Automatic CD30 scoring method for whole slide images of primary cutaneous CD30$^+$ lymphoproliferative diseases

Tingting Zheng,[1] Song Zheng,[2,3,4] Ke Wang,[1] Hao Quan,[1] Qun Bai,[1] Shuqin Li,[1] Ruiqun Qi,[2,3,4] Yue Zhao,[1,3] Xiaoyu Cui ,[1] Xinghua Gao[2,3,4]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jcp-2022-208344).

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, Liaoning, China
[2]Department of Dermatology, The First Hospital of China Medical University, Shenyang, Liaoning, China
[3]National and Local Joint Engineering Research Center of Immunodermatological Theranostics No, Heping District, Liaoning Province, China
[4]NHC Key Laboratory of Immunodermatology, Heping District, Liaoning Province, China

Correspondence to
Dr Xiaoyu Cui, Northeastern University, Shenyang, Liaoning Province, China; cuixy@bmie.neu.edu.cn
Dr Yue Zhao; zhaoyue@bmie.neu.edu.cn
Dr Xinghua Gao; gaobarry@hotmail.com

TZ and SZ contributed equally.

Received 18 April 2022
Accepted 7 July 2022

To cite: Zheng T, Zheng S, Wang K, *et al*. *J Clin Pathol* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jclinpath-2022-208344

ABSTRACT

Aims Deep-learning methods for scoring biomarkers are an active research topic. However, the superior performance of many studies relies on large datasets collected from clinical samples. In addition, there are fