



## ORIGINAL ARTICLE

# Deep learning-based classification of dermatological lesions given a limited amount of labelled data

S. Kramer,<sup>1,\*</sup>  Y. Li,<sup>1</sup> N. Jakob,<sup>1</sup> A.S. Boehm,<sup>1</sup> H. Wolff,<sup>1</sup> P. Tang,<sup>2</sup> T. Lasser,<sup>2</sup> L.E. French,<sup>1</sup> D. Hartmann<sup>1</sup> 

<sup>1</sup>Department of Dermatology and Allergy, University Hospital, LMU Munich, Munich, Germany

<sup>2</sup>Department of Informatics, School of Computations, Information, and Technology, and Munich Institute of Biomedical Engineering, Technical University of Munich, Munich, Germany

\*Correspondence: S. Kramer. E-mail: sebastian.kramer@med.uni-muenchen.de

## Abstract

**Background** Artificial intelligence (AI) techniques are promising in early diagnosis of skin diseases. However, a precondition for their success is the access to large-scaled annotated data. Until now, obtaining this data has only been feasible with very high personnel and financial resources.

**Objectives** The aim of this study was to overcome the obstacle caused by the scarcity of labelled data.

**Methods** To simulate the scenario of label shortage, we discarded a proportion of labels of the training set. The training set consisted of both labelled and unlabelled images. We then leveraged a self-supervised learning technique to pretrain the AI model on the unlabelled images. Next, we fine-tuned the pretrained model on the labelled images.

**Results** When the images in the training dataset were fully labelled, the self-supervised pretrained model achieved 95.7% of accuracy, 91.7% of precision and 90.7% of sensitivity. When only 10% of the data were labelled, the model could still yield 87.7% of accuracy, 81.7% of precision and 68.6% of sensitivity. In addition, we also empirically verified that the AI model and dermatologists are consistent in visually inspecting the skin images.

**Conclusions** The experimental results demonstrate the great potential of the self-supervised learning in alleviating the scarcity of annotated data.

Received: 8 February 2022; Accepted: 10 June 2022

## Conflicts of interest

None.

## Funding source

This work was supported by the German Federal Ministry of Health (2520DAT920).

## Introduction

Skin diseases have caused an enormous burden to individuals. Many of them are listed in the most prevalent diseases in various literature.<sup>1–5</sup> To facilitate the early diagnosis of skin diseases, different rules such as the ABCD-rule<sup>6,7</sup> and the three-point<sup>8</sup> or seven-point<sup>9,10</sup> checklist are developed and widely accepted by dermatologists. However, it has been shown that there exists a significant gap in diagnosis sensitivity or specificity between experienced dermatologists and general practitioners<sup>11–13</sup> when they use tools like dermoscopy or epiluminescence microscopy. The problem can become more severe in the countries which face a shortage of dermatologists. This motivates us to assist dermatologists in skin disease diagnosis using AI techniques, which are becoming increasingly important in medical imaging, such as disease classification, pathological findings and quantification of disease extent.<sup>14,15</sup> In dermatology, several works suggest that

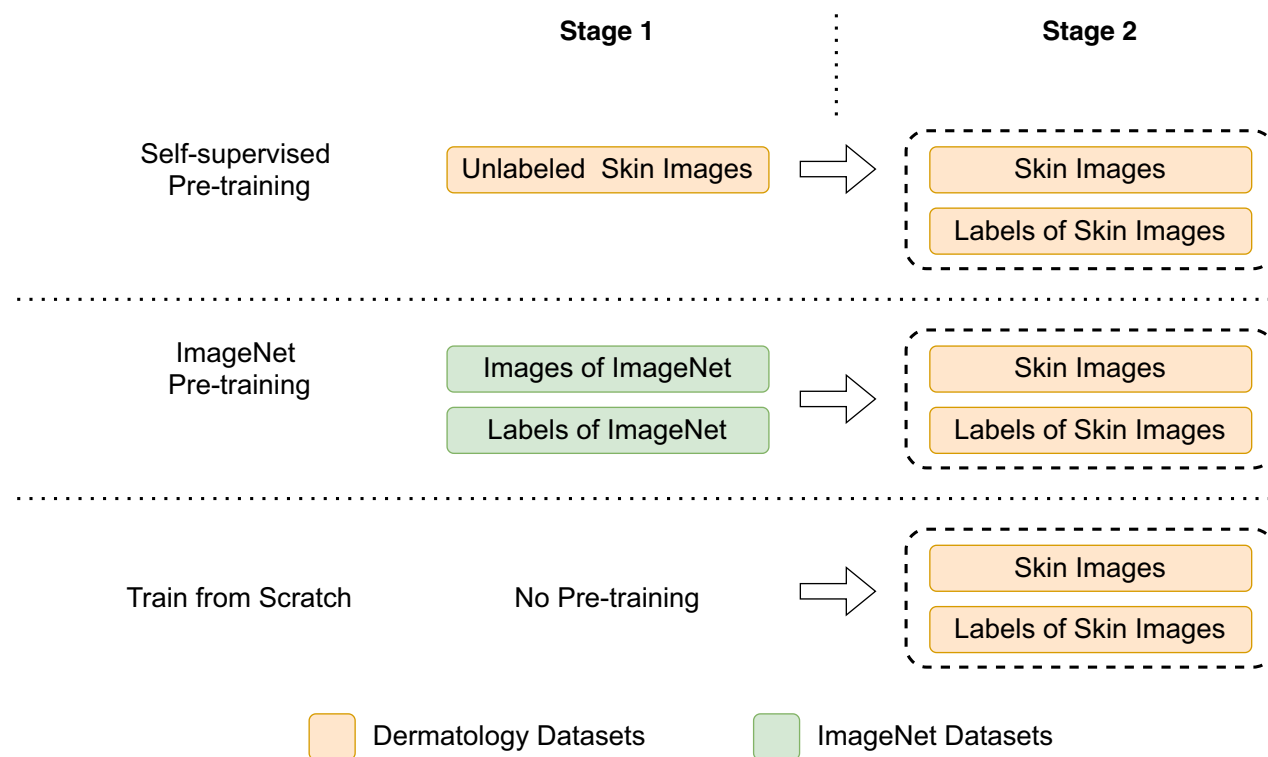
AI models can achieve comparable performance as dermatologists on recognition of skin diseases.<sup>16–20</sup> Nevertheless, one essential prerequisite of these remarkable achievements is large-scale annotated datasets. Most AI models, for example deep neural networks (DNNs), are data-driven. In order to learn the patterns from the skin diseases with significant visual variations, they require to be trained on an extensive collection of skin images, which need to be annotated by experienced dermatologists.

