A Multimodal Deep Learning-based Platform that Assists in the Classification of Lupus Subtypes and Differential Diagnosis Among Lupus and Other Similar Skin Diseases

Qianjin Lu (qianlu5860@pumcderm.cams.cn)
The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Qianwen Li
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Zhi Yang
Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University

Kaili Chen
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Ming Zhao
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Hai Long
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Yueming Deng
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Haoran Hu
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Chen Jia
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Meiyu Wu
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics

Zhidan zhao
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics
Huan Zhu
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics
Suqing Zhou
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics
Mingming Zhao
Department of Dermatology, The Second Xiangya Hospital of Central South University; Hunan Key Laboratory of Medical Epigenomics
Yiqun Jiang
Institute of Dermatology, Chinese Academy of Medical Sciences and Peking Union Medical College
Wei Zhang
Institute of Dermatology, Chinese Academy of Medical Sciences and Peking Union Medical College
Bo Zhang
Second Xiangya Hospital of Central South University
Yanling HE
Department of Dermatology, Beijing Chao-yang Hospital, Capital Medical University
Liwei RAN
Department of Dermatology, Beijing Chao-yang Hospital, Capital Medical University
Chunlei Zhang
Department of Dermatology, Peking University Third Hospital
Wenting Wu
Department of Dermatology, Peking University Third Hospital
Quzong Suolang
Department of Dermatology, People's Hospital of Tibet Autonomous Region
Hanhuan Luo
Department of Dermatology, People's Hospital of Tibet Autonomous Region
Xiaojing Kang
Department of Dermatology, People's Hospital of Xinjiang Uygur Autonomous Region
Caoying Wu
Department of Dermatology, People's Hospital of Xinjiang Uygur Autonomous Region
Hongzhong Jin
Department of Dermatology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College
Lei Chen

Department of Dermatology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College

Qing Guo

Department of Dermatology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Guangji Gui

Department of Dermatology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Shanshan Li

Department of Dermatology, The First Bethune Hospital of Jilin University

Henan Si

Department of Dermatology, The First Bethune Hospital of Jilin University

Shuping Guo

Department of Dermatology, The First Hospital of Shanxi Medical University

Hongye Liu

Department of Dermatology, The First Hospital of Shanxi Medical University

Xiguang Liu

Department of Dermatology, The Hei Long Jiang Provincial Hospital

Guozhang Ma

Department of Dermatology, The Hei Long Jiang Provincial Hospital

Danqi Deng

Second Affiliated Hospital of Kunming Medical College

Limei Yuan

Second Affiliated Hospital of Kunming Medical College

Jianyun Lu

Department of Dermatology, The Third Xiangya Hospital, Central South University

Jinrong Zeng

Department of Dermatology, The Third Xiangya Hospital, Central South University

Xian Jiang

Department of Dermatology, West China Hospital, Sichuan University

Xiaoyan Lyu

Department of Dermatology, West China Hospital, Sichuan University

Liuqing Chen

Department of Dermatology, Wuhan No.1 Hospital

Bin Hu

Department of Dermatology, Wuhan No.1 Hospital

Juan Tao

Union Hospital, Tongji Medical College, HUST

Yuhao Liu

Union Hospital, Tongji Medical College, HUST

Gang Wang

Department of Dermatology, Xijing Hospital, Fourth Military Medical University, Xi'an
https://orcid.org/0000-0002-5842-8080

Guannan Zhu
University of California, Davis

Zhirong Yao
Department of Dermatology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

Qianyue Xu
Department of Dermatology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

Bin Yang
Dermatology Hospital of Southern Medical University

Yu Wang
Dermatology Hospital of Southern Medical University

Yan Ding
Hainan Provincial Hospital of Skin Disease

Xianxu Yang
Hainan Provincial Hospital of Skin Disease

Kai Hu
Xiangtan University

Haijing Wu
Hunan Key Laboratory of Medical Epigenomics, The Second Xiangya Hospital of Central South University

---

Article

Keywords: Lupus erythematosus, deep learning, multimodal DLS, multi-immunohistochemistry

Posted Date: October 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-939556/v1

Abstract

Lupus erythematosus (LE) has the reputation of being *the Great Imitator* due to its various cutaneous manifestations, which often lead to difficulties in accurate diagnosis and classification. Even with the assistance of deep learning systems (DLSs), clinical skin image-based single modalities still experience difficulty distinguishing subtypes of LE. Here, we present an application of multimodal DLS (MMDLS), in which the training database consists of clinical skin images, multi-immunohistochemistry (multi-IHC) images and the 2019 European League Against Rheumatism/American College of Rheumatology (EULAR/ACR) scores, to assist in the classification of LE subtypes (discoid LE (DLE), annular subacute cutaneous LE (A-SCLE), papulosquamous subacute cutaneous LE (P-SCLE), systemic LE (SLE)) and to aid in the differential diagnosis of other similar skin diseases. A total of 386 cases with 580 clinical skin images, 3380 multi-IHC images and 2019 EULAR/ACR scores from 25 institutions in China were included, and EfficientNet-B3 and ResNet-18 were utilized in this study. By comparing the single-modal and dual-modal DLS performances in multiple classifications, we demonstrate the superiority of the MMDLS (Ave-Sen = 0.9116, Ave-Spe = 0.9921, Ave-Pre = 0.9281). Moreover, the performance of MMDLS on 13 classifications was superior to that of dermatologists and pathologists (F1 score: dermatologists = 0.4471, pathologists = 0.6691, MMDLS = 0.9582) with an average area under the curve (AUC) of 0.9956. These results highlight the potential of the combined application of MMDLS and multi-IHC images to assist pathologists and dermatologists in diagnosing LE subtypes and similar skin diseases.

Introduction

Lupus erythematosus (LE) is an autoimmune disease that exhibits widespread tissue damage in the affected organs that manifest as life-threatening lung or renal injuries. Approximately 80% of LE patients have skin involvement, and most patients exhibit skin lesions from first onset or during the course of disease [1]. LE skin lesions are highly variable and heterogeneous and can be divided into three main subtypes: discoid lupus erythematosus (DLE), subacute cutaneous LE (SCLE), and acute cutaneous LE (ACLE)[2]. Clinically, LE patients with different skin lesions show diverse disease development and prognoses. Patients with DLE lesions are characterized by scarring and disfiguring skin lesions with rare systemic involvement. ACLE lesions represent the primary skin features of SLE patients and are characterized by reversible skin damage and multiple organ impairment. SCLE lesions can be further divided into papulosquamous SCLE (P-SCLE) with high risk progression to SLE and annular SCLE (A-SCLE), which represents a stable clinical condition [3]. The treatment varies based on lupus subtypes, and early intervention can potentially stop the progression of cutaneous lupus into systemic lupus [4,5]. Therefore, accurate classification of LE is important for dermatologists and rheumatologists to develop appropriate therapeutic strategies for patients. However, an efficient tool for LE subtype classification is lacking.

