Learning Feature Extractors for AMD Classification in OCT Using Convolutional Neural Networks

Dafydd Ravenscroft[1] Jingjing Deng[1] Xianghua Xie[1]
1. Department of Computer Science, Swansea University, United Kingdom
http://csvision.swansea.ac.uk
Louise Terry[2] Tom H. Margrain[2] Rachel V. North[2] Ashley Wood[2]
2. School of Optometry and Vision Sciences, Cardiff University, United Kingdom

***Abstract*—In this paper, we propose a two-step textural feature extraction method, which utilizes the feature learning ability of Convolutional Neural Networks (CNN) to extract a set of low level primitive filter kernels, and then generalizes the discriminative power by forming a histogram based descriptor. The proposed method is applied to a practical medical diagnosis problem of classifying different stages of Age-Related Macular Degeneration (AMD) using a dataset comprising long-wavelength Optical Coherence Tomography (OCT) images of the choroid. The experimental results show that the proposed method extracts more discriminative features than the features learnt through CNN only. It also suggests the feasibility of classifying different AMD stages using the textural information of the choroid region.**

I. Introduction

Texture analysis is a fundamental problem in computer vision and image processing, and it has been widely applied to surface defection discovery [1] and image-based medical diagnosis [2], which generally involve classifying input based on the textural appearance of an object. There are a wide range of different methods which can be used in feature extraction, such as wavelets, co-occurrence matrices, local binary patterns, Markov random fields. Most works focus on designing or applying hand-crafted features to present the appearance structures for specific problems [3]. Gabor filters that extract impulse responses from multi-scales and different orientations often outperform other methods, particularly in texture discrimination. This is due to the similarity between its feature descriptors and the stimulation mechanism of human visual system. However, designing hand-crafted filters is time consuming, and it is extremely difficult to find the right features for some challenging problems. For example, in this paper, we are going to investigate the feasibility of classifying different stages of AMD based on the texture information of choroidal OCT images. Age-Related Macular Degeneration (AMD) is a progressive eye disease which is the leading cause of vision loss in the developed world [4]. AMD is a progressive disease; an early stage, where vision loss is minimal, can develop to one of two end stages; dry (geographic atrophy) or wet (neovascular AMD) [5]. The choroidal vascular structure changes with this progression; loss of vessels and connectivity alters the shape and texture compared to healthy eyes. Fig. 1 shows examples of the choroid OCT image in different AMD categories. There are two major difficulties of applying traditional hand-crafted methods to this problem. Firstly, for the choroid section, the variations of local textural appearance are very subtle and nearly random in high frequency bands, while such variations transmute across slides in low frequency bands, which makes feature designing a non-trivial task. Secondly, the choroid sections are irregular in shape across different subjects, which results in arbitrary length of feature descriptors.

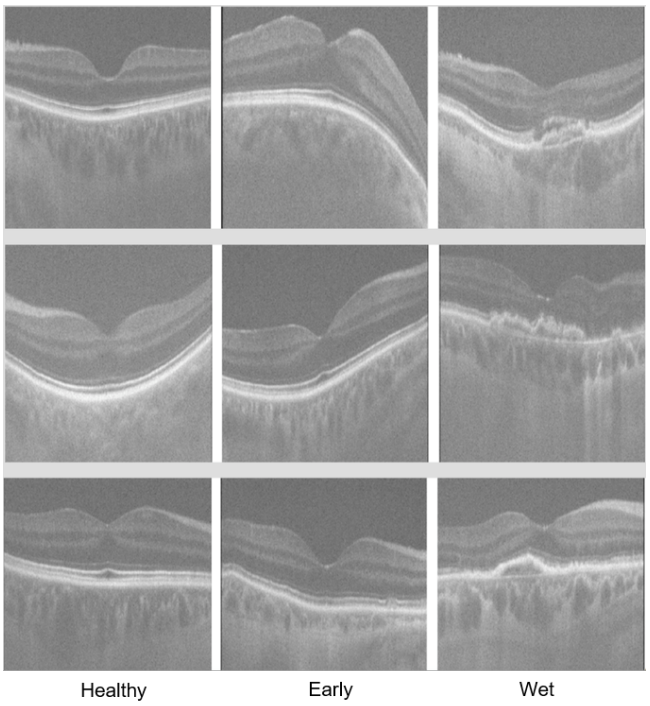

Fig. 1. Examples of choroidal OCT scans for healthy, early AMD and wet AMD.

