A Machine Learning Based Quantitative Data Analysis for Screening Skin Abnormality Based on Optical Coherence Tomography Angiography (OCTA)

Ji, Yubo; Yang, Shufan; Zhou, Kanheng; Li, Chunhui; Huang, Zhihong

Published in:
IEEE International Ultrasonics Symposium, IUS

DOI:
10.1109/IUS52206.2021.9593642

Publication date:
2021

Document Version
Peer reviewed version

A Machine Learning Based Quantitative Data Analysis for Screening Skin Abnormality Based on Optical Coherence Tomography Angiography (OCTA)

Yubo Ji
Dpt. Biomedical Engineering
University of Dundee
Dundee, UK
y.ji@dundee.ac.uk

Shufan Yang
Dpt. Computing
Edinburgh Napier University
Edinburgh UK
s.yang@napier.ac.uk

Kanheng Zhou
Dpt. Biomedical Engineering
University of Dundee
Dundee, UK
KZhou001@dundee.ac.uk

Chunhui Li
Dpt. Biomedical Engineering
University of Dundee
Dundee, UK
c.li@dundee.ac.uk

Zhihong Huang
Dpt. Biomedical Engineering
University of Dundee
Dundee, UK
z.y.huang@dundee.ac.uk

***Abstract*—Lack of accurate and standard quantitative evaluations limit the progress of applying the OCTA technique into skin clinical trials. More systematic research is required to investigate the possibility of using quantitative OCTA techniques for screening skin diseases. This prospective study included 88 participants (60 normal and 28 abnormal skin samples). In total, 40 OCTA quantitative parameters (3 for epidermis feature, 27 for dermis feature, 10 for vascular feature) were obtained by each OCT and OCTA data volumes. The proposed method relies on linear support vector machines (SVM), while the coefficient of multiple linear regression is also employed to select seven most significant features. Result shows that the proposed method can improve the classification accuracy which can arrive at 93%. Moreover, selected features provide us with direction to determine which biomarker is potential for clinical diagnosis of specific skin abnormalities.**

Keywords—OCTA, Skin disease, Linear Regression, SVM, Machine Learning

I. INTRODUCTION

Visual inspection is a fundamental component of the assessment of a suspicious skin lesion; however, the evidence suggests that melanoma will be missed if visual inspection is used on its own (accuracy = 60%) [1]. Even when dermoscopy are used, experienced dermatologists and surgeons can only reach a diagnostic sensitivity of 68% for melanoma due to its inherent depth limit [2]. In order to confirm diagnosis, a biopsy may be performed. While biopsy is currently the gold standard diagnostic method for suspicious skin lesions, it may be inconvenient for the patient and caused complications, such as bleeding, infection, and scarring [3].

An effective high-resolution cutaneous imaging technique has been introduced which is the optical coherence tomography (OCT) that has a moderate penetration depth for non-invasive inspection of the skin. OCT and OCTA are performed as 'optical biopsy' in clinical practice which allows in situ, safe, immediate investigation of micromorphology, microvascular and pathology without tissue removal. OCT has been studied to evaluate a variety of skin disorders including tumors [4], various inflammatory [5] as well as surgical interventions [6].

OCT therefore has a close relationship within dermatology originating from building structural cross-section skin images to track epidermis thickness. For example, thickening of the epidermis sometimes imply in the irritant dermatitis and psoriasis [7] while disarrangement of the epidermis always happened over the BCC (basal cell carcinoma) nodule [8]. The dermal-epidermal junction (DEJ) region is also a key biomarker inferred from OCT image as the scattering property of the skin can be reflected by the sharpness of DEJ. The investigation of specific structure feature in the dermis has enhanced the possibility for cancerous diagnose (depicted as honeycomb structure) [9] and actinic porokeratosis diagnosis (depicted as hyperreflective/white streaks and dots) [10]. The second-order texture assessment model (GLCM) provides desirable objective information for skin structure imaging which is published by Saba Adabi [11]. On a parallel scale, quantitative approach for vascular imaging gives the complementary information in the detection of changing status taking place inside the tissue. The technique proposed in [12] determines vessel density (VD) measurements of arm skin obtained in 15 healthy Asian subjects. Measures of VD and vessel thickness (diameter) have typically dominated to be a useful parameter in more and more preliminary study, i.e. monitoring process of cutaneous wound healing [13], separating inflammatory skin condition from normal [14]. However, most of these investigations are not provided quantitatively analyses or have very limited sample size. Hence, they are not statistical meaningful. Additionally, as far as we know, none of researchers reported characterizes normal and abnormal skin tissue through comprehensive and systematic quantitative indicators obtained by both of structural and its functional information.

