



Development and validation of the interpretability analysis system based on deep learning model for smart image follow-up of nail pigmentation

Yanqing Chen^{1,2#^}, Haofan Liu^{3#^}, Zhaoying Liu^{1^}, Yang Xie^{4^}, Yingxue Yao^{3^}, Xiaofen Xing^{3^}, Han Ma^{1,2^}

¹Department of Dermatology, Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China; ²Guangdong Provincial Key Laboratory of Biomedical Imaging, Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China; ³School of Electronic and Engineering, South China University of Technology, Guangzhou, China; ⁴Department of Dermatology, Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Contributions: (I) Conception and design: Y Chen, Z Liu, X Xing, H Ma; (II) Administrative support: X Xing, H Ma; (III) Provision of study materials or patients: Y Chen, Z Liu, Y Xie, H Ma; (IV) Collection and assembly of data: Y Chen, H Liu, Z Liu, Y Xie; (V) Data analysis and interpretation: Y Chen, H Liu, Y Yao, X Xing, H Ma; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Han Ma. Fifth Affiliated Hospital, Sun Yat-sen University, 52 Meihua East Road, Xiangzhou District, Zhuhai 519000, China. Email: drmahan@sina.com; Xiaofen Xing. South China University of Technology, 381 Wushan Road, Tianhe District, Guangzhou 510640, China. Email: xfxing@scut.edu.cn.

Background: Nail pigmentation can be a clinical manifestation of malignant melanoma and a variety of benign diseases. Nail matrix biopsy for pathologic examination, the gold standard for diagnosis of subungual melanoma, is a painful procedure and may result in nail damage. Therefore, there is a great need for non-invasive methods and long-term follow-up for nail pigmentation. The objective of this study is to establish an intelligent precursor system to provide convenient monitoring for nail pigmentation, early warning subungual melanoma, and reduce nail biopsies to the minimum necessary.

Methods: Dermoscopic images of nail lesions were obtained from outpatients between 2019 and 2020. The images were divided into the training set and the test set using k-fold cross validation at a ratio of 10:1. The deep learning model is developed based on the Pytorch framework. The model structure is optimized using the image segmentation model U-Net. An image segmentation module analyzed the contours of the whole nail plate and pigmented area according to the boundary features of the input images and a rule calculation module used the output information of the segmentation model to automatically analyze specific indicators referring to the “ABCDEF” rule. The model’s results were compared with those of clinical experts.

Results: From 550 dermoscopic images of nail lesions obtained, 500 were selected randomly as the training set, and the remaining 50 as the test set. Our image segmentation module realized automatic segmentation of the pigmented area and the whole nail plate with dice coefficient to be 0.8711 and 0.9652, respectively. Five qualitative indicators were selected in the interpretability analysis system and the models showed a certain degree of consistency with the evaluation by clinical experts, i.e., R^2 for area ratio *vs.* breadth score was 0.8179 ($P < 0.001$), for mean pixel value *vs.* pigment score was 0.7149 ($P < 0.001$), for evenness *vs.* pigment score was 0.5247 ($P < 0.001$).

Conclusions: The proposed system made accurate segmentation of the nail plate and pigmented area and achieved medically interpretable index analysis. It is potentially a primer of an intelligent follow-up system that will enable convenient and spatially unaffected management and monitoring of nail pigmentation. It may provide clinicians with understandable auxiliary information for diagnosis.

[^] ORCID: Yanqing Chen, 0000-0003-3092-1823; Haofan Liu, 0000-0002-6848-1475; Zhaoying Liu, 0000-0001-9189-1184; Yang Xie, 0000-0001-5729-6082; Yingxue Yao, 0000-0001-5250-281X; Xiaofen Xing, 0000-0002-5939-5815; Han Ma, 0000-0003-0101-495X.

Keywords: Deep learning; dermoscopy; interpretability; melanoma; nail pigmentation

Submitted Jan 19, 2022. Accepted for publication May 10, 2022.

doi: 10.21037/atm-22-1714

View this article at: <https://dx.doi.org/10.21037/atm-22-1714>

Introduction

Brown or black pigmentation within the fingernails or the toenails is a common presenting problem in dermatology. Common causes of nail pigmentation are benign disorders such as melanocytic nevus, subungual hemorrhage, onychomycosis, exogenous staining, and systemic diseases. However, it may also be a sign of a life-threatening malignant disease: nail apparatus melanoma.

Nail apparatus melanoma accounts for about 2% of all melanomas, and up to 20% of melanomas in dark-skinned and Asian populations (1,2). Early diagnosis and treatment greatly improve the prognosis, but unfortunately, melanoma of the nail unit is usually diagnosed at a later stage, which leads to poor prognosis. The gold standard for the diagnosis of nail apparatus melanoma is histopathological examination of a biopsy specimen of the nail matrix, which is usually a painful procedure and may result in nail dystrophy, permanent deformity, and paresthesia (3). Therefore, it is often rejected by patients.

Non-invasive methods to assist in the diagnosis and follow-up of nail apparatus melanoma can balance missed diagnosis and excessive nail biopsies. A clinical analysis system, named the “ABCDEF” rule, was developed to evaluate the risk of suspected cases of nail apparatus melanoma (4,5). “A” stands for age, African Americans, Asians, and native Americans; “B” stands for brown to black band with breadth of 3 mm or more and variegated borders; “C” stands for change in the nail band; “D” stands for the digit involved; “E” stands for extension of the pigment onto the nailfold; and “F” stands for family or personal history (4,5). To further establish the diagnosis, dermoscopy, a non-invasive procedure obtaining images by a dermoscope, has been used to evaluate the characteristics of nail lesions, and helps to dynamically observe changes in the pigmentation of the nail (4).

Remote intelligent system is a promising tool for improving patient compliance with disease surveillance. Artificial intelligence in medical imaging has made remarkable progress in recent years and has the potential to greatly affect clinical disease management (6,7).

However, deep learning models usually work similar to a black box, resulting in a lack of interpretability. Most studies of melanoma based on deep learning algorithms are classification, segmentation, and detection (8,9). Compared with traditional image algorithms, such end-to-end convolutional neural network (CNN) models usually work without feature engineering with the support of high-quality data (8,9). The poor interpretability of CNN models in clinical practice is risky and makes it difficult to gain the trust of clinicians and patients (10).

In order to improve this situation, we developed and validated an indicator analysis system based on a deep learning image segmentation model of pigmentations in the nail. The proposed analysis system first identifies pigmented lesions on nails, and then extract relevant indicators with reference to the established “ABCDEF” rule to provide interpretable data. The aim of the systems is to assist in dynamic monitoring of nail pigmentation, early warning of subungual melanoma, and prompting the best timing from non-invasive follow-up to necessary invasive examination in an interpretable manner. We present the following article in accordance with the TRIPOD reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-1714/rc>).

Methods

Image data set

We included patients who presented to the outpatient for abnormal pigmentation within the nail at the Department of Dermatology in Fifth Affiliated Hospital of Sun Yat-sen University in China between March 1, 2019 and October 31, 2020. The diagnostic categories included: nail matrix nevus, malignant melanoma, subungual hemorrhage, onychomycosis, subungual glomus tumor, etc. Dermoscopic images of the nail lesions were taken with the same dermascope. The average image size was $\approx 1,500 \times 2,400$ dpi. The dermoscopic images were randomly divided into the training set and the test set for the proposed deep learning

segmentation model using k-fold cross validation at a ratio of 10:1.

To develop the deep learning model, the dermoscopic images were input as the source samples. Each image was annotated by the image labeling tool Labelme to mark the contours of the nail area and pigmented spots or lines respectively. A set of binary gray images was generated as the training label for the deep learning segmentation model. In order to reduce the over-fitting problem during the training process of the segmentation model and enhance the adaptability of the model in clinical scenarios, we performed data enhancement operations such as rotation, translation, cropping and noise addition on the training set. The test set was used to evaluate the effect of the segmentation model.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Fifth Affiliated Hospital of Sun Yat-sen University (No. [2021](K225-1)) and all patients gave informed consent.

Automatic analysis system

Our nail pigmentation automatic segmentation system consisted of two modules: image segmentation and rule calculation module. A deep learning image segmentation model was used in the automatic image segmentation module because it can accurately recognize the contours of the whole nail plate and pigmented line area, respectively, according to the boundary features of the input images. The rule calculation module is an intelligent analysis module connected with the segmentation module. It makes use of the output information of the segmentation model to automatically analyze specific indicators of the dermoscopic nail images. The system was developed using the Python open-source deep learning framework PyTorch. Two different U-Net models were used as the image segmentation model. For each dermoscopic nail image, the two U-Net models output the contours of the whole nail and dark spots or black lines areas respectively.

Referring to “ABCDEF” rule, breadth of nail band, border, digit involvement, and pigment extension of the nail band were features which can be reflected through images. Four individual experienced dermatologists scored the breath, pigment, border, extension (on the nailfold) and the general risk (to be malignant) for the nail images. Each item has a 10-step score, ranging from 0 to 9, whereby the score 9 reflects the most consistent with malignant.

Using the results of the image segmentation model, the

rule analysis module analyzed five qualitative indicators which reflect the above features. These indicators include area ratio, mean pixel value, evenness, irregularity, and skin invasion. The area ratio indicator is the ratio of the number of pixels in dark spots or black lines and the whole nail area. Mean pixel value indicator was obtained by calculating the average pixel value ratio of the dark spots or black lines relative to the whole nail area. Evenness indicator was represented by the pixel value standard deviation. The irregularity indicator, which means the irregularity of the shape of dark spots and black lines, was evaluated it by computing the Hu moments gap between the target area contour and its minimum circumscribed rectangle. The skin invasion indicator was calculated as the percentage of the non-intersecting part of nail area and black line area to the black line area. We then analyzed the consistency of these indicators with features of the nail pigmentation evaluated by dermatologists.

Statistical analysis

To evaluate the performance of the image segmentation model, pixel accuracy (PA), Jaccard index, and the dice coefficient (DC) were used. The U-Net model predicts the classification probability of the pixels of the input image, which is divided into foreground (target area) and background, and outputs a pixel probability matrix with the same size as the original image. Each value in the matrix corresponds to the classification probability of the pixel at the same position in the source image. The value is between 0 and 1. Pixels with a probability value >0.5 are classified as foreground according to prior knowledge.

Linear regression analysis was used to study the relationship between the representative indicators and clinical score evaluated by experts. R square (R^2 , coefficient of determination) was used to determine the strength of the correlation between the indicators and the clinical features assessed by experts. R^2 value ranges from 0 to 1, with over 0.7 considered a good fit, between 0.5–0.7 was considered a moderate fit for the given model. The statistical analyses were performed by using GraphPad Prism 6 and $P < 0.05$ was considered significant.

Results

Accurate segmentation of nail images in the proposed automatic segmentation system

A general flowchart of the indicator analysis system for

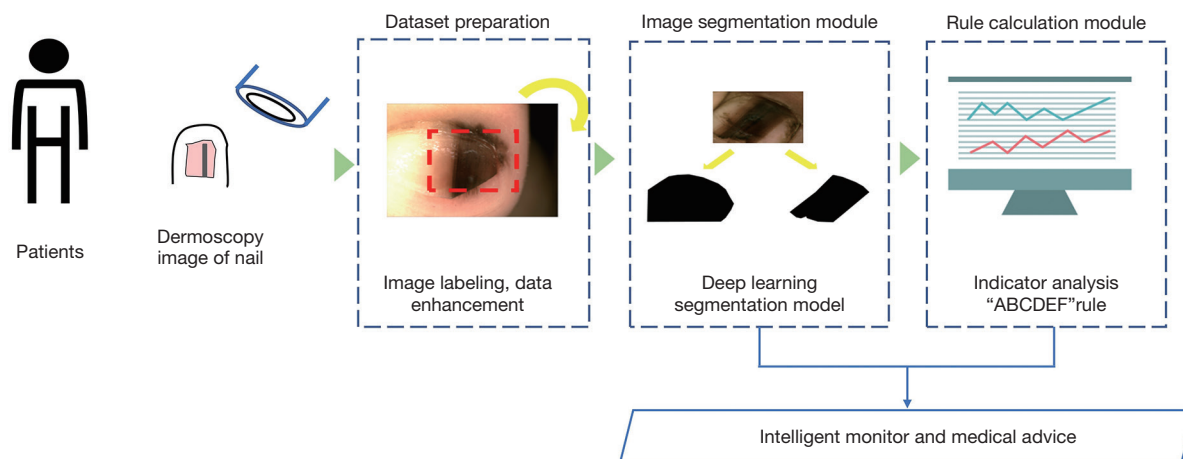


Figure 1 Flowchart of the indicator analysis system for images of nail pigmentation.

