

Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography

ATEFEH ABDOLMANAFI,^{1,*} LUC DUONG,² NAGIB DAHDAH,³ AND FARIDA CHERIET⁴

¹Dept. of Software and IT Engineering, École de Technologie Supérieure, Montréal, Canada

²Dept. of Software and IT Engineering, École de Technologie Supérieure, Montréal, Canada

³Div. of Pediatric Cardiology and Research Center, Centre Hospitalier Universitaire Sainte-Justine, Montréal, Canada

⁴Dept. of Computer Engineering, École Polytechnique de Montréal, Montréal, Canada

*Atefeh.abdolmanafi.1@ens.etsmtl.ca

Abstract: Kawasaki disease (KD) is an acute childhood disease complicated by coronary artery aneurysms, intima thickening, thrombi, stenosis, lamellar calcifications, and disappearance of the media border. Automatic classification of the coronary artery layers (intima, media, and scar features) is important for analyzing optical coherence tomography (OCT) images recorded in pediatric patients. OCT has been known as an intracoronary imaging modality using near-infrared light which has recently been used to image the inner coronary artery tissues of pediatric patients, providing high spatial resolution (ranging from 10 to 20 μm). This study aims to develop a robust and fully automated tissue classification method by using the convolutional neural networks (CNNs) as feature extractor and comparing the predictions of three state-of-the-art classifiers, CNN, random forest (RF), and support vector machine (SVM). The results show the robustness of CNN as the feature extractor and random forest as the classifier with classification rate up to 96%, especially to characterize the second layer of coronary arteries (media), which is a very thin layer and it is challenging to be recognized and specified from other tissues.

© 2017 Optical Society of America

OCIS codes: (100.0100) Image processing; (100.2960) Image analysis; (100.4996) Pattern recognition, neural networks; (110.0110) Imaging systems; (110.2960) Image analysis; (110.4500) Optical coherence tomography.

References and links

1. A. Dionne, R. Ibrahim, C. Gebhard, M. Bakloul, J.-B. Selly, M. Leye, J. Déry, C. Lapierre, P. Girard, A. Fournier, N. Dahdah, "Coronary wall structural changes in patients with kawasaki disease: new insights from optical coherence tomography (oct)," *J. Am. Heart Assoc.* **4**, e001939 (2015).
2. E. Regar, J. Ligthart, N. Bruining, G. van Soest, "The diagnostic value of intracoronary optical coherence tomography," *Herz.* **36**, 417–429 (2011).
3. B. Preim and D. Bartz, *Visualization in Medicine: Theory, Algorithms, and Applications* (Morgan Kaufmann, 2007).
4. G. Ferrante, P. Presbitero, R. Whitbourn, and P. Barlis, "Current applications of optical coherence tomography for coronary intervention," *Int. J. Cardiol.* **165**, 7–16 (2013).
5. H. G. Bezerra, M. A. Costa, G. Guagliumi, A. M. Rollins, and D. I. Simon, "Intracoronary optical coherence tomography: a comprehensive review: clinical and research applications," *JACC: Cardiovascular Interventions* **2**, 1035–1046 (2009).
6. R. A. Costa, M. Skaf, L. A. Melo, D. Calucci, J. A. Cardillo, J. C. Castro, D. Huang, and M. Wojtkowski, "Retinal assessment using optical coherence tomography," *Prog. Retinal Eye Res.* **25**, 325–353 (2006).
7. A. M. Zysk, F. T. Nguyen, A. L. Oldenburg, D. L. Marks, and S. A. Boppart, "Optical coherence tomography: a review of clinical development from bench to bedside," *J. Biomed. Opt.* **12**, 051403 (2007).
8. J. W. Newburger, M. Takahashi, M. A. Gerber, M. H. Gewitz, L. Y. Tani, J. C. Burns, S. T. Shulman, A. F. Bolger, P. Ferrieri, R. S. Baltimore, W. R. Wilson, L. M. Baddour, M. E. Levison, T. J. Pallasch, D. A. Falace, K. A. Taubert "Diagnosis, treatment, and long-term management of kawasaki disease a statement for health professionals from the committee on rheumatic fever, endocarditis and kawasaki disease, council on cardiovascular disease in the young, american heart association," *Circulation* **110**, 2747–2771 (2004).
9. J. M. Orenstein, S. T. Shulman, L. M. Fox, S. C. Baker, M. Takahashi, T. R. Bhatti, P. A. Russo, G. W. Mierau, J. P. de Chadarévian, E. J. Perlman, C. Trevenen, A. T. Rotta, M. B. Kalelkar, A. H. Rowley "Three linked vasculopathic

- processes characterize kawasaki disease: a light and transmission electron microscopic study," *PloS one* **7**, e38998 (2012).
10. K. C. Harris, A. Manouzi, A. Y. Fung, A. De Souza, H. G. Bezerra, J. E. Potts, and M. C. Hosking, "Feasibility of optical coherence tomography in children with kawasaki disease and pediatric heart transplant recipients," *Circulation: Cardiovascular Imaging* **7**, 671–678 (2014).
 11. S. Celi and S. Berti, "In-vivo segmentation and quantification of coronary lesions by optical coherence tomography images for a lesion type definition and stenosis grading," *Med. Image Anal.* **18**, 1157–1168 (2014).
 12. H. Yabushita, B. E. Bouma, S. L. Houser, H. T. Aretz, I.-K. Jang, K. H. Schlendorf, C. R. Kauffman, M. Shishkov, D.-H. Kang, E. F. Halpern, G. J. Tearney "Characterization of human atherosclerosis by optical coherence tomography," *Circulation* **106**, 1640–1645 (2002).
 13. D. Levitz, L. Thrane, M. Frosz, P. Andersen, C. Andersen, S. Andersson-Engels, J. Valanciunaite, J. Swartling, and P. Hansen, "Determination of optical scattering properties of highly-scattering media in optical coherence tomography images," *Opt. Express* **12**, 249–259 (2004).
 14. C. Xu, J. M. Schmitt, S. G. Carlier, and R. Virmani, "Characterization of atherosclerosis plaques by measuring both backscattering and attenuation coefficients in optical coherence tomography," *J. Biomed. Opt.* **13**, 034003 (2008).
 15. G. Van Soest, T. Goderie, E. Regar, S. Koljenović, G. L. van Leenders, N. Gonzalo, S. van Noorden, T. Okamura, B. E. Bouma, G. J. Tearney, J. W. Oosterhuis, P. W. Serruys, A. F. W. Van der Steen "Atherosclerotic tissue characterization in vivo by optical coherence tomography attenuation imaging," *J. Biomed. Opt.* **15**, 011105 (2010).
 16. G. J. Ughi, T. Adriaenssens, P. Sinnaeve, W. Desmet, and J. D'ÄŽhooge, "Automated tissue characterization of in vivo atherosclerotic plaques by intravascular optical coherence tomography images," *Biomed. Opt. express* **4**, 1014–1030 (2013).
 17. 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.
 18. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 (2014).
 19. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision* (Springer, 2014), pp. 818–833.
 20. D. Eigen, J. Rolfe, R. Fergus, and Y. LeCun, "Understanding deep architectures using a recursive convolutional network," arXiv preprint arXiv:1312.1847 (2013).
 21. S.-C. B. Lo, J.-S. Lin, M. T. Freedman, and S. K. Mun, "Computer-assisted diagnosis of lung nodule detection using artificial convolution neural network," in "Medical Imaging 1993," (International Society for Optics and Photonics, 1993), pp. 859–869.
 22. H.-P. Chan, S.-C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Med. Phys.* **22**, 1555–1567 (1995).
 23. B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images," *IEEE Trans. Medical Imaging* **15**, 598–610 (1996).
 24. H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data," *IEEE Trans. Pattern Analysis and Machine Intelligence* **35**, 1930–1943 (2013).
 25. H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in ct imaging," in "SPIE Medical Imaging," (International Society for Optics and Photonics, 2015), pp. 94131G.
 26. F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. van Ginneken, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box," *Med. Image Anal.* **26**, 195–202 (2015).
 27. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.* (2016).
 28. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation* **9**, 1735–1780 (1997).
 29. H. Azizpour, A. Sharif Razavian, J. Sullivan, A. Maki, and S. Carlsson, "From generic to specific deep representations for visual recognition," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops," (2015), pp. 36–45.
 30. B. van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in "2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)," (IEEE, 2015), pp. 286–289.
 31. Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in "SPIE Medical Imaging," (International Society for Optics and Photonics, 2015), pp. 94140V.
 32. J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Convolutional neural networks for mammography mass lesion classification," in "2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)," (IEEE, 2015), pp. 797–800.
 33. H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE J. Biomed. Health Informatics* **19**, 1627–1636 (2015).

