DOI: 10.1111/jdv.15965 JEADV

ORIGINAL ARTICLE

Smart identification of psoriasis by images using convolutional neural networks: a case study in China

S. Zhao,[1,2,3,†] B. Xie,[4,†] Y. Li,[4,†] X. Zhao,[4] Y. Kuang,[1,2,3] J. Su,[1,2,3] X. He,[4] X. Wu,[5] W. Fan,[5] K. Huang,[1,2,3] J. Su,[6] Y. Peng,[6] A.A. Navarini,[7,8,*] W. Huang,[9,*] X. Chen[1,2,3,*]

[1]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China
[2]Hunan Key Laboratory of Skin Cancer and Psoriasis, Changsha, China
[3]Hunan Engineering Research Center of Skin Health and Disease, Changsha, China
[4]School of Information Science and Engineering, Central South University, Changsha, China
[5]Tencent Medical AI Lab, Beijing, China
[6]Faculty of Computer Science, University of Sunderland, Sunderland, UK
[7]Department of Dermatology, University Hospital of Basel, Basel, Switzerland
[8]Department of Biomedical Engineering, University Hospital of Basel, Basel, Switzerland
[9]Mobile Health Ministry of Education – China Mobile Joint Laboratory, Xiangya Hospital, Central South University, Changsha, China
*Correspondence: X. Chen, W. Huang and A.A. Navarini. E-mails: chenxiangck@126.com, whuangcn@hotmail.com and alexander.navarini@usb.ch

Abstract

Background Psoriasis is a chronic inflammatory skin disease, which holds a high incidence in China. However, professional dermatologists who can diagnose psoriasis early and correctly are insufficient in China, especially in the rural areas. A smart approach to identify psoriasis by pictures would be highly adaptable countrywide and could play a useful role in early diagnosis and regular treatment of psoriasis.

Objectives Design and evaluation of a smart psoriasis identification system based on clinical images (without relying on a dermatoscope) that works effectively similar to a dermatologist.

Methods A set of deep learning models using convolutional neural networks (CNNs) was explored and compared in the system for automatic identification of psoriasis. The work was carried out on a standardized dermatological dataset with 8021 clinical images of 9 common disorders including psoriasis along with full electronic medical records of patients built over the last 9 years in China. A two-stage deep neural network was designed and developed to identify psoriasis. In the first stage, a multilabel classifier was trained to learn the visual patterns for each individual skin disease. In the second stage, the output of the first stage was utilized to distinguish psoriasis from other skin diseases.

Results The area under the curve (AUC) of the two-stage model reached 0.981 ± 0.015, which outperforms a single-stage model. And, the classifier showed superior performance (missed diagnosis rate: 0.03, misdiagnosis rate: 0.04) than 25 Chinese dermatologists (missed diagnosis rate: 0.19, misdiagnosis rate: 0.10) in the diagnosis of psoriasis on 100 clinical images.

Conclusions Using clinical images to identify psoriasis is feasible and effective based on CNNs, which also builds a solid technical base for smart care of skin diseases especially psoriasis using mobile/tablet applications for teledermatology in China.

Received: 11 April 2019; Accepted: 8 August 2019

Conflict of interests

The authors declared that they have no conflicts of interest to this work.

Funding source

This study was supported by Specialized Basic Work of Science and Technology (2015FY111100) and National Key R&D Program of China (2018YFC0117000).

Introduction

Skin conditions affect 30–70% of all people and thus represent a heavy burden on global health.[1] Some disease such as

[†]These authors contributed equally to the work.

psoriasis can be difficult to manage. This chronic autoimmune condition is a frequent problem in dermatology practices, partly because it is not diagnosed and treated sufficiently early by family physicians. According to statistics,[2] over 125 million patients with psoriasis have been documented worldwide, corresponding to a prevalence of 1–3%. Chronic psoriasis is difficult to control, and its predominant place in dermatology has led to a high volume of research, both clinical and bench-based.[3–7] However, without an accurate diagnosis of the condition, proper management cannot be initiated and severe forms can result in cardiovascular, hepatic and renal complications. In China, the dermatologist and patient ratio is as large as 1:60 000. Also, most experienced dermatologists are located in large cities, further adding to scarcity of dermatology in the rural areas of China. Due to the limited skin-specific medical knowledge of family physicians, psoriasis is often missed or labelled with wrong diagnoses. As a result, many patients are unaware of having psoriasis and may not consult in a tertiary hospital for proper treatment. Therefore, building an automatic psoriasis identification system is becoming an urgency.

Deep learning, especially CNNs, and its medical applications have made great progress in recent years.[8–10] In the field of computer-aided diagnosis for skin diseases, especially for tumours,[11–15] the monumental breakthroughs have been made in Ref.,[16] where CNN surpassed senior dermatologists on both classification tasks (keratinocytes cancer vs. benign seborrhoeic keratosis and malignant melanoma vs. benign moles). Han *et al.*[17] proposed to use the ResNet-152 model on clinical images to classify benign and malignant cutaneous tumours. However, because of the inadequate dataset and complex visual characteristics of psoriasis, there are few works on artificial intelligence-based diagnosis of psoriasis. Pal *et al.* have made a lot of advanced research progress by using CNNs such as segmentation of histopathologic scans of psoriasis skin biopsies,[18] severity grading of psoriasis skin disease[19] and detection of munro's microabscesses.[20] However, it is inconvenient for both dermatologists and patients to diagnose psoriasis with histopathology-derived images. And, there is no work on visual identification of psoriasis based on clinical image datasets derived mainly of Chinese patients at present.

In this study, for better recognizing psoriasis, we present a new clinical image dataset named XiangyaDerm-Pso9, which is customized for psoriasis classification and is a subset of the large-scale Chinses dataset XiangyaDerm where 8021 clinical images of psoriasis and 8 other types of skin diseases are selected. Two-stage neural network model as another contribution is made in this paper. Our experimental results indicate the advantages for the two-stage approach over the single ones and demonstrate good generalization performance in simulated clinical reality.

Material and methods

Dataset

We used the XiangyaDerm-Pso9 dataset to train the psoriasis classifier. The standardized dataset was built by following two steps.

Step one was the collection. All images were collected by dermatologists with informed consent of patients from Xiangya hospital. This prospective study was approved by the Ethics Committee of Xiangya Hospital, Central South University and informed consent was obtained from all participants. To achieve superb quality of images, a professional digital camera (Canon, Resolution: 350dpi) was used to capture each visual feature of skin lesion with maintaining sufficient illumination. In addition, images of different zoom levels were also collected for capturing both spatial distribution characteristics and fine features such as texture. Finally, XiangyaDerm-Pso9 containing 8021 images was accumulated by selecting nine disorders (900 images of psoriasis). Figure 1 shows a set of sample images in XiangyaDerm-Pso9. For better generalization performance, the diversity of dataset is necessary. These images are clinical dermatological images containing multiple body parts of patients.

Step two was data filter and annotation. After the data collection was completed, we filtered out the unqualified images and labelled the remaining ones. We removed four categories of images to obtain a clean dataset during data preprocessing: Case 1: skin lesions were obscured or altered by visible topical treatments or any other coloured residues, which can incur deep adverse impact on the training process. Case 2: special parts, such as fingers, where lesions are smaller and appear darker than normal. Case 3: skin lesions covered by hair or any other visual elements, which makes it hard to extract interpretable images. Case 4: Excessive exudate, which leads to loss surface appearance and disease-specific texture. Subsequently, three professional dermatologists who have been engaged in dermatology for more than 10 years from Xiangya Hospital annotated each image according to the corresponding medical record and pathology results to generate the final standardized experimental dataset. The data filter and annotation process as shown in Fig. 1.

Based on the above steps, our dataset has superior properties, which can be summarized as follows:

- High diagnostic credibility: all the images are verified by pathological examination and medical history, and the labelling was completed by experienced dermatologists.
- Skin disease selection characteristics: among the 9 selected common disorders, four 'erythematous' diseases have similar appearance to psoriasis (lichen planus, parapsoriasis, lupus erythematosus and eczema). The similarities can be described as follows: other than the red colour, the lesion primary is papular or plaque-like. On the other hand, the visual appearance of the other four lesions is clearly distinct from psoriasis (basal cell carcinoma, squamous cell

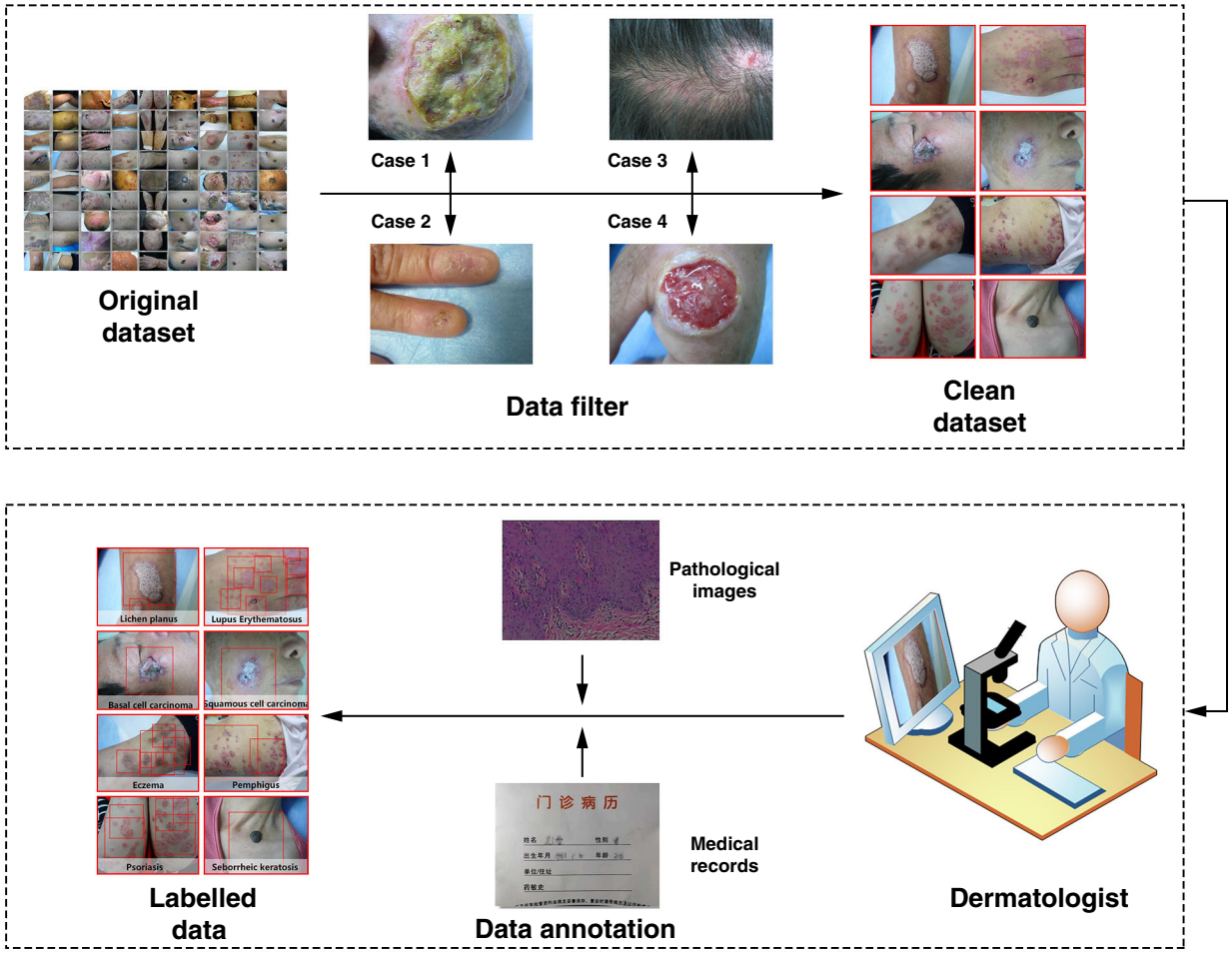

Figure 1 The data filter and annotation process. Case 1: skin lesions coated with coloured medicine. Case 2: special parts. Case 3: skin lesions covered by hair. Case 4: excessive exudate. 9 categories of skin diseases are abbreviated as lichen planus (LP), parapsoriasis (Par), lupus erythematosus (LE), basal cell carcinoma (BCC), squamous cell carcinoma (SCC), eczema (Ecz), pemphigus (Pem), psoriasis (Pso) and seborrhoeic Keratosis (SK).

carcinoma, pemphigus and seborrhoeic keratosis). This segmented dataset allowed us to pursue an iterative approach of first identifying psoriasis within four clearly distinct disease classes (the latter 4) and subsequently to focus on the former 4 diseases with more subtle visual differences to psoriasis.

- Balanced distribution: to keep a balanced instance distribution, we allocate almost the same instance count to each disease: 832 images for basal cell carcinoma, 889 for seborrhoeic keratosis and 900 each for all other classes, equalling 8021 in total. Moreover, we compared the number of clinical images of these nine skin diseases to existing large-scale publicly available datasets (>1000 images), as shown in Table 1.

Methods

Intuitively, psoriasis identifying task can be seen as binary classification of the data as being psoriasis or 'non-psoriasis' data. However, model convergence became difficult because the eight kinds of 'non-psoriasis' diseases had different image characteristics. Hence, for psoriasis identifying task, we proposed two-stage modelling method, which is shown in Fig. 2.

In stage one, we trained a 9-label classification after pretraining on ImageNet. We firstly fed the images of nine skin diseases tagged with 9 labels into the CNN model for the prediction results. This stage can avoid the large intraclass variation caused by mixing label and effectively improve the fitting efficiency.

In stage two, we made psoriasis prediction based on the output of stage one. This concrete strategy is to add up the

Table 1 Statistics of existing clinical skin disease datasets

Dataset	Pso	LP	Par	LE	Ecz	BCC	SCC	Pem	SK	Total
Derm101	470	60	\	59	806	3474	310	13	1220	**6412**
Dermnet	985	271	\	207	993	539	135	120	398	**3648**
DermIS	239	212	9	160	244	243	103	134	48	**1392**
AtlasDerm	130	175	15	141	53	277	121	118	11	**1041**
Danderm	43	17	\	17	8	\	\	\	\	**85**
Dermofit	\	\	\	\	\	239	88	\	257	**584**
XiangyaDerm-Pso9	900	900	900	900	900	832	900	900	889	8021

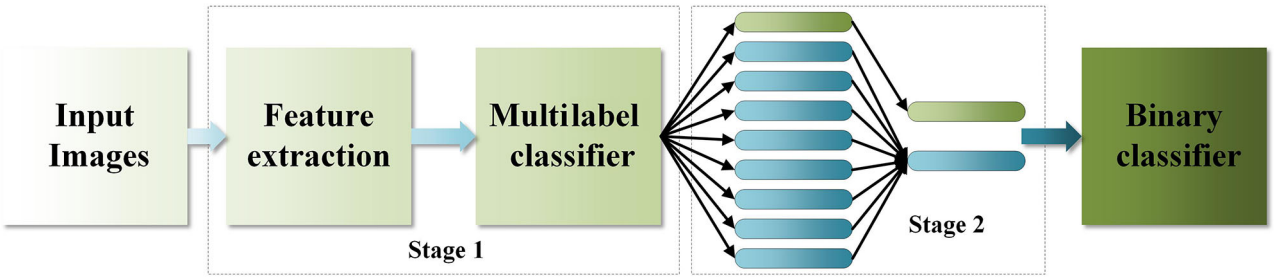

Figure 2 The proposed two-stage model.

probabilities of eight 'non-psoriasis' predictions and compare it with the probability of psoriasis to get the final classification result.

We applied four CNN architectures: DenseNet,[21] InceptionV3,[22] InceptionResNetV2[23] and Xception[24] in this paper. 9-label classifiers of stage one were trained through TensorFlow, which is a software tool for deep learning, and the calculation of the whole networks was driven by TITAN Xp graphics card. For better training effect, some parameters have been fine tuned. Firstly, data transformation such as rotation, zoom and horizontal flip was taken for data diversity, data loading was shuffled to keep similar images from being trained in same batches, and then, the batch size was empirically set as 25 according to the fitting efficiency. Secondly, RMSprop was chosen as the optimizer, which is strategy of updating weight, after we compared with other optimizers through several groups of experiments.

In our task, we denoted 'psoriasis' as positive class, 'non-psoriasis' as negative class and set a threshold value to separate the positive or negative cases. By continuously adjusting this threshold, we were able to get multiple sets of different sensitivity and specificity. Then, we marked these sets in coordinates and drew the receiver operating characteristic (ROC) curves. Finally, we calculated the AUC, which reflect the quality of training and chose the optimal sensitivities and specificity by the Youden index. The formula can be expressed as follows:

$$\begin{cases} \text{Youden's index} = \text{sensitivity} + \text{specificity} - 1 \\ \text{sensitivity} = \text{true positive/positive} \\ \text{specificity} = \text{true negative/negative} \end{cases} \quad (1)$$

For generalization performance analysis, we built two sets by selecting diseases that are different from training, validation and test sets in this paper to simulate a real-world clinical situation. We believed that these diseases could also be well identified, if the network indeed have learned deep feature of psoriasis.

Results

In this section, we conducted two experiments to explore optimal method from two perspectives. Firstly, we fitted different CNN modules into the proposed model to evaluate the performance respectively, including DenseNet121, InceptionResNetV2, Xception and InceptionV3. Secondly, we compared the performance of the proposed two-stage model against the one-stage one. In addition, to evaluate the proposed model, we compared our model with human dermatologists and tested the generalization performance on a simulated real-world clinical situation.

Two-stage model's performance with different CNN modules

We divide this dataset into three subsets: training, validation and testing in the ratio of 6:2:1 and ensure that there is no

overlap between these sets. As a result, the ROC curves were shown in Appendix S1. The point with the maximum Youden index value was highlighted red in the ROC curve.

To further demonstrate the overall performance of the strategy, the AUC was calculated, and the range values in the 95% confidence interval were also calculated by the DeLong method. The specificity and sensitivity with the optimal Youden index are also shown in Table 2.

We can draw a conclusion that, the AUC value of InceptionV3 0.981 ± 0.015 was the best, followed by Xception, InceptionResNetV2 and DenseNet. These experiments demonstrate the capability of the proposed model in identifying psoriasis and InceptionV3 was better than other architectures based on our dataset.

Two-stage model vs one-stage model

To demonstrate the advantage of the two-stage model, we compared it with the one-stage one. In the one-stage model, we removed the multilabel classifier layer from the two-stage model and directly train a binary classifier. The division of dataset was same as that in two-stage model. Here, we chose InceptionV3 as the CNN module, which outperforms other CNN modules. To avoid the occasionality, we applied a bootstrap strategy and repeated the process for 8 times.

As shown in Appendix S2 and Table 2, the proposed two-stage model (AUC: 0.981 ± 0.015) outperformed the one-stage model (AUC: 0.966 ± 0.004). In the one-stage model, we allocate all cases from 8 disorders into the same category 'non-psoriasis'. However, the visual patterns of these 8 disorders are various. For instance, the visual patterns of basal cell carcinoma and parapsoriasis are quite different, which leads to a fuzzier classifying boundary of 'non-psoriasis' class.

Human versus machine

To compare the performance of our classifier with human dermatologists in identifying clinical images of psoriasis, we chose 100 images that are not included in training set as a new test set and InceptionV3 as the module for competition. This set contains nine categories of diseases involved in this paper. The results are shown in Fig. 3. It indicates that our classifier (missed diagnosis rate: 0.03, misdiagnosis rate: 0.04) has capability to identify psoriasis with a level of competence comparable to 25 dermatologists (missed diagnosis rate: 0.19, misdiagnosis rate: 0.10) and outperforms all of them (accuracy of classifier: 0.96, mean accuracy of 25 dermatologists: 0.87). However, the doctor's diagnosis often takes into account some other information such as patient's history, so we suppose this system will help them a lot only in images recognition.

Table 2 The AUC for the prediction of psoriasis based on the four networks

	AUC (95%CI)	Specificity	Sensitivity
InceptionV3 (One-stage)	0.966 ± 0.004	0.96	0.91
DenseNet121 (Two-stage)	0.954 ± 0.024	0.97	0.83
InceptionResNetV2 (Two-stage)	0.975 ± 0.022	0.97	0.95
Xception (Two-stage)	0.976 ± 0.022	0.98	0.93
InceptionV3 (Two-stage)	0.981 ± 0.015	0.98	0.92

Generalization performance analysis

The aforementioned experimental results revealed the good performance of proposed method. However, in clinical reality, 'non-psoriasis' is not limited to the selected kinds of diseases in our training set. We built two new test sets obtained from Xiangya Hospital. The first test set consists of 50 skin disease images. These diseases have different appearances to psoriasis, including scar tissue, Bowen's disease, molluscum contagiosum, lichenoid amyloidosis, porokeratosis, melanoma, condyloma acuminatum and verruca vulgaris etc. The second test set again consisted of 50 images, which have somewhat similar appearance with psoriasis, such as pityriasis rosea. The experimental result showed that overall accuracy is 0.88, which indicates that our proposed method still has good performance on both new test set simulating a real-world selection of dermatoses not occurring in our original training set and is suitable for practical computer-aided diagnosis system.

Next, we chose to quantify the performance of classifiers for each category of 'non-psoriasis'. In order to check the outputs of the CNN tend to bias towards one class may causing by small data scale,[25] we presented the confusion matrices among 9 disorders (Appendix S3) for four types of typical CNN models, respectively. Each column represents the prediction category, and each row represents the true category of the data. From the confusion matrices, we found that there was no obvious bias among the nine disorders and the accuracy of psoriasis prediction was significantly higher than other diseases, namely more than 0.9; with one exception of DenseNet that achieved 0.8.

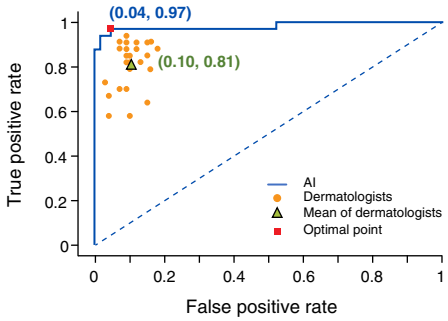

Figure 3 Performance comparison between AI system and 25 dermatologists. AI system using InceptionV3 (dodger blue curve) and 25 dermatologists (orange dot: 25 dermatologists; green triangle: average value of 25 dermatologists) were tested with 100 chosen images from the Xiangya dataset.

Moreover, it could be found that the misdiagnosis rate of psoriasis in each network was very low. The highest misdiagnosis rate was only 5.2% (InceptionV3) and the lowest was only 2.5% (DenseNet). We concluded that the CNNs could well extract the deep features of skin images for distinguishing psoriasis from other skin diseases.

To further analyse the performance between selected eight other skin diseases, we calculated the average accuracy of four networks according to Appendix S3. For the four disorders, which had similar clinical manifestation of skin lesions to psoriasis visually, they showed poor average accuracy (0.51). This indicates that CNNs easily confused them, which reflects the reality in the clinic. Even professional dermatologist cannot readily tell these diseases apart without taking information into account that is derived from the patient's history.

Discussion

Using a large dataset of >8000 images of 9 skin conditions and a two-stage approach, a CNN was trained to identify psoriasis with an accuracy at least in the range of 25 dermatologists. Increasing the difficulty with the use of visually similar pictures of close differential diagnoses resulted in suboptimal accuracy. However, these results indicate that, for visually similar skin conditions, CNNs may reach a classification potential limit. Visually distinct diseases have readily identifiable visual characteristics compared to psoriasis, thus deep features can be extracted more easily. Indeed, distinguishing visually similar diseases would be of much higher clinical significance, but may not be within reach with the so far still quite limited datasets.

The performance of deep learning model heavily relies on the properties of dataset, under the current technology level. Yang *et al.*[26] presented medical representations inspired by dermatological criteria for diagnosing clinical skin lesions based on proposed dataset[27] with superior performance, which was collected from Derm101. Many other innovative researches[16,17,28] are also attributed to these publicly available datasets, although not all of them are pathologically supported like our purposed dataset. Hence, building higher performance datasets through doctor–patient cooperation become an important way to effectively apply artificial intelligence to dermatology diagnosis. However, the shortcoming of the dataset presented in this paper is still exists. For instance, the scalp can sometimes be the only location of psoriasis and there are several signs to recognize psoriasis by looking at the nails. But the small number and low quality of such images make us abandon this point. In future study, diagnostic models using data from more body parts and making comprehensive decisions will be explored. On the other hand, skin disease diagnosis modelling with image data alone may not meet the accuracy requirements of actual clinical diagnosis, therefore we may need to collect more types of data such as question about whether itching or not to construct multiple dimensions diagnosis model in future studies.

We are confident that the construction of psoriasis classification model in this paper will become an important basis for further development and construction of psoriasis intelligent aided diagnosis and treatment platform as shown in Appendix S4. At present, the treatment of psoriasis is varied, including methotrexate, oral retinoic acid, cyclosporin, apremilast and biological agents. Our team is working on data collation and expects Artificial intelligence (AI) system to play an ideal role in recommending treatment options for psoriasis, predicting efficacy, and predicting whether psoriasis is susceptible to arthritis and other diseases.

In conclusion, we built a new standardized dataset with professional annotations, namely XiangyaDerm-Pso9, and used a two-stage model based on CNNs to achieve high accuracy and generalization performance in the identification task for psoriasis, which is a skin disease with high incidence and recurrent nature. The result shows that our classifier has superior performance than 25 dermatologists on 100 selected clinical images. This is the first time that the application of artificial intelligence in visual identification of psoriasis has explored in depth based on clinical image dataset mainly of Chinese patients. We plan to pursue and motivate follow-up studies address the urgency of building an automatic identification system for additional skin diseases.

Acknowledgements

Authors are grateful to all the doctors and nurses in Department of Dermatology, Xiangya Hospital, Central South University. And, the research is support by National Key R&D Program of China (2018YFC0117000) and Specialized Basic Work of Science and Technology (2015FY111100).

References

1 Hay RJ, Johns NE, Williams HC *et al.* The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *J Invest Dermatol* 2014; **134**: 1527–1534.
2 Langan SM, Seminara NM, Shin DB *et al.* Prevalence of metabolic syndrome in patients with psoriasis: a population-based study in the United Kingdom. *J Invest Dermatol* 2012; **132**: 556–562.
3 Chen W, Zhang X, Zhang W, Peng C, Zhu W, Chen X. Polymorphisms of SLCO1B1 rs4149056 and SLC22A1 rs2282143 are associated with responsiveness to acitretin in psoriasis patients. *Sci Rep* 2018; **8**: 13182.
4 Gao M, Liao L, Jia X, Kuang Y, Chen X. Role of soluble CD 147 in psoriatic patients: a preliminary study. *J Dermatol* 2018; **45**: e266–e267.
5 Hu Z, Xiong Z, Xu X *et al.* Loss-of-function mutations in filaggrin gene associate with psoriasis vulgaris in Chinese population. *Hum Genet* 2012; **131**: 1269–1274.
6 Wu LS, Li FF, Sun LD *et al.* A miRNA-492 binding-site polymorphism in BSG (basigin) confers risk to psoriasis in central south Chinese population. *Hum Genet* 2011; **130**: 749–757.
7 Lu H, Kuang YH, Su J *et al.* CD147 is highly expressed on peripheral blood neutrophils from patients with psoriasis and induces neutrophil chemotaxis. *J Dermatol* 2010; **37**: 1053–1056.
8 Gulshan V, Peng L, Coram M *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; **316**: 2402–2410.

9 Lindsey R, Daluiski A, Chopra S *et al.* Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci* 2018; **115**: 11591–11596.
10 Gurovich Y, Hanani Y, Bar O *et al.* Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med* 2019; **25**: 60.
11 Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans Med Imaging* 2016; **36**: 994–1004.
12 Haenssle HA, Fink C, Schneiderbauer R *et al.* Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018; **29**: 1836–1842.
13 Fujisawa Y, Otomo Y, Ogata Y *et al.* Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol* 2019; **180**: 373–381.
14 Walker BN, Rehg JM, Kalra A *et al.* Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: laboratory and prospective observational studies. *EBioMedicine* 2019; **1**: 176–183.
15 Zhao XY, Wu X, Li FF *et al.* The application of deep learning in the risk grading of skin tumors for patients using clinical images. *J Med Syst* 2019; **43**: 283.
16 Esteva A, Kuprel B, Novoa RA *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115.
17 Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol* 2018; **138**: 1529–1538.
18 Pal A, Garain U, Chandra A, Chatterjee R, Senapati S. Psoriasis skin biopsy image segmentation using Deep Convolutional Neural Network. *Comput Methods Programs Biomed* 2018; **1**: 59–69.
19 Pal A, Chaturvedi A, Garain U, Chandra A, Chatterjee R. Severity grading of psoriatic plaques using deep CNN based multi-task learning. In 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, 1478–1483.
20 Pal A, Chaturvedi A, Garain U, Chandra A, Chatterjee R, Senapati S. Caps-DeMM: Capsule network for detection of Munro's microabscess in skin biopsy images. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2018: 389–397.
21 Chollet F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017: 1251–1258.
22 Szegedy C, Liu W, Jia Y *et al.* Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015: 1–9.
23 Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. The Impact of Residual Connections on Learning. In AAAI Conference on Artificial Intelligence. 2017.
24 Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017: 4700–4708.
25 Han SS, Lim W, Kim MS, Park I, Park GH, Chang SE. Interpretation of the outputs of a deep learning model trained with a skin cancer dataset. *J Invest Dermatol* 2018; **138**: 2275.
26 Yang J, Sun X, Liang J, Rosin PL. Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018: 1258–1266.
27 Sun X, Yang J, Sun M, Wang K. A benchmark for automatic visual classification of clinical skin disease images. In European Conference on Computer Vision. Springer, Cham, 2016: 206–222.
28 Liao H, Li Y, Luo J. Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks. In 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016: 355–360.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. ROC curve of four different CNN modules.
Appendix S2. ROC curves of using binary classifiers (InceptionV3).
Appendix S3. Confusion Matrices of the four CNN models on test set, which contains 900 images consisted of 100 images of each skin disease.
Appendix S4. The flow of the whole AI platform which we suppose.