In this study, we proposed an automated classification model to distinguish abnormal skin from normal skin via abundant features from each OCT and OCTA volumes. The method is developed by combining the linear-regression-based feature selection method and SVM classifier. Using the proposed model, it is shown that extracted features (epidermis and dermis structural feature and microvascular features) have potential ability to distinguish the normal and abnormal skin, the accuracy can arrive at 93.3%.

Author Accepted Manuscript: Ji, Yubo et al. "A Machine Learning Based Quantitative Data Analysis for Screening Skin Abnormality Based on Optical Coherence Tomography Angiography (OCTA)". IEEE International Ultrasonics Symposium, IUS. 2021. https://doi.org/10.1109/IUS52206.2021.9593642

II. METHOD

The block diagram of the proposed pipeline for skin classification model in OCT and its corresponding OCTA datasets is illustrated in Fig. 1. The procedure is consisted of data collection, feature acquiring, feature selection based on linear regression, SVM-based classification model.

A. Participants

The research protocol was approved by the School of Science and Engineering Research Ethics Committee (SSEREC) of University of Dundee, which also conformed to the tenets of the Declaration of Helsinki. In the work package, 53 healthy volunteers have been investigated in this study. We elected to scan two sites (face and dorsal forearm) for normal skin detection. Diseased images in this study were taken from 23 confirmed diagnosis of skin lesion. Each participates had one skin lesion; 13 with skin wound and 5 with skin inflammation and 5 for Lipoma. All the skin lesions are in dorsal forearm and face, which are belonged to thin skin. Their age, gender, Fitz-Patrick skin type, skin site, has been recorded confidentially and their informed consent was acquired.

B. Handheld Swept-Source Optical Coherence Tomography System (SS-OCT)

A homebuilt handheld SS-OCT system was used in this study for data acquisition. The system utilized a 200 kHz swept source laser (SL132120, Thorlabs Inc., Newton, NJ, USA) with a center wavelength of 1300 nm and a bandwidth of 100 nm, which provided an axial resolution of ~7 μm and a penetration depth up to 1.5 mm in skin tissue. The handheld probe was designed for comfortable handling while imaging, which was composed by a 2D galvo-scanner, an objective lens (LSM03, Thorlabs Inc., Newton, NJ, USA), a built-in charge coupled device (CCD) camera (UC20MPE, Spinel USA LLC, Newport Beach, CA, USA) and a 6-inch liquid-crystal display (LCD). The physical lateral resolution of the system was approximately 20 μm in air.

C. Imaging Protocol and Data Processing Method

Imaging protocol followed in this research was same for every single voluntary participant. Mineral oil (extremely allergy rate) was applied on the area of imaging evenly. Moreover, in the purpose of flattening the focal plane, a 4mm thickness of glass mirror (disinfect using alcohol in advance) is set on the head of the probe and attached to the skin gently. Eight hundred A-lines formed one B-frame and 3200 B-frames (800 locations×4 repeats/location) formed one complete OCTA volume, which covered a field of view of 10mm×10mm. The total scanning times for each volunteer will was around 3 minutes including preparing. Refractive index n of 1.35 was applied in this study.

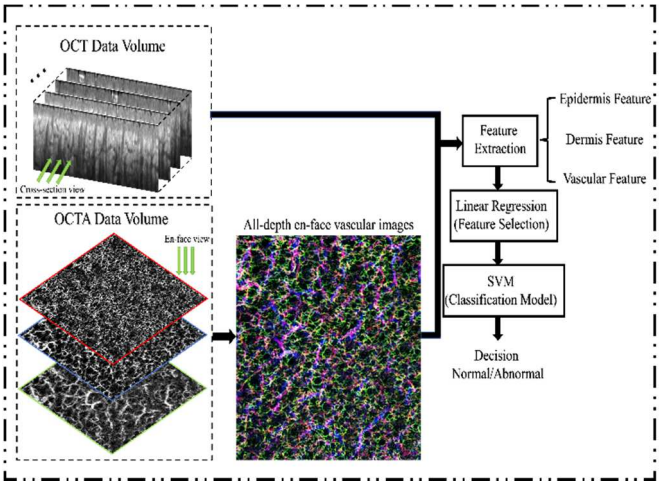

Fig. 1. Pipeline for SVM-based classification

The structure image at each location was generated by averaging 4 repeated by averaging repeated cross-section structure images. Microvascular information can be sifted out through an appropriate eigen decomposition (ED) filter by eliminating the lowest two principal components [15]. Sorted maximum intensity projection (sMIP) algorithm is used for all-depth volumes. The representative normal and abnormal OCT and OCTA result are shown in Fig.2.