34. G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 652–660.
35. M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, Z. Xu, D. J. Mollura "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* pp. 1–6 (2016).
36. J. Margeta, A. Criminisi, R. Cabrera Lozoya, D. C. Lee, and N. Ayache, "Fine-tuned convolutional neural nets for cardiac mri acquisition plane recognition," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* pp. 1–11 (2015).
37. N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imaging* **35**, 1299–1312 (2016).
38. A. Abdolmanafi, A. S. Prasad, L. Duong, and N. Dahdah, "Classification of coronary artery tissues using optical coherence tomography imaging in kawasaki disease," in "SPIE Medical Imaging," (International Society for Optics and Photonics, 2016), pp. 97862U.
39. H. Azarnoush, S. Vergnole, V. Pazos, C.-É. Bisailon, B. Boulet, and G. Lamouche, "Intravascular optical coherence tomography to characterize tissue deformation during angioplasty: preliminary experiments with artery phantoms," *J. Biomed. Opt.* **17**, 096015 (2012).
40. N. Foin, J. M. Mari, S. Nijjer, S. Sen, R. Petraco, M. Ghione, C. Di Mario, J. E. Davies, and M. J. Girard, "Intracoronary imaging using attenuation-compensated optical coherence tomography allows better visualisation of coronary artery diseases," *Cardiovascular Revascularization Medicine* **14**, 139–143 (2013).
41. D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *J. Physiol.* **148**, 574–591 (1959).
42. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (Springer, 2012), pp. 1097–1105.
43. A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis* (Springer Science & Business Media, 2013).
44. M. Kuhn and K. Johnson, *Applied Predictive Modeling* (Springer, 2013).
45. Z. Wang and X. Xue, "Multi-class support vector machine," in *Support Vector Machines Applications* (Springer, 2014), pp. 23–48.
46. C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Trans. Intelligent Systems and Technology (TIST)* **2**, 27 (2011).

1. Introduction

1.1. Coronary artery structure

A normal arterial wall is composed of three layers. The first layer, the intima, is a transparent, achromatic, and extremely elastic structure comprised of endothelial cells in direct contact with circulating blood. It is characterized by a signal-rich pattern in OCT images, and a normal intima has a reported thickness of $61.7 \pm 17.0 \mu\text{m}$. The next layer of the arterial wall, the media, is homogeneous and composed of smooth muscle cells, lined by the inner and the outer elastica layers which are composed of elastic fibers. The media is specified by a signal-poor pattern in OCT images with a normal thickness of $61.4 \pm 16.7 \mu\text{m}$. Finally, the adventitia is the outermost layer of the artery, surrounding the media and characterized by a signal-rich layer in OCT images [1, 2].

One of the most important abnormalities resulting from Kawasaki disease is intimal hyperplasia, which can be eccentric or concentric and is followed by a mean intima thickening of $390.8 \pm 166.0 \mu\text{m}$. Mean media thickness in the case of aneurysmal artery is $30.2 \pm 56.9 \mu\text{m}$. Changes to the normal structure of the vessel wall as a consequence of severe intimal hyperplasia leads to a partial disappearance of the media. Composition of calcified nodules, white thrombus, fibrosis, and macrophage accumulation are additional abnormalities [1].

1.2. Optical coherence tomography (OCT)

Most of the traditional cardiac imaging modalities, such as X-ray angiography and computed tomography, are effective to visualize the outline of lumen, where the contrast agent flows in the

coronary artery. However, they do not characterize the internal structure of the tissues, such as the vessel wall layers and plaque accumulation [3]. The inner vessel wall geometry and visualization of the morphology of both plaques and arterial wall layers allow for detection and evaluation of the thickness of each layer of the coronary artery, various thrombi, and calcifications, enabling improvements to the process of diagnosis, treatment, and follow-up the patients with coronary complications. Although intravascular ultrasound (IVUS) may be used to assess the inner part of the vessels, its application is restricted due to its suboptimal resolution of 100 - 150 μm [4].

Optical Coherence Tomography (OCT), an intracoronary imaging modality that uses near-infrared light, has many clinical applications because of its high resolution (ranging from 10 to 20 μm). OCT is a promising method for quantifying all information about the inner parts of the vessels by producing a sequence of cross-sectional images of coronary arteries with high resolution about 10 times higher than IVUS [5]. At first, OCT has been widely used in retinal imaging as one of the significant diagnostic technologies of retinal diseases and glaucoma [6]. Then, it has been developed to other medical applications, specifically in cardiology [7]. The introduction of OCT for intravascular imaging was found to be an interesting alternative for intravascular ultrasound (IVUS) imaging.

1.3. *Kawasaki disease (KD)*

Kawasaki disease is an acute childhood inflammatory disease characterized by fever, rash, bilateral nonexudative conjunctivitis, erythema of the lips and oral mucosa, changes in the extremities, and cervical lymphadenopathy. While a high dose of Intravenous Immune Globulin (IVIG) infusion decreases the occurrence of coronary abnormalities, about 15% to 25% of untreated children suffer a risk of experiencing coronary artery aneurysms or ectasia [8], which may be followed by intima hyperplasia, thrombi, stenosis, lamellar calcifications, disappearance of the media border, and significantly the stiffness of the arterial wall [9].

In vivo intravascular visualization of coronary arteries and diagnostic assessment of coronary artery abnormalities are feasible in children and may provide highly valuable progressive information [1, 10]. In this work, we focused on the segments of coronary arteries of patients which are identified as normal by cardiologist to extract all attributes describing each coronary artery layer. This information can be used to distinguish the normal versus abnormal coronary arteries to detect intima hyperplasia, fibrous, and calcification.

1.4. *Related works*

Considering intima hyperplasia, thrombi, stenosis, lamellar calcifications, and disappearance of the media border, tissue classification and specifically classification of the coronary artery layers is very important to evaluate the thickness of the layers. Furthermore, intima thickening and disappearance of the media complicate the classification task. Therefore, manual segmentation of the coronary artery layers is tedious, time-consuming, and particularly error-prone from one observer to another.

Automatic lumen segmentation to assess the stenosis grading and characterization of the plaque types in OCT images of coronary arteries have been performed by Celi et al. [11]. Yabushita et al. have proposed a method of plaque characterization by correlating OCT images with histology [12]. Other studies focused on atherosclerosis plaque characterization using optical properties of tissues [13]. The actual back scattering and attenuation coefficients were measured by Xu et al. [14]. The automatic quantification of optical attenuation coefficients have been proposed by Van Soest et al. [15]. Ughi et al. [16] have proposed their classification method using texture features and attenuation coefficients to characterize atherosclerosis tissues. However, these studies did not address the challenging task of classification of coronary artery layers.