Collecting skin images to construct the datasets is expensive, especially for those diseases with low prevalence. Furthermore, manually labelling the images is highly time-consuming and laborious. As a result of a shortage of dermatologists, many clinics can only provide partially labelled datasets. In this case, the AI models only get little supervision when being trained on the labelled subset. Consequently, the models often fail to

capture the actual patterns of diseases and only yield suboptimal performances.

A solution to this dilemma is to transfer a pretrained model from a large-scale dataset to a small target dataset. It is to note that the domains of the two datasets are not necessarily the same. Since the pretrained model has already learned to capture low-level features like edges and colours, they can be easily fine-tuned on the labelled target dataset. A pretrained model is often trained on a benchmark dataset like ImageNet,<sup>21</sup> which contains over one million images from natural scenes. Another way to pretrain a model is to perform self-supervised learning,<sup>22–27</sup> which has enjoyed huge success recently. One of the most appealing advantages of self-supervised learning is that it does not require image labels. Thus, one can use unlabelled images instead of discarding them when training the model. In addition, previous works<sup>22–27</sup> also suggest a good generalization ability of self-supervised pretrained models. Moreover, it has been shown that self-supervised pretrained models suffer substantially less from the class imbalance of diseases,<sup>28</sup> which is considered a common challenge in the medical field.

Some existing works in the dermatological field have shown the performance gain from self-supervised learning. Kwasigroch *et al.*<sup>29</sup> reported that models pretrained with self-supervised techniques yield higher ROC AUC in binary classification. However, this work focuses on dermoscopic images. Moreover, in the benchmark evaluation experiments,<sup>22–27</sup> the self-supervised techniques adopted in their experiments yield suboptimal performance compared to the recently developed ones.<sup>22–27</sup> Chaves *et al.*<sup>30</sup> empirically proved that self-supervised pretraining is particularly useful in low training data scenarios, but it mainly considers distinguishing melanoma from other diseases. Azizi *et al.*<sup>31</sup> proposed a novel self-supervised technique called Multi-Instance Contrastive Learning and demonstrated its effectiveness on both skin image and chest X-ray classification. Nevertheless, as its name suggests, this algorithm is specifically designed for multi-instance learning, which has a quite different problem setting. Specifically, multi-instance learning aims to learn a shared representation for multiple instances that depict the same object. For example, the instances of an object can be pictures taken from different perspectives. However, each object is associated

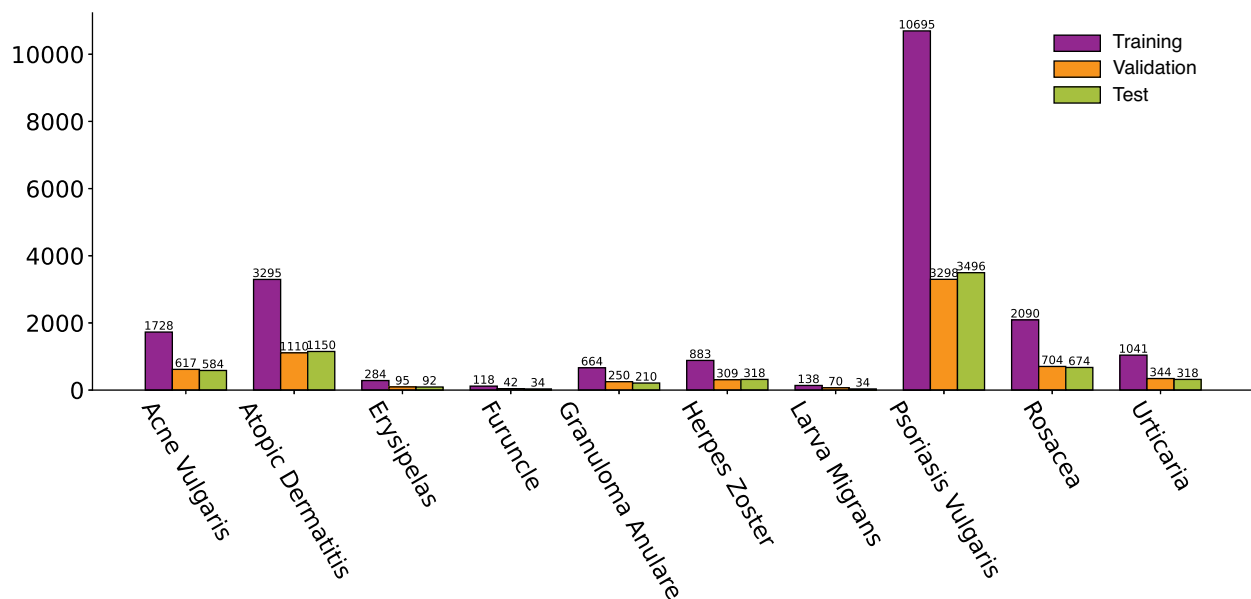


**Figure 1** Training strategies. Blocks in the diagram denote the data sources used for training. The labelled images used at the second stage are sampled from the training set, and the sampling process is described in Section 2. The unlabelled images used by self-supervised pretraining are the full training set.