Priya et al. [6] proposed a machine learning approach for classifying AMD using color retinal photographs, where the hand-crafted features were extracted, such as retinal vessel density, and average retinal vessel thickness. Similarly Koprowski et al. [7] proposed a random forests based method to classify choroidal OCT images into pre-defined clinical conditions by extracting high level features, such as number of detected objects, and average position of the center of gravity,

from low level texture information. However, those high-level features heavily rely on high quality detection results of blood vessels and other anatomical structures, which normally requires extra human resource. In [8] blood bubbles detected in the Retinal Pigment Epithelium (RPE) layer are used to construct a binary classifier which discriminates between AMD and Diabetic Macular Edema (DME). In [9] thickness measurements of the RPE Drusen Complex and the Total Retina are used to train a generalized linear classifier for detection of AMD. In [10] Deng examines using hand crafted Gabor filters for feature extraction on OCT images with machine learning classifiers being used for categorizing the images into stages of AMD. Current advances in Deep Neural Networks shows a superior performance boost in visual recognition tasks. Especially, CNN based methods that enable joint end-to-end learning for both feature representation, and decision making. In this paper, we present a fully automatic method using CNNs to learn the textural features via categorizing the OCT choroid images into broadly defined AMD disease stages of increasing severity.

Our major contributions are twofold as follows: (1) In traditional texture recognition methods hand-crafted feature extractors, such as Gabor filters or wavelets, are used to produce the feature vector before being passed onto machine learning classifiers. When choosing hand-crafted filters, assumptions have to be made about the features expected to occur. By using CNNs to learn a bank of filters for feature extraction we overcome this issue. It allows the development of the optimal set of filters for the data rather than having to make assumptions about what features exist and choosing filters accordingly. (2) We propose an AMD dataset that includes 75 scans obtained using a long-wavelength OCT imaging technique, where both the annotation of the choroidal region, and pathology category are provided. The experimental results show the feasibility of diagnosing the stage of AMD via investigating the textural appearance of choroid images. The rest of the paper is organized as follows: Section II presents our proposed method for filter training, feature extraction and classification, Section III presents and discusses the dataset and experimental results.

II. Proposed Method

In this paper we propose a fully automated machine learning approach for the classification of the stages of AMD using the textural appearance of OCT choroidal images. In traditional texture recognition techniques, hand-picked feature extractors such as Gabor filters or wavelets are used for feature extraction with the resultant feature descriptor being passed onto machine learning classifiers. This study presents a method to fully automate the process of texture recognition using learnable feature extractors rather than hand-picked ones. A CNN is used to automatically train the filters to be used for feature extraction rather than using predefined filters. This allows the development of a set of filters which are best suited to the data rather than having to make assumptions about what features exist and choosing filters accordingly. The set of learnt filters are convolved across the input images with histograms computed from the resultant output to produce the feature descriptors. These are passed onto the machine learning techniques for supervised classification.

A. Filter Training

CNNs combine both feature representation learning and supervised discrimination into a uniform end-to-end training framework, which have become very popular in recent years and produce the top results for many machine vision problems [11], [12], [13]. In this paper, a CNN is introduced to hierarchically learn the textural features in a supervised manner. In convolutional layers, a bank of locally receptive filters convolve across the input image to form visual evidences for prediction layers at the forward pass stage. At the backward pass stage, these filters are automatically optimized via back-propagating the prediction error of the forward pass. Fully connected layers are also included in the network; in these all nodes from one layer are connected to all nodes in the next with weightings updated in the same way. This allows pertinent localized features to be more easily identified. Fig. 2 and Table I show the architecture details of the proposed CNN. Due to the irregular shape of the choroidal region (see Fig. 1), it is difficult to extract large local patches without including other structures. As such, local patches with size of 48×48 are cropped randomly from the annotations with overlap. In order to interpret the low level textural features learnt through the discrimination task, only one convolutional layer is used. The networks that are used for natural image recognition tasks generally have small kernel sizes, such as 3×3, as natural images have much sharper corners and higher contrast compared to medical imaging. In our case, 40 filter kernels with size of 9×9 are used in order to identify the discriminative patterns in low frequency bands.