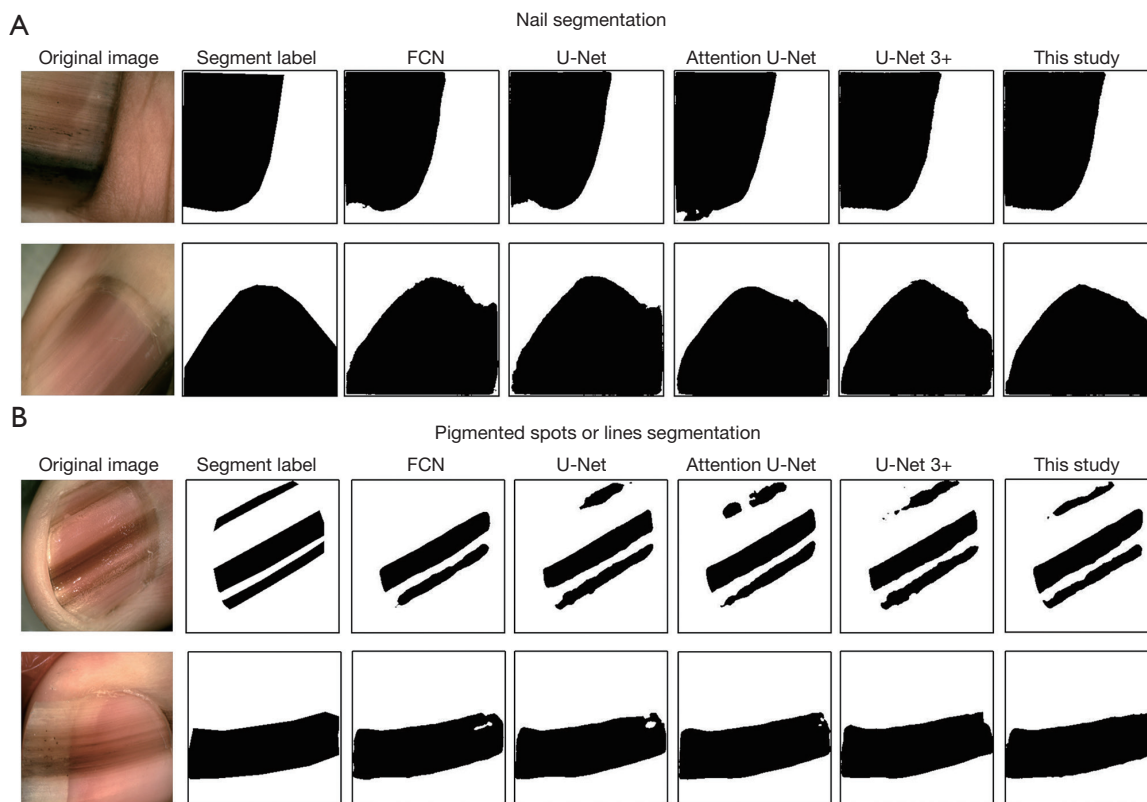


Figure 2 Segmentation label and results of the algorithms. (A) Nail area segmentation; (B) pigmented spot or line area segmentation. FCN, fully convolutional network.

images of nail pigmentation is presented in *Figure 1*. Among the data set of 550 dermoscopic images collected from outpatients with nail pigmentation, 500 images were classified randomly as the training set and the remaining

50 comprised the test set. Accurate segmentation ensures the accuracy and reliability of the rule calculation module. With the image labeling tool, the contours of the nail area (*Figure 2A*) and pigmented spots or lines (*Figure 2B*) were