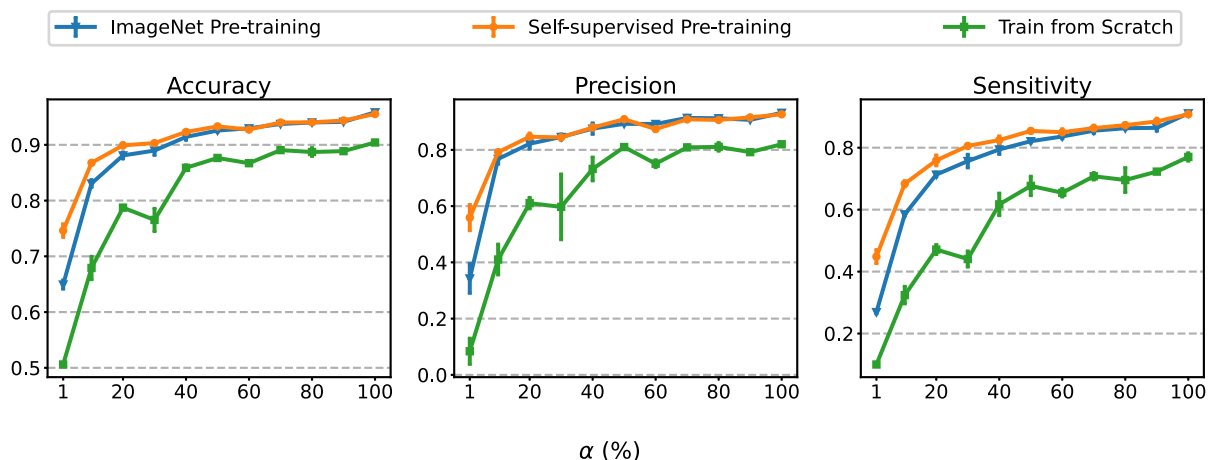
with only one instance (i.e., one image) in our skin image datasets. Hence, the algorithm proposed in Azizi *et al.*<sup>31</sup> does not apply to our task.

In this work, we aim to use self-supervised learning techniques to enhance the model's performance in skin disease classification, on the condition that only a limited amount of skin images is labelled. The rest of the paper is structured as follows: Subsection 2.1 demonstrates three training strategies used in the

experiments. Next, in Subsection 2.2, we show the datasets of skin images that we have collected. We then demonstrate the classification performance in Subsection 3.1. In Subsection 3.2, we visualize the samples using their features extracted by the model and show how the model's discrimination ability changes with the size of the labelled train set. In Subsection 3.3, we explain the model using saliency maps. In the end, we present our conclusion in Section 4.



**Figure 2** Class distribution in the skin image datasets. Training, validation, and test set contain 20 926, 6829 and 6910 samples, respectively. Numbers above the bars denote the number of samples of the corresponding disease.



**Figure 3** Accuracy, precision and sensitivity on the test set. The error bars denote  $\pm 1$  standard deviation of the results from three trials.

## Materials and methods

### Training strategies

We attempt to overcome the data shortage with the help of self-supervised learning. We consider three training strategies for comparison, as shown in Fig. 1.

(a) **Self-supervised pretraining:** This strategy consists of two stages: we perform self-supervised learning on the unlabelled skin images at the first stage. After that, we transfer the pre-trained model to the labelled dataset of skin images.

(b) **ImageNet pretraining:** This strategy is also composed of two stages: We pretrain the model in a supervised manner on ImageNet. After that, we transfer the pretrained model to the labelled dataset of skin images. There are two key differences between the strategy (a) and (b). Firstly, the dataset (ImageNet) for pretraining in (b) is from a different domain. Secondly, class labels are used during pretraining in (b).

(c) **Train from scratch:** We do not perform pretraining in this strategy. Instead, we skip the first stage and directly train the model on the labelled dataset of skin images.

We use EfficientNet-B4<sup>32</sup> as the classification model in all the experiments because this model is small but accurate. For self-supervised pretraining, we adopt MoCo-V2<sup>23</sup> due to its compelling performance. After making a trade-off between training time and accuracy, we train the classification model for 27 000 iterations in all the experiments.

### Datasets of skin images

A total of 34 665 skin images was collected from the photography laboratory at the *Department of Dermatology and Allergy, University Hospital, LMU Munich*. For all images used, written informed consent was obtained from the patient. Additionally, we anonymized all the images by manually erasing biometric features to preserve privacy. The spatial resolution of the raw photographs is 3840 by 2160. The locations of skin lesions were annotated with rectangular bounding boxes. Then, the skin lesions in the bounding boxes were cropped from the raw images and used as input for classification. Considering that the computation resources are limited, we re-scaled all the cropped images such that their longest side is maximal 512-pixel. During training, the images are randomly cropped, flipped and rotated. Next, the images are resized to 512 by 512 and fed into the neural network.

