Applications and Development of Large Language Models in Dermatology

Qingyue Yu

{ qy63@sussex.ac.uk }

Sussex Artificial Intelligence Institute, Zhejiang Gongshang University, Zhejiang, 310000, China

Abstract. Dermatology diseases have multi-faceted and complicated clinical presentations, with the distribution of the available medical resources being skewed, which is very problematic in terms of diagnosis. The development of large visual models with convolutional neural networks and Transformer architectures alongside multimodal large models increases diagnostic challenges in dermatology with new methods of overcoming them due to the rapid growth of deep learning technologies. In this paper, attention is paid to four representative large-scale models, PanDerm, HOT-AI, MF Early Diagnosis, and SAMCL, and a systematic comparison is performed in relation to the dimensions, such as the scale of data, situations where they are applicable, the benefits, and limitations of the models. The analysis shows that both types of models focus on the dissimilarity of generalizability and specialization. To show promising results in few-shot learning, multimodal fusion, and robustness, however, such models usually lack representativeness within data, lack generalization, and adequate clinical validation. The theoretical context of model choice and implementation in practice, by assessing model performance and clinical adaptability, will identically indicate the direction to which technical optimization, clinical integration, and standardization methods should be taken in the future, which has a lot of implications in developing dermatological diagnosis to precision, efficiency, and inclusiveness.

Keywords: Dermatological Diagnosis, Convolutional Neural Networks, Visual Transformer, Multimodal Fusion

1 Introduction

A clinical field that requires a high degree of visual traits, dermatology has more than 2,000 unique diseases whose clinical presentation is complicated and diverse, which makes it difficult to diagnose. At the same time, medical resources are not equally distributed, and they cannot get prompt and correct diagnosis by most of the patients. The recent development of large visual models and multimodal large models has presented new avenues in technology that can be used in the past few years to improve the quality, efficiency and affordability of dermatological diagnosis.

Sulthana et al. introduced the S-MobileNet model, which is aimed at fulfilling the clinical

ICMEEA 2026, April 24-26, Singapore, Singapore

DOI 10.4108/eai.24-4-2026.2364855

requirement of skin lesion classification, and attaining high-quality seven-class classification on the HAM10000 dataset [1]. As a result, SkinGPT-4 by Zhou et al. is based on a visual encoder with Llama-2, which allows the interactive diagnosis of dermatological images [2]. Yanagisawa et al. paid attention to the DeepLabv3+ segmentation model and the Inception-V3 classification model, which were combined with each other to determine atopic dermatitis and atopic dermatitis complications [3]. Despite the fact that there are studies that have investigated the use of large models in dermatological image classification, there are not many systematic reviews. Moreover, shortcomings in current models, e.g. diagnostic difficulty in rare diseases, cross-institutional generalisation, etc., have been insufficiently summarized and critically evaluated.

The paper is a review of the technical features, clinical performance, and main application scenarios of four representative large models, dermatology, in terms of their practical values related to key clinical outcome measures, like diagnostic accuracy and the risk of missed or misdiagnosed cases. This paper will discuss the above analysis to compose the strengths and limitations of current models, the development and research directions that should be taken in the future, and the theoretical underpinnings and technical sources of future studies and clinical practice.

2 Fundamental Theory

2.1 Convolutional Neural Network

Convolutional neural networks(CNN) achieve local feature extraction in images through a combination of convolutional layers, activation layers, pooling layers, and fully connected layers, demonstrating outstanding performance in image classification tasks. The main benefit of CNNs is that they have local connections and a sharing mechanism, which simplifies the complexity of model parameters without much loss of locally important features of skin lesions, including texture and edges. Such benefits render CNNs suitable in determining normal skin lesions in common dermatological complications such as eczema and acne. Nonetheless, CNNs have limited feature extraction capabilities on the basis of their kernel size, and thus, it is difficult to determine global connections among lesions and skin.

2.2 Visual Transformer (ViT)

The self-attention mechanism in the ViT architecture enables the model to adaptively capture correlations between image patches without relying on local constraints, addressing the shortcomings of CNNs in capturing long-range features. This feature enables the ViT architecture to effectively learn global patterns like the morphology and distribution patterns of skin lesions on the skin with high precision so that the architecture can identify diseases that have similar clinical appearances but whose distribution patterns differ. But, ViT architecture requires training data volumes; thus, it is not applicable in the situation with low computing power like in mobile devices.

2.3 Multimodal Fusion

The diagnosis of skin diseases commonly demands the combination of the data related to different sources such as images, medical history, and pathology. Multimodal fusion method combines the information of multiple modalities, and it could be done either at feature level or decision level, which goes beyond the constraints of a single-modality information. Nonetheless, the heterogeneity of modalities in data brings difficulty in designing fusion strategies. It is imperative to synchronously match modalities with information and weight between modalities to improve the performance of models.

3 Introduction to Existing Models

3.1 PanDerm

PanDerm is a multimodal visual foundation model that targets dermatology, i.e., founded on Vision ViT-Large architecture and trained on a sizable dataset by a self-supervised learning paradigm. This model combines masked latent modeling and CLIP feature alignment to combine into a technical principle, which allows it to learn rich, semantic, and generalizable representations of dermatological images without making any manual annotations.

The PanDerm training data is highly extensive and varied, having more than 2.1 million real dermatological images of 11 clinical locations across the world. They include four primary imaging modalities, namely full-body photography, clinical images, dermoscopy images, and skin pathological sections [4], which allows the PanDerm model to be robust to varying imaging conditions and provide rich visual prior knowledge to be used later in tasks of the downstream [5].

PanDerm is the first to combine four fundamentals of dermatological imaging into one model, allowing the use of any of the four modalities to study all other modalities, achieving the ability to fuse results and monitor patient workflow. The best results were seen in PanDerm in 28 different benchmark tests, which include skin cancer errors, laser segmentation, prediction of metastasis, and others. The diagnostic accuracy of PanDerm was higher by 10.2 in the early melanoma detection task compared to clinicians; it increased the diagnostic accuracy of dermatologists by 11 in the task of diagnosing skin cancer and by 16.5 in the task of differentiating 128 types of skin diseases in non-specialist doctors. Moreover, PanDerm outperforms or exceeds the performance of prior special-purpose models with comparatively small fractions of labeled data on most tasks [4]. Another important aspect of medical AI is the limited availability of labeled data in liquidity.

PanDerm has an advantage over traditional models in terms of multimodal fusion and the ability to perform full tasks and it can support 28 clinical tasks, such as skin cancer screening, lesion segmentation, and metastasis prediction. At the same time, the model is characterized by a high speed of calculations and a decrease in training time, which does not require the use of extensive amounts of annotated data. Nevertheless, because PanDerm is implemented with ViT-Large architecture, the large scale of the parameters requires the use of high-performance GPUs to make inferences, which is not feasible to execute in primary-care hospitals with limited resources. Moreover, there is the stability of the model in rare diseases, other than those it has evaluated, which needs to be clinically validated over a long period. Furthermore, more objective measures of fairness of the model and research of possible bias in human-machine cooperation are the aspects of future research.

3.2 Hierarchical-Out of Distribution-Clinical Triage (HOT) AI Model

HOT AI model is a multi-task artificial AI model, which is specifically developed for skin lesions diagnosis. It has convolutional neural networks and Transformer encoder-decoder modules at its core, taking clinical images as the key input. According to an end-to-end hierarchical prediction, the HOT AI model can concurrently provide three important chunks of information: hierarchical diagnosis, out-of-distribution (OOD) detection warning, and clinical triage recommendation. In the case of the triage, the model will combine clinical and dermoscopy images to perform multimodal analysis, which will considerably increase the strength of the diagnosis.

The HOT AI model has been trained and validated on the Molemap large-scale dermatology dataset, which contains 208,287 clinical and dermoscopic images of 141 communities in Australia and New Zealand that concern 78,760 patients [6]. The labels in the data have a hierarchical three tier annotation hierarchy, which has a common long-tail distribution.

The key contribution of the HOT AI model is that, through the multi-task integrated design and clinical-oriented functional modules, the hierarchical prediction can simulate the thinking of clinical doctors and thus avoid the misdiagnosis caused by a single diagnostic output. Meanwhile, the model introduces the Mixed and Prototype Learning (MPL) strategy, which is tailored to address the scarcity issue of middle and tail category data, and meanwhile enhances the OOD detection performance while maintaining/improving the classification accuracy of lesions in distribution [6]. In addition, the triage recommendation module of the HOT AI model can automatically detect the cases that need further dermoscopy examination, and thus optimize the medical procedure in the limited-resource environment and reduce unnecessary image acquisition. Clinical evaluation shows that the diagnostic accuracy of the HOT AI model in triage recommendation cases has been improved [7], and it could identify over 78% target samples to be referral.

The HOT AI model demonstrates significant advantages in enhancing diagnostic transparency and aiding clinical decision-making through multifunctional integration and hierarchical output, particularly for primary triage scenarios involving non-specialist physicians. However, since the training data of the HOT AI model primarily originates from European populations in Australia and New Zealand, a lack of skin tone diversity and metadata annotation may limit its generalization capabilities. Additionally, the performance of the model relies on standardized imaging equipment and remains susceptible to domain shift under non-standard acquisition conditions. Future external validation across diverse populations and real-world clinical studies are necessary to further evaluate the practical application value.

3.3 Early Diagnosis Model of MF

Peking Union Medical College Hospital and Beihang University have jointly developed a multimodal convolutional neural network-based model for the early diagnosis of mycosis fungoides. This model integrates patients' clinical information, clinical images, and dermatoscopic images to enable automated differential diagnosis between early-stage mycosis fungoides (MF) and five common inflammatory skin diseases. The core architecture of the model employs RegNetY400MF as the backbone feature extraction network, processing clinical images and dermatoscopic images separately. Compared with the ResNet18 network used in the team's previous work [8], RegNetY400MF is designed based on neural architecture search technology, achieving a better balance between computational efficiency and feature representation capability.

The data set to be used in constructing the model was obtained between 2016 and 2020 among 1,157 patients with dermatology at Peking Union Medical College Hospital, including 2,452 clinical and 6,550 dermatoscopic images [9], and the clinical data. In preprocessing data, the data augmentation methods were used to improve the models resistance to difference in imaging conditions such as random contrast enhancement and gamma adjustment.

The model of mycosis fungoides early diagnosis is the first systematic use of multimodal learning techniques to diagnose mycosis fungoides at its early stages according to the literature, a gap in the research on this specialized discipline. In a comparative test against 23 dermatologists, the model achieved an independent diagnostic sensitivity of 83.5%, surpassing the average physician accuracy by 8.9 percentage points. When used as an auxiliary tool, it can elevate the sensitivity of physician diagnoses from 74.6% to 86.2% [9]. The model is also shown to have a very high clinical utility, where primary care physicians can avoid the possibility of missing and misdiagnosing mycosis fungoides early in patients, thus ensuring that extended time can be used to treat patients in a primary care setting.

The early diagnosis model of MF uses a lightweight and efficient architecture RegNetY400MF that gives an ideal feature extraction base to the image classification task in a particular disease classification. The model does not necessitate to have a high-end computing equipment, and therefore, it can be used in a clinical implementation. The sample size and representativeness of the model are however limited and the training data is based on a single center, thus this could limit its applicability to other populations and device conditions. Also, interpretability of this model is lacking and to a certain degree, it influences the level of trust between clinicians. It is necessary to establish universality by means of large-scaled prospective studies using a multicenter approach in the future to increase reliability in clinical practice.

3.4 Spatial Alignment Multimodal Contrastive Learning (SAMCL) Model

SAMCL model is a Transformer architecture-based multimodal deep learning model. It can combine clinical images and dermatoscopic images effectively using a spatial alignment module which removes misalignment of features across modalities. At the same time, it works with contrast learning methods to increase the discriminability of its features, which increases diagnostic accuracy on inflammatory skin diseases.

The model was trained and tested with the self-constructed PUMCH-ISD. This data consists of 1,950 clinical photographs and 7,798 dermatoscopic photographs of 1,174 patients with eight types of inflammatory skin diseases [10]. All photos have been taken through highly qualified physicians in standard conditions and diagnosis has been mutually accepted by two senior dermatologists to ascertain the quality and consistency of the data. Also, the model was cross-data tested on the publicly available Derm7pt dataset to determine its ability to generalize.

SAMCL is one of the first, which suggests a spatial alignment mechanism and balanced contrast learning strategy of multimodal dermatological images that solves the inherent problems of inconsistency of scales in multimodal models and imbalanced data. The help of SAMCL model has improved the mean diagnostic accuracy of 20 dermatologists up to 89.0%

compared to 77.5% in the clinical diagnostic testing [10]. In addition, the visual heatmaps of the model show that the attention regions of the model are highly consistent with the lesion areas studied by the physicians, which contributes to the increased readability and clinical validity.

In the analysis of several experiments, the SAMCL model is superior to existing mainstream multimodal methods, and its leverage is that it efficiently incorporates image data of various scales and sources to promote the discriminative ability of the model in complicated inflammatory diseases. But because the data used in the training of the model is mostly Chinese, the generalization power of the model to other skin types of ethnicities is yet to be authenticated. In some diseases where testing is not very difficult to perform a diagnosis, the auxiliary effectiveness of the model is small. The model can also cause misclassifications in rare cases and this can adversely affect clinical decision-making. The further study should focus on the combination of other modalities, which could assist in diagnosis more effectively, such as pathology and genetics.

4 Challenges and Development

Table 1. Multidimensional Comparative Analysis of Four AI Models for Skin Diseases.

Model Dimension	PanDerm	HOT-AI Model	Early Diagnosis Model of MF	SAMCL Model
Core Positioning	Comprehensive Dermatology Workflow	Primary Care Dermatology Triage	Differential Diagnosis of MF and Inflammatory Skin Diseases	Differential Diagnosis of Inflammatory Skin Diseases
Data Scale	Multi-center large-sample studies	Single-institution long-tail samples	Single-center small-sample studies	Single-center medium-sample studies
Modal Coverage	TBP Total Body Photography + Clinical Images + Dermoscopy Images + Pathology Images	Clinical Images + Dermoscopy Images	Clinical Images + Dermoscopy Images + Patient Metadata	Clinical Images + Dermoscopy Images
Key Advantages	Multi-task versatility with superior performance on sparse data	Robust OOD detection with layered outputs enhancing transparency	Lightweight models delivering high MF classification accuracy	High-precision modality fusion with strong discriminative capability
Applicable Scenarios	General Hospital/Cancer Center	Primary Care Facilities	Tertiary Hospital/Specialty Clinic	Tertiary Hospital
Challenges and Limitations	High computational resource requirements;	Inadequate data representativeness; Poor device adaptability	Limited sample size; Insufficient model interpretability	Data limitations; Limited efficacy for certain diseases

	Insufficient coverage of rare diseases; Lack of population equity assessment	

As shown in Table 1, the big models of various forms of dermatological conditions are capable of adjusting to a variety of clinical contexts to create a space of complementary applications. Nevertheless, in general, big models at present experience typical difficulties in the field of dermatology. At the data level, the sample sizes, representativeness, and diversity are still inadequate, and the coverage is not done on various ethnicities and skin types. Such technological problems that models have are excessive consumption of computational resources and a lack of device flexibility in practical implementation. The clinical aspect does not have adequate real-life scenario validation and collaborative clinical assessment that creates a chance of misdiagnosis and lack of interpretability, thus restricting their use in clinical practice [11].

In further development, dermatology large models need to concentrate on the creation of more balanced, multi-centre, and cross-population high-quality datasets to make the model more generalisable and fair. The multimodal fusion mechanisms and space alignment methods need further optimization in order to enhance the discriminative power and strength of the model in few-shot conditions. Clinical collaboration and practical validation should, simultaneously, be more robust in order to foster transparent and interpretable model outputs. More graded and adjusted design and adaptation to the real clinical practice are necessary, to make the application safer and more effective in primary care institutions and special medical institutions.

5 Conclusion

There are four representative large models in the field of dermatology reviewed systematically in this paper, and comparative analyses across various dimensions are done. Findings have shown that these models have specific benefits on data scale, coverage of modalities, and use conditions, and are an ecosystem of applications that cover a variety of clinical requirements with potential synergy. These models, however, also have similar challenges such as lack of data representativeness, lack of generalization abilities, lack of model interpretability, and lack of clinical validation. The importance of this study is that it offers a systematic guided clinical practice and model selection. It outlines the potential of all the types of models to increase the level of effectiveness and accuracy in dermatological diagnosis and reveals a number of general challenges, which lead to the direction of optimization in the future. The future world of development entails further technological integration and evolution of models, further development of clinical applications and system integration to realize more value out of the large models on skin health diagnosis, and eventually more accurate delivery of skin healthy services.

References

[1] Sulthana, R., Chamola, V., Hussain, Z., Albalwy, F., Hussain, A.: A novel end-to-end deep convolutional neural network based skin lesion classification framework. Expert Systems with Applications 246, 123056 (2024)

[2] Zhou, J., He, X., Sun, L., Xu, J., Chen, X., Chu, Y., Gao, X.: Pre-trained multimodal large language model enhances dermatological diagnosis using SkinGPT-4. Nature Communications 15(1), 5649 (2024)

[3] Yanagisawa, Y., Shido, K., Kojima, K., Yamasaki, K.: Convolutional neural network-based skin image segmentation model to improve classification of skin diseases in conventional and non-standardized picture images. Journal of Dermatological Science 109(1), 30–36 (2023)

[4] Yan, S., Yu, Z., Primiero, C., Vico-Alonso, C., Wang, Z., Yang, L., Ge, Z.: A multimodal vision foundation model for clinical dermatology. Nature Medicine, 1–12 (2025)

[5] Mahbod, A., Ecker, R., Woitek, R.: Fusion of Foundation and Vision Transformer Model Features for Dermatoscopic Image Classification. arXiv preprint arXiv:2505.16338 (2025)

[6] Mehta, D., Primiero, C., Betz-Stablein, B., Nguyen, T.D., Gal, Y., Bowling, A., Ge, Z.: Multi-task AI models in dermatology: Overcoming critical clinical translation challenges for enhanced skin lesion diagnosis. Journal of the European Academy of Dermatology and Venereology (2025)

[7] Mehta, D., Betz-Stablein, B., Nguyen, T.D., Gal, Y., Bowling, A., Haskett, M., Ge, Z.: Revamping AI models in dermatology: Overcoming critical challenges for enhanced skin lesion diagnosis. arXiv preprint arXiv:2311.01009 (2023)

[8] Xie, F., Zhao, D., Wang, K., Liu, Z., Wang, Y., Zhang, Y., Liu, J.: Early Mycosis Fungoides Recognition Based on Multimodal Image Fusion. Journal of Data Acquisition and Processing 38(4) (2023)

[9] Liu, Z., Zhang, Y., Wang, K., Xie, F., Liu, J.: Early diagnosis model of mycosis fungoides and five inflammatory skin diseases based on multi-modal data-based convolutional neural network. British Journal of Dermatology, ljaf212 (2025)

[10] Wang, J., Zhang, Y., Xie, F., Liu, J.: Enhancing Diagnosis of Psoriasis and Inflammatory Skin Diseases: A Spatially Aligned Multimodal Model Integrating Clinical and Dermoscopic Images. Journal of Investigative Dermatology (2025)

[11] Khan, S., Ali, H., Shah, Z.: Identifying the role of vision transformer for skin cancer—A scoping review. Frontiers in Artificial Intelligence 6, 1202990 (2023)