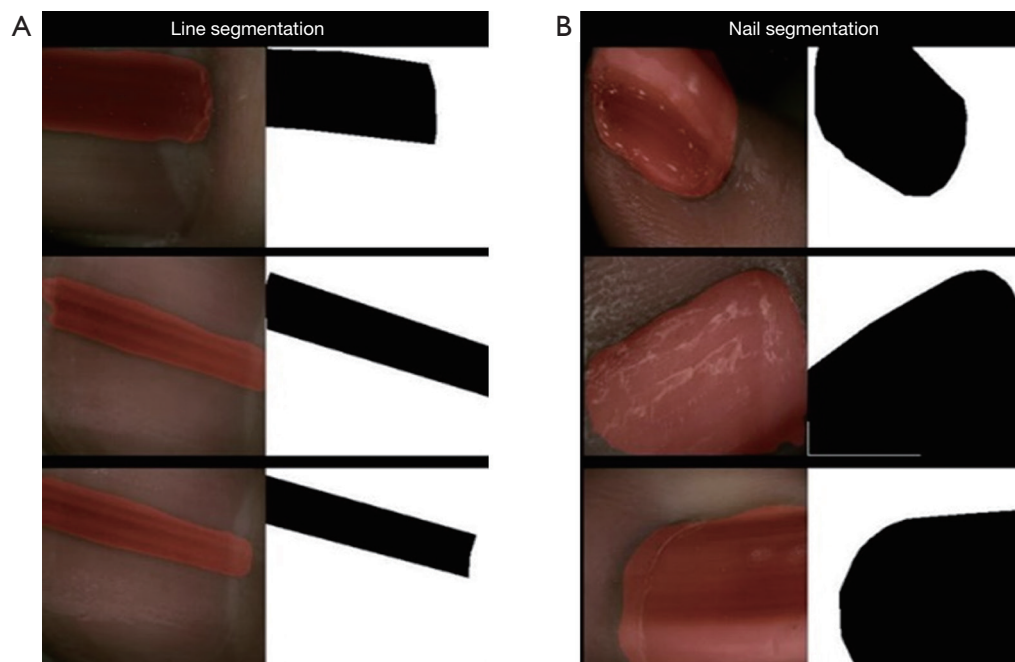


Figure 3 Segmentation results of the nail images in the proposed model. (A) Pigmented spot or line area segmentation; (B) Nail area segmentation. The right-hand image is the original segmentation label in each sample pair. Specifically, the left-hand image is the visualization of the segmentation result, and red area in the left-hand image is the segmentation result outputted by the model. The result shows that the nail image segmentation model was robustness and showed significant segmentation of the target area.

Table 1 Accuracy evaluation of the segmentation of the nail area and the pigmented spot or line area

Target area	PA	Jaccard	DC
Nail area	0.9785	0.9332	0.9652
Pigmented spot or line area	0.9753	0.7842	0.8711

PA, pixel accuracy; DC, dice coefficient.

accurately marked. *Figure 3* is a representative example of the results of the segmentation of the nail images, including segmentation of a pigmented line area (*Figure 3A*) and the whole nail plate (*Figure 3B*). Applying the trained image segmentation model to the test set, accuracy of the model was evaluated using PA, Jaccard index, and DC (*Table 1*). The results showed that the proposed nail image segmentation model had a good or acceptable segmentation effect on the target area. Notably, the DC reached 96.52% and PA reached 97.85% in the segmentation of the nail area. In addition, the segmentation of the nail area achieved higher PA, Jaccard index, and DC than the pigmented spot or line area. In summary, the results suggested that accurate segmentation can be achieved using the proposed model.

Interpretability of the quantitative indicators of the proposed model

To achieve clinical interpretability, we devised some digitally comparable features that can be extracted according to the “ABCDEF” rule, especially “B” and “E”, which are closely related to the morphological features of the lesions. Five important indicators were selected and named the area ratio, mean pixel value, evenness, irregularity, and skin invasion respectively (*Figure 4*). To analyze the clinical significance of these indicators, five representative nail pigmentation images were presented for index analysis (*Tables 2,3*). The breadth, border, and nailfold extension of the pigmented lesions, which are clinical aspects represented by “B” and “E” in the “ABCDEF” rule, were scored by experienced dermatologists and quantitatively reflected in these indicators (*Tables 2,3*). Linear regression analysis showed that the numerical value output by the computer had a similar change trend as the evaluation value of clinical experts. Further, we found that the indicator area ratio and mean pixel value were of good consistency with the judgments of clinical experts, with the R^2 for area ratio *vs.* breadth score to be 0.8179 ($P < 0.001$), for mean pixel

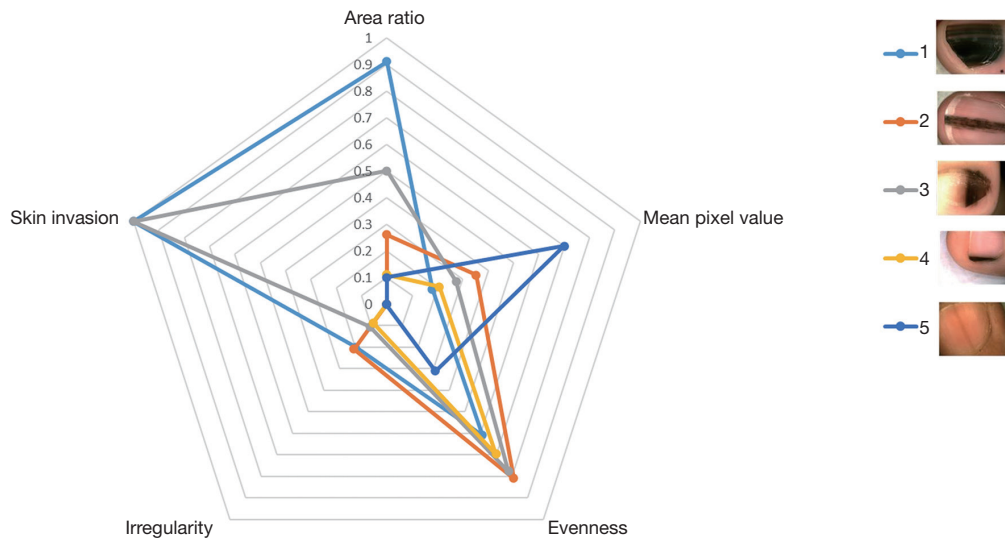


Figure 4 Radar chart showing the features of the 5 qualitative indicators indexed by the proposed model.

Table 2 Expert scores and advice for the characteristics of five representative nail pigmentation images

Image No.	Breadth score	Border score	Pigment score	Extension score	Malignant/benign	Medical advice
1	9	7.5	9	8.5	9	Biopsy as soon as possible
2	2	9	8	1	7	Biopsy
3	4	9	8	9	9	Biopsy as soon as possible
4	1	3	6	0	2	Follow up
5	0.5	0	2	0	1	No intervention required

Table 3 Index analysis results for representative nail images with pigmented line

Item	Image No.				
	1	2	3	4	5
Representative image					
Area ratio	0.91	0.26	0.50	0.11	0.10
Mean pixel value	35.86	70.43	54.81	41.45	140.05
Evenness	18.27	24.28	23.31	20.90	9.33
Irregularity	2.01	2.10	1.07	0.88	0.03
Skin invasion	True	False	True	False	False

value *vs.* pigment score to be 0.7149 ($P < 0.001$) (Figure 5). Acceptable consistency was observed between evenness and pigment score ($R^2 = 0.5247$, $P < 0.001$) (Figure 5). Therefore,

from the comparison of the sample images with quantitative indicators, we concluded that our intelligent analysis system has the potential to develop into an interpretable intelligent

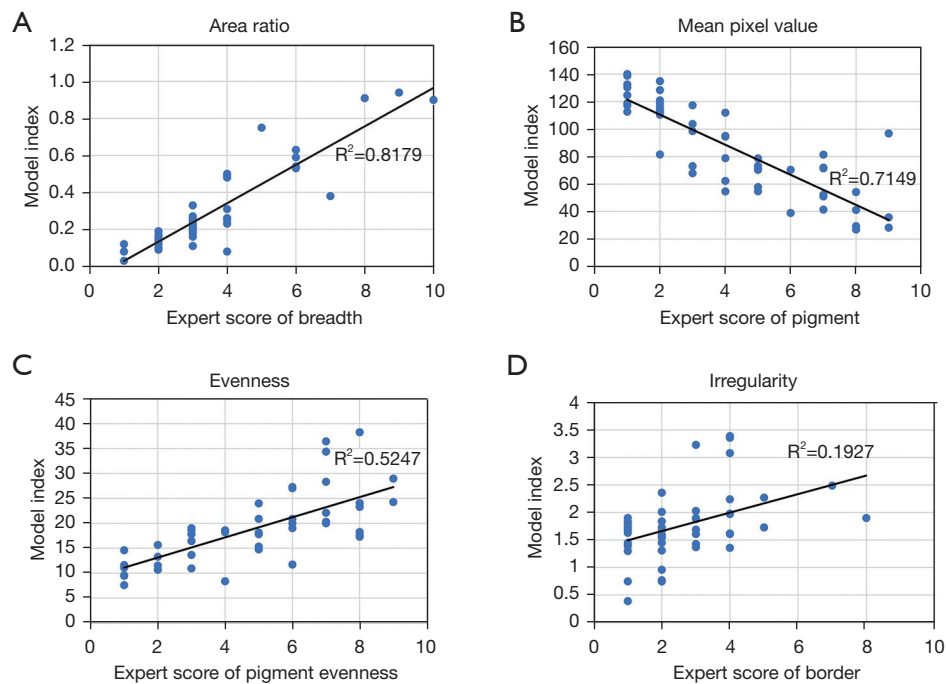


Figure 5 Regression analysis of the representative qualitative indicators and clinical expert assessment. (A) Area ratio *vs.* breadth score ($P<0.001$); (B) mean pixel value *vs.* pigment score ($P<0.001$); (C) evenness *vs.* pigment score ($P<0.001$); and (D) irregularity *vs.* border score ($P=0.0014$).

nail pigmentation follow-up system.

Discussion

Deep learning models have been applied to various scenarios of medical image recognition and classification (11-13). For example, Esteva *et al.* proposed a dermatologist-level classification model that can distinguish keratinocyte carcinomas and malignant melanomas for binary classification data and the performance of this system is comparable to that of clinical experts (11); Han *et al.* established a deep learning model of classification of clinical images of 12 skin diseases (12); and Yu *et al.* used a CNN with dermoscopic images of acral melanoma and benign nevi (13). In view of the different structures of nail and skin, the current artificial intelligence diagnosis system is currently insufficient for nail images specifically. As a risk evaluation for malignant melanoma, the “ABCDE” rule for nails is also different from that for other body parts. Furthermore, it is increasingly recognized that the so-called “black box” working method of deep learning models is an important obstacle to their widespread use in clinical practice (10,14). Having unexplainable outcomes will always

make it difficult to gain the confidence of clinicians and patients on health issues (10,14).

Long-term clinical follow-up is an important aspect of early diagnosis of nail apparatus melanoma, and for reducing the number of nail biopsies to the minimum necessary. Therefore, an interpretability analysis system based on a deep learning model of nail images will have positive clinical benefit if it can provide reliable and interpretable information of the characteristics and risk of nail pigmentation.

The proposed segmentation model showed good potential because of it performed well in segmentation of the nail images. An accurate segmentation model will first extract the target area and eliminate the interference of background shadows. Next, a connected rule-based module will give an intuitive analysis of the nail-specific “ABCDE” rule as a clinical evaluation.

For the proposed model in this study, some digitally comparable characteristics that were consistent with “B” and “E” of the “ABCDE” rule were extracted. The model outcomes have good or acceptable consistency with the expert evaluation. Area ratio corresponded to breadth, mean pixel value and evenness corresponded to pigment,

irregularity corresponded to border, and skin invasion corresponded to extension. With long-term follow-up comparing these characteristics at different times, the “C”, which means changes, would eventually be estimated with image series. Therefore, with the proposed system, we can perform the index analysis and features interpretation based on dermoscopic images and finally achieve dynamic, intelligent follow-up and warning of high risk. Compared with a simple classification model, our intelligent analysis system conforms to the actual clinical diagnosis process, and the specific indicator analysis is interpretable, which is more suitable for practical applications.

