | 0.959 ± 0.002 | 0.993 ± 0.001 | 0.980 ± 0.001 |
| Sensitivity | 0.868 ± 0.001 | 0.976 ± 0.001 | 0.996 ± 0.002 | 0.988 ± 0.001 |
| Specificity | 0.826 ± 0.001 | 0.971 ± 0.012 | 0.996 ± 0.002 | 0.984 ± 0.001 |
| AUC      | 0.868 ± 0.001 | 0.971 ± 0.012 | 0.996 ± 0.002 | 0.984 ± 0.001 |

Table 3 showed that local description features from different scaling levels were complementary to each other and their integrations increased the macular disease classification accuracy. The data suggested that LBP features didn’t separate the AMD and DME samples well, and the three class classification problem only got an overall accuracy of 78.6%. The LCP features improved the classification performance by 17.7% in the overall accuracy. But another 3.0% improvement was achieved by the 493 features of the SP:O setting. Another round of feature selection was carried out on the 493 features using the CSE strategy with SMO as the embedded classifier. Only 23 features were selected for further analysis, and this small feature subset achieved 98.0% in the overall accuracy, a slight decrease (1.3%) compared with the feature subset SP:O in Table 3.

A smaller feature subset takes less time to build a classification model, and avoids the possibility of over-fitting. Table 4 showed that it took only 0.10 second to build a SMO classification model based on the 23 features of SP:O + CSE, while the 493 features of SP:O took 3.08 seconds.

Table 4. Time to build a SMO classification model using the two feature subsets. There are 493 and 23 features for the settings SP:O and SP:O + CSE, respectively.

|          | SP:O | SP:O + CSE |
|----------|------|------------|
| Building time (s) | 3.08 | 0.10 |

3.3. Evaluation of classification algorithms

A comprehensive comparative study was conducted to evaluate how different classification algorithms perform on this data set, as shown in Table 5. SMO performed the best on the
AMD samples with an overall accuracy 97.8%, and was ranked 2nd on the DME samples with 0.3% decreased accuracy compared with LR. LR performed the best on detecting the DME samples, with an overall accuracy 94.3%. And SMO performed the best on both Normal samples and the overall data set. In Table 6, SMO performed the best on the accuracy, specificity and sensitivity, and was ranked 2nd on the AUC performance with 1.3% decreased area under roc curve compared with BP and LR. The decision tree J48 performed the worst while RF performed much better. This suggests that the OCT LCP features harbor complicate associations and the simple decision rules trained by the J48 algorithm does not fit well with the inner patterns of this data set. So SMO was proposed as the default classification algorithm in this study.

Table 5. The classification accuracy of the seven classification algorithms on the data set. The abbreviations of the classification algorithms were defined in the section Material and methods. The first three rows gave the prediction accuracies for the sample classes AMD, DME and Normal, respectively. The last row Acc gave the overall accuracies of the seven classification algorithms on the data set.

|       | SMO       | BP         | LR         | RF         |
|-------|-----------|------------|------------|------------|
| AMD   | 0.978 ± 0.008 | 0.975 ± 0.005 | 0.965 ± 0.003 | 0.963 ± 0.004 |
| DME   | 0.940 ± 0.004 | 0.936 ± 0.005 | 0.943 ± 0.004 | 0.912 ± 0.004 |
| Normal| 0.996 ± 0.001 | 0.986 ± 0.002 | 0.990 ± 0.001 | 0.994 ± 0.002 |
| Acc   | 0.980 ± 0.001 | 0.973 ± 0.002 | 0.973 ± 0.001 | 0.970 ± 0.002 |

Table 6. The classification accuracy of the seven classification algorithms on the data set. The abbreviations of the classification algorithms were defined in the section Material and methods. The rows gave the accuracy, specificity, sensitivity and AUC, respectively.

|       | SMO       | BP         | LR         | RF         |
|-------|-----------|------------|------------|------------|
| Acc   | 0.980 ± 0.001 | 0.973 ± 0.002 | 0.973 ± 0.001 | 0.970 ± 0.002 |
| Sensitivity | 0.980 ± 0.004 | 0.973 ± 0.002 | 0.974 ± 0.001 | 0.970 ± 0.002 |
| Specificity | 0.988 ± 0.001 | 0.985 ± 0.001 | 0.985 ± 0.001 | 0.980 ± 0.001 |
| AUC   | 0.984 ± 0.001 | 0.997 ± 0.001 | 0.997 ± 0.001 | 0.996 ± 0.001 |