TABLE I
The parameters of the proposed CNN architecture.

No	Type	Parameter
0	Input	48×48×3 images scaled to [0,1]
1	Conv.	40 9×9 filters with stride 1
2	ReLU	Rectified linear unit
3	F.C.	Fully connected with 128 outputs
4	ReLU	Rectified linear unit
5	F.C.	Fully connected with 128 outputs
6	ReLU	Rectified linear unit
7	F.C.	Fully connected with 3 outputs
8	Softmax	Softmax probability for multi-classes

B. Feature Generalization

CNNs are usually powerful discriminators but we introduce histograms for feature generalization. The CNNs are only trained on patches extracted from the choroid; using a relatively small area presents a challenge for developing an

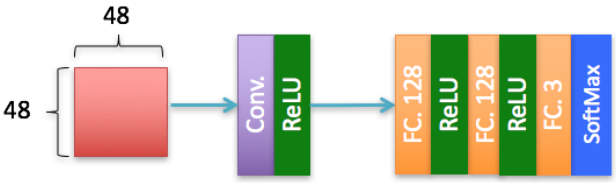

Fig. 2. The network architecture of the proposed CNN.

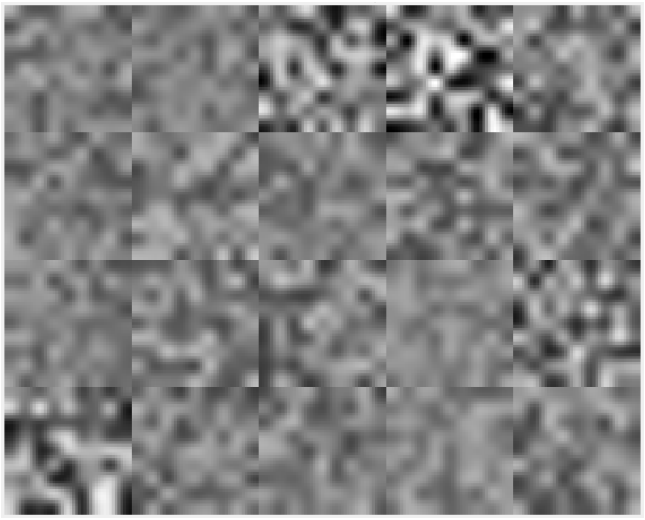

Fig. 3. Examples of self-learnt filter kernels using the CNN.

accurate model. By introducing histograms, we can adapt the method to use the whole choroidal region rather than patches. Convolving the filters across the entire slice and then taking a histogram of the choroidal region allows a description of the whole of the annotated region rather than a subsection of it.

Fig. 3 shows examples of learnt filter kernels from the convolutional layer. The bank of learnt filters are, in turn, convolved across the images of the extracted choroids. Convolving the filters across the images is a form of linear filtering which provides the responses of the kernel patterns. The annotated choroidal regions have variations in size and shape, therefore, outputs of the convolutional response vectors will be of inconsistent length. It is necessary to have the same sized feature descriptors for all slices in order to train the classifiers. To achieve this, the histogram of filter kernel responses of the annotated region is computed to be used as the feature descriptors rather than using local features directly. In addition, to allow images of different sizes and shapes to have the same length of feature output, histogram based descriptors produce a representation of the distribution of kernel responses which greatly improves the generalization ability. Especially, in our case, the kernel filters are self-learnt through performing discrimination tasks, which generally is difficult to link to the pathological changes, and interpret their physical meanings (see Fig. 3). The histogram descriptor provides the quantified statistical measurements of the response distribution of given kernel patterns, which helps to identify the discriminative features. The histogram descriptor is calculated for each of the different filters with the results concatenated to produce one feature vector for each image. Fig. 4 shows examples of histogram descriptors of different AMD classes produced by the top 3 filter kernels. This image demonstrates the differences between each class with the healthy and wet AMD classes being most distinct, whereas the early AMD and wet AMD are most similar.