Advances in machine learning and pattern classification have lead to significant advances

for automatic image recognition. For instance, Convolutional Neural Networks (CNNs) have been demonstrated as very powerful techniques in broad range of tasks and in various fields of studies such as computer vision, language processing, image processing, and medical image analysis [17–20]. A wide range of image detection problem using CNNs can be traced back to the 90's, including lung nodule detection, micro-calcification, mass tissues detection on mammography [21–23]. Also, unsupervised deep learning for multiple organ detection using 4D data is performed by Shin et al [24].

The recent applications of CNNs in medical image analysis include pancreas segmentation using CT images of the abdomen [25], classification of pulmonary peri-fissural nodules [26], and brain tumor segmentation [27]. The strength of CNNs originates from its deep structure which permits to extract the features from various abstraction layers [17–20]. Basically, all CNNs consist of a series of layers defined by a specific number of filters or kernels that mainly have the role of feature detector from a set of input images. Sliding the filters on the input images and calculating the convolution of these filter matrices and input image matrix creates a set of convolved features or simply feature maps. The exact meaning of learning a CNN is to train the CNN architecture by the values of these convolutional operations [28].

The transferability of the information preserved on pre-trained CNNs which is one of the most significant characteristic of CNNs, has been demonstrated by the work of Azizpour et al. [29]. The recent studies show the significant applications of transfer learning in medical imaging to extract features from a new dataset using pre-trained CNNs or to use CNNs as the classifier by fine-tuning a pre-trained CNN [30–36]. Van Ginneken et al. have used penultimate layer of OverFeat pre-trained CNN as feature extractor for pulmonary nodules detection in CT acquisition. Then, extracted features have been used in a SVM classifier [30]. Other studies used pre-trained CNNs as a feature extractor on various medical image datasets such as chest radiograph data in the work of Bar et al. [31] and mammography images in the work of Arevalo et al. [32]. Recently, pre-trained CNNs have been also used for classification tasks. For instance, Chen et al. have applied pre-trained CNNs on ImageNet dataset to detect the fetal abdominal standard plane in ultrasound images by keeping the low-level representations extracted from natural images and modify the parameters of the last layers based on the characteristics of ultrasound images [33]. The application of transfer learning using pre-trained CNNs in medical image analysis have been shown in other studies [34–36]. Also, it has been demonstrated in the work of Tajbakhsh et al. that using pre-trained CNNs with adequate fine-tuning works better than training a CNN from scratch in medical image analysis applications [37]. In their experiments, they have considered, different categories of medical images in radiology, cardiology, and gastroenterology using different medical imaging systems such as colonoscopy images for polyp detection, CT pulmonary angiography (CTPA) for PE diagnosis, and Carotid intima-media thickness (CIMT) which is a noninvasive ultrasonography method in cardiology [37].

In this study, our main contribution is automatic classification of coronary artery layers in pediatric patients using the images obtained from OCT system. Our work contributes to identify the features, which perfectly describe both intima and media layers in OCT images using a pre-trained CNN as feature extractor. We also determine if it is better to fine-tune a pre-trained network and use it as the classifier or applying pre-trained CNNs as feature extractor and using the activations of the last fully connected layer to train other classifiers in our application. Finally, we analyze the performance of the classifiers using CNN features and compare the results against tissue classification results of coronary arteries using texture analysis, which is recently done by our group [38].

2. Material and methods

2.1. Pre-processing

We started the pre-processing by automatic recognition and removal of the guide-wire from the images. This step is applied on all the sequences obtained from all the patients. The images are subsequently converted to a planar representation by transformation from Cartesian coordinates to polar coordinates, where the vertical and horizontal axes correspond to radial distance and polar angle, respectively. The approximate region of interest, which consists of the lumen, the arterial wall layers, and the catheter, is extracted by applying active contour. Finally, the catheter and unwanted red blood cells are removed from the images by finding the smallest connected components. All the pre-processing steps are shown in Fig.1 and Fig.2.

2.2. Initial segmentation

Since OCT is a new imaging modality recently introduced in cardiology, to the best of our knowledge, there is no ground-truth available in the literature. Manual segmentation to create the ground-truth is tedious, time-consuming, and imprecise when the size of the tissues is very small. Moreover, the impacts of the disease on these small tissues make them more complicated to be recognized by trained operators. To improve the precision of our classification, we developed an automated approach based on the work of Azarnoush et al. on the intravascular images obtained from phantoms using OCT system [39]. This initial segmentation is using peak information and image quantization to perform tissue segmentation to create ground-truth. The segmentation results were validated by an expert cardiologist.

For each frame of a sequence, we assess the image profile at different penetration depths by scanning the images from 0 to 359 degrees along the radius in planar representation. It is possible to recognize the number of layers present for each frame of a sequence by considering the fact that passing the light from one tissue of the sample to another creates a peak in the A-scan obtained from the backscattered light in the OCT system [40].

Accordingly, a single peak signifies that just the intima is present; media disappearance is obvious in these parts of the images (see Fig.3(a)). Regardless of the level of noise, the presence of two peaks would correspond to two layers (intima and media) and correspondingly three borders (intima, intima-media, and media borders). In realistic laboratory conditions, however, three other possibilities must be considered:

- If the two peaks lie close to one another, then both peaks correspond to one layer, the only layer present in this case is intima (see Fig.3(c)).
- considering the fact that the tissue surrounding the media (adventitia) is characterized by a signal-rich pattern in OCT images. Therefore, passing from media results in a high peak value. A low value of the second peak also describes a case where just a single layer, the intima, is present (see Fig.3(b)).

When there are more than two peaks, they could all belong to one layer (e.g., if peak values are very close to each other), or they could belong to two different layers (see Fig.3(d), Fig.3(e), and Fig.3(f)). To quantify what we mean by low peak values and how close peak values must be to one another to be considered as belonging to the same layer, we apply image quantization to segment each frame to 9 gray levels using 8 threshold values (see Fig.3), the first threshold for each frame determines the minimum difference between two peak values to consider them as belonging to one layer. Also, the peak values less than the first threshold value for each frame is considered as noise. By mapping points that are recognized as intima, intima-media, and media borders using peak values and image quantization to the real images, we can see the precision of the method As illustrated in Fig.3(a), a single peak corresponds to the presence of just intima. In