3.4. Comparison with the existing studies

The best model achieved in this study was compared with the existing model proposed by Liu et al. [21], as shown in Table 7. The 10-fold cross validations were repeated 10 times with different random seeds. The same classification performance measurement was calculated for the comparison. The measurement Area Under the receiver operator characteristic Curve (AUC) was calculated for the model proposed in this study. AUC evaluates a classification problem independently of different cutoff parameters. The proposed model in this study
outperformed the study Liu et al. in all the three sample classes, and achieved 0.064 improvement in the measure AUC.

Table 7. The AUC results of the two methods. The SMO based classifier trained over the data set SP:O + CSE was compared with the results in the study Liu et al.

|              | SP:O + CSE    | Liu et al. |
|--------------|---------------|------------|
| **AMD**      | 0.995 ± 0.001 | 0.926 ± 0.009 |
| **DME**      | 0.970 ± 0.003 | 0.846 ± 0.011 |
| **Normal**   | 0.987 ± 0.001 | 0.969 ± 0.002 |
| **AUC**      | 0.984 ± 0.001 | 0.920 ± 0.001 |

Another study evaluated the classification problem by individual subjects based on the same data set [22]. Srinivasan et al. extracted the multi-scale Histogram Of Oriented descriptors (HOG) as the feature vector of an OCT image, and the model was trained using the Support Vector Machine (SVM) algorithm [22]. 45 experiments were carried out using the leave-three-subjects-out cross validation strategy. That is to say, for an experiment, one subject was randomly selected from each of the three sample classes, and the rest 42 subjects were used as the training data set. Srinivasan et al. achieved 100% in accuracies for the two disease classes AMD and DME, but incorrectly predicted two normal samples, as shown in Table 8. A rule of majority was employed to generate the patient level prediction. That is to say, a patient is defined to belong to a class label, which is the predicted class label of most OCT images of this patient. The best model achieved in this study accurately detected all the DME samples and Normal samples, but missed one AMD sample.

Table 8. Subject based classification performances of two studies. The accuracy was calculated based on whether the majority of a subject's OCT images were correctly predicted. Since there are three classes, the measurement accuracy is defined as the true positive rate for each class.

|              | This study     | Srinivasan et al. |
|--------------|----------------|-------------------|
| **AMD**      | 14/15 = 93.3%  | 15/15 = 100%      |
| **DME**      | 15/15 = 100%   | 15/15 = 100%      |
| **NORMAL**   | 15/15 = 100%   | 13/15 = 86.7%     |

4. Conclusion and future scopes

This study demonstrated that a carefully selected set of features from different scaling levels were essential for the classification performances of two macular diseases compared against the normal controls using the SD-OCT images. The proposed model also outperformed the existing models on the levels of both individual OCT images and patients. Firstly, the LCP features of different pyramid settings worked very well for the investigated classification problem. The original version of non-overlapping LCP blocks may be improved by allowing additional overlapping blocks between the neighboring ones. Secondly, the multi-scaling pyramid setting introduced a major improvement for both LBP and LCP features. And two rounds of feature selection procedures significantly reduced the feature numbers, while retaining similar accuracies. Lastly, the classification algorithms SMO and LR achieved very good performances, and SMO performed the best overall performance. The best model in this study also outperformed the existing studies.

In conclusion, the proposed model may represent an automatic macular disease detection system based on the SD-OCT images, and provided useful strategies for other imaging-based clinical diagnosis modeling. Further investigations of how to optimize the scaling and aligning steps, and the parameters of the machine learning algorithms are planned. Confirmation in a larger population with more diversity and a user-friendly program interface will also facilitate the clinical application of the proposed model.
Funding
This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040400) and the startup grant from the Jilin University. Thanks for the help from the major research projects of Xi’an Siyuan University (XASY-B1601).

Acknowledgments
Help on image coding from Mr. Jikui Liu is appreciated. The helpful comments from the anonymous reviewers are acknowledged.