In addition, LE skin lesions are easily confused with many other skin diseases, such as lichen planus (LP), erythema annulare centrifugum (EAC), rosacea (Ros), erythema multiforme (EM), psoriasis (Pso), eczema

(Ecz), vasculitis (Vas) and dermatomyositis (DM); this confusion has caused LE to be referred to as the *great imitator* [6–14]. The heterogeneity and imitating features of LE represent difficulties and challenges in determining an accurate diagnosis and classification of LE.

At present, LE diagnosis and classification is mainly based on comprehensive evaluation from many aspects, including the first clinical impression, pathological analyses and laboratory tests [15]. Both are challenging for most inexperienced dermatologists or pathologists, which leads to low diagnostic accuracy along with delayed or improper treatments. In a study of clinical and pathological diagnoses of 29987 skin biopsies, only 43.94% (406/924) of patients with clinically diagnosed LE were eventually confirmed to have LE skin lesions, suggesting a high misdiagnosis ratio in real-world clinical work [16]. Although pathological biopsy is generally considered the cornerstone of diagnosis for cutaneous LE, a deficiency in sensitivity has been reported in a previous study [17]. This evidence suggests that a more effective tool to assist dermatologists and pathologists in diagnosing LE subtypes and similar skin diseases and subsequently reduce the misdiagnosis rate is urgently needed.

The use of artificial intelligence tools represents an effective method to address the abovementioned problems in LE diagnosis. Recent advances in deep learning have facilitated the development of artificial intelligence tools to assist in diagnosing skin disorders based on images [18]. Most current studies on DLS are single models that use history, clinical skin images, or pathological images, such as hematoxylin and eosin (HE) staining [19,20]. In a recent study, a high confusion rate of Pso with Ecz was reported given the similar appearance of these skin lesions. Therefore, the diagnostic efficiency of single-model DLS of clinical skin images may be limited by the *imitating nature* of LE skin lesions [21]. Given the lack of characteristic and well-directed pathological changes in LE lesions, the application of single-model DLS of HE staining images in LE may not be as effective as its application in cancer tissue [22]. Considering that the differences in immune cell components is an important factor in the differential diagnosis of lupus from other diseases, we have developed a skin in situ test kit to provide more information about the in situ distributions and quantities of immune cells in LE lesions, and this method exhibits enhanced diagnostic efficiency compared with HE staining [17,23–25].

Here, we further improved our previous skin in situ test kit from duplex detection ($CD4^+$ and $CD19^+$) to quadruple detection (CD4+, CD8+, CD19+, and CD11b+), which provides more details for various diseases (Fig. 1). Our results verify the diagnostic advantage of multi-IHC image-based DLS compared to clinical skin image-based DLS. Combining clinical skin images, multi-IHC images, and the 2019 EULAR/ACR score, we first developed an MMDLS for LE diagnosis and classification. The superiority of the MMDLS is demonstrated by the comparison with single-modal DLS performance in LE diagnosis. Moreover, our MMDLS shows a better performance on 13 classification diagnoses compared with dermatologists and pathologists. Based on these efforts, we hope to improve the total classification accuracy for LE to provide benefits for early intervention and individualized therapies.

Results

Data preparation and development of the MMDLS program.

Our overall program mainly consists of two parts. The first part involves the input of multimodal data, including three datasets consisting of clinical skin images, multi-IHC images and modified 2019 EULAR/ACR scores. Initially, we collected 884 cases with a diagnosis of suspected lupus from 25 hospitals in China. The final diagnosis of all cases was further confirmed based on the diagnosis of two senior dermatologists and follow-up clinical information. Any case with incomplete data or with a final diagnosis beyond the 13 classifications was excluded. The final dataset included 386 cases with 580 clinical skin images, 3380 multi-IHC images and clinical data (Extended Data Table 1).

For the clinical skin image dataset, we retrospectively collected images of skin lesions on the face, trunk, and limbs captured by smartphones. For each case, we ensured that the images were clear and representative for training and testing procedures. Then, we performed multi-IHC staining of CD4, CD8, CD11b, and CD19 on the skin sections. In addition, the 2019 EULAR/ACR score is included in this study because the diagnosis of SLE in particular demands the comprehensive evaluation of overall clinical data. To expand the differences among LE subtypes and other similar skin diseases, we obtained the 2019 EULAR/ACR score at 100× magnification. Therefore, in this study, the 2019 EULAR/ACR score was modified and utilized for all SLE and non-SLE cases to evaluate and quantify systemic involvement.

The second part includes the algorithm and construction of the MMDLS. To prepare the MMDLS, we randomly selected 310 cases from all the cases used for training, and 76 of the remaining cases served as an independent testing dataset (training vs. testing = 8:2). Notably, for pattern recognition, a suitable feature encoder plays a significant role in the diagnosis results. Currently, advanced feature encoders include VGG-16, Inception-V3, ResNet-18 and EfficientNet-B3, and these encoders are widely used in feature extraction and pattern classification[26]. This study selected the most efficient feature extractor for images of different modalities according to the training time, accuracy, and number of parameters. The results showed that EfficientNet-B3 exhibited the highest accuracy of multi-IHC images (top-1 accuracy=77.21%, Extended Data Figure 1) and the lowest number of parameters (40.88 MB). For clinical skin images, ResNet-18 achieved the highest accuracy (top-1 accuracy=70.76%, Extended Data Figure 1). In addition, ResNet-18 showed the shortest training time of only 428 seconds, whereas other networks required at least 700 seconds to achieve similar results. Finally, we employed EfficientNet-B3 and ResNet-18 for deep feature extraction of multi-IHC images and clinical skin images, respectively. The 2019 EULAR/ACR scores were then integrated into the diagnostic system by binding neurons to the fully connected layer. The comparison of the different encoders is detailed in Extended Data Figure 1, and the overall workflow is shown in Figure 2.

Comparison of MMDLS performance with single-modal DLS and dual-modal DLS

To obtain a qualified MMDLS, we conducted 10 repeated experiments with training dataset, and testing dataset were used to evaluate the capacity of each MMDLS. The testing dataset with the highest classification accuracy composed of 76 cases was chosen to generate the final MMDLS and perform the comparison experiment (Fig. 3). To verify the superiority of MMDLS, we compared the overall performance of MMDLS with single-modal DLS including clinical skin image modality, multi-IHC image modality, and dual-modal DLS (i.e., clinical skin image and 2019 EULAR/ACR score; multi-IHC image and 2019 EULAR/ACR score; and clinical skin image and multi-IHC image).

First, the efficiencies of different DLS models were compared based on the average sensitivity, specificity, and precision using the testing dataset (Fig. 3a). By comparison, we found that the data of the 3 abovementioned evaluation indices steadily improved when the modalities were used in combination. Of all the models, the MMDLS developed based on the fusion of 3 modalities showed the best performance (Ave-Sen=0.9116, Ave-Spe=0.9921, Ave-Pre=0.9281), revealing the advanced efficiency of MMDLS in classification. The multiclassification models exhibited specificities ranging from 0.9751 to 0.9921, and the differences among these models were insignificant. The average sensitivity and precision of MMDLS showed the most significant progress compared with single- and dual-modal DLS models, suggesting a low misdiagnosis rate and high accuracy of MMDLS in the multiclassification task.

To visually assess the differences, we further applied the ROC curve to evaluate the classification efficiency of all DLS models (Fig. 3b). The ROC curve is universally used in diagnostic tests and is estimated by the area under the ROC curve (AUC), which is consistent with previous indices. In the comparison of different DLS models, we found that the MMDLS still had the highest AUC of 0.9956. From the perspective of deep learning, we visualized the advantages of the multimodal strategy by box plots, which are generated from the values of top-1 accuracy in the repeated experiments (n =10). The box plot simultaneously highlights the diagnostic accuracy and the stability of different models. As shown in Figure 3c, in single-modal DLS, the top-1 accuracy of the multi-IHC image modality was greater and the distribution was more concentrated (IHC: top-1 acc=69.00±1.13%) compared with that of the clinical skin image modality. In other words, outliers and variable performances were more frequently noted in the clinical skin image modality. Given that the multi-IHC image showed better diagnostic accuracy and better stability as a single-modal DLS, we then fused the 2019 ELUAR/ACR score into the multi-IHC image. The average top-1 accuracy was increased by 3.6% (IHC+ACR: top-1 acc=80.73±0.45%). Together, the MMDLS exhibited the highest accuracy and the best stability (CLI+IHC+ACR: top-1 acc=91.08±0.44%). This finding indicates that single or dual models are inadequate for decision-making.

Classification efficiency of 13 categories by proposed MMDLS

In this section, we employed the same testing dataset mentioned in the comparison of MMDLS with single-modal DLS and dual-modal DLS. To identify the special features of different models in specific diseases, we compared the classification parameters for 13 categories in detail.

Here, the F1 score was used to compare the diagnostic and classifying efficiency and accuracy of the modalities for each subtype (Fig. 4a, Extended Data Table 2). For the clinical skin image modality, the classification efficiency for 4 LE subtypes was unsatisfactory (average F1 score=0.4583), which is largely due to the similar appearances between LE skin lesions and other skin diseases. In contrast, the multi-IHC imaging modality exhibits remarkable predominance in diagnosing the LE subtypes, especially for DLE and P-SCLE, for which the F1 score was increased by 0.4645 (101.84%) and 0.4728 (224.61%), respectively. Although the diagnostic efficiency of the multi-IHC modality was superior to that of the clinical skin image modality in general, the clinical skin image modality had its own advantages in the EM, Ros, DM, and Ecz groups. Regarding the MMDLS, the results show that the F1 score of all 13 categories increased dramatically. Compared with the clinical skin image modality, the F1 score of MMDLS for DLE, A-SCLE, P-SCLE and SLE obviously improved (DLE: from 0.4561 to 0.9844, A-SCLE: from 0.5287 to 0.8644, P-SCLE: from 0.2105 to 0.7961, SLE: from 0.6378 to 0.8765, Fig. 4a). The F1 scores of EAC, Vas, DM, LP and EM experienced high rates of growth from clinical skin images to MMDLS, ranging from 0.1248 (15.08%) to 0.2773 (39.46%). Overall, the MMDLS showed the best performance on the diagnosis and classification of all 13 categories (average F1 score=0.9128).

Classification is visualized based on the t-SNE plots shown in Figure 4b-e. In t-SNE plots, the more convergent and independent the point clouds are, the easier it is to distinguish the diseases.
For the clinical skin image modality, the t-SNE plot showed uniform and scattered points, and most of the 13 categories could not be accurately grouped. Although the dots of the multi-IHC image modality showed a clustering tendency, some groups could not be properly separated from other diseases, especially A-SCLE and parts of P-SCLE, SLE, EM and Ecz. After the inclusion of the 2019 EULAR/ACR score, the t-SNE graph showed a more obvious clustering distribution of approximately 13 categories, but the differentiation of SLE, A-SCLE and P-SCLE remained challenging. In MMDLS, the data of 13 categories were positioned relatively far away from each other, demonstrating the advanced diagnostic ability of the algorithm. MMDLS performed well in distinguishing the four LE subtypes, especially DLE and SLE. Our results showed that the clustering result constantly improved with the fusion of multimodal information. However, although we made progress compared to previous studies[21,27], the differentiation of A-SCLE and P-SCLE remains challenging.

A confusion matrix was also constructed to further demonstrate the performance of MMDLS (Fig. 4f). First, in the clinical skin image modality, the classification efficiency of P-SCLE was

unsatisfactory, and it was easily misjudged as DLE (57%), SLE (14%), or Ecz (16%). In contrast, in the multi-IHC imaging modality, the misdiagnosis rate of P-SCLE was significantly reduced from 88% to 27% (Extended Figure 2). Finally, in MMDLS, the accuracy of P-SCLE increases to 71%, which is far higher than that noted in other DLS models. The diagnostic accuracy of DLE in the multi-IHC imaging modality was also greater than that obtained with the clinical skin imaging modality (approximately 41% higher). In MMDLS, the accuracy of DLE was up to 98%. As shown in Figure 4f, the MMDLS exhibited the greatest ability to distinguish conditions based on all 13 groups, achieving an average accuracy of 92.54%.

Comparison of MMDLS to dermatologists and pathologists.

To further evaluate the performance of MMDLS in the real world, 76 cases in the testing dataset with the best diagnostic accuracy were selected, and we retrospectively collected the clinical diagnosis and pathological diagnosis of each case. Here, the clinical diagnosis refers to the first impression diagnosis based on features of skin lesions and clinical manifestations. Pathological diagnosis was made by pathologists based on HE staining. The MMDLS diagnosis is also modified to one diagnosis for each case. The gold standard for diagnosis, which acts as a true label, is the diagnosis made by two senior dermatologists who reached a consensus based on all clinical data.

To compare the MMDLS performance with that of the dermatologists and pathologists, we evaluated the average sensitivity, specificity, precision and F1 score (Fig. 5.a). Considering the nature of multiclassification models, the specificities were insignificantly different. Thus, further discussion is unnecessary. Compared with the clinical diagnosis and pathological diagnosis, MMDLS achieved marked increases in sensitivity (Sen: dermatologists=0.4033 pathologists=0.6516, MMDLS=0.9542), indicating that MMDLS can effectively reduce misdiagnosis. Regarding precision, our results suggested that the clinical diagnosis (Pre=0.5022) and pathological diagnosis (Pre=0.6876) appear to have relatively low precision, whereas the precision of MMDLS can reach 0.9622, further illustrating the high diagnostic accuracy of MMDLS. As our findings show, the classifying ability of MMDLS is superior to that of dermatologists and pathologists (F1 score: dermatologists=0.4471, pathologists=0.6691, MMDLS=0.9582). MMDLS maintained an overall performance advantage compared to dermatologists and pathologists (Fig. 5a).

In addition, the accuracy of the classification result was further evaluated based on the confusion matrix, which intuitively displays the performance of MMDLS, dermatologists and pathologists based on 13 classifications (Fig. 5.b). Notably, in clinical practice, an individual and definite diagnosis is unable to be made for many cases, or it is difficult for clinicians to determine the subtype. Dermatologists usually diagnose lupus skin lesions as “LE” instead of clearly classifying the lesion. Therefore, we added several

prediction labels, including undefined diagnosis (UD), other skin diseases that do not belong to the 13 categories (OSD), and uncertain lupus subtype (ULE), to the clinical and pathological diagnosis. We used this method to obtain a diagnosis that was close as possible to the actual clinical diagnosis. As shown in Figure 5b, up to 62.96% of LE patients could not be accurately diagnosed in the clinical diagnosis. Some cases were misdiagnosed as uncertain lupus subtype or other skin diseases. In other cases, multiple "probable" diagnoses were taken into account. Although pathological diagnosis is regarded as the gold standard for diagnosing LE skin lesions, only 44.44% of LE skin samples could be correctly classified by pathologists. In contrast, both in the diagnosis and classification of LE, MMDLS exhibited the best performance, achieving an average accuracy of 88.98%. With the exception of A-SCLE, the diagnosis correction rate for other groups approached or even achieved 100%.

Discussion

In this study, we aimed to solve the difficulties in the diagnosis and classification of LE subtypes and other skin diseases confused with LE. Our previous research identified the different morphological features of $CD4^+$ and $CD19^+$ cells in skin tissues from LE patients with different skin lesions [17]. Here, we further improved our previous skin in situ test kit by expanding it from duplex detection (CD4 and CD19) to quadruple detection (CD4, CD8, CD19, and CD11b) for more comprehensive observation of different cell types in various skin diseases. This multi-IHC staining panel is able to locate most of the inflammatory cells and therefore offers greater potential to assist pathologists and dermatologists in differentiating skin diseases from a novel perspective.

Given the substantial development in artificial algorithms in recent years, we hope that skin in situ multi-IHC staining, a new and effective method for the identification and typing of CLE, will represent a better approach with the help of AI algorithms to extract the key features for diagnosis. However, in clinical practice, the diagnosis of skin disease, particularly LE subtypes, should never rely on a single skin examination. Therefore, we developed a new multimodal algorithm called MMDLS, which includes clinical skin images, multi-IHC images, and 2019 EULAR/ACR scores. The characteristics of multi-IHC images combined with the pictures of patients' skin lesions and the 2019 EULAR/ACR scores constitute a multimodal diagnostic system for a comprehensive diagnosis and an attempt to imitate the diagnostic process of a real dermatologist. In this algorithm, the EfficientNetB3 network and the RestNet18 network are used for feature extraction from the multi-IHC image modality and the clinical skin image modality, respectively.

First, we compared the performance of single-modal DLS, dual-modal DLS and MMDLS based on different combinations of multi-IHC images, clinical skin images and 2019 EULAR/ACR scores for all 13 classifications. Compared with the clinical skin image modality, the multi-IHC image modality exhibited better performance for the differential diagnosis of LE subtypes, suggesting that the feature extraction of in situ immune cells might represent an effective tool to discriminate lesions based on appearance. Additionally, the MMDLS (AUC = 0.9956, top-1 accuracy = 91.08 ± 0.44%, Pre = 0.9281) outperformed any of the other combinations in the multiple classification task.

Our multiple classification task examined 13 subgroups consisting of 4 LE subtypes (SLE, A-SCLE, P-SCLE, and DLE), 8 disease controls (LP, EAC, Pso, DM, Vas, EM, Ecz, and Ros) and a healthy control to better assess the performance of DLS. The MMDLS offers advanced performance in the differential diagnosis among disease controls and LE subtypes given that the classification accuracy ranges from 89–100%. Additionally, MMDLS precisely distinguishes DLE and SLE with a classification accuracy of greater than 98%. However, the classification of A-SCLE and P-SCLE remains difficult. The confusion between P-SCLE and SLE could be attributed to the clinical and pathological overlap, as some patients with SCLE have a tendency to develop SLE. Additionally, the limited number of patients with SCLE used for algorithm training may also explain the challenges encountered for the differentiation of LE subtypes.

It is difficult to distinguish LE subtypes, especially A-SCLE and P-SCLE, based on the first clinical impression. A-SCLE is typically misdiagnosed as EAC or other skin diseases, such as granuloma annulare. In contrast, P-SCLE is always diagnosed as suspected LE, but clinicians cannot easily decide which subtype it belongs to. Although pathological examination is considered the gold standard for many skin diseases, the similar manifestations of LE subtypes on H&E staining consistently causes problems for pathologists. Compared to the clinical diagnosis and pathological diagnosis, MMDLS exhibits the highest diagnostic accuracy, especially in the classification of LE subtypes. Hence, MMDLS exhibits potential clinical utility.

In this study, we tried that the number of cases in 13 categories was balanced. However, the number of clinical skin images and multi-IHC images are quite difficult to balance because the number of skin lesions is generally less than the number of views you can capture under the microscope. In a previous study that included 9241 clinical images, the diagnostic accuracies of SLE, SCLE, and DLE were 65.9%, 54.6% and 72.9%, respectively [27]. Although we did not include a plethora of clinical images, our conclusion of clinical skin image-based single-modal imaging obtained similar results, whereas the MMDLS performed much better.

in this study, we first tried a multimodal system that showed good performance in the diagnosis and classification of LE. As the first study to integrate clinical information and multi-IHC images, there are some limitations due to the limited sample size and incomplete clinical information. In addition, algorithms and AI systems can be further modified in the future. However, our study is a good start to advance the use of AI and medicine as part of a much deeper collaboration, and the real “precise diagnosis” might be obtained in the near future.

Declarations

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 81830097, 81972943, 81773332, and 82173425), Key Research and Development Program of Hunan Province (2018XK2304), Hunan Talent Young Investigator (No. 2019RS2012), Hunan Outstanding Young

Investigator (No. 2020JJ2055), CAMS Innovation Fund for Medical Sciences (CIFMS) (2019-I2 M-5–033), and the Nonprofit Central Research Institute Fund of the Chinese Academy of Medical Sciences (No. 2020-RC320–003).

Author contributions

Qianwen Li and Kaili Chen performed the IHC staining, collected and arranged data for usage and worked on manuscript writing. Zhi Yang performed the experiments, analyzed the results, developed the MMDLS and helped with manuscript writing. Kai Hu, Haijing Wu, and Qianjin Lu initiated the project, guided the study and revised the manuscript. Ming Zhao provided administrative support. Hai Long offered clinical expertise and guidance. Yueming Deng, Haoran Hu, Chen Jia, Meiyu Wu helped with image processing and data collections. Zhidan Zhao, Suqing Zhou, Huan Zhu, Mingming Zhao helped search the medical history of patients. Yiqun Jiang, Bo Zhang, Wei Zhang, Yanling HE, Liwei RAN, Chunlei Zhang, Wenting Wu, Suolangquzong, Hanhuan Luo, Xiaojing Kang, Caoying Wu, Hongzhong Jin, Lei Chen, Qing Guo, Guangji Gui, Shanshan Li, Henan Si, Shuping Guo, Hongye Liu, Xiguang Liu, Guozhang Ma, Danqi Deng, Limei Yuan, Jianyun Lu, Jinrong Zeng, Xian Jiang, Xiaoyan Lyu, Liuqing Chen, Bin Hu, Juan Tao, Yuhao Liu, Gang Wang, Guannan Zhu, Zhirong Yao, Qianyue Xu, Bin Yang, Yu Wang, Yan Ding, and Xianxu Yang provided data for multicenter cases.

Competing interests

The authors have no conflicts of interest to declare.

References

1. Wenzel, J. Cutaneous lupus erythematosus: new insights into pathogenesis and therapeutic strategies. *Nat Rev Rheumatol* **15**, 519–532 (2019).
2. Li, Q., *et al.* A comprehensive review of immune-mediated dermatopathology in systemic lupus erythematosus. *J Autoimmun* **93**, 1–15 (2018).
3. Li, Q., Wu, H., Zhou, S., Zhao, M. & Lu, Q. An Update on the Pathogenesis of Skin Damage in Lupus. *Curr Rheumatol Rep* **22**, 16 (2020).
4. Joseph, A.K., Abbas, L.F. & Chong, B.F. Treatments for disease damage in cutaneous lupus erythematosus: A narrative review. *Dermatol Ther*, e15034 (2021).
5. Cooper, E.E., Pisano, C.E. & Shapiro, S.C. Cutaneous Manifestations of "Lupus": Systemic Lupus Erythematosus and Beyond. *Int J Rheumatol* 2021, 6610509 (2021).

6. Grabbe, S. & Kolde, G. Coexisting lichen planus and subacute cutaneous lupus erythematosus. *Clin Exp Dermatol* **20**, 249–254 (1995).
7. Lospinoso, D.J., Fernelius, C., Edhegard, K.D., Finger, D.R. & Arora, N.S. Lupus erythematosus/lichen planus overlap syndrome: successful treatment with acitretin. *Lupus* **22**, 851–854 (2013).
8. Chander, R., Yadav, P., Singh, A. & Nangia, A. Systemic lupus erythematosus presenting as erythema annulare centrifugum. *Lupus* **23**, 1197–1200 (2014).
9. Chasset, F., *et al.* Rare diseases that mimic Systemic Lupus Erythematosus (Lupus mimickers). *Joint Bone Spine* **86**, 165–171 (2019).
10. Shteyngarts, A.R., Warner, M.R. & Camisa, C. Lupus erythematosus associated with erythema multiforme: does Rowell's syndrome exist? *J Am Acad Dermatol* **40**, 773–777 (1999).
11. Gunawan, H., Awalia, A. & Soeroso, J. The Coexistence of Systemic Lupus Erythematosus and Psoriasis: Is It Possible? *Acta Med Indones* **50**, 144–150 (2018).
12. Higashi, N. & Kawana, S. Atopic eczema complicated by systemic lupus erythematosus. *Eur J Dermatol* **15**, 500–502 (2005).
13. Calamia, K.T. & Balabanova, M. Vasculitis in systemic lupus erythematosis. *Clin Dermatol* **22**, 148–156 (2004).
14. Schempp, C.M., Schauer, F., Huhn, C.K., Venhoff, N. & Finzel, S. Skin inflammation associated with arthritis, synovitis and enthesitis. Part 2: rheumatoid arthritis, reactive arthritis, Reiter's syndrome, Lyme borreliosis, dermatomyositis and lupus erythematosus. *J Dtsch Dermatol Ges* **17**, 167–181 (2019).
15. Fanouriakis, A., Tziolos, N., Bertsias, G. & Boumpas, D.T. Update omicronn the diagnosis and management of systemic lupus erythematosus. *Ann Rheum Dis* **80**, 14–25 (2021).
16. Analysis of clinical and pathological diagnoses of 29 987 skin biopsy samples in Peking Union Medical College Hospital. *Chinese Journal of Dermatology* **53**(2020).
17. Xu, Y., *et al.* The point-of-care-testing of nucleic acids by chip, cartridge and paper sensors. *Chinese Chemical Letters* (2021).
18. Liu, Y., *et al.* A deep learning system for differential diagnosis of skin diseases. *Nat Med* **26**, 900–908 (2020).
19. Esteva, A., *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
20. Hashimoto, N., *et al.* Multi-scale Domain-adversarial Multiple-instance CNN for Cancer Subtype Classification with Unannotated Histopathological Images. 3851–3860 (2020).
21. Wu, H., *et al.* A deep learning, image based approach for automated diagnosis for inflammatory skin diseases. *Ann Transl Med* **8**, 581 (2020).
22. Hollon, T.C., *et al.* Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nature medicine* **26**, 52–58 (2020).

23. Li, Q., *et al.* A skin in situ immune cell detection kit for the diagnosis and classification of cutaneous lupus erythematosus. *Annals of Translational Medicine* **9**, 1062–1062 (2021).
24. Hussein, M.R., Aboulhagag, N.M., Atta, H.S. & Atta, S.M. Evaluation of the profile of the immune cell infiltrate in lichen planus, discoid lupus erythematosus, and chronic dermatitis. *Pathology* **40**, 682–693 (2008).
25. Brown, T.T., Choi, E.Y., Thomas, D.G., Hristov, A.C. & Chan, M.P. Comparative analysis of rosacea and cutaneous lupus erythematosus: histopathologic features, T-cell subsets, and plasmacytoid dendritic cells. *J Am Acad Dermatol* **71**, 100–107 (2014).
26. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science* (2014).
27. Wu, H., *et al.* A deep learning-based smartphone platform for cutaneous lupus erythematosus classification assistance: Simplifying the diagnosis of complicated diseases. *Journal of the American Academy of Dermatology* (2021).

Figures

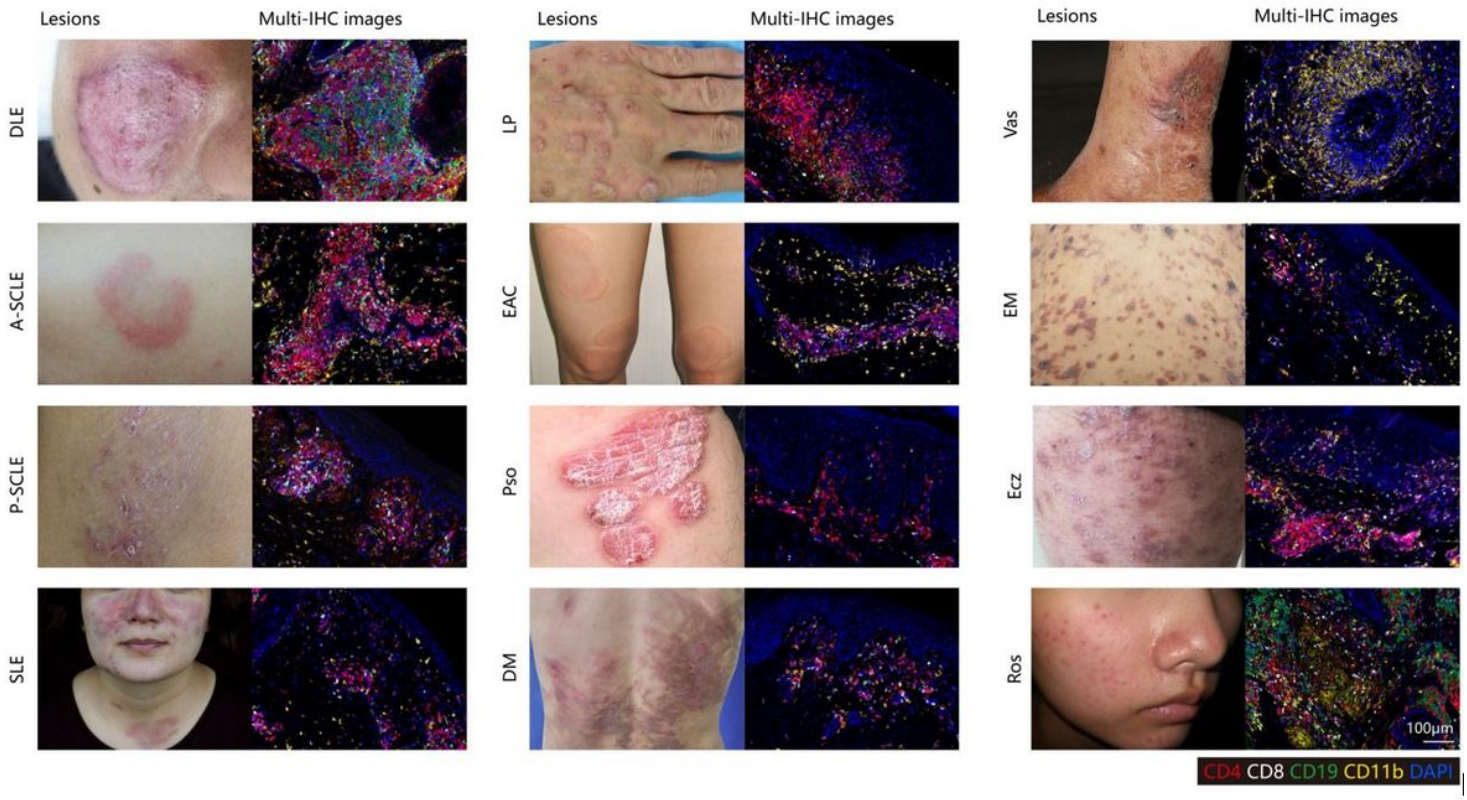

Figure 1

Typical skin images and multi-IHC images in samples from patients with lupus and other similar skin diseases. For each IHC image, positive staining of CD4 are shown in red, CD8 in white, CD11b in yellow, CD19 in green and DAPI in blue.

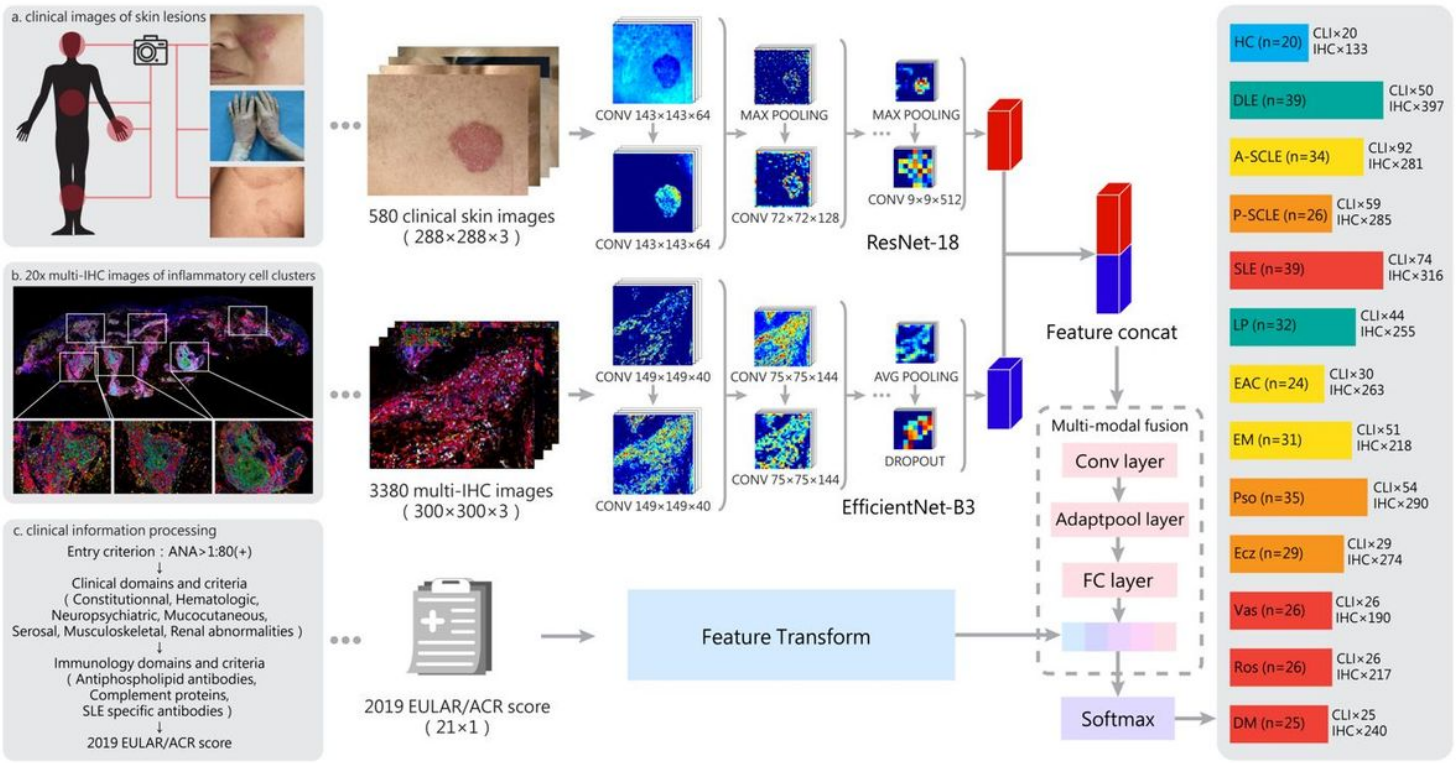

Figure 2

The overall workflow of the data preparation and the construction of MMDLS architecture. Conv stands for convolution operation, Ave pooling, Max pooling, and Adaptpool refer to pooling operations in different ways, Dropout meaning a regularization operation, FC is fully connected layer, and Softmax is an activation function. For each patient, the input data includes clinical skin images (CLI), multi-IHC images (IHC) and 2019 EULAR/ACR scores for a comprehensive assessment. The DLS employed the ResNet-18 algorithm to extract the deep features for the skin image modality and the EfficientNet-B3 algorithm for the multi-IHC image modality. The fused features of skin and multi-IHC images are then combined with the 2019 EULAR/ACR scores modality for a final automatic classification

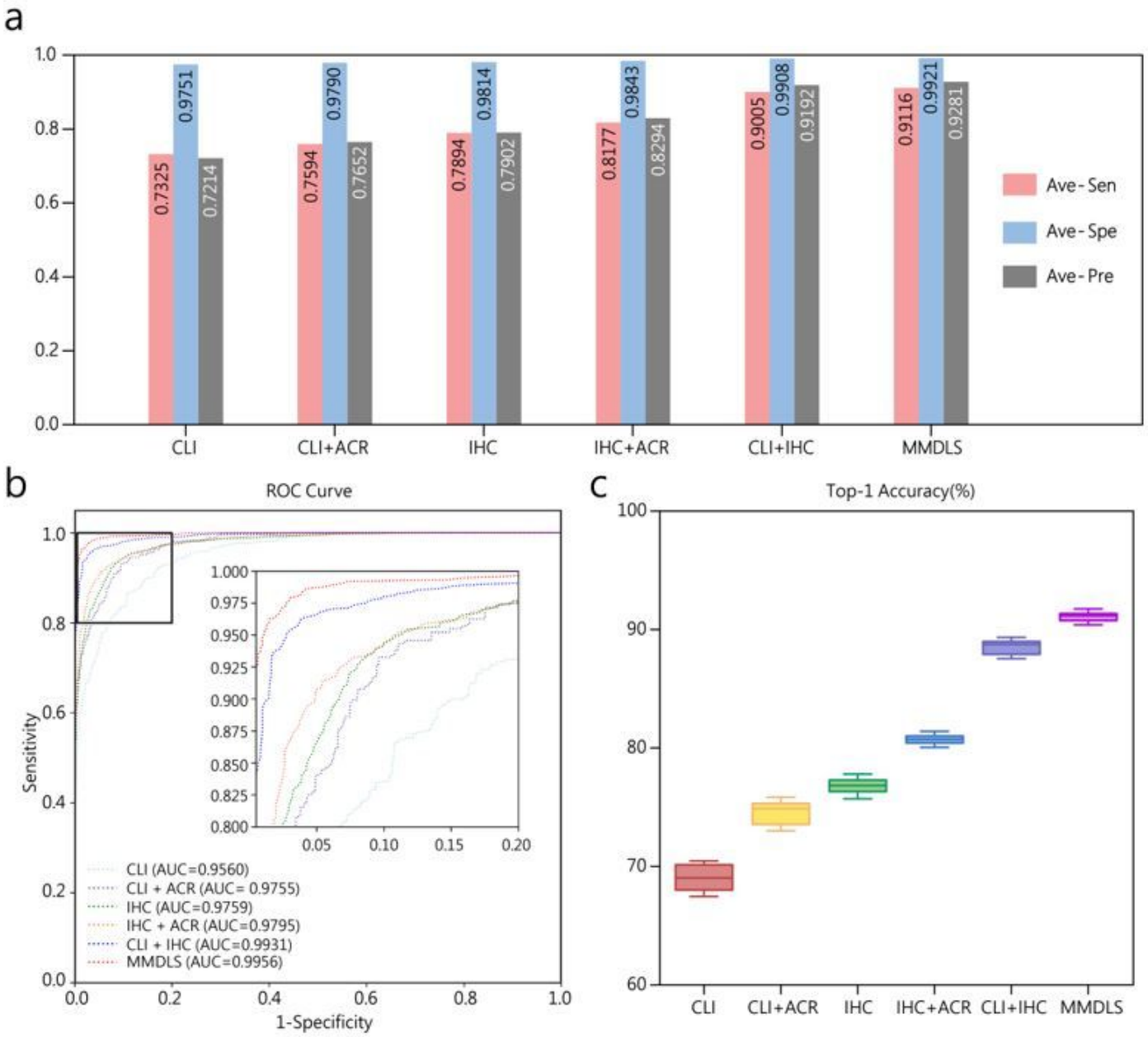

Figure 3

Comparison of single-modal, dual-modal and multi-modal DLS performance. a. Bar chart of Average sensitivity, specificity and precision of different combinations of modalities. Different combinations of clinical skin image modality (CLI), multi-IHC image modality (IHC) and 2019 EULAR/ACR score modality (ACR) were tested. MMDLS refers to the combination of all 3 modalities. b. Receiver operating characteristic (ROC) curve for the combination of various modalities. The area under the ROC curve (AUC) was calculated to evaluate the performances. c. Box plot of the top-1 accuracy for different combinations of modalities. Each box represents a series of top-1 accuracy values in different repeating experiments. The box plot intuitively shows the distribution of the values and the diagnostic accuracy of can also be discerned.

