

## Original Paper

# Artificial Intelligence–Aided Diagnosis System for the Detection and Classification of Private-Part Skin Diseases: Decision Analytical Modeling Study

Wei Wang<sup>1</sup>, PhD; Xiang Chen<sup>2</sup>, Prof Dr Med; Licong Xu<sup>3</sup>, MM; Kai Huang<sup>2</sup>, MD; Shuang Zhao<sup>2</sup>, Prof Dr Med; Yong Wang<sup>1</sup>, Prof Dr

<sup>1</sup>School of Automation, Central South University, Changsha, China

<sup>2</sup>Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China

<sup>3</sup>Jinhua Fifth Hospital, Jinhua, China

**Corresponding Author:**

Yong Wang, Prof Dr

School of Automation

Central South University

932 South Lushan Road

Changsha, 410083

China

Phone: 86 18507313729

Email: [ywang@csu.edu.cn](mailto:ywang@csu.edu.cn)

## Abstract

**Background:** Private-part skin diseases (PPSDs) can cause a patient's stigma, which may hinder the early diagnosis of these diseases. Artificial intelligence (AI) is an effective tool to improve the early diagnosis of PPCDs, especially in preventing the deterioration of skin tumors in private parts such as Paget disease. However, to our knowledge, there is currently no research on using AI to identify PPCDs due to the complex backgrounds of the lesion areas and the challenges in data collection.

**Objective:** This study aimed to develop and evaluate an AI-aided diagnosis system for the detection and classification of PPCDs: aiding patients in self-screening and supporting dermatologists' diagnostic enhancement.

**Methods:** In this decision analytical modeling study, a 2-stage AI-aided diagnosis system was developed to classify PPCDs. In the first stage, a multitask detection network was trained to automatically detect and classify skin lesions (type, color, and shape). In the second stage, we proposed a knowledge graph based on dermatology expertise and constructed a decision network to classify seven PPCDs (condyloma acuminatum, Paget disease, eczema, pearly penile papules, genital herpes, syphilis, and Bowen disease). A reader study with 13 dermatologists of different experience levels was conducted. Dermatologists were asked to classify the testing cohort under reading room conditions, first without and then with system support. This AI-aided diagnostic study used the data of 635 patients from two institutes between July 2019 and April 2022. The data of Institute 1 contained 2701 skin lesion samples from 520 patients, which were used for the training of the multitask detection network in the first stage. In addition, the data of Institute 2 consisted of 115 clinical images and the corresponding medical records, which were used for the test of the whole 2-stage AI-aided diagnosis system.

**Results:** On the test data of Institute 2, the proposed system achieved the average precision, recall, and  $F_1$ -score of 0.81, 0.86, and 0.83, respectively, better than existing advanced algorithms. For the reader performance test, our system improved the average  $F_1$ -score of the junior, intermediate, and senior dermatologists by 16%, 7%, and 4%, respectively.

**Conclusions:** In this study, we constructed the first skin-lesion-based dataset and developed the first AI-aided diagnosis system for PPCDs. This system provides the final diagnosis result by simulating the diagnostic process of dermatologists. Compared with existing advanced algorithms, this system is more accurate in identifying PPCDs. Overall, our system can not only help patients achieve self-screening and alleviate their stigma but also assist dermatologists in diagnosing PPCDs.

(*J Med Internet Res* 2024;26:e52914) doi: [10.2196/52914](https://doi.org/10.2196/52914)

**KEYWORDS**

artificial intelligence-aided diagnosis; private parts; skin disease; knowledge graph; dermatology; classification; artificial intelligence; AI; diagnosis

**Introduction**

Skin diseases affect 1.9 billion people worldwide and place a burden on patients' mental health and quality of life [1-3]. Private-part skin diseases (PPSDs) are a group of high-incidence skin diseases that occur in the private parts of the human body (such as breasts, genitals, or anus), including syphilis, genital herpes, Paget disease, and condyloma acuminatum [4-7]. In clinical practice, a patient needs to expose the affected area sufficiently to dermatologists for visual inspection. Under this condition, patients with PPSDs may become nervous and embarrassed due to stigma and may be reluctant to see a doctor, which hinders the early diagnosis of PPSDs [8-11]. Patients may also be unaware of PPSDs, which is another important reason for the delay of the PPSD diagnosis. For example, sometimes patients easily mistake Paget disease for eczema, which may lead to cancer metastasis. Moreover, patients with PPSDs may hide some medical records because PPSDs may be related to sexual infidelity. Therefore, in order to promote the early diagnosis of PPSDs, it is necessary to develop a novel and private artificial intelligence (AI)-aided diagnosis technology.

Among many advanced AI algorithms, convolutional neural networks (CNNs) have developed rapidly and have shown remarkable performance on many computer vision-related tasks [12-16]. CNNs can learn meaningful and robust features directly from data and have been widely used in AI-aided diagnosis of skin diseases [17-22]. For example, Esteva et al [23] trained an end-to-end CNN with a dataset of 129,450 clinical images. The performance of this CNN is comparable to that of dermatologists in two binary classification tasks (keratinocyte carcinoma vs benign seborrheic keratosis and melanoma vs benign nevus) [23]. It is a seminal work in applying AI to skin disease diagnosis. Fink et al [24] used a pretrained GoogleNet Inception\_v4 architecture to classify combined naevi and melanomas and achieved better performance than 11 trained dermatologists [24].

Although CNN-based algorithms have shown promising results in the field of skin disease diagnosis, most of them focus on the classification of skin tumors, and none has been developed for the diagnosis of PPSDs. It may be a feasible solution to introduce existing CNN-based algorithms in skin diseases to assist diagnosis of PPSDs. However, directly applying them to PPSDs has the following two issues. First, the skin representation in clinical images of private parts is complex. For example, a large area in clinical images of private parts is occupied by irregular tissues or other visual obstacles (eg, hair and skinfolds) that are irrelevant to disease recognition, which has a negative effect on the recognition performance. Second, the data on PPSDs is scarce because it is difficult to collect the data, which limits the application of CNN-based algorithms in PPSD diagnosis. Because of these two issues, the classification of the clinical images of PPSDs is usually more difficult than that of images in natural scenes [25]. Therefore, it is necessary

to develop more effective techniques to improve the classification performance of PPSDs.

In this paper, we make the first attempt at the assistant diagnosis of PPSDs and develop a two-stage AI-aided diagnosis system by simulating the diagnostic process of dermatologists. This system builds a dermatological knowledge graph from skin lesions and medical records to disease diagnosis, aiming to reduce the model's dependence on data. Therefore, the system can render the diagnosis of PPSDs with only a small amount of data. Unlike other algorithms that directly learn the mapping from original images to diseases, this system recognizes the type, color, and shape of all skin lesions in original images, and combines the recognition results of skin lesions with the medical records to determine diseases. This system simplifies the disease classification problem into the skin lesion classification problem, which eliminates irregular tissues and visual obstacles in original images to a certain degree, thus alleviating the issue of complex skin representation in clinical images.

When our system is applied to dermatological assistant diagnosis, patients with PPSDs can avoid face-to-face diagnosis with dermatologists, thereby relieving their sense of shame and embarrassment. In addition, patients with PPSDs can carry out disease screening in time and fill in the medical records truthfully.

**Methods****Ethical Considerations**

This study is a secondary analysis based on the data generated during the diagnosis and treatment of PPSDs, all of which were approved by the Ethics Committee of Xiangya Hospital (review number 202308636). Before taking photos, all patients signed informed consent forms for imaging data shooting and scientific research analysis. Of course, all of these data were anonymized and the information that exposed the patients' identities was removed. As it is a necessary step in the diagnosis and treatment of PPSDs, patients who provide data have no additional financial compensation. We ensure that all the images in the written materials published by this study do not contain the identity information of individual participants.

**Dataset Collection and Annotation**

For this study, ethics review and institutional review board approval were obtained from 2 participating institutes (Institute 1: Xiangya Hospital of Central South University, and Institute 2: Wuhan No.1 Hospital). This study followed the Helsinki protocol and good clinical practice guidelines [26,27]. For the development and evaluation of our AI-aided diagnosis system, we retrospectively collected clinical images of PPSDs from the two institutes between July 2019 and April 2022. The collection criteria include (1) private parts skin diseases, (2) images do not affect judgment, and (3) informed consent holder. Exclusion criteria include (1) the inability to capture clear images and (2) those who do not agree to disclose images of skin lesions. These

processes were reported in accordance with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guidelines. The data collection process can be divided into 2 steps.

The first step was the collection of clinical images and their corresponding medical records for 7 categories of PPSDs: condyloma acuminatum, Paget disease, eczema, pearly penile papules, genital herpes, syphilis, and Bowen disease. Most of the collected cases had pathological confirmation and the remaining cases were unanimously recognized by 3 dermatologists from Institute 1. The medical records included the location of skin lesions, subjective symptoms, medication history, and history of high-risk sexual activities. Through data collection, we have collected 635 clinical images, of which 520 clinical images from Institute 1 were used for system development, and 115 clinical images and their corresponding medical records from Institute 2 were used for system evaluation. Detailed information on the PPSD dataset is provided in Table S1 in [Multimedia Appendix 1](#).

In the second step, we conducted data labeling and cropping on the collected clinical images. To simulate the visual diagnosis process of dermatologists, three dermatologists from Institute 1 were asked to label the clinical images at the bounding box level, where each bounding box represents a skin lesion (see examples later). These labels characterize the appearance of the skin lesions, including type (papule, plaque, nodule, erosion, and vesicle), color (red, brown, and skin color), and shape (rice shape, round, irregular, papillary, and cauliflower). By cropping the bounding boxes from the original images (each may have multiple bounding boxes), we finally obtained 2701 skin lesion images. It is worth noting that among the five types of skin lesions, the amount of data for erosion and vesicle is relatively small, and we added some images of these 2 types of skin lesions from other human body parts.

### AI-Aided Diagnosis System Overview

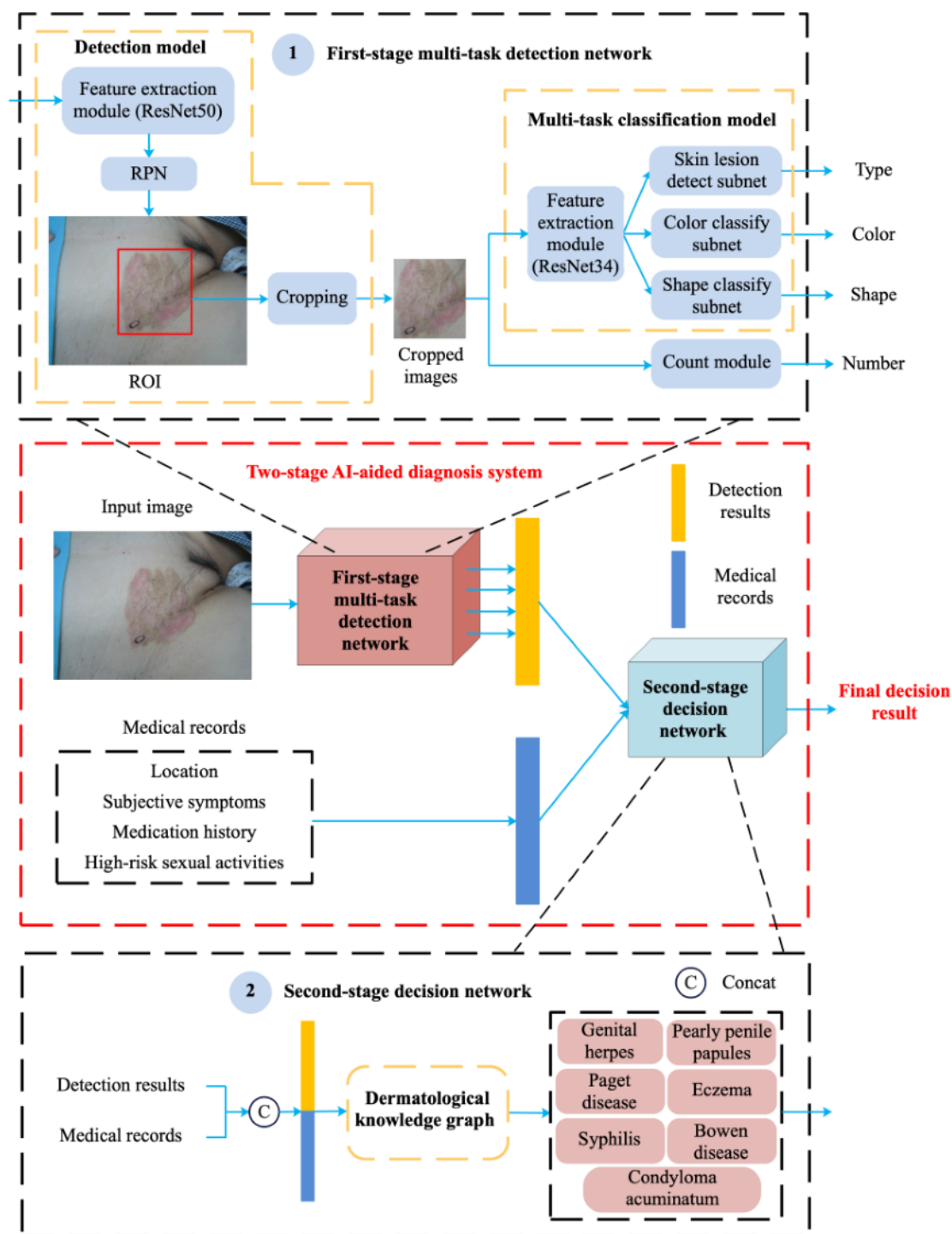
In the first stage, the multitask detection network consists of a detection model and a multitask classification model. For an input original image, the detection model first uses the feature extraction module together with the region proposal network

to extract the region of interests in the image, that is, the skin lesions. Then, we perform the cropping operation. Note that an image may have multiple bounding boxes of skin lesions. Under this condition, the results of the detection model include multiple cropped images of skin lesions. Then, these cropped images are fed into the multitask classification model to further predict the type, color, and shape of skin lesions, together with their confidence scores (probabilities). Afterward, we count the number of skin lesions in each type. The detailed architecture of the multitask detection network can be seen in Section S1 in [Multimedia Appendix 1](#).

In the second stage, to imitate the dermatologists' diagnostic process (visual diagnosis and inquiry diagnosis), we designed a decision network based on a dermatological knowledge graph (Figure S1 in [Multimedia Appendix 1](#)), which was constructed by dermatologists referencing medical guidelines and professional textbooks. The dermatological knowledge graph is composed of 30 decision paths. For each decision path, the decision network encodes the detection results of the first stage with medical records into a vector using predefined encoding rules. Subsequently, the inference result for each decision path is derived by multiplying the elements within the vector. Finally, the probabilities of all decision paths belonging to the same disease are summed and normalized to infer the final decision result. The inference process is shown in Figure S2 and Section S2 in [Multimedia Appendix 1](#).

Inspired by how dermatologists diagnose disease, we proposed a 2-stage AI-aided diagnosis system capable of processing multimodal information, as shown in [Figure 1](#). In the first stage, we trained a multitask detection network to jointly learn the type, color, and shape of skin lesions. In the second stage, we designed a decision network embedded with dermatology expertise, which combined the detection results of the first stage with medical records to obtain the final decision result. Overall, our two-stage system leverages multitask learning to explore skin lesion representations at multiple granularities and introduce dermatology expertise to alleviate the data dependency of deep learning models. The details of the two stages are presented as follows.

**Figure 1.** Flowchart of the proposed two-stage AI-aided diagnosis system for PPSDs. AI: artificial intelligence; PPSD: private-part skin disease; ROI: region of interest; RPN: region proposal network.



## Reader Study Protocol

In order to test whether our system can assist dermatologists in diagnosing PPSDs, we conducted a reader study. The participants in this study included 3 senior dermatologists, 5 intermediate dermatologists, and 5 junior dermatologists. Senior dermatologists refer to dermatologists who have been in medical practice for more than 10 years. Intermediate dermatologists

refer to those who have been licensed for more than 3 years. Junior dermatologists are those who have been in medical practice for less than 3 years. The participants were provided with 115 pairs of clinical images and their corresponding medical records as the test set to evaluate the system. They were asked to give the final disease diagnosis for the test set in 2 scenarios. In scenario one, the participants made diagnoses based on the pairs of clinical images and medical records. In

scenario two, the participants were additionally given the skin lesion detection results and disease prediction results (with PPSD confidence score) provided by our system, and they were asked to remake their diagnoses. We recorded the performance of the participants in both scenarios and calculated the performance difference between these two scenarios for each PPSD.

### Performance Evaluation Index

To evaluate the first-stage multitask detection network, we used the confusion matrix, receiver operating characteristic (ROC) curve, and area under the ROC curve (AUROC) to test the performance of skin lesion classification. The confusion matrix is used to observe the performance of the multitask classification model in various categories. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class. The ROC curve is the plot of the true positive rate against the false positive rate, at various threshold settings. AUROC measures the entire 2D area underneath the entire ROC curve, and a larger AUROC value represents better classification performance. In addition, for system evaluation, we calculated precision, recall, and  $F_1$ -score to measure the PPSD classification performance. These indices are defined as:  $\text{precision} = \text{TP} / (\text{TP} + \text{FP})$ ,  $\text{recall} = \text{TP} / (\text{TP} + \text{FN})$ , and  $F_1\text{-score} = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$ , where TP, TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively. During the evaluation process, for a positive sample, if the classifier predicts it as a positive sample, it is marked as true positive; otherwise, it is marked as false negative. For a negative sample, if the

classifier predicts it as a negative sample, it is marked as true negative; otherwise, it is marked as false positive.

## Results

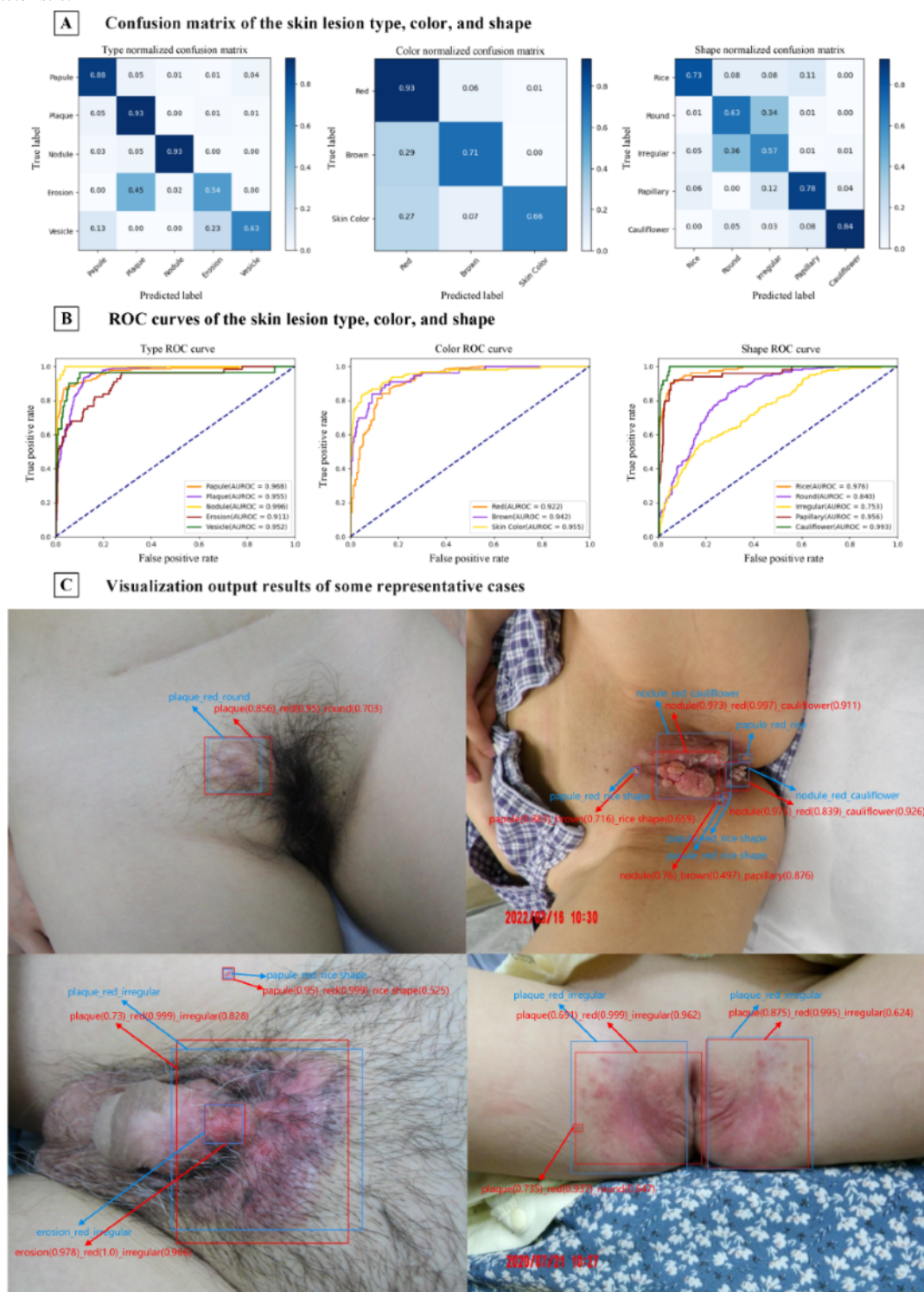
### Performance Report of the Multitask Detection Network

To verify the performance of the multitask detection network in the first stage, we carried out the classification of the type, color, and shape of skin lesions. From Figure 2A, the results of the confusion matrices show that the network has the highest accuracy on both plaque (0.93) and nodule (0.93) in lesion type recognition. In color and shape recognition, our network achieved the highest accuracy on red (0.93) and cauliflower (0.84), respectively. The results in Figure 2B suggest that the AUROCs of all categories in type and color recognition are higher than 0.9. In particular, the AUROCs of nodule in type recognition and skin color in color recognition are 0.996 and 0.955, respectively. In shape recognition, our network performed well on cauliflower (0.993), rice shape (0.976), and papillary (0.956), but performed relatively poorly on round (0.840) and irregular (0.753). The reason for this observation is elaborated in the discussion.

Figure 2C visualizes the output results of the first-stage multitask detection network in some representative cases. Here, the blue bounding boxes and texts represent the ground truths of the locations of the skin lesions and the ground truths of the labels of the skin lesion type, color, and shape, respectively. The red bounding boxes and texts represent the detection results of this network, and the numbers in brackets denote the confidence scores of the predicted categories.



**Figure 2.** Results of multitask classification model. (A) The confusion matrix. (B) The ROC curve. (C) The visualization results. ROC: receiver operating characteristic.



## Comparison of Existing Algorithms and Proposed System

To demonstrate the effectiveness of the system, we compared it with seven existing advanced deep learning algorithms. Among the 7 algorithms, ResNet101, SENet, SKNet, and Convnext are CNN-based networks, and Deit, Levit, and Swin-twins are transformer-based networks. The results in Table 1 show that the average precision, recall, and  $F_1$ -score of our system on seven PPSDs are 0.81, 0.86, and 0.83, respectively,

better than those of all the competitors. The system performance during training and validation is shown in Table S2 in Multimedia Appendix 1. In addition, our system achieved the best  $F_1$ -score among all the compared algorithms on condyloma acuminatum (0.90), genital herpes (0.94), Bowen disease (0.63), eczema (0.82), Paget disease (0.82), and syphilis (0.78). The comparison of the seven existing methods and the proposed system in computation time is given in Table S3 and Section S3 in Multimedia Appendix 1.

**Table 1.** Precision, recall, and  $F_1$ -score of our system and some advanced algorithms for each PPSD<sup>a</sup>.

PPSDs	ResNet101	SKNet	Convnext	Deit	SENet	Levit	Swin-twins	Ours
<b>Average</b>								
F1 <sup>b</sup>	0.52	0.59	0.57	0.59	0.59	0.60	0.69	0.83 <sup>c</sup>
R <sup>d</sup>	0.57	0.62	0.59	0.60	0.63	0.61	0.69	0.86
P <sup>e</sup>	0.55	0.62	0.72	0.62	0.64	0.64	0.72	0.81
<b>SP<sup>f</sup></b>								
F1	0.21	0.51	0.22	0.57	0.39	0.40	0.44	0.78
R	0.13	0.38	0.13	0.50	0.25	0.29	0.33	0.73
P	0.60	0.82	1.00	0.67	0.86	0.64	0.67	0.84
<b>PD<sup>g</sup></b>								
F1	0.64	0.64	0.77	0.59	0.66	0.71	0.71	0.82
R	0.74	0.91	0.87	0.65	0.87	0.87	0.74	0.78
P	0.57	0.49	0.69	0.54	0.53	0.61	0.68	0.86
<b>EZ<sup>h</sup></b>								
F1	0.32	0.17	0.36	0.48	0.36	0.34	0.67	0.82
R	0.35	0.12	0.35	0.41	0.35	0.35	0.82	0.94
P	0.29	0.29	0.38	0.58	0.38	0.33	0.56	0.73
<b>BD<sup>i</sup></b>								
F1	0.10	0.31	0.45	0.13	0.19	0.33	0.38	0.63
R	0.06	0.25	0.44	0.13	0.13	0.25	0.38	0.56
P	0.20	0.40	0.47	0.14	0.40	0.50	0.38	0.71
<b>PPP<sup>j</sup></b>								
F1	0.91	1.00	1.00	0.91	1.00	1.00	0.89	0.91
R	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00
P	0.83	1.00	1.00	0.83	1.00	1.00	1.00	0.83
<b>GH<sup>k</sup></b>								
F1	0.80	0.78	0.55	0.77	0.78	0.71	0.93	0.94
R	0.75	0.88	0.38	0.63	0.88	0.63	0.88	1.00
P	0.86	0.70	1.00	1.00	0.70	0.83	1.00	0.89
<b>CA<sup>l</sup></b>								
F1	0.67	0.71	0.64	0.70	0.75	0.71	0.79	0.90
R	0.95	0.82	0.95	0.86	0.95	0.91	0.86	1.00
P	0.51	0.62	0.48	0.59	0.62	0.59	0.73	0.81

<sup>a</sup>PPSD: private-part skin disease.  
<sup>b</sup>F1:  $F_1$ -score.  
<sup>c</sup>Values in italics format indicate optimal values under the corresponding indicators.  
<sup>d</sup>R: recall.  
<sup>e</sup>P: precision.  
<sup>f</sup>SP: syphilis.  
<sup>g</sup>PD: Paget disease.  
<sup>h</sup>EZ: eczema.  
<sup>i</sup>BD: Bowen disease.

<sup>j</sup>PPP: pearly penile papules.

<sup>k</sup>GH: genital herpes.

<sup>l</sup>CA: condyloma acuminatum.

### Reader Study Performance

In this study, the diagnostic performance (including precision, recall, and  $F_1$ -score) of the 3 groups of dermatologists was evaluated in 2 scenarios (first without and then with system support). From Table 2, in terms of the average  $F_1$ -score, our system (0.83) outperformed the junior (0.57) and the intermediate dermatologists (0.78) in scenario 1, and was comparable to the senior dermatologists (0.85). Furthermore, with the support of the system (scenario 2), the dermatologists

improved their diagnostic performance. Among the 3 groups of dermatologists, the junior group showed the greatest improvement, and their average precision, recall, and  $F_1$ -score increased by 17%, 14%, and 16%, respectively. In particular, the  $F_1$ -score of genital herpes had the largest increase of 41%. For the intermediate group, the average precision, recall, and  $F_1$ -score all increased by about 7%. The senior group showed the smallest improvement in diagnostic performance, with the average precision, recall, and  $F_1$ -score increasing by 2%, 5%, and 4%, respectively.



**Table 2.** Classification performance of the three levels of dermatologists in the first and second scenarios<sup>a</sup>.

Disease	Junior	Junior+AI <sup>b</sup>	Junior- $\Delta^c$	Intermediate	Intermedi- ate+AI	Intermediate- $\Delta$	Senior	Senior+AI	Senior- $\Delta$	Ours
<b>Average</b>										
F1 <sup>d</sup>	0.57	0.73	0.16	0.78	0.85	0.07	0.85	0.89	0.04	0.83
R <sup>e</sup>	0.61	0.75	0.14	0.78	0.85	0.07	0.85	0.90	0.05	0.86
P <sup>f</sup>	0.58	0.75	0.17	0.80	0.88	0.08	0.88	0.90	0.02	0.81
<b>SP<sup>g</sup></b>										
F1	0.67	0.75	0.08	0.78	0.81	0.03	0.87	0.88	0.01	0.78
R	0.65	0.68	0.03	0.75	0.74	-0.01	0.88	0.85	-0.03	0.73
P	0.70	0.85	0.15	0.83	0.96	0.13	0.85	0.92	0.07	0.84
<b>PD<sup>h</sup></b>										
F1	0.46	0.64	0.18	0.75	0.83	0.08	0.86	0.88	0.02	0.82
R	0.37	0.55	0.18	0.67	0.82	0.15	0.83	0.84	0.01	0.78
P	0.63	0.78	0.15	0.86	0.85	-0.01	0.91	0.93	0.02	0.86
<b>EZ<sup>i</sup></b>										
F1	0.68	0.71	0.03	0.68	0.77	0.09	0.79	0.89	0.10	0.82
R	0.82	0.87	0.05	0.82	0.89	0.07	0.88	0.98	0.10	0.94
P	0.58	0.61	0.03	0.58	0.70	0.12	0.72	0.81	0.09	0.73
<b>BD<sup>j</sup></b>										
F1	0.37	0.61	0.24	0.66	0.75	0.09	0.77	0.74	-0.03	0.63
R	0.34	0.56	0.22	0.62	0.67	0.05	0.70	0.67	-0.03	0.56
P	0.43	0.67	0.24	0.76	0.85	0.09	0.90	0.86	-0.04	0.71
<b>PPP<sup>k</sup></b>										
F1	0.64	0.77	0.13	0.95	0.98	0.03	1.00	1.00	0.00	0.91
R	0.88	0.80	-0.08	1.00	0.96	-0.04	1.00	1.00	0.00	1.00
P	0.51	0.78	0.27	0.91	1.00	0.09	1.00	1.00	0.00	0.83
<b>GH<sup>l</sup></b>										
F1	0.36	0.77	0.41	0.73	0.91	0.18	0.75	0.94	0.19	0.94
R	0.33	0.83	0.50	0.70	0.88	0.18	0.67	0.96	0.29	1.00
P	0.48	0.76	0.28	0.77	0.96	0.19	0.90	0.93	0.03	0.89
<b>CA<sup>m</sup></b>										
F1	0.80	0.88	0.08	0.89	0.90	0.01	0.94	0.92	-0.02	0.90
R	0.89	0.98	0.09	0.93	0.98	0.05	1.00	1.00	0.00	1.00
P	0.73	0.80	0.07	0.86	0.83	-0.03	0.88	0.85	-0.03	0.81

<sup>a</sup>The dermatologists included 3 senior dermatologists, 5 intermediate dermatologists, and 5 junior dermatologists.<sup>b</sup>AI: artificial intelligence.<sup>c</sup> $\Delta$  indicates the difference in each evaluation index between two scenarios (dermatologist and dermatologist + AI).<sup>d</sup>F1: F<sub>1</sub>-score.<sup>e</sup>R: recall.<sup>f</sup>P: precision.<sup>g</sup>SP: syphilis.<sup>h</sup>PD: Paget disease.

<sup>i</sup>EZ: eczema.

<sup>j</sup>BD: Bowen disease.

<sup>k</sup>PPP: pearly penile papules.

<sup>l</sup>GH: genital herpes.

<sup>m</sup>CA: condyloma acuminatum.

## Discussion

### Principal Results

In our study, to help the early diagnosis of PPSDs, we established the first PPSD dataset and developed the first AI-aided diagnosis system of PPSDs. The main finding was that compared with existing advanced classification algorithms, our 2-stage AI-aided diagnosis system achieved more accurate classification performance of PPSDs and detected all the skin lesions in images.

Interpretability is an important factor in AI-aided diagnosis, as it can help clinicians understand the predictions of AI algorithms and enable better human-machine collaboration in clinical practice. In the first stage, our system could detect all the skin lesions in images and give visualization results, which indicates the interpretability of our system to predict skin lesions. In addition, in the second stage, our system simulates the diagnosis process of dermatologists, inferring PPSDs based on the dermatological knowledge graph. Since the dermatological knowledge graph is in line with the clinical routine of dermatologists, our system is interpretable in the prediction of PPSDs.

AI has been proven to be applicable in the diagnosis, treatment, and efficacy prediction of skin diseases. Huang et al [28] used an AI-based approach for assessing the severity of psoriasis, which trains end-to-end with only images and severity scores [28]. They also demonstrated that the severity score predicted by the AI model is close to the Psoriasis Area and Severity Index score diagnosed by experienced dermatologists. Philips et al [29] reported that AI-based classification methods not only outperform human experts in the diagnosis of pigmented skin diseases but also further improve diagnostic accuracy.

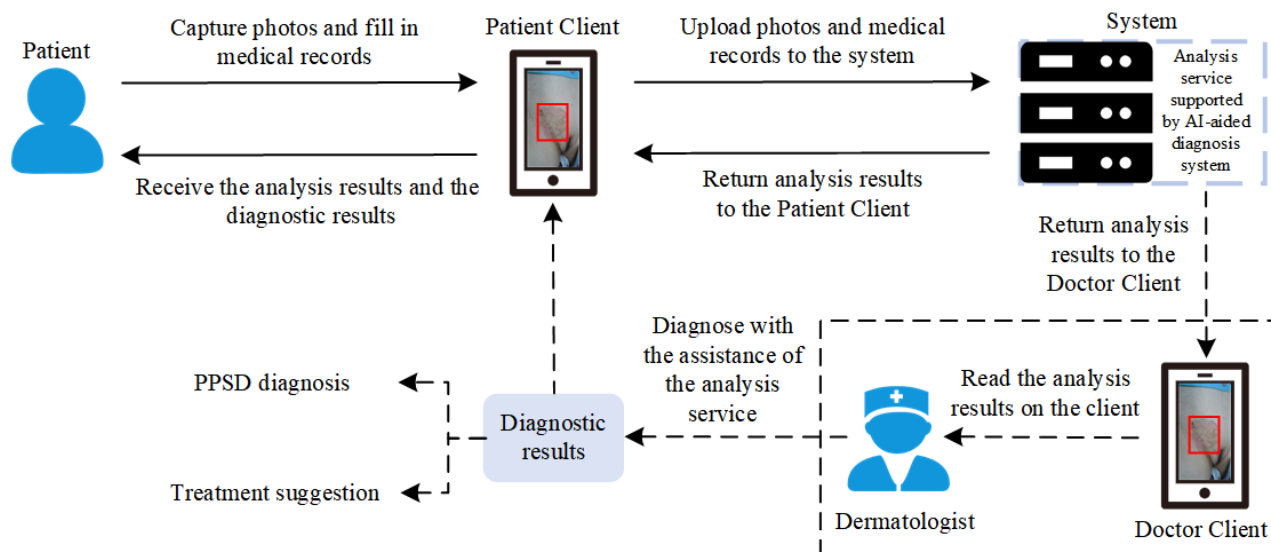
During patient consultations, visual inspection is the first step of skin disease diagnosis. The lesion-affected private parts need to be fully exposed to dermatologists, which causes embarrassment for patients and further delays medical treatment, thereby significantly impacting patients' mental health [30,31]. The system developed in this study avoids the stigma of exposing private parts to strangers, allowing patients to undergo dermatological inspections with more comfort and ease. It

should be noted that the purpose of this system is to assist patients in self-screening, not to replace dermatologists. Additionally, in the future, remote medical treatment can be used for digital PPSD treatment, further reducing the degree of medical delays [32,33].

It can be seen from Table 2 that in terms of  $F_1$ -score, the system performed poorly for Bowen disease (0.63). The reason is that the dermatological knowledge graph in the second stage considers the shape of the plaque (irregular or round) as the key to the diagnosis of Bowen disease, while the multitask detection network in the first stage has poor recognition performance on irregular and round classifications. It is not surprising that our system performed poorly in recognizing the shape of plaque. In fact, dermatologists also find it challenging to accurately determine the shape of lesions when annotating images of Bowen disease. When assisting dermatologists in diagnosing PPSDs, our system improved the diagnostic performance of all the dermatologists, demonstrating the feasibility of our system in AI-supported image analysis. In particular, with the assistance of our system, the diagnostic performance of dermatologists in the junior group has been improved, basically helping them to reach the ability of dermatologists in the intermediate group. Regarding Paget disease, an extremely serious skin tumor, the  $F_1$ -score of dermatologists in the junior group is improved by 18%. The above results preliminarily suggest that our system can help dermatologists in primary hospitals to diagnose PPSDs, and it can also be used as a reference when dermatologists have disagreements.

Our 2-stage AI-aided diagnosis system is helpful to assist patients in self-screening, and the results of the system are sent to dermatologists for further confirmation. The workflow is shown in Figure 3. First, patients with PPSDs can take a photo of the skin lesion, fill in the medical records through the patient client in the mobile phone, and upload them to the system. Then, the system detects all the skin lesions (type, color, and shape) in the photo and combines them with the medical records to generate the preliminary analysis results. Afterward, the preliminary analysis results are sent to dermatologists for further diagnosis. By combining patient information with the analysis results given by the system, dermatologists make the final diagnostic results and treatment suggestions.

**Figure 3.** The routine clinical workflow of our two-stage AI-aided diagnosis system. AI: artificial intelligence; PPSD: private-part skin disease.



## Limitations

This study has some limitations. First, the volume of our data was limited due to the difficulty of data collection on PPSDs. Second, we only included seven categories of PPSDs, so some other diseases were not considered. Third, the decision network in the second stage of our system was designed based on a knowledge graph of dermatology expertise, which consists of a series of directed decision paths. Since a single clinical image cannot fit all decision paths, certain PPSD confidence scores will be 0. As a result, the value of AUROC cannot be properly calculated. Fourth, the variety of datasets was insufficient because we could only collect data from 2 sources. Finally, the diagnostic performance of our system for Bowen disease was not satisfactory due to the difficulty of data labeling. Overall, further efforts should be made to overcome these limitations.

## Conclusions

In this paper, we developed a two-stage AI-aided diagnosis system for PPSDs. Different from existing methods that directly learn image-to-disease mapping, the developed system simulated a dermatologist's diagnostic process by first identifying skin lesions, and then inferring diseases based on the skin lesion identification results and medical records. The system addressed the issues of complex skin representation in the images of PPSDs and data dependence. Compared with existing advanced algorithms, this system was more clinically relevant and performed better. In addition, the results of the reader study suggest that our system can improve the performance of dermatologists in diagnosing PPSDs. In practical applications, our system has the potential to alleviate the stigma of patients with PPSD and avoid treatment delays.

## Acknowledgments

The authors thank the doctors and nurses in the Department of Dermatology, Xiangya Hospital, Central South University. Special thanks for the support and appreciation of 13 dermatologists who participated in the reader study.

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional materials.

[DOCX File, 2556 KB-Multimedia Appendix 1]

## References

1. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med*. 2020;26(6):900-908. [doi: [10.1038/s41591-020-0842-3](https://doi.org/10.1038/s41591-020-0842-3)] [Medline: [32424212](https://pubmed.ncbi.nlm.nih.gov/32424212/)]
2. Wu JH, Cohen BA. The stigma of skin disease. *Curr Opin Pediatr*. 2019;31(4):509-514. [doi: [10.1097/MOP.0000000000000792](https://doi.org/10.1097/MOP.0000000000000792)] [Medline: [31188167](https://pubmed.ncbi.nlm.nih.gov/31188167/)]

3. Germain N, Augustin M, François C, Legau K, Bogoeva N, Desroches M, et al. Stigma in visible skin diseases—a literature review and development of a conceptual model. *J Eur Acad Dermatol Venereol*. 2021;35(7):1493-1504. [doi: [10.1111/jdv.17110](https://doi.org/10.1111/jdv.17110)] [Medline: [33428316](https://pubmed.ncbi.nlm.nih.gov/33428316/)]
4. Peeling RW, Mabey D, Kamb ML, Chen X, Radolf JD, Benzaken AS. Syphilis. *Nat Rev Dis Primers*. 2017;3:17073. [FREE Full text] [doi: [10.1038/nrdp.2017.73](https://doi.org/10.1038/nrdp.2017.73)] [Medline: [29022569](https://pubmed.ncbi.nlm.nih.gov/29022569/)]
5. Gupta R, Warren T, Wald A. Genital herpes. *Lancet*. 2007;370(9605):2127-2137. [doi: [10.1016/S0140-6736\(07\)61908-4](https://doi.org/10.1016/S0140-6736(07)61908-4)] [Medline: [18156035](https://pubmed.ncbi.nlm.nih.gov/18156035/)]
6. Kibbi N, Owen JL, Worley B, Wang JX, Harikumar V, Downing MB, et al. Evidence-based clinical practice guidelines for extramammary paget disease. *JAMA Oncol*. 2022;8(4):618-628. [doi: [10.1001/jamaoncol.2021.7148](https://doi.org/10.1001/jamaoncol.2021.7148)] [Medline: [35050310](https://pubmed.ncbi.nlm.nih.gov/35050310/)]
7. Tying SK, Cauda R, Baron S, Whitley RJ. Condyloma acuminatum: epidemiological, clinical and therapeutic aspects. *Eur J Epidemiol*. 1987;3(3):209-215. [doi: [10.1007/BF00149726](https://doi.org/10.1007/BF00149726)] [Medline: [3308509](https://pubmed.ncbi.nlm.nih.gov/3308509/)]
8. Valentine JA, Delgado LF, Haderxhanaj LT, Hogben M. Improving sexual health in U.S. rural communities: reducing the impact of stigma. *AIDS Behav*. 2022;26(Suppl 1):90-99. [FREE Full text] [doi: [10.1007/s10461-021-03416-4](https://doi.org/10.1007/s10461-021-03416-4)] [Medline: [34436713](https://pubmed.ncbi.nlm.nih.gov/34436713/)]
9. Chime PE, Okoli PC, Chime EN, Anekpo CC, Ozougwu AO, Ofojebe PC. Diseases associated with stigma: a review. *OJPsych*. 2022;12(02):129-140. [doi: [10.4236/ojpsych.2022.122011](https://doi.org/10.4236/ojpsych.2022.122011)]
10. Liu H, Detels R, Li X, Ma E, Yin Y. Stigma, delayed treatment, and spousal notification among male patients with sexually transmitted disease in China. *Sex Transm Dis*. 2002;29(6):335-343. [doi: [10.1097/00007435-200206000-00005](https://doi.org/10.1097/00007435-200206000-00005)] [Medline: [12035023](https://pubmed.ncbi.nlm.nih.gov/12035023/)]
11. Lee AS, Cody SL. The stigma of sexually transmitted infections. *Nurs Clin North Am*. 2020;55(3):295-305. [doi: [10.1016/j.cnur.2020.05.002](https://doi.org/10.1016/j.cnur.2020.05.002)] [Medline: [32762851](https://pubmed.ncbi.nlm.nih.gov/32762851/)]
12. Zhao Z, Wu CM, Zhang S, He F, Liu F, Wang B, et al. A novel convolutional neural network for the diagnosis and classification of rosacea: usability study. *JMIR Med Inform*. 2021;9(3):e23415. [FREE Full text] [doi: [10.2196/23415](https://doi.org/10.2196/23415)] [Medline: [33720027](https://pubmed.ncbi.nlm.nih.gov/33720027/)]
13. Naylor CD. On the prospects for a (deep) learning health care system. *JAMA*. 2018;320(11):1099-1100. [doi: [10.1001/jama.2018.11103](https://doi.org/10.1001/jama.2018.11103)] [Medline: [30178068](https://pubmed.ncbi.nlm.nih.gov/30178068/)]
14. Sun X, Yang J, Sun M. A benchmark for automatic visual classification of clinical skin disease images. 2016. Presented at: European Conference on Computer Vision; January 15-16, 2025; London, UK. [doi: [10.1007/978-3-319-46466-4\\_13](https://doi.org/10.1007/978-3-319-46466-4_13)]
15. 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(7):1529-1538. [FREE Full text] [doi: [10.1016/j.jid.2018.01.028](https://doi.org/10.1016/j.jid.2018.01.028)] [Medline: [29428356](https://pubmed.ncbi.nlm.nih.gov/29428356/)]
16. Li Z, Jiang Y, Li B, Han Z, Shen J, Xia Y, et al. Development and validation of a machine learning model for detection and classification of tertiary lymphoid structures in gastrointestinal cancers. *JAMA Netw Open*. 2023;6(1):e2252553. [FREE Full text] [doi: [10.1001/jamanetworkopen.2022.52553](https://doi.org/10.1001/jamanetworkopen.2022.52553)] [Medline: [36692877](https://pubmed.ncbi.nlm.nih.gov/36692877/)]
17. Huang K, Jiang Z, Li Y, Wu Z, Wu X, Zhu W, et al. The classification of six common skin diseases based on xiangya-derm: development of a Chinese database for artificial intelligence. *J Med Internet Res*. 2021;23(9):e26025. [FREE Full text] [doi: [10.2196/26025](https://doi.org/10.2196/26025)] [Medline: [34546174](https://pubmed.ncbi.nlm.nih.gov/34546174/)]
18. 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*. 2017;36(4):994-1004. [doi: [10.1109/TMI.2016.2642839](https://doi.org/10.1109/TMI.2016.2642839)] [Medline: [28026754](https://pubmed.ncbi.nlm.nih.gov/28026754/)]
19. Kawahara J, Hamarneh G. Fully convolutional neural networks to detect clinical dermoscopic features. *IEEE J Biomed Health Inform*. 2019;23(2):578-585. [doi: [10.1109/JBHI.2018.2831680](https://doi.org/10.1109/JBHI.2018.2831680)] [Medline: [29994053](https://pubmed.ncbi.nlm.nih.gov/29994053/)]
20. Han SS, Moon IJ, Lim W, Suh IS, Lee SY, Na JJ, et al. Keratinocytic skin cancer detection on the face using region-based convolutional neural network. *JAMA Dermatol*. 2020;156(1):29-37. [FREE Full text] [doi: [10.1001/jamadermatol.2019.3807](https://doi.org/10.1001/jamadermatol.2019.3807)] [Medline: [31799995](https://pubmed.ncbi.nlm.nih.gov/31799995/)]
21. Xie F, Fan H, Li Y, Jiang Z, Meng R, Bovik A. Melanoma classification on dermoscopy images using a neural network ensemble model. *IEEE Trans Med Imaging*. 2017;36(3):849-858. [doi: [10.1109/TMI.2016.2633551](https://doi.org/10.1109/TMI.2016.2633551)] [Medline: [27913337](https://pubmed.ncbi.nlm.nih.gov/27913337/)]
22. Winkler JK, Blum A, Kommos K, Enk A, Toberer F, Rosenberger A, et al. Assessment of diagnostic performance of dermatologists cooperating with a convolutional neural network in a prospective clinical study: human with machine. *JAMA Dermatol*. 2023;159(6):621-627. [doi: [10.1001/jamadermatol.2023.0905](https://doi.org/10.1001/jamadermatol.2023.0905)] [Medline: [37133847](https://pubmed.ncbi.nlm.nih.gov/37133847/)]
23. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. [FREE Full text] [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)] [Medline: [28117445](https://pubmed.ncbi.nlm.nih.gov/28117445/)]
24. Fink C, Blum A, Buhl T, Mitteldorf C, Hofmann-Wellenhof R, Deinlein T, et al. Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *J Eur Acad Dermatol Venereol*. 2020;34(6):1355-1361. [doi: [10.1111/jdv.16165](https://doi.org/10.1111/jdv.16165)] [Medline: [31856342](https://pubmed.ncbi.nlm.nih.gov/31856342/)]
25. Li H, Pan Y, Zhao J, Zhang L. Skin disease diagnosis with deep learning: a review. *Neurocomputing*. 2021;464:364-393. [doi: [10.1016/j.neucom.2021.08.096](https://doi.org/10.1016/j.neucom.2021.08.096)]
26. Ringquist EJ, Kostadinova T. Assessing the effectiveness of international environmental agreements: the case of the 1985 Helsinki protocol. *Am J Polit Sci*. 2004;49(1):86-102. [doi: [10.1111/j.0092-5853.2005.00112.x](https://doi.org/10.1111/j.0092-5853.2005.00112.x)]

27. Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The good clinical practice guideline: a bronze standard for clinical research. *Lancet*. 2005;366(9480):172-174. [doi: [10.1016/S0140-6736\(05\)66875-4](https://doi.org/10.1016/S0140-6736(05)66875-4)] [Medline: [16005342](#)]
28. Huang K, Wu X, Li Y, Lv C, Yan Y, Wu Z, et al. Artificial intelligence-based psoriasis severity assessment: real-world study and application. *J Med Internet Res*. 2023;25:e44932. [FREE Full text] [doi: [10.2196/44932](https://doi.org/10.2196/44932)] [Medline: [36927843](#)]
29. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Netw Open*. 2019;2(10):e1913436. [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.13436](https://doi.org/10.1001/jamanetworkopen.2019.13436)] [Medline: [31617929](#)]
30. Link BG, Phelan JC. Stigma and its public health implications. *Lancet*. 2006;367(9509):528-529. [doi: [10.1016/S0140-6736\(06\)68184-1](https://doi.org/10.1016/S0140-6736(06)68184-1)] [Medline: [16473129](#)]
31. Wittkowski A, Richards HL, Griffiths CEM, Main CJ. The impact of psychological and clinical factors on quality of life in individuals with atopic dermatitis. *J Psychosom Res*. 2004;57(2):195-200. [doi: [10.1016/S0022-3999\(03\)00572-5](https://doi.org/10.1016/S0022-3999(03)00572-5)] [Medline: [15465076](#)]
32. Martin-Gonzalez M, Azcarraga C, Martin-Gil A, Carpena-Torres C, Jaen P. Efficacy of a deep learning convolutional neural network system for melanoma diagnosis in a hospital population. *Int J Environ Res Public Health*. 2022;19(7):3892. [FREE Full text] [doi: [10.3390/ijerph19073892](https://doi.org/10.3390/ijerph19073892)] [Medline: [35409575](#)]
33. Giavina-Bianchi M, de Sousa RM, Paciello VZA, Vitor WG, Okita AL, Prôa R, et al. Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting. *PLoS One*. 2021;16(9):e0257006. [FREE Full text] [doi: [10.1371/journal.pone.0257006](https://doi.org/10.1371/journal.pone.0257006)] [Medline: [34550970](#)]

## Abbreviations

**AI:** artificial intelligence

**AUROC:** area under the receiver operating characteristic curve

**CNN:** convolutional neural network

**PPSD:** private-part skin disease

**ROC:** receiver operating characteristic

**TRIPOD:** Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

*Edited by A Mavragani; submitted 19.09.23; peer-reviewed by H Zhu, Z Li; comments to author 09.02.24; revised version received 04.04.24; accepted 12.11.24; published 27.12.24*

*Please cite as:*

Wang W, Chen X, Xu L, Huang K, Zhao S, Wang Y

Artificial Intelligence–Aided Diagnosis System for the Detection and Classification of Private-Part Skin Diseases: Decision Analytical Modeling Study

*J Med Internet Res* 2024;26:e52914

URL: <https://www.jmir.org/2024/1/e52914>

doi: [10.2196/52914](https://doi.org/10.2196/52914)

PMID: [39729353](https://pubmed.ncbi.nlm.nih.gov/39729353/)

©Wei Wang, Xiang Chen, Licong Xu, Kai Huang, Shuang Zhao, Yong Wang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 27.12.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.