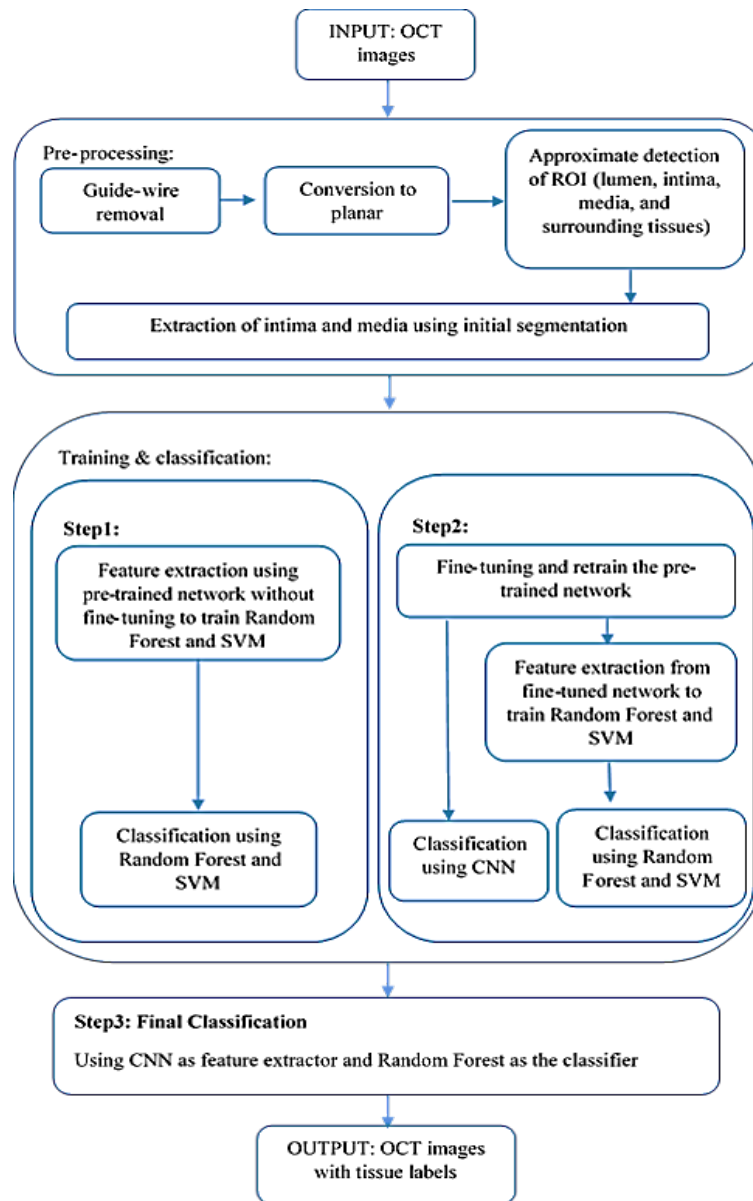


Fig. 1. Flowchart of the tissue classification algorithm. The process of training, feature extraction, and classification using pre-trained CNN just as feature generator is shown in step 1 and fine-tuning the network to use it as the classifier as well as feature extractor to train Random Forest and SVM is demonstrated in step2. Step 3 show our final decision to select the optimal classification algorithm based on measured classification accuracy, sensitivity, and specificity at each step of the work and for each classifier.

Fig.3(b), according to the thresholds obtained by image quantization, the second peak value is too low to be considered as corresponding to any of the tissues. Correspondingly, two borders are recognized in this part of the image (intima and intima-media). In Fig.3(c), there are two peaks, but the two peaks are very close to each other (the difference between the peak values is less than the first threshold level). This is in agreement with the image, which shows no media border.

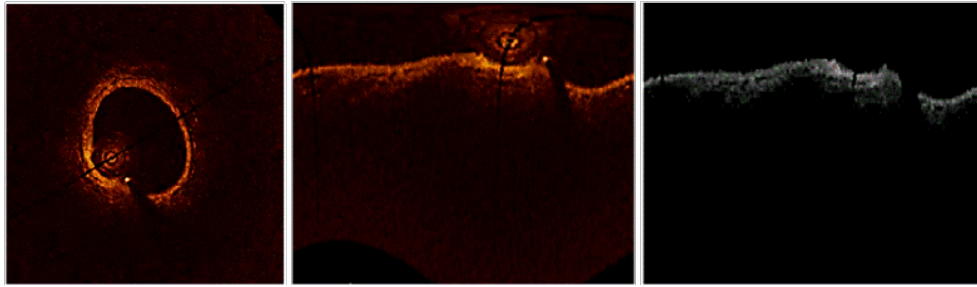


Fig. 2. Pre-processing steps in order from left to right: original image, converting to planar representation, and extracting the region of interest by removing all the background.

In Fig.3(d), there are three peaks, but the last two peak values are too low to be considered as corresponding to any of the tissues. In Fig.3(e), the first two peaks are not close to each other to be considered as belonging to the same layer, and the second peak value is greater than the first threshold level. We have three borders of intima, intima-media, and media. Considering the width of the first peak, intima thickening is obvious in this part. In Fig.3(f), an example of more than two peaks is depicted; the first two peaks belong to intima, the last two peak values are very low to be considered as corresponding to any of the tissues. The results of initial segmentation of one frame for four different patients are shown in Fig.4.

2.3. Feature extraction and classification

Referring to the work of Tajbakhsh et al. [37], it is already determined that using pre-trained CNNs with proper fine-tuning works better in practice than full training of a CNN on scarcely available medical image datasets. In our application, we use pre-trained AlexNet model as feature generator by removing the top output layers (classification layers) and using the activations of the last fully connected layer as training input for Random Forest and Support Vector Machine (SVM). Also, we are interested to fine-tune the AlexNet model by finding the optimal learning rates for the weights at each layer and prepare the network for classification task.

2.3.1. Convolutional neural networks (CNNs)

Generally, every Convolutional Neural Network architecture which is applicable in image processing builds on four main operations: convolution, non linearity (ReLU), pooling or sub-sampling, and classification. In CNN, each convolutional filter creates one feature map when it moves through the whole image with a defined stride. Therefore, the size of the kernel determines the depth of the network. After every convolutional operation, a Rectified Linear Unit (ReLU) has been applied. Since, convolution is a linear operator, it is required to introduce the non-linearity by storing non-negative values in the feature map and replacing the negative values by zero. The pooling or sub-sampling is used for dimensionality reduction by keeping the most important information [28, 41].

In detail, a CNN is trained by minimizing the cost function with respect to the weights at each layer using stochastic gradient descent. The cost function is defined as follows:

$$L = -(1/|X|)\sum_j^{X^j} \ln(p(y^j|X^j)) \quad (1)$$

Where X is the size of the training set and $\ln(p(y^j|X^j))$ denotes the probability of j^{th} image to be classified correctly with the corresponding label y. for each layer of the network, the weights are updated at each iteration i as follow:

$$V_{i+1} = \mu V_i - \gamma_i \alpha \partial L / \partial W \quad (2)$$

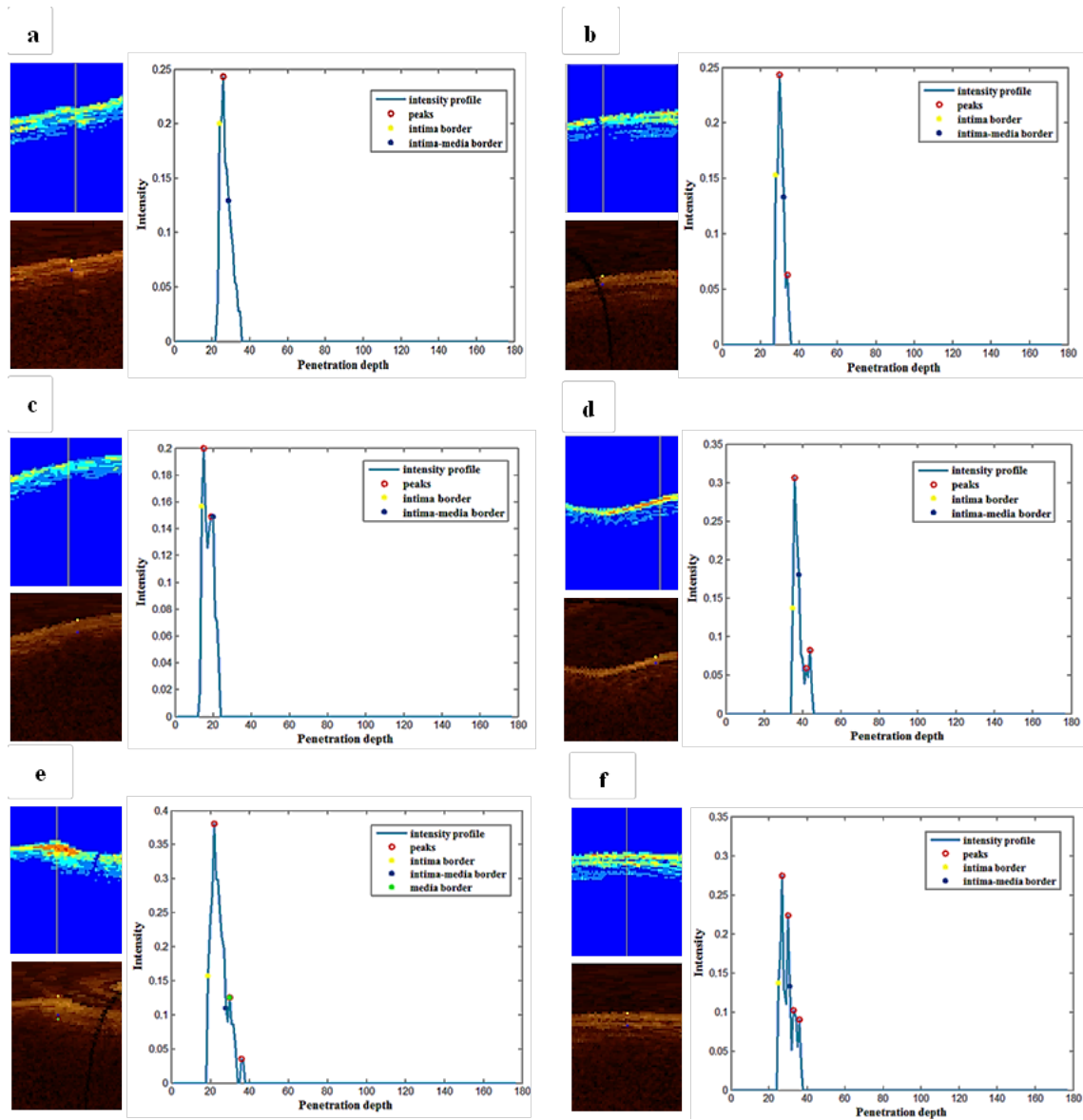


Fig. 3. Peak detection and image quantization. Red circles show the peaks in the image profile; yellow, blue, and green are used to display intima, intima-media, and media borders, respectively.

$$W_{i+1} = W_i + V_{i+1} \quad (3)$$

Where μ is the momentum, α is the learning rate, γ is the scheduling rate which reduces the learning rate at the end of iterations, and W is the weight at each iteration i for each layer [37, 42]. The training process starts with initialized weights for each convolutional layer from a zero-mean Gaussian distribution and standard deviation.

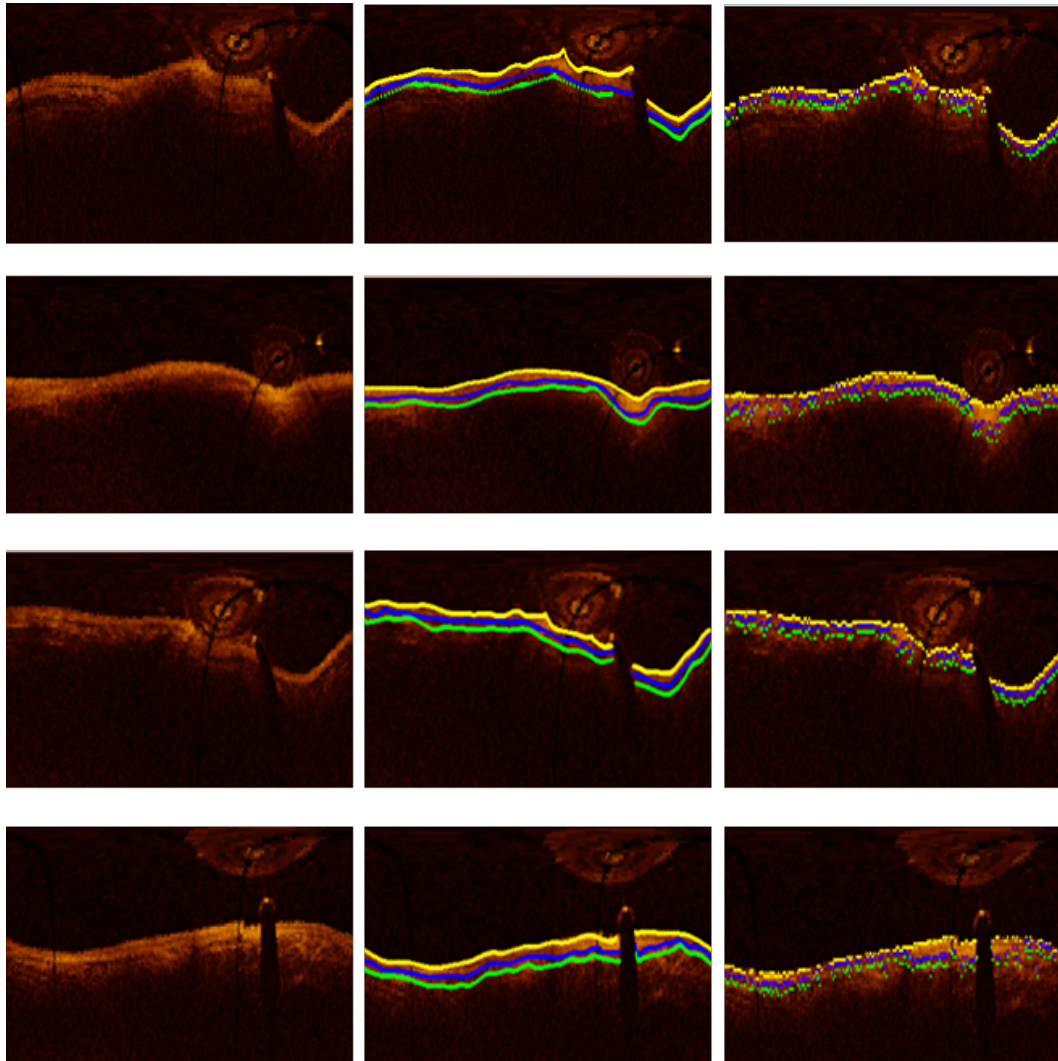


Fig. 4. Initial segmentation of one frame for four different patients. From left to right: Planar representation of the original image, manual segmentation, initial segmentation. Yellow, blue, and green dots show intima, intima-media, and media borders, respectively.

Pre-trained AlexNet model

For both feature extraction and classification using CNN, the pre-trained AlexNet model [42] is used in our experiments. AlexNet is trained on 1.2 million images from the ImageNet dataset, which are labeled with 1000 semantic classes. The network consists of 60 million parameters and 650000 neurons. It composed of eight learned layers, five convolutional layers, and three fully-connected layers with a final 1000-way softmax with GPU implementation of convolutional operation to speed up the training process of such a huge network. The architecture of the AlexNet used in our experiments is shown in Table 1. The model is trained using stochastic gradient descent with the batch size of 128, momentum of 0.9, and weight decay of 0.0005 to reduce the training error of the model [42].