The classes of lesions considered in this study include: (1) acne vulgaris, (2) atopic dermatitis, (3) erysipelas, (4) furuncle, (5) granuloma anulare, (6) herpes zoster, (7) larva migrans, (8) psoriasis vulgaris, (9) rosacea, (10) urticaria. It is to note that the morbidity rates of these diseases are much different, and they are not equally frequently documented in the clinic. Thus, the number of collected samples varies widely. As shown in Fig. 2, the dominant class: psoriasis

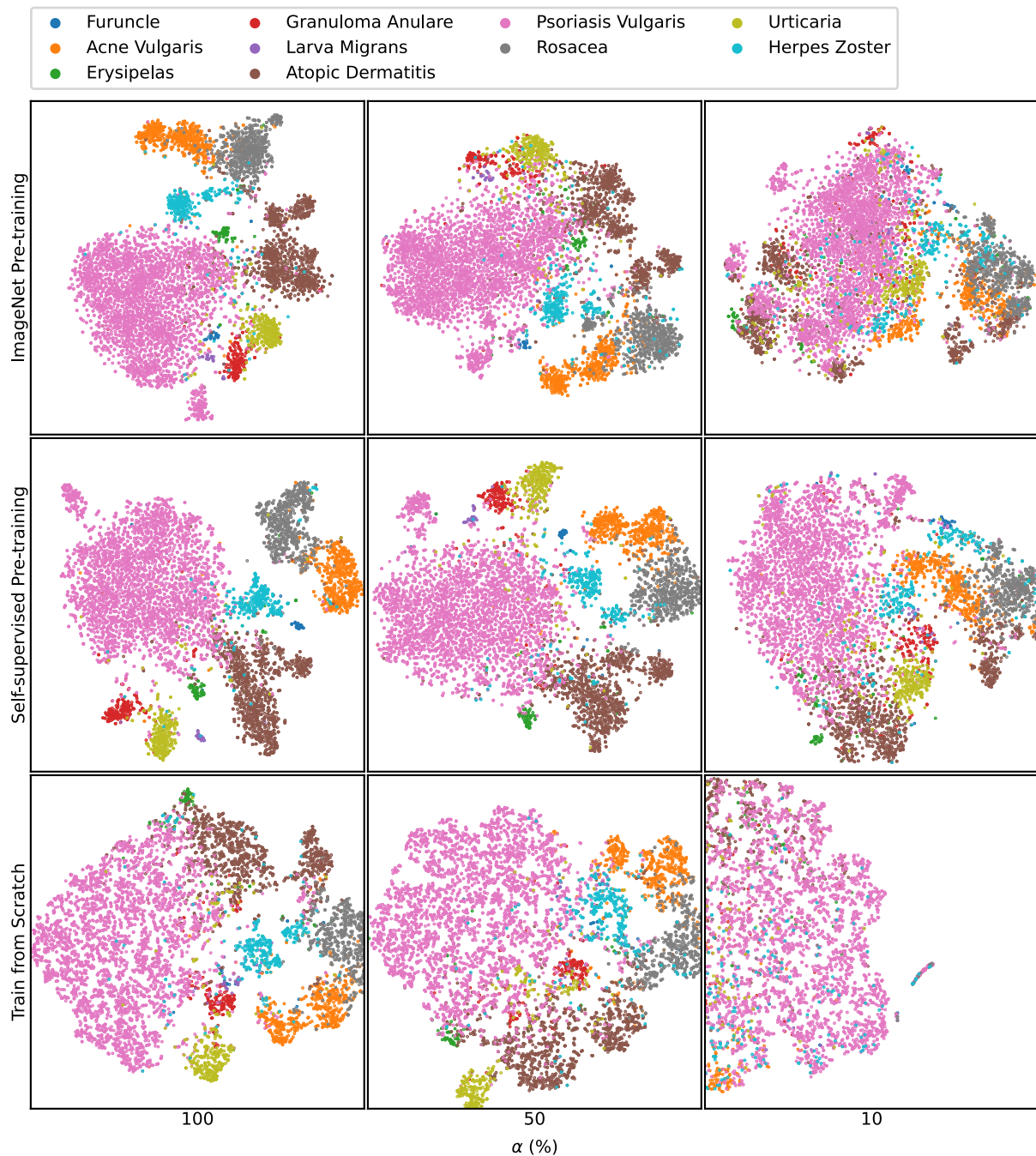
**Table 1** Class-wise classification performance