D. Feature Acquiring

The features used in our work are tabulated in Table.1. The 40 features can be divided into three categories: epidermis structural features, dermis structural features and vascular features.

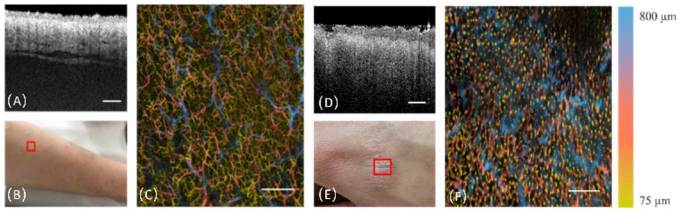

Fig. 2. Representative normal and abnormal OCT and OCTA Result (A) cross-sectional OCT structure images for normal skin. (B) Photograph of the healthy skin located in arm, where the region of imaging is indicated by a red rectangular frame. (C)All-depth microvascular image of healthy skin. (D) cross-sectional OCT images for a patient with psoriasis. (E) Photographs of the psoriasis located in arm, where the red rectangular frame indicates the region of imaging. (F) All-depth microvascular image of the psoriasis. Scale bar is represents 200 μm

TABLE 1. List of Features Used in Skin Classification Study

Feature	# Dimension
Epidermis Structural Feature	
Epidermis thickness	1
DEJ layer variance	1
Range of Max position and Min position of DEJ	1
Dermis Structural Feature	
Energy, Correlation, Inertia, Entropy (GLCM)	4
Inverse difference moment (GLCM)	1
4 directions of above features (GLCM)	20
Inter-layer contrast (epidermis-papillary dermis layer)	1
Inter-layer contrast (papillary dermis-Reticular dermis layer)	1
Vascular Feature (all depth)	
Mean Vascular density	1
Mean Fractal dimension	1
Mean Vessel length density	1
Mean Vessel diameter index	1
Mean Branch Density	1
Variance of above feature	5

E. Feature Selection

To reduce the computational burden of classifier and avoid overfitting, feature selection or feature pruning is a very crucial step in the pipeline. In this paper, coefficients of linear regression models are applied for selecting and interpreting features. The most important features shall have the highest coefficients in the model while the features which are not correlated with the output variables shall have coefficient values close to zero which is based on the idea that when all features are on the same scale. If the value of OCT/OCTA features are $f1, f2, \dots fn$ and the predicted output are $o1, o2, \dots on$ then the coefficient is given by the formula:

$$C = \frac{S_{OF}}{\sqrt{S_O S_F}} \tag{1}$$

where $S_{OF} = \frac{\sum_i (o_i - \bar{o})(f_i - \bar{f})}{n-1}$, $S_F = \frac{\sum_i (f_i - \bar{f})^2}{n-1}$, $S_O = \frac{\sum_i (o_i - \bar{o})^2}{n-1}$

Although dimensionality reduction can cause some loss of information, the transformed data can be more compact and reasonably decrease the time for data processing and feature capture. In this research, the best seven features are selected according to the proposed feature selection algorithm.

F. Classification Model

A method that is used to supervise the machine learning technique and widely used in the classification problems and pattern recognition is called the Support Vector Machine (SVM). The algorithms of the SVM performs a classification by creating multi-dimensional hyperplane that optimally discriminate between the two classes and by maximizing the margins between two data clusters. By using a special nonlinear function called kernel it achieves a high discriminative power and transform the input space into a multi-dimensional space. In order to generate the SVM models, LibSVM is used which is freely available in SVM software library [16]. Among them, the linear SVM is applied in this study.

Sensitivity, specificity and accuracy were calculated based on the following formulas when the cutoff value was set to 0.2 in the SVM model. The IOU of the ROC was also provided.

$$Sensitivity = \frac{TP}{TP + FP} \tag{2}$$

$$Specificity = \frac{TN}{TN + FN} \tag{3}$$

$$Accuracy = \frac{CPD}{TTD} \tag{4}$$

Where TP, FP, TN, FN, CPD, TTD represents number of true positive, false positive, true negative, false negative, corrected predict data, total testing data, respectively.