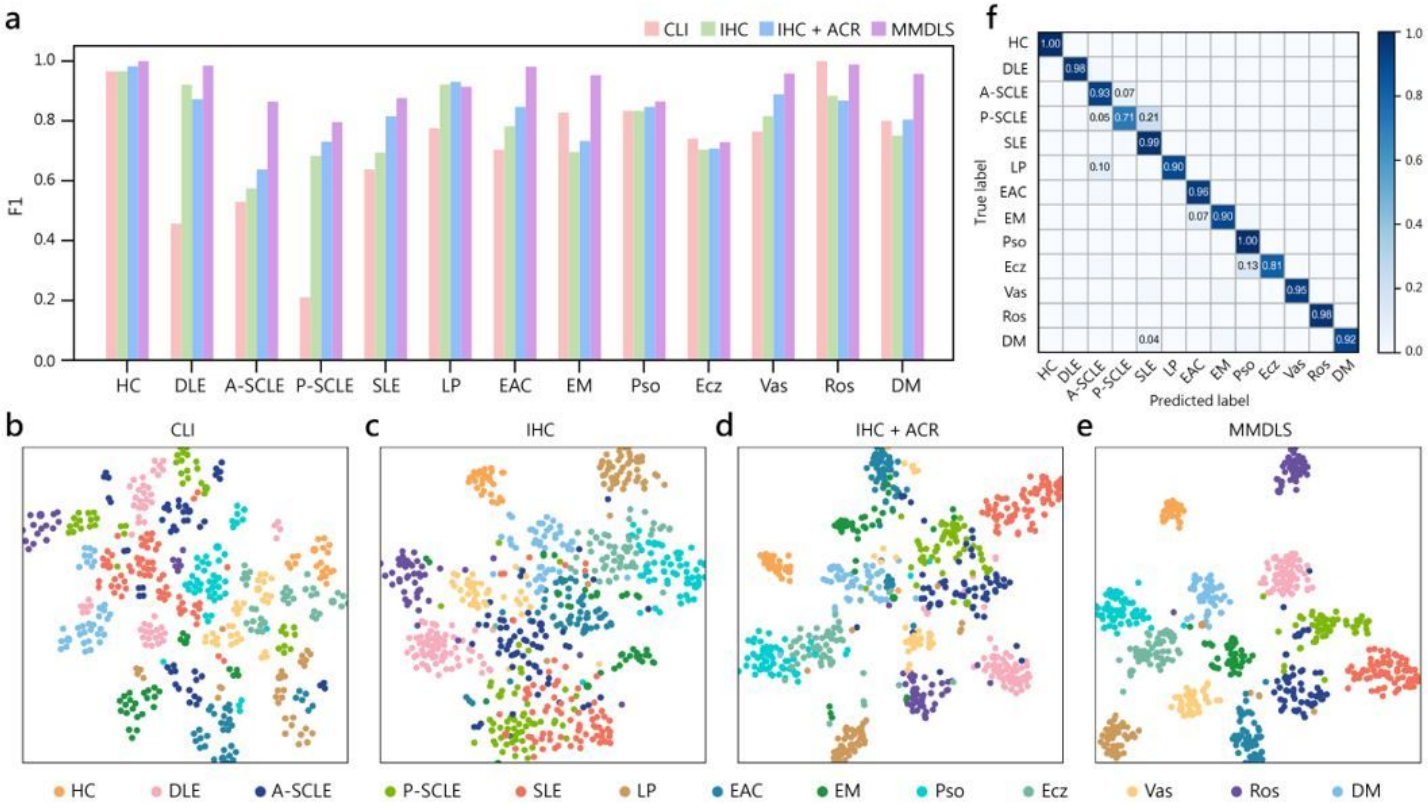

Figure 4

The performance of MMDLS on the diagnosis and classification of 13 categories. a. F1 score of single, dual and multi-modal classification for all 13 categories. b-e. t-SNE visualization of clustering 13 categories by the algorithm. Point clouds with different colors represent different diseases subgroups. The distance between different point clouds represents the degree of overlap between different groups. f. The confusion matrix shows the multiclass confusion matrix for multimodal classification. The gradation bar from light blue to deep blue represents a variation from low diagnostic accuracy to high diagnostic accuracy.

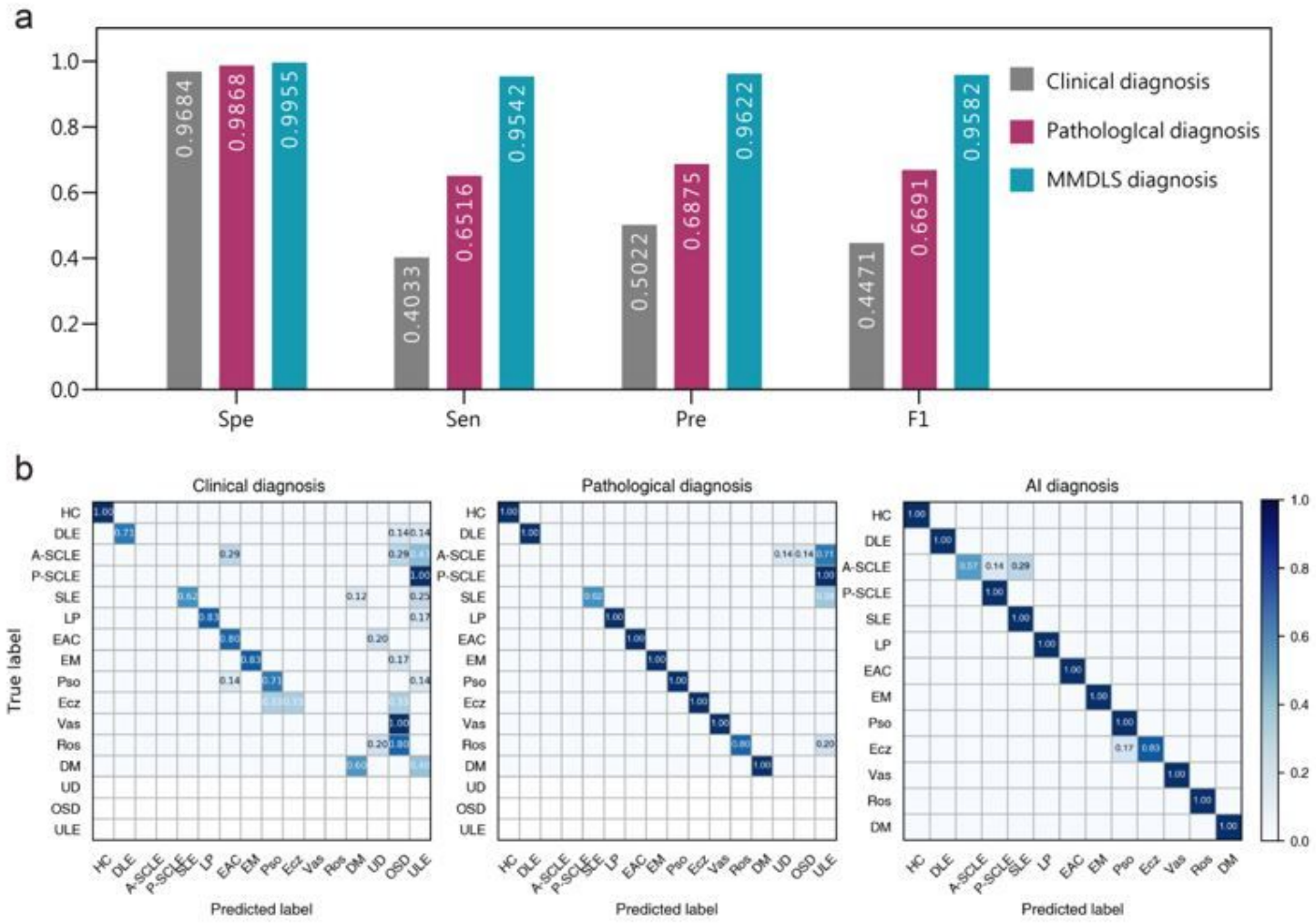

Figure 5

Comparison among clinical diagnosis, pathological diagnosis and MMDLS a. The bar chart compares performances represented by the average specificity, sensitivity, precision and F1 value of clinical, pathological and MMDLS diagnosis. b. The confusion matrix shows in detail of diagnosis made by doctors and MMDLS via comparison of predicted diagnosis and true diagnosis on 13 categories.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineSupplementalTables.docx
- OnlineSupplementalFigures.docx
- OnlineSupplementalTables.docx
- OnlineSupplementalMethods.docx
- MICLAIMChecklist.pdf