(a)	$\alpha = 100\%$			$\alpha = 50\%$			$\alpha = 10\%$		
	IN	SS	FS	IN	SS	FS	IN	SS	FS
Acne Vulgaris	<b>0.95</b>	0.94	0.87	<b>0.92</b>	<b>0.92</b>	0.84	0.82	<b>0.87</b>	0.61
Atopic Dermatitis	<b>0.96</b>	0.95	0.89	0.90	<b>0.92</b>	0.86	0.73	<b>0.83</b>	0.57
Erysipelas	<b>0.95</b>	0.89	0.76	<b>0.97</b>	0.88	0.77	<b>0.80</b>	<b>0.80</b>	0.01
Furuncle	0.86	<b>0.88</b>	0.61	0.88	<b>1.00</b>	0.99	0.59	<b>0.76</b>	0.00
Granuloma Anulare	0.92	<b>0.93</b>	0.81	0.82	<b>0.94</b>	0.77	0.68	<b>0.74</b>	0.67
Herpes Zoster	0.94	<b>0.95</b>	0.83	<b>0.89</b>	0.88	0.75	0.67	<b>0.77</b>	0.00
Larva Migrans	<b>0.88</b>	0.81	0.78	<b>0.91</b>	0.90	0.53	0.73	<b>0.80</b>	0.00
Psoriasis Vulgaris	<b>0.98</b>	0.97	0.94	0.95	<b>0.96</b>	0.91	0.88	<b>0.90</b>	0.79
Rosacea	<b>0.92</b>	<b>0.92</b>	0.85	0.88	<b>0.92</b>	0.86	0.77	<b>0.83</b>	0.53
Urticaria	<b>0.95</b>	0.93	0.90	0.86	<b>0.88</b>	0.89	0.77	<b>0.87</b>	0.59
(b)	$\alpha = 100\%$			$\alpha = 50\%$			$\alpha = 10\%$		
Disease	IN	SS	FS	IN	SS	FS	IN	SS	FS
Acne Vulgaris	<b>0.94</b>	0.93	0.89	0.89	<b>0.91</b>	0.84	0.69	<b>0.80</b>	0.37
Atopic Dermatitis	<b>0.95</b>	<b>0.95</b>	0.81	0.91	<b>0.92</b>	0.86	0.80	<b>0.83</b>	0.63
Erysipelas	0.85	<b>0.88</b>	0.71	0.64	<b>0.80</b>	0.40	0.35	<b>0.49</b>	0.01
Furuncle	<b>0.91</b>	0.85	0.47	0.68	<b>0.79</b>	0.15	0.29	<b>0.55</b>	0.00
Granuloma Anulare	<b>0.87</b>	0.86	0.70	0.76	<b>0.79</b>	0.60	0.40	<b>0.65</b>	0.11
Herpes Zoster	<b>0.89</b>	0.87	0.74	0.77	<b>0.81</b>	0.64	0.55	<b>0.59</b>	0.00
Larva Migrans	<b>0.88</b>	0.85	0.73	<b>0.85</b>	0.82	0.24	0.32	<b>0.47</b>	0.00
Psoriasis Vulgaris	0.98	<b>0.99</b>	0.96	<b>0.97</b>	<b>0.97</b>	0.96	0.93	<b>0.96</b>	0.86
Rosacea	<b>0.97</b>	<b>0.97</b>	0.94	<b>0.94</b>	<b>0.94</b>	0.91	0.88	<b>0.90</b>	0.92
Urticaria	0.89	<b>0.92</b>	0.86	0.82	<b>0.87</b>	0.73	0.58	<b>0.62</b>	0.35

(a) Class-wise precision; (b) Class-wise sensitivity. For compactness of the tables, we abbreviate ImageNet pretraining to 'IN', self-supervised pretraining to 'SS', and training from scratch to 'FS', respectively. Note that the accuracy is defined over all the samples and cannot be computed in a per-class manner. For  $\alpha = 50\%$ , the IN and SS are both 92%, therefore they are both highlighted with bold font.

vulgaris has 17 487 samples, while the least frequent class: furuncle has only 192 samples. We partitioned the images into training, validation and test set with approximate percentages of 60%, 20% and 20%, respectively. We also kept the class distributions in each set as close as possible. Furthermore, we ensured that no images from the same patient appeared in two different sets. To enhance the reliability of our experiments, a board-certified dermatologist with over 30-year professional experience validated and curated all images in the test set. All the images with incorrect or uncertain diagnoses were removed.

In order to simulate the scenarios where only a limited number of images is labelled, we randomly sample a subset of the training set and only provide this subset to the model at the second stage of all training strategies. The size of the sampled subset is controlled by a parameter  $\alpha$  ranging from 10% to 100%. For example,  $\alpha = 100\%$  means the subset contains the entire training set, while  $\alpha = 10\%$  means the



**Figure 4** Visualization of the embedding vectors produced by the EfficientNet-B4. Each scatter point denotes a sample from the test set. Clusters of different diseases are shown in different colours. If each cluster of disease is compact and the clusters are distant from each other, it then suggests that the classifier is discriminative.

subset contains only 10% of the training set. In addition, we approximately maintain the class distribution when sampling the subset.

The study followed the Declaration of Helsinki and was approved by the Ethical Committee of the Ludwig Maximilian University (LMU), Munich, Germany (reference no: 21–0257).

## Results

### Classification performance

For an imbalanced data set, merely using accuracy to assess the classification model is often misleading and not sufficient. Therefore, we also adopt precision along with sensitivity for performance evaluation. To encounter uncertainty introduced by randomly sub-sampling the train set, we perform cross-validation for three times and report the standard deviation in the evaluation results. As shown in Fig. 3, when using 100% of the train set at the second stage, the ImageNet pretrained model yields 95.5% accuracy, 93.1% precision and 91.3% sensitivity. The results are close to the self-supervised pretrained model's performance, which is 95.7% accuracy, 91.7% precision and 90.7% sensitivity. However, the model trained from scratch shows substantially lower performance (90.4% accuracy, 82.5% precision and 78.1% sensitivity).

The advantages of the self-supervised pretrained model become more apparent with the decrease of  $\alpha$ . When  $\alpha = 10\%$ , we can still achieve 87.7% accuracy, 81.7% precision and 68.6% sensitivity with the self-supervised pretrained model. In addition, we also show the class-wise precision and sensitivity in Table 1. Although the self-supervised pretrained model also suffers from great performance degradation when  $\alpha$  drops, it still outperforms the other two models by a large margin on the minor classes when  $\alpha = 10\%$ . In short, the classification metrics nicely demonstrates the effectiveness of the self-supervised pre-training.