III. Result

According to the correlation coefficient value of linear regression, seven feature was selected for reduced dimensional feature subset. which are 1) epidermis thickness, 2) variance of DEJ layer, 3) GLCM entropy 90°, 4) vascular density, 5) fractal dimension, 6) variance of vascular density, 7) variance of fractal dimension. Among them, 1)-2) are epidermis features, 3) is dermis feature while 4)-7) are vascular features.

These results showed below were obtained when the classifier was used with the seven selected features. The ROC curve for different combinations of selected features is summarized in Fig.3. The ROC curve of using all the selected features (AUC=0.987) is significantly better than other combination of features. Similarly, in accordance with the quantitative analysis for accuracy prediction (Table 2), when all the selected features are employed, experimental results reveal the benefits of combining epidermis, dermis and vascular features. On the other hand, using dermis features has the lowest accuracy compared to others (accuracy=61.5%). Applying epidermis and vascular features for the classifier also provided satisfying results, 91.1% accuracy for abnormal skin classification.

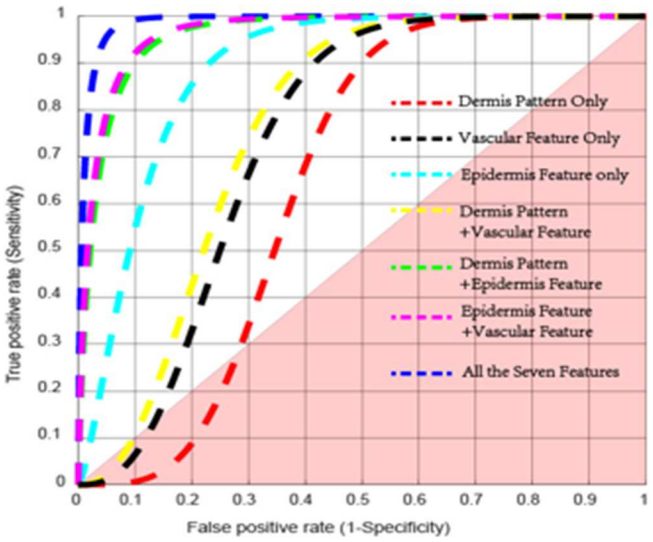

Fig. 3. ROC curve result for different combination of feature

TABLE 2. Classification Accuracy Using DifferentCombination of Features

	AUC	Sensitivity	Specificity	Accuracy
Dermis Pattern Only	0.64924	0.5142	0.65	0.6148
Vascular Feature only	0.7391	0.6	0.72	0.6889
Epidermis Feature Only	0.88708	0.7428	0.79	0.7777
Dermis Pattern+Vascular Feature	0.7648	0.6285	0.74	0.7111
Dermis pattern + epidermis Feature	0.95995	0.8285	0.93	0.9037
Epidermis Feature + Vascular Feature	0.9657	0.8857	0.92	0.9111
All the seven features	**0.98725**	**0.9429**	**0.93**	**0.9333**

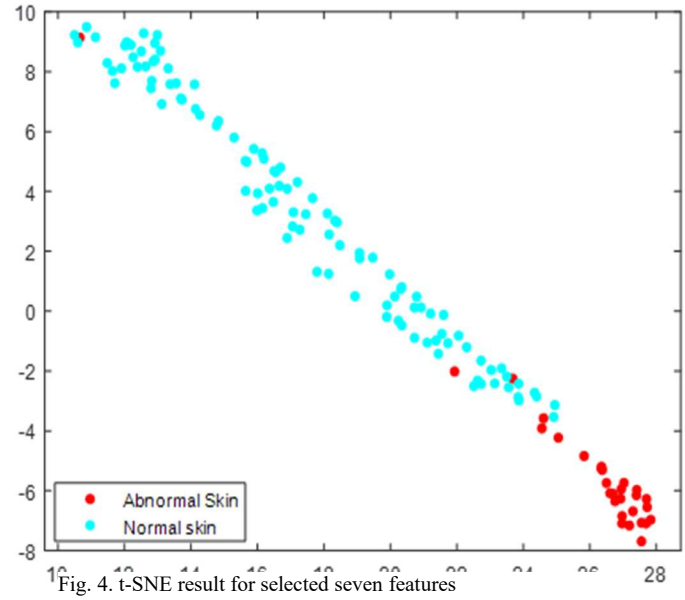

Fig. 4. t-SNE result for selected seven features

The result of t-SNE in Fig.4 showed that the seven features extracted from OCT and OCTA can separate the normal and abnormal skin completely, with the exception of a few outliers. t-SNE produces a feature map in which there is clear boundary between normal and abnormal skin.

IV. Conclusion

This research is aiming to build a database based on versatile quantitative pathology and non-pathology information based on OCT and OCTA. Not only is the epidermis and dermis information employed, but also vascular information is extracted and utilised as features in a classifier. Also, we elected a pipeline for automated abnormal skin classification using OCT and OCTA images. The method is developed using coefficient of linear regression for feature selection and SVM for classification method. This is a supervised learning algorithm used for separating feature space of normal and abnormal skin by analysing data and constructing a set of hyperplanes. According to the quantitative analysis, we demonstrated that the pipeline could assist in diagnosis of abnormalities of skin based on microstructure and microvascular information. Additionally, the proposed workflow can be generalized for detection of other skin structural or functional abnormalities. Hence, the outcome of this study suggests that it can be extended by an interactive machine learning kernel interface which shall be added to the OCTA devices.

References

[1] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *J The lancet oncology,* vol. 3, no. 3, pp. 159-165, 2002.

[2] M. Binder *et al.*, "Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *%J Archives of dermatology,* vol. 131, no. 3, pp. 286-291, 1995.

[3] U. Nischal, K. Nischal, and U. Khopkar, "Techniques of skin biopsy and practical considerations," *J Journal of cutaneous aesthetic surgery,* vol. 1, no. 2, p. 107, 2008.

[4] M. Mogensen, L. Thrane, T. M. Jørgensen, P. E. Andersen, and G. B. J. J. o. b. Jemec, "OCT imaging of skin cancer and other dermatological diseases," vol. 2, no. 6-7, pp. 442-451, 2009.

[5] A. J. Deegan *et al.*, "Optical coherence tomography angiography of normal skin and inflammatory dermatologic conditions," vol. 50, no. 3, pp. 183-193, 2018.

[6] M. E. Brezinski, G. J. Tearney, S. A. Boppart, E. A. Swanson, J. F. Southern, and J. G. Fujimoto, "Optical biopsy with optical coherence tomography: feasibility for surgical diagnostics," *J Journal of Surgical Research,* vol. 71, no. 1, pp. 32-40, 1997.

[7] J. Welzel, M. Bruhns, and H. H. Wolff, "Optical coherence tomography in contact dermatitis and psoriasis," *J Archives of dermatological research,* vol. 295, no. 2, pp. 50-55, 2003.

[8] T. Gambichler *et al.*, "In vivo optical coherence tomography of basal cell carcinoma," *J Journal of dermatological science,* vol. 45, no. 3, pp. 167-173, 2007.

[9] A. A. Hussain, L. Themstrup, and G. B. E. Jemec, "Optical coherence tomography in the diagnosis of basal cell carcinoma," *J Archives of dermatological research,* vol. 307, no. 1, pp. 1-10, 2015.

[10] K. Friis, L. Themstrup, and G. Jemec, "Optical coherence tomography in the diagnosis of actinic keratosis—A systematic review," *J Photodiagnosis photodynamic therapy,* vol. 18, pp. 98-104, 2017.

[11] S. Adabi *et al.*, "Universal in vivo textural model for human skin based on optical coherence tomograms," vol. 7, no. 1, pp. 1-11, 2017.

[12] S. Men *et al.*, "Repeatability of vessel density measurement in human skin by OCT-based microangiography," vol. 23, no. 4, pp. 607-612, 2017.

[13] A. J. Deegan *et al.*, "Optical coherence tomography angiography monitors human cutaneous wound healing over time," *J Quantitative imaging in medicine surgery,* vol. 8, no. 2, p. 135, 2018.

[14] A. J. Deegan *et al.*, "Optical coherence tomography angiography of normal skin and inflammatory dermatologic conditions," *J Lasers in surgery medicine* vol. 50, no. 3, pp. 183-193, 2018.

[15] S. Yousefi, Z. Zhi, and R. K. Wang, "Eigendecomposition-based clutter filtering technique for optical microangiography," *J IEEE transactions on biomedical engineering,* vol. 58, no. 8, pp. 2316-2323, 2011.

[16] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *J ACM transactions on intelligent systems technology* vol. 2, no. 3, pp. 1-27, 2011.