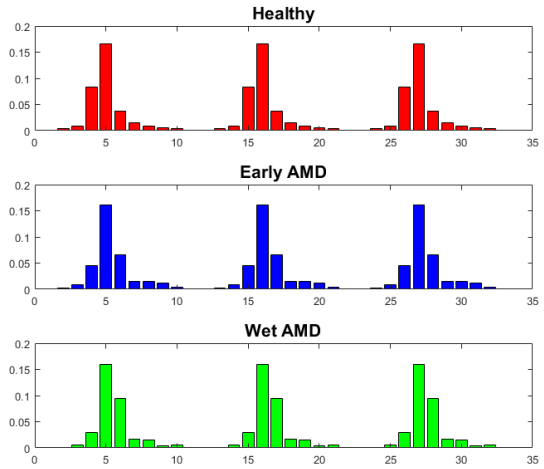

Fig. 4. Examples of histogram descriptors of different AMD classes from top 3 filter kernels.

C. Supervised Classification

To evaluate the discriminative power of proposed feature descriptors, both parametric and non-parametric classifiers are employed to distinguish different AMD stages, such as: neural networks, random forests, and k-nearest neighbor. In order to compare the discriminative power of histogram descriptors with the features of filter kernel responses, the traditional neural network (NN) is used, which has the same architecture of the fully connected layers of the kernel learning CNN in Fig. 2. Random forest (RF) is an ensemble method which combines a number of weak classifiers to create an accurate predictive model [14], [15]. It averages the results of multiple decision trees, each of which consists of a set of recursive binary splits with leaf nodes assigning a probability of the training sample belonging to each class. The variable importance is evaluated during the training process through permutation, which ranks the discriminative power of learnt filter kernels. k-nearest neighbor (k-NN) is a non-parametric method which predicts the testing samples based on the votes of k nearest training samples in the feature space, which is used as a benchmark against other classifiers [16].

III. Experimental Result

The dataset consists of 25 healthy eye scans from the control group, and 50 scans from AMD patients classified into one of two categories: early AMD and wet AMD. Therefore, for

each category the dataset contains 25 eye scans. In order to obtain high quality images, the long-wavelength (1040nm) OCT imaging technique is used to provide sufficient light penetration into the choroid structure. For each eye, a volume of 512×1024×512 pixels is produced. Each eye has its axial eye length (AEL) measured, and the images were scaled accordingly; this was done to control for errors in image scaling [17]. All samples were collected by the same operator and classified by three experienced optometrists into the pathological categories. Classifications were made by examining the shape and appearance of the retina based on an adapted version of an accepted and widely used clinical classification system [18]. We take these classifications to be the ground truth. In preprocessing, for each eye, the outline of the choroidal region was manually labelled on every tenth slice, hence meaning the dataset consisted of over 3,800 labelled slices. Automatic image segmentation has been shown to work in medical examples [19], [20], [21] but we chose manual segmentation to ensure accuracy and consistency. Fig. 5 shows examples of labelled OCT scans of the three categories. From each image the closed curve created by the labels was extracted leaving just the choroidal layer for each slice.

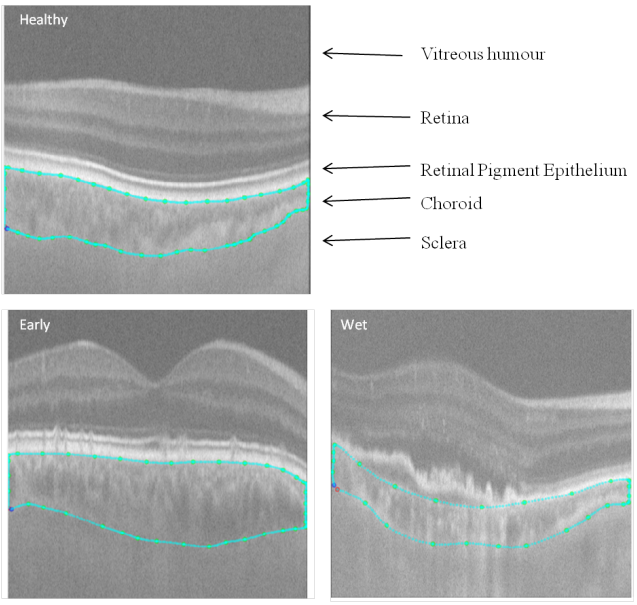

Fig. 5. Examples of labelled OCT scans for each of the three classes with visible signs of pathology within the retina.