### Visualization of the feature vectors

An accurate classification model can often extract discriminative features from the input images. Discriminative features have two critical characteristics. Firstly, feature vectors of the same disease class are clustered together. Secondly, the feature vector clusters of different diseases are distant from each other. We exploit the t-SNE<sup>33</sup> technique to visualize the feature vectors of images in the test set, as demonstrated in Fig. 4. When using the entire training set (i.e.  $\alpha = 100\%$ ), both ImageNet pretraining and self-supervised pretraining strategies can produce highly discriminative feature vectors, for which the boundaries between different diseases are clear. As  $\alpha$  decreases, the feature vectors of different diseases tend to overlap, making it increasingly difficult for the model to determine the boundaries to classify the diseases. This phenomenon is very conspicuous when  $\alpha = 10\%$ . The feature vectors of atopic dermatitis extracted by the ImageNet pretrained model are overlapped with the clusters of other diseases. However, this problem is much alleviated by exploiting the self-supervised pretraining. The visual observation is well aligned with the classification performance on atopic dermatitis: 73.1% of precision and 80.0% of sensitivity yielded by ImageNet pretraining, while 83.3% of precision and 83.1% of sensitivity by self-supervised pretraining. In contrast, the model trained from

scratch cannot distinguish the dominant disease (psoriasis vulgaris) from others when  $\alpha$  is low (10%). In this case, the model predicts the most images as psoriasis, indicating that it fails to learn the deep features of the diseases. As shown in Fig. 5, we also make a similar observation when experimenting with another model architecture: ResNet50,<sup>34</sup> one of the most widely used architectures in computer vision. Although there is a clear margin between the psoriasis vulgaris embeddings and other disease embeddings, it is still difficult to distinguish between the embeddings of the rare diseases.

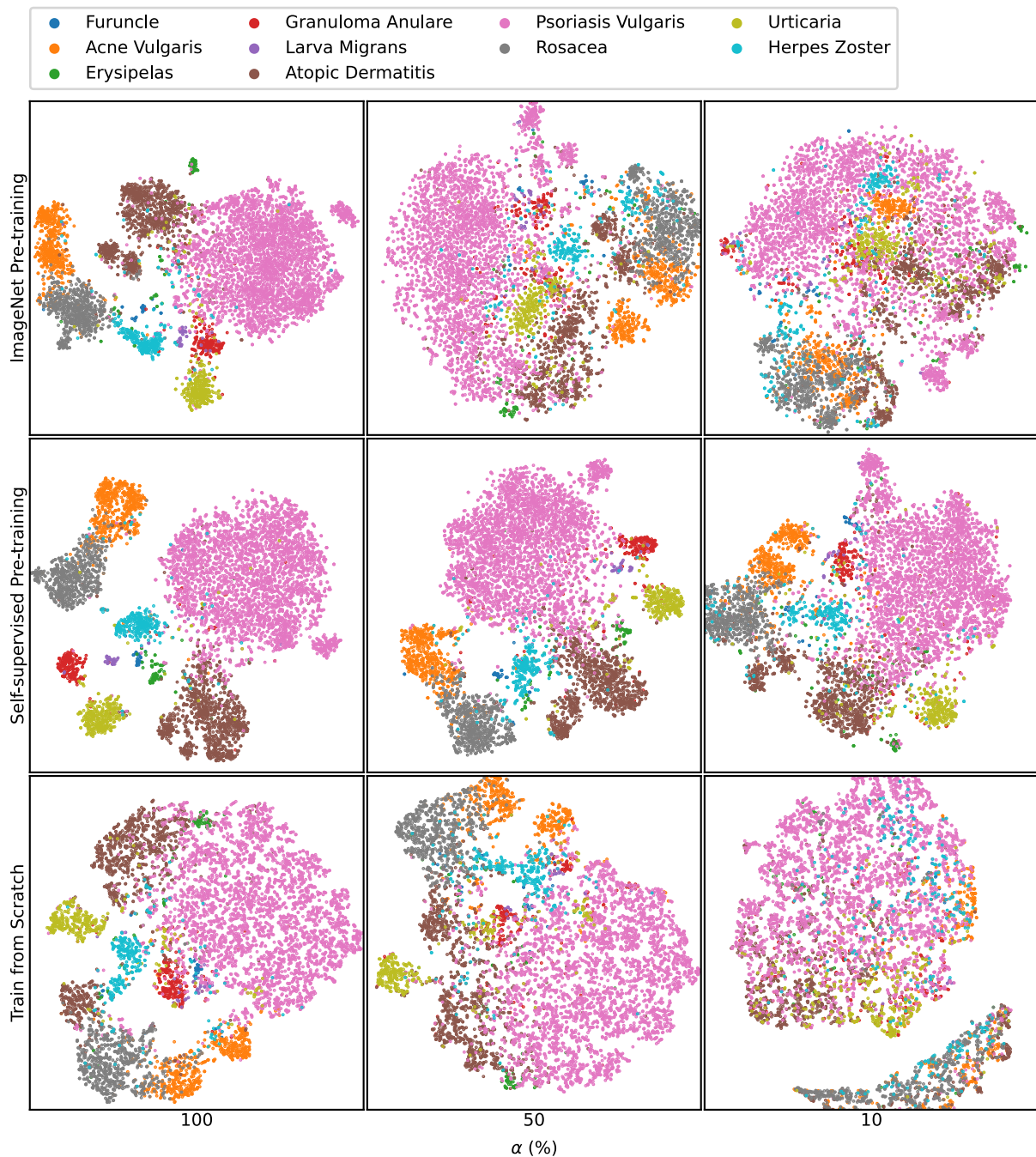
### Explanation of the model

Recent years have seen a surge in research activity in explaining neural networks. For image classification, the saliency maps are widely used for visualizing the explanation results. They are two-dimensional heat maps, in which each pixel value indicates the amount of information that the corresponding pixel provides to the model. Recently, a novel explanation approach<sup>35</sup> for neural networks has been proposed to explain at the input image level, which is more suitable for dermatology. We utilize this technique to generate the saliency maps and explore how consistent the neural network's vision is with the dermatologists.