Deep learning models usually have a mass of parameters that lead to slow calculation speed. We performed convolutional layer compression on the feature extraction module of the segmentation model, which saved computing resources and memory resources while maintaining the same segmentation performance.

In order to make the intelligent analysis system more sensitive and robust for images of nail pigment lesions, the dataset has to be enlarged to further increase the accuracy and reliability of the segmentation and index analysis results.

Our study has provided a prototype artificial intelligence model with good potential for clinical application. Further studies are needed to enrich the index types, optimize the index analysis formula, incorporate more modal information, and build a multimodal analysis system for nail pigmentation. As the first step, the analysis system based on a dermoscopic image database is still insufficient to adapt to mobile phone images. In the future, the improvement and application of this analysis system to images obtained by mobile phones will help to monitor diseases more conveniently, without being restricted by location and dermoscopic equipment. In addition, the training of this model is based on the judgement of experts and therefore there may be bias. However, bias could be reduced in the future by collecting opinion from more experienced clinicians and referring to updated recognized standards. In addition, as the indicators in this study are interpretable and have similar logic to clinical practice, when bias occurs, the model can be corrected retrospectively.

Acknowledgments

The authors thank the dermatology team of the Fifth Hospital and the Third Hospital of Sun Yat-sen University for clinical advice and help in the study.

Funding: This study was funded by the Medical Science and Technology Research Project of Guangdong Province (No. A2021365) and Guangdong Provincial Key Laboratory of Human Digital Twin (2022B1212010004).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-1714/rc>

Data Sharing Statement: Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-1714/dss>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-1714/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Fifth Affiliated Hospital of Sun Yat-sen University (No. [2021](K225-1)) and informed consent was obtained from all patients.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Kim JH, Park JH, Lee DY. Site distribution of cutaneous melanoma in South Korea: a retrospective study at a single tertiary institution. *Int J Dermatol* 2015;54:e38-9.
2. Zhang J, Yun SJ, McMurray SL, et al. Management of Nail Unit Melanoma. *Dermatol Clin* 2021;39:269-80.
3. Ronger S, Touzet S, Ligeron C, et al. Dermoscopic examination of nail pigmentation. *Arch Dermatol*

- 2002;138:1327-33.
4. Levit EK, Kagen MH, Scher RK, et al. The ABC rule for clinical detection of subungual melanoma. *J Am Acad Dermatol* 2000;42:269-74.
 5. Littleton TW, Murray PM, Baratz ME. Subungual Melanoma. *Orthop Clin North Am* 2019;50:357-66.
 6. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer* 2018;18:500-10.
 7. Gore JC. Artificial intelligence in medical imaging. *Magn Reson Imaging* 2020;68:A1-4.
 8. Zunair H, Ben Hamza A. Melanoma detection using adversarial training and deep transfer learning. *Phys Med Biol* 2020;65:135005.
 9. Jayalakshmi GS, Kumar VS. Performance analysis of Convolutional Neural Network based Cancerous Skin Lesion Detection System. 2019 International Conference on Computational Intelligence in Data Science, 2019:1-6.
 10. Ras G, van Gerven M, Haselager P. Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges. In: Escalante HJ, Escalera S, Guyon I, et al. editors. *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Cham: Springer, 2018:19-36.
 11. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
 12. Han SS, Kim MS, Lim W, et al. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. *J Invest Dermatol* 2018;138:1529-38.
 13. Yu C, Yang S, Kim W, et al. Acral melanoma detection using a convolutional neural network for dermoscopy images. *PLoS One* 2018;13:e0193321.
 14. Chang J, Lee J, Ha A, et al. Explaining the Rationale of Deep Learning Glaucoma Decisions with Adversarial Examples. *Ophthalmology* 2021;128:78-88.
- (English Language Editor: K. Brown)

Cite this article as: Chen Y, Liu H, Liu Z, Xie Y, Yao Y, Xing X, Ma H. Development and validation of the interpretability analysis system based on deep learning model for smart image follow-up of nail pigmentation. *Ann Transl Med* 2022;10(10):551. doi: 10.21037/atm-22-1714