The CNN was trained with weight decay of 5×10^{-4}, a batch size of 128 and was trained for 20 epochs with learning rates logarithmically spaced vectors between 10^{-2} and 10^{-5}. Patches of consistent dimension are extracted from the slices to train the network. Ten patches of 48×48 pixels are extracted from each annotated slice, providing over 500 patches per eye. Each patch is given the same classification as the slice to which it belongs. The convolution of each of the learnt filters was calculated for each image and grouped into 11 bins to produce the histogram. The histograms across the different filters are concatenated to produce a single feature vector for each image. As each filter bank consisted of 40 different filters this produces a feature vector for each image consisting of 440 values. Then, each of the classifiers were applied independently. The random forest consisted of 50 random decision trees, the k-NN used a k size of 5 and the neural networks contained two hidden layers with 100 and 25 nodes respectively. For each method of validation the training and testing process was iterated 10 times with the demonstrated results the combination of these.

To perform AMD classification, 10-fold and 2-fold cross validations were used. For 10-fold cross validation the whole dataset was split into ten randomly sampled, evenly sized groups with an equal numbers of slices from each eye. One subset was held for testing whilst the other nine were used for training. This training set was used for learning the filters in the CNN and to train the classifiers. Table III shows the results of 10-fold classification for the three classifiers. Neural networks, random forests and k-NN achieved correct classification accuracies of 83.3%, 66.2% and 52.9% respectively. k-NN is outperformed by the other classifiers which indicates the non-linear decision boundary is necessary to distinguish different AMD stages. Neural networks were decidedly the most accurate. They are useful for learning the hierarchical structure of features. Table II shows the result of using the CNN directly for feature learning and classification, where on average 33.6% is achieved which is significantly lower than the histogram descriptor (83.3% on average in Table III). It strongly suggests that the histogram descriptor improves the discriminative power by a large margin. The primitive filter kernels shown in Fig. 3 is rather noisy and tends to appear random, however, in Fig. 4, the distributions of their responses are discriminative.

The results of 2-fold cross validation of our proposed method are summarized in Table V, with Table IV showing the results through classifying using the CNN only. The CNN only method is unable to distinguish between stages of AMD whereas the respective prediction accuracies for the NNs, RFs and k-NNs were 76.9%, 54.3% and 50.8%. The accuracy was expected to decline across all classifiers due to the relative decrease in the size of the training set. However, a similar pattern occurs in which using the neural network as a classifier produces greater accuracy than the random forest and k-NN. The random forests and, in particular, the neural networks produced a significantly better accuracy than achieved from the k-NN classifier. The results suggest the feasibility of our approach for detecting textural changes in the choroid from which stages of AMD can be classified.

TABLE II
CONFUSION MATRIX OF 10-FOLD CNN CLASSIFICATION (%)

	Healthy	**Early AMD**	**Wet AMD**
Healthy	**80.7**	**80.0**	**80.4**
Early AMD	19.2	20.0	19.4
Wet AMD	0.06	0	0.20

TABLE III
CONFUSION MATRICES OF CLASSIFIERS USING HISTOGRAM FEATURE DESCRIPTORS FOR 10-FOLD CROSS VALIDATION (%)

		Healthy	Early AMD	Wet AMD	Avg.
NN	**Healthy**	**81.2**	10.6	5.2	**83.3**
	Early AMD	11.1	**80.9**	7.0	
	Wet AMD	7.7	8.6	**87.8**	
RFC	**Healthy**	**59.8**	22.0	11.9	**66.2**
	Early AMD	24.3	**63.0**	12.4	
	Wet AMD	15.8	15.0	**75.7**	
k-NN	**Healthy**	**57.7**	37.2	27.8	**52.9**
	Early AMD	28.2	**48.2**	19.3	
	Wet AMD	14.2	14.6	**52.9**	

TABLE IV
CONFUSION MATRIX OF 2-FOLD CNN CLASSIFICATION (%)

	Healthy	Early AMD	Wet AMD
Healthy	**100**	**100**	**100**
Early AMD	0	0	0
Wet AMD	0	0	0

TABLE V
CONFUSION MATRICES OF CLASSIFIERS USING HISTOGRAM FEATURE DESCRIPTORS FOR 2-FOLD CROSS VALIDATION (%)

		Healthy	Early AMD	Wet AMD	Avg.
NN	**Healthy**	**73.9**	15.1	9.0	**76.9**
	Early AMD	12.7	**72.2**	6.3	
	Wet AMD	13.3	12.7	**84.6**	
RFC	**Healthy**	**49.6**	28.1	20.1	**54.5**
	Early AMD	30.3	**52.6**	18.6	
	Wet AMD	20.1	19.4	**61.4**	
k-NN	**Healthy**	**55.2**	40.0	24.9	**50.8**
	Early AMD	30.0	**45.5**	23.3	
	Wet AMD	14.8	14.5	**51.8**	

IV. CONCLUSIONS

In this paper, we proposed an automatic texture feature extraction method, where the low-level primitive filters were self-learnt by CNNs through discrimination tasks, and then high level feature descriptors were constructed by computing the regional histogram of kernel responses. It shows the feasibility on an AMD classification problem, where it outperforms solely using the feature learnt by the CNN, and promising quantitative results were reported.

REFERENCES

[1] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *ELCVIA: electronic letters on computer vision and image analysis*, vol. 7, no. 3, pp. 1–22, 2008.

[2] G. Castellano, L. Bonilha, L. Li, and F. Cendes, "Texture analysis of medical images," *Clinical radiology*, vol. 59, no. 12, pp. 1061–1069, 2004.

[3] M. Mirmehdi, *Handbook of texture analysis*. Imperial College Press, 2008.

[4] Y. Kanagasingam, A. Bhuiyan, M. D. Abramoff, R. T. Smith, L. Goldschmidt, and T. Y. Wong, "Progress on retinal image analysis for age related macular degeneration," *Progress in retinal and eye research*, vol. 38, pp. 20–42, 2014.

[5] C. W. Spraul, G. E. Lang, and H. E. Grossniklaus, "Morphometric analysis of the choroid, bruch's membrane, and retinal pigment epithelium in eyes with age-related macular degeneration." *Investigative ophthalmology & visual science*, vol. 37, no. 13, pp. 2724–2735, 1996.

[6] R. Priya and P. Aruna, "Automated diagnosis of age-related macular degeneration from color retinal fundus images," in *Technology (ICECT), 3rd International Conference on*, E. Computer, Ed. IEEE: 2, 2011, pp. 227–230.

[7] R. Koprowski, S. Teper, Z. Wróbel, and E. Wylegala, "Automatic analysis of selected choroidal diseases in oct images of the eye fundus," *BioMedical Engineering OnLine*, vol. 12, p. 117, 2013.

[8] J. Sugmk, S. Kiattisin, and A. Leelasantitham, "Automated classification between age-related macular degeneration and diabetic macular edema in oct image using image segmentation," in *Biomedical Engineering International Conference (BMEiCON), 2014 7th*. IEEE, 2014, pp. 1–4.

[9] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, C. A. Toth *et al.*, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, 2014.

[10] J. Deng, X. Xie, L. Terry, A. Wood, N. White, T. Margrain, and R. North, "Age-related macular degeneration detection and stage classification using choroidal oct images," in *International Conference on Image Analysis and Recognition*, 2016.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[14] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[15] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[16] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[17] L. Terry, N. Cassels, K. Lu, J. H. Acton, T. H. Margrain, R. V. North, J. Fergusson, N. White, and A. Wood, "Automated retinal layer segmentation using spectral domain optical coherence tomography: Evaluation of inter-session repeatability and agreement between devices," *PloS one*, vol. 11, no. 9, p. e0162001, 2016.

[18] Age-Related Eye Disease Study Research Group and others, "The age-related eye disease study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the age-related eye disease study report number 6," *American journal of ophthalmology*, vol. 132, no. 5, pp. 668–681, 2001.

[19] E. Essa, X. Xie, I. Sazonov, P. Nithiarasu, and D. Smith, "Shape prior model for media-adventitia border segmentation in ivus using graph cut," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2012, pp. 114–123.

[20] J.-L. Jones, X. Xie, and E. Essa, "Combining region-based and imprecise boundary-based cues for interactive medical image segmentation," *International journal for numerical methods in biomedical engineering*, vol. 30, no. 12, pp. 1649–1666, 2014.

[21] E. Essa, X. Xie, and J.-L. Jones, "Minimum s-excess graph for segmenting and tracking multiple borders with hmm," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 28–35.