We illustrate the saliency maps along with the skin images in Fig. 6. We have the following findings by visual inspection: Firstly, the neural network might exploit information only from a part instead of the entire area of lesions. Because the information from that part is sufficiently discriminative. For instance, if we look closely at the psoriasis vulgaris examples in Fig. 6, we find that only a representative part of lesions on the skin is highlighted. Secondly, the neural network might also utilize the auxiliary information from the non-lesion area. In particular, the edges of extremities, lips and faces can help the model recognize the body location, which is highly informative when making a diagnosis. As shown in Fig. 6, the model recognizes human faces from the lips, and it classifies the images as rosacea by combining the information from the rosacea lesions with this auxiliary information. In other words, it has learned the high co-occurrence between the human faces and rosacea.

### Conclusion

In this study, we applied self-supervised learning to classification of dermatological diseases. We showed that self-supervised pre-training can alleviate the shortage of labelled data. We also demonstrated that the neural network's vision is well aligned with dermatologists. These empirical findings held great promise for deploying the AI techniques to clinics to assist dermatologists with early diagnosis of skin diseases and decision making. However, there is still much room for improvement and future work. For example, the datasets can be extended to include more images and pathologies. In addition, techniques tackling the class imbalance can be adopted to improve the model's performance on the diseases that are more challenging to diagnose.

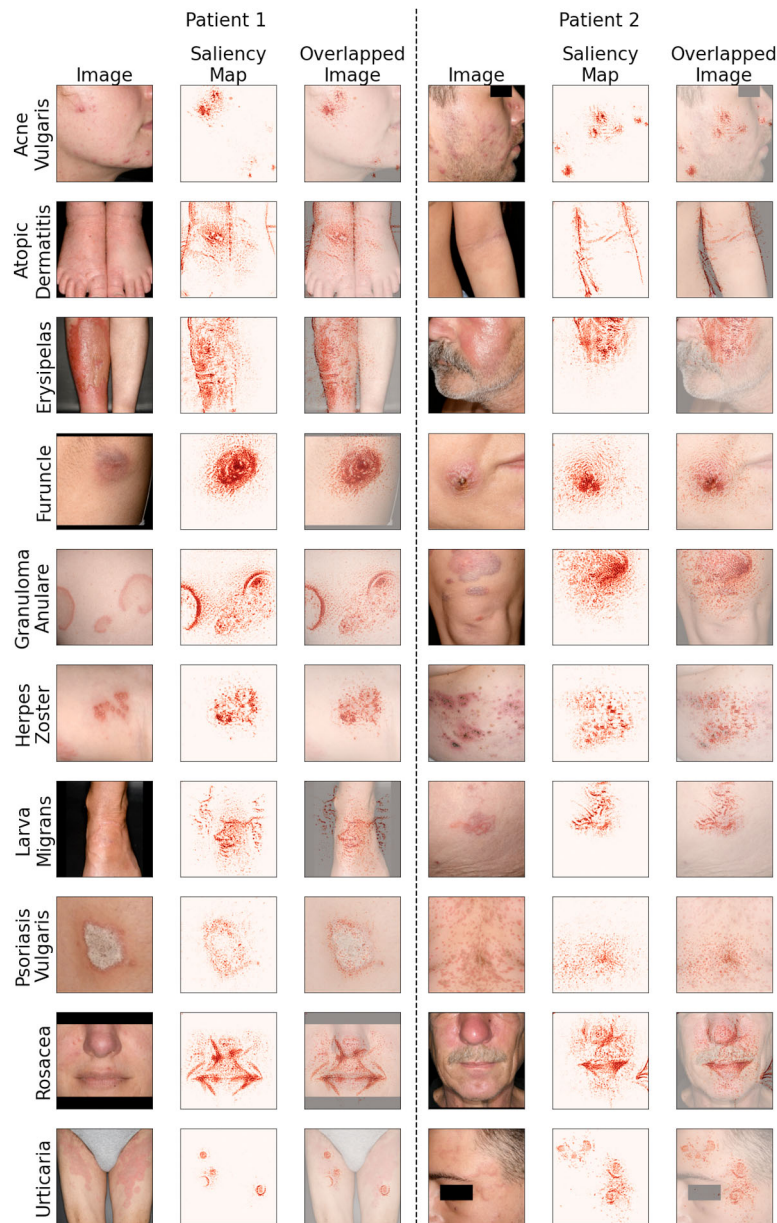


**Figure 5** Visualization of the embedding vectors produced by the ResNet-50. A similar phenomenon is observed: self-supervised pre-training yields much more distinguishable embeddings than ImageNet pretraining or training from scratch in low-label scenarios.

### Acknowledgements

This work was funded by the German Federal Ministry of Health (2520DAT920). The patients in this manuscript have

given written informed consent to publication of their case details. Open Access funding enabled and organized by Projekt DEAL.



**Figure 6** Visualization of the saliency maps. Two samples are demonstrated for each disease. For each sample, skin image (left), saliency map (middle) and overlapped image (right) are shown. If a region in the saliency map is dark red, then it suggests that the region provides most of the information for the neural network to classify the input image.

### Data availability statement

Pictures can not be shared due to privacy regulations of the hospital.

### References

- 1 Hay RJ, Johns NE, Williams HC *et al*. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *J Invest Dermatol* 2014; **134**: 1527–1534.
- 2 Chu S, Mehrmal S, Uppal P, Giesey RL, Delost ME, Delost GR. Burden of skin disease and associated socioeconomic status in Europe: an ecologic study from the global burden of disease study 2017. *Jaad Int* 2020; **1**: 95–103.
- 3 Bridgman AC, Fitzmaurice C, Dellavalle RP *et al*. Canadian burden of skin disease from 1990 to 2017: results from the global burden of disease 2017 study. *J Cutan Med Surg* 2020; **24**: 161–173.
- 4 Svensson A, Ofenloch RF, Bruze M *et al*. Prevalence of skin disease in a population-based sample of adults from five European countries. *Brit J Dermatol* 2018; **178**: 1111–1118.