Table 1. AlexNet architecture consists of five convolutional layers, and three fully connected layers.

layer	layer type	input of each layer	output
conv1	convolution	227x227x3, stride [4 4], padding [0 0]	55x55x96
pool1	max pooling	55x55x96, stride [2 2], padding [0 0]	27x27x96
conv2	convolution	27x27x96, stride [1 1], padding [2 2]	27x27x256
pool2	max pooling	27x27x256, stride [2 2], padding [0 0]	13x13x256
conv3	convolution	13x13x256, stride [1 1], padding [1 1]	13x13x384
conv4	convolution	13x13x384, stride [1 1], padding [1 1]	13x13x384
conv5	convolution	13x13x384, stride [1 1], padding [1 1]	13x13x256
pooling5	max pooling	13x13x256, stride [2 2], padding [0 0]	6x6x256
fc6	fully connected	6x6x256	1x4096 feature vector
fc7	fully connected	1x4096 feature vector	1x4096 feature vector
fc8	fully connected	4096D feature vector	1x1000

Transfer learning and fine-tuning

In transfer learning, we use the same architecture of the pre-trained CNN. The last fully connected layer (fc8 in this network), is designed based on the number of classes. Therefore, for the first step, the last 3 layers of the pre-trained network (fc8, prob, and classification layer) are replaced by a set of layers which are designed for multi-class classification to classify intima and media. Accordingly, the number of neurons in the last fully connected layer is set based on the number of classes in our dataset. The next step is fine-tuning; which means to initialize the weights of each layer in our network by transferring the weights from a pre-trained CNN and using the same structure of the pre-trained architecture.

Since low-level features are related to more generic characteristics of the images such as edge orientation detectors, or color blob detectors that should be applicable to many tasks and overfitting concerns of deep fine-tuning considering the small size of our dataset for each patient, we prefer just to fine-tune the weights of the last few layers of the network. But, from another perspective, our dataset is completely different from the original dataset of the pre-trained CNN. Therefore, it is more reliable if we fine-tune the pre-trained network by continuing the back-propagation and change the network slightly. We started fine-tuning from the last fully connected layer (the new fc8 that we replaced based on our classification task). The weights of all other layers remain constant by forcing the learning rates to zero for those layers. The parameters are selected based on grid searching for an extensive interval of values. We keep μ and γ at 0.9 and 0.95 respectively and change the learning rate for the last fully connected layer by setting the learning rate to 0.1. For the next steps, we continue fine-tuning by changing the learning rates of the last two layers, last three layers and so on, to reach the best performance of the network to stop fine-tuning. Table 2 shows the learning rates for each step. We started decreasing the learning rate to 0.01 from fc6 (first fully connected layer in the network). In this way, the weights of the last layers, which are more dataset specific are changing faster than the rest of the network.

2.3.2. Random forest

Generating an ensemble of trees using random vectors, which control the growth of each tree in the ensemble significantly increases the classification accuracy. Random Forest works efficiently

Table 2. Learning rates at each step of fine-tuning the AlexNet model in our experiments. μ and γ are fixed at 0.9 and 0.95 respectively at all the steps of fine-tuning. We started fine-tuning from the last fully connected layer by setting the learning rate to 0.1 for this layer and zero for other layers. We continue changing the network slightly. We started decreasing the learning rates during fine-tuning from fc6. So, the weights of the last layers which are more dataset specific change faster than the rest of the network.

layers	step 1	step 2	step 3	step 4	step 5	step6	step7	step8
fc8	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
fc7	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1
fc6	0.0	0.0	0.01	0.01	0.01	0.01	0.01	0.01
conv5	0.0	0.0	0.0	0.01	0.01	0.01	0.01	0.01
conv4	0.0	0.0	0.0	0.0	0.01	0.01	0.01	0.01
conv3	0.0	0.0	0.0	0.0	0.0	0.01	0.01	0.01
conv2	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.01
conv1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01

on large data sets, carries a very low risk of overfitting, and is a robust classifier for noisy data. The trees are grown based on the CART methodology to maximum size without pruning. Two important factors which affect the Random Forest accuracy are the strength, s , of each tree and the correlation, ρ , between them. Generalization error for Random Forest classifier is proportional to the ratio ρ/s^2 . Hence, the smaller this ratio, the better functioning of Random Forest will be concluded. The correlation between trees is reduced by random selection of subset of features at each node to split on [43, 44]. To improve the performance of the classifier in our experiments, we started from 100 trees and increase the number of trees to 1000. The optimal number of trees is chosen by considering the Out Of Bag (OOB) error rate. By setting the number of trees to 241, the error rate is low, almost close to the minimum error rate, and fewer number of trees reduces the computational burden; so, classifier performance is faster. The number of randomly selected predictors (another tuning parameter in Random Forest) is set to 7. Random Forest training and validation is described in section 2.3.4.

2.3.3. Support vector machine (SVM)

SVM is robust against large number of variables, large data sets, and noisy data, which are the principal challenges of medical images [45]. Non-linear decision boundary is obtained using SVM by means of a kernel function. We employed two-class SVM classifier with gaussian Radial Basis Function (RBF) as the kernel using C-Support Vector Classifier (C-SVC) available in the LIBSVM library [46]. Using grid searching, we found the optimal values of regularization and gaussian kernel bandwidth parameters, C and γ , which are set to 10 and 0.5 respectively. γ is related to the inverse of the RBF kernel extent. Therefore, the smaller γ , the wider kernel will be resulted. The trade-off between the SVM complexity and the number of non-separable samples is controlled by C . The larger C , the higher training accuracy is obtained [45]. SVM training and validation is described in section 2.3.4.

2.3.4. Training and validation

A total of 26 sequences of intracoronary cross-sectional OCT images are obtained from patients with the history of KD using the ILUMIEN OCT system (St. Jude Medical Inc., St. Paul, Minnesota, USA) with the axial and lateral resolutions of 12-15 μm and 20-40 μm respectively.

In the experiments, for each patient, the ROIs (intima and media) are extracted from each frame of the sequence using the initial segmentation and they are labeled as one and two for

intima and media respectively. We have a total of 4800 ROIs adapted to the pre-trained network for all the 26 patients (with an average of 180 ROIs per patient).

The experiments are performed separately for each patient and the classifiers retrained for each sequence of images per patient. For each experiment, the ROIs are divided randomly into three equal parts to create the training, validation, and test sets. To ensure that there is no correlation between these three sets, $2/3$ of the ROIs are randomly selected as the training set and the remaining $1/3$ is considered as the test set. To create the validation set, half of the training set is randomly selected as the validation set and the remaining built the final training set. Then, the accuracy is calculated on the validation set and the training process is stopped when the highest accuracy on the validation set is obtained. By terminating the training process, the features are extracted from the last fully connected layer just before the classification layer (fc7) of the fine-tuned network. Extracted features are used to train Random Forest and SVM. Classification is performed on the test set using CNN, Random Forest, and SVM (see Fig.1). We also apply the pre-trained network as feature extractor without fine-tuning. The extracted features from the layer fc7 of the network are used to train Random Forest and SVM. Then, the classification of the layers is performed on the test set using Random Forest and SVM (see Fig.1).

In the first part of this work, we performed experiments for each patient, separately, as it is reported in the literature because of the variety of textures in coronary artery tissues from patient to patient. We randomly selected different frames of the sequence for each patient to ensure that consecutive frames are not chosen and our system does not become biased. To show the generalization of our method, another experiment is performed with the same configuration as the previous steps but with different selections of training, validation, and test sets. We train our algorithm using the images obtained from various patients. To create the training set, we select the OCT sequences from ten different patients. The remaining eighteen patients are split into two equal sets to create the validation and test sets.

Classification accuracy at each step of the work and for each ROI is calculated by comparing the predicted labels and the the ground-truth for both intima and media. By considering the intima as the positive class, and media as the negative class, the sensitivity is measured as the true positive rate for intima and specificity is calculated as true negative rate for media.

3. Results and discussion

The results of fine-tuning at each step are shown in Table 3 as measured values of accuracy, sensitivity, and specificity. Also, Fig.5 shows the accuracy of tissue classification for both classes, intima and media, for all 26 patients at each step of fine-tuning. The classification rate up to 94% shows that the learning rates listed in Table 2 are selected properly. From the sixth step of fine-tuning (fc8 to third convolutional layer), the results are almost the same compared to steps seven and eight. Figure 5 also shows a very close results for the last three steps of fine-tuning. Therefore, it is reasonable to stop fine-tuning at step six (fc8 to the third convolutional layer).