- 5 Giesey RL, Mehrmal S, Uppal P, Delost G. Global burden of skin and subcutaneous disease: a longitudinal analysis from the global burden of disease study from 1990–2017. *Ski J Cutan Medicine* 2021; **5**: 125–136.
- 6 Nachbar F, Stolz W, Merkle T et al. The ABCD rule of dermatoscopy high prospective value in the diagnosis of doubtful melanocytic skin lesions. *J Am Acad Dermatol* 1994; **30**: 551–559.
- 7 Robinson JK, Turrisi R. Skills training to learn discrimination of ABCDE criteria by those at risk of developing melanoma. *Arch Dermatol* 2006; **142**: 447–452.
- 8 Zalaudek I, Argenziano G, Soyer HP et al. Three-point checklist of dermatoscopy: an open internet study. *Brit J Dermatol* 2006; **154**: 431–437.
- 9 Healsmith MF, Bourke JF, Osborne JE, Graham-Brown RAC. An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma. *Brit J Dermatol* 1994; **130**: 48–50.
- 10 Johr RH. Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist. *Clin Dermatol* 2002; **20**: 240–247.
- 11 Menzies SW, Bischof L, Talbot H et al. The performance of SolarScan: an automated Dermoscopy image analysis instrument for the diagnosis of primary melanoma. *Arch Dermatol* 2005; **141**: 1388–1396.
- 12 Binder M, Schwarz M, Winkler A et al. Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch Dermatol* 1995; **131**: 286–291.
- 13 Burrioni M, Corona R, Dell'Eva G et al. Melanoma computer-aided diagnosis: reliability and feasibility study. *Clin Cancer Res Official J Am Assoc Cancer Res* 2004; **10**: 1881–1886.
- 14 Zhou SK, Greenspan H, Davatzikos C et al. A review of deep learning in medical imaging: imaging traits, technology trends, case studies with Progress highlights, and future promises. *P IEEE* 2021; **109**: 820–838.
- 15 Esteva A, Chou K, Yeung S et al. Deep learning-enabled medical computer vision. *Npj Digital Medicine* 2021; **4**: 5.
- 16 Liu Y, Jain A, Eng C et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med* 2020; **26**: 900–908.
- 17 Maron RC, Weichenthal M, Utikal JS et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer* 2019; **119**: 57–65.
- 18 Bajaj L, Kumar H, Hasija Y. Automated system for prediction of skin disease using image processing and machine learning. *Int J Comput Appl* 2018; **180**: 9–12.
- 19 Tschandl P, Codella N, Akay BN et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* 2019; **20**: 938–947.
- 20 Esteva A, Kuprel B, Novoa RA et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–118.
- 21 Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. *2009 IEEE Conf Comput Vis Pattern Recognit* 2009: 248–255.
- 22 He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. *2020 IEEE Conf Comput Vis Pattern Recognit Cvpr* 2020: 9726–9735.
- 23 Chen X, Fan H, Girshick R, He K. Improved baselines with momentum contrastive learning. *Arxiv* 2020.
- 24 Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. Proceedings of the 37th International Conference on Machine Learning. PMLR, 2020: 1597–1607. <https://dl.acm.org/doi/abs/10.5555/3524938.3525087>
- 25 Chen T, Kornblith S, Swersky K, Norouzi M, Hinton GE. Big Self-Supervised Models Are Strong Semi-Supervised Learners. Advances in Neural Information Processing Systems. Curran Associates, Inc., 2020: 22243–22255. <https://arxiv.org/abs/2006.10029>
- 26 Grill J-B, Strub F, Altché F et al. Bootstrap your Own Latent - A New Approach to Self-Supervised Learning. Advances in Neural Information Processing Systems. Curran Associates, Inc., 2020: 21271–21284. <https://arxiv.org/abs/2006.07733>
- 27 Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. Advances in Neural Information Processing Systems. Curran Associates, Inc., 2020: 9912–9924. <https://arxiv.org/abs/2006.09882?context=cs>
- 28 Yang Y, Xu Z. Rethinking the Value of Labels for Improving Class-Imbalanced Learning. Advances in Neural Information Processing Systems. Curran Associates, Inc., In, 2020: 19290–19301.
- 29 Kwasigroch A, Grochowski M, Mikołajczyk A. Self-supervised learning to increase the performance of skin lesion classification. *Electronics* 2020; **9**: 1930.
- 30 Chaves L, Bissoto A, Valle E, Avila S. An evaluation of self-supervised pre-training for skin-lesion analysis. arXiv preprint arXiv: 210609229 2021. <https://arxiv.org/abs/2106.09229>
- 31 Azizi S, Mustafa B, Ryan F, et al. Big self-supervised models advance medical image classification. arXiv preprint arXiv:210105224 2021.
- 32 Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019: 6105–6114. <https://arxiv.org/abs/1905.11946>
- 33 van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; **9**: 2579–2605.
- 34 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Ieee Conf Comput Vis Pattern Recognit Cvpr* 2016; **2016**: 770–778.
- 35 Zhang Y, Khakzar A, Li Y, Farshad A, Kim ST, Navab N. Fine-grained neural network explanation by identifying input features with predictive information. Thirty-Fifth Conference on Neural Information Processing Systems, 2021. [https://www.researchgate.net/publication/355060350\\_Fine-Grained\\_Neural\\_Network\\_Explanation\\_by\\_Identifying\\_Input\\_Features\\_with\\_Predictive\\_Information](https://www.researchgate.net/publication/355060350_Fine-Grained_Neural_Network_Explanation_by_Identifying_Input_Features_with_Predictive_Information)