By fine-tuning the pre-trained network from the classification layer (fc8) to the third convolutional layer, the classification of the coronary artery layers are performed using the deep fine-tuned network, Random Forest, and SVM by applying the features extracted from fc7 (the last fully connected layer just before the classification layer) for each patient separately. The reported results in Table 4 and Fig.6 show the good performance of Random Forest and CNN than SVM for classification of the layers.

For the next step of the work, CNN is used as the feature extractor for each sequence of images per patient. Then, Random Forest and SVM are trained using the activations of the last fully connected layer just before the classification layer. The results demonstrate that using the pre-trained CNN as a feature generator and employing the extracted CNN features to train Random Forest compete against using CNN as the classifier even with deep fine-tuning the network (see Table 5 and Fig.7). To show the generalization of our method, the training process is performed

Table 3. Measured values of accuracy, sensitivity, and specificity to find the optimal depth of fine-tuning based on the performance of the network to classify intima and media at each step of fine-tuning. Values are reported as means \pm standard deviation for 26 sequences.

depth of fine-tuning	accuracy	sensitivity	specificity
fc8	0.947 \pm 0.054	0.927 \pm 0.089	0.968 \pm 0.056
fc7-fc8	0.944 \pm 0.067	0.929 \pm 0.117	0.958 \pm 0.070
fc6-fc8	0.959 \pm 0.043	0.947 \pm 0.060	0.971 \pm 0.043
conv5-fc8	0.945 \pm 0.072	0.915 \pm 0.105	0.974 \pm 0.049
conv4-fc8	0.950 \pm 0.056	0.957 \pm 0.050	0.943 \pm 0.095
conv3-fc8	0.968 \pm 0.040	0.967 \pm 0.053	0.969 \pm 0.050
conv2-fc8	0.951 \pm 0.048	0.946 \pm 0.050	0.955 \pm 0.074
conv1-fc8	0.971 \pm 0.039	0.957 \pm 0.068	0.985 \pm 0.039

Table 4. Measured values of accuracy, sensitivity, and specificity to evaluate the performance of CNN, Random Forest, and SVM to classify intima and media. Values are reported as mean \pm standard deviation for 26 sequences. In this experiment, fine-tuning is performed from fc8 to the third convolutional layer for CNN. Features are extracted from fc7 (the last fully connected layer just before the classification layer) to train Random Forest and SVM.

classifier	accuracy	sensitivity	specificity
CNN	0.97 \pm 0.04	0.97 \pm 0.05	0.97 \pm 0.05
Random Forest	0.96 \pm 0.04	0.97 \pm 0.05	0.96 \pm 0.07
SVM	0.94 \pm 0.07	0.94 \pm 0.07	0.95 \pm 0.11

Table 5. Measured values of accuracy, sensitivity, and specificity. Values are reported as mean \pm standard deviation for 26 sequences. In this experiment, CNN is used as feature extractor for our dataset. Features are extracted from fc7 (the last fully connected layer just before the classification layer) to train Random Forest and SVM. The performances of Random Forest and SVM are compared against the best performance of the CNN as classifier in our experiments when the network is fine-tuned from fc8 to the third convolutional layer.

classifier	accuracy	sensitivity	specificity
CNN	0.97 \pm 0.04	0.97 \pm 0.05	0.97 \pm 0.05
Random Forest	0.96 \pm 0.06	0.95 \pm 0.08	0.95 \pm 0.06
SVM	0.90 \pm 0.10	0.87 \pm 0.13	0.93 \pm 0.11

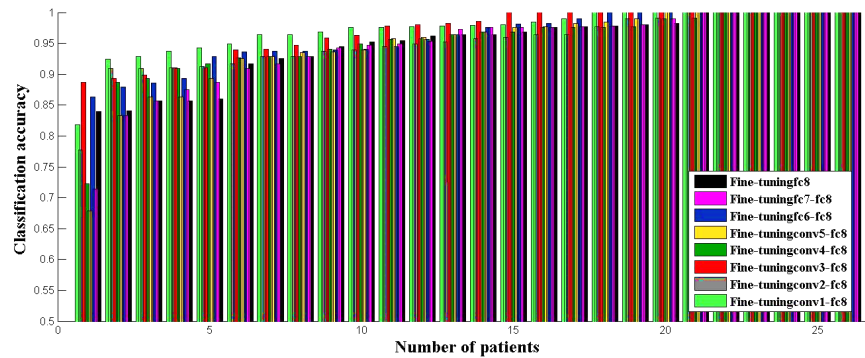


Fig. 5. Tissue classification accuracy for all 26 sequences of intravascular OCT images at each step of fine-tuning the network from fc8 to the first convolutional layer to find the optimal depth of fine-tuning.

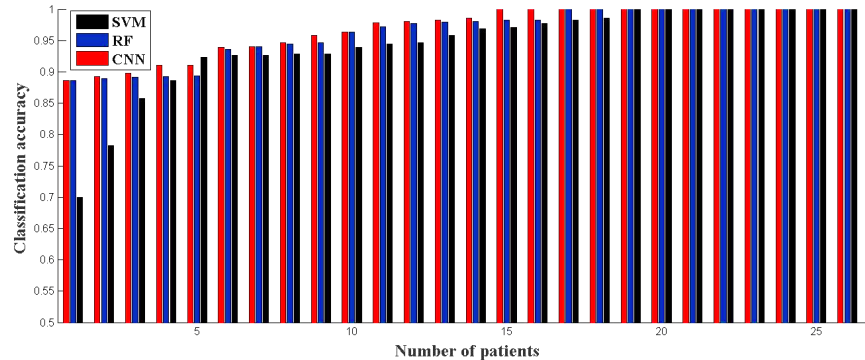


Fig. 6. Performance of CNN, Random Forest, and SVM based on classification accuracy for each patient. Fine-tuning is performed from fc8 to the third convolutional layer for CNN. Features are extracted from fc7 (the last fully connected layer just before the classification layer) of the pre-trained and fine-tuned network to train Random Forest and SVM.

using different OCT images obtained from various patients. Measured accuracy, sensitivity, and specificity for this step of the work are shown in Table 6 and Table 7. Table 6 shows the tissue classification results of fine-tuned CNN, Random Forest, and SVM. Table 7 represents the tissue classification results of Random forest and SVM using the features generated from pre-trained CNN without fine-tuning. The results in Tables 6 and 7 show the capability of our method to be generalized to other future cases.

In this study, to validate our results, the experiments are performed on 26 different sequences of intracoronary OCT images obtained from various patients. Our findings show that although CNNs are robust to be used as classifier and using pre-trained CNNs significantly decreases the computational burden, training time, and convergence issues. But, retraining the network during fine-tuning requires a considerable amount of time. Overfitting concerns in deep fine-tuning the network and finding proper learning rates for each layer are other issues of using CNNs as the classifier. Our results show that it is more efficient to use pre-trained CNNs as feature generators for our application by removing the classification layer and using the activations of the last fully connected layer to train Random Forest. By comparing the results of tissue classification using CNN features against our previous work [38], CNN features are substantially robust to

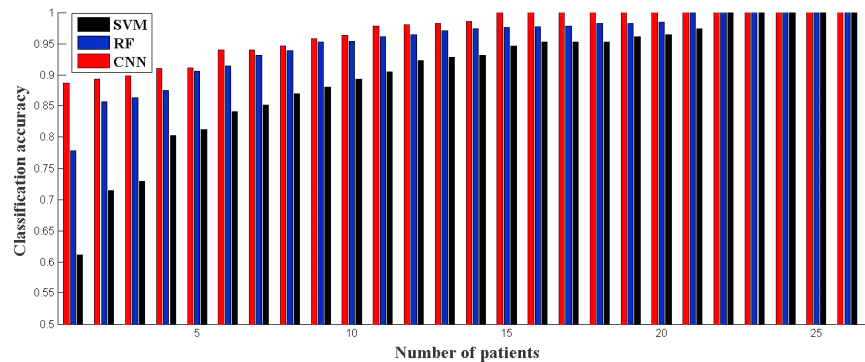


Fig. 7. Performance of Random Forest, and SVM based on classification accuracy for each patient. CNN is used as feature extractor for our dataset. Features are extracted from fc7 (the last fully connected layer just before the classification layer) of the pre-trained network to train Random Forest and SVM. The performance of RF and SVM compared against the best performance of the CNN as the classifier in our experiments when the network is fine-tuned from fc8 to the third convolutional layer.

Table 6. Measured values of accuracy, sensitivity, and specificity to evaluate the performance of CNN, Random Forest, and SVM to classify intima and media for the next step of the work when our algorithm is trained on different patients. In this experiment, fine-tuning is performed from fc8 to the third convolutional layer for CNN. Features are extracted from fc7 (the last fully connected layer just before the classification layer) of the fine-tuned network to train Random Forest and SVM.

classifier	accuracy	sensitivity	specificity
CNN	0.92	0.85	0.99
Random Forest	0.92	0.90	0.94
SVM	0.92	0.88	0.96

describe the characteristics of objects of interest than textural features. Also, comparing the tissue classification accuracy measured in our experiments with the work of Ughi et al. [16] show that using the same classifier (Random Forest) CNN features are more discriminant than optical properties and texture features for tissue classification.

As the classifier, Random Forest works efficiently on large data sets, carries a very low risk of overfitting, and training the model using Random Forest is considerably fast compared to CNN. In our experiments, it takes 5 hours to fine-tune and retrain the pre-trained network. Feature extraction using CNN takes 45 minutes, and training Random Forest takes 4 minutes. Therefore, training the model using Random Forest for all 26 patients was 75 times faster than fine-tuning the pre-trained network exclusive of the time which has been spent to find the optimal values for learning rates and depth of fine-tuning. The measured values of accuracy, sensitivity, and specificity using Random Forest as the classifier also compete against the same values obtained by deep fine-tuning the CNN and using that as the classifier.

4. Conclusion

The main contribution of this study is classification of the coronary artery layers using OCT imaging in pediatric patients. Fully automated tissue classification method is proposed in this work by using a pre-trained CNN as feature extractor by removing the classification layers and using the activations of the last fully connected layer to train Random Forest and SVM. The

Table 7. Measured values of accuracy, sensitivity, and specificity for the next step of the work when our algorithm is trained on different patients. In this experiment, CNN is used as feature extractor for our dataset. Features are extracted from fc7 (the last fully connected layer just before the classification layer) to train Random Forest and SVM. The performances of Random Forest and SVM are compared against the best performance of the CNN as the classifier in our experiments when the network is fine-tuned from fc8 to the third convolutional layer.

classifier	accuracy	sensitivity	specificity
CNN	0.92	0.85	0.99
Random Forest	0.88	0.98	0.78
SVM	0.83	0.94	0.71

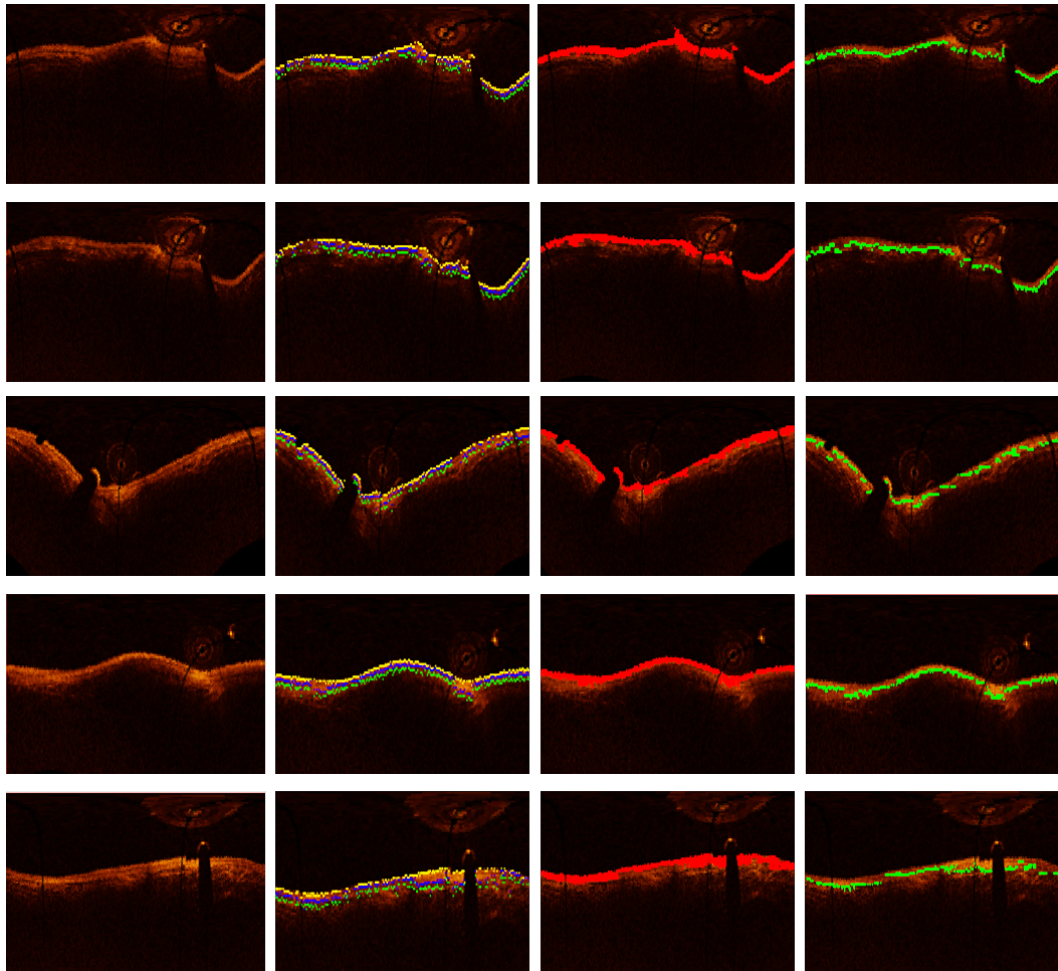


Fig. 8. Classification results for one frame of five different patients. From left to right for each patient: original image converted to planar representation, initial segmentation, intima (red), and media (green)

results confirm the robustness of CNN features to describe the tissue characteristics and Random Forest as the classifier considering the small size of the arteries in children and infants, which is followed by very thin layers in the structure of coronary arteries, and OCT artifacts. Considering

the results obtained from different steps of this work, two major points can be noted: 1. Training the algorithm on a specific patient and classify the layers. 2. Training the algorithm on a set of patients and generalize it for future cases.

This will contribute to estimate intima-media thickening to evaluate the functionality of coronary arteries in patients suffering from Kawasaki disease. The stiffness of the coronary artery tissues (distensibility) as a result of calcium deposits and fibrous scarring while the layers have normal thickness is another significant abnormality caused by KD. Future works will be focused on detecting other abnormalities, evaluating distensibility, dynamic, and geometry of the vessels using stationary OCT imaging.

Funding

This study was supported by the *Fonds de Recherche du Québec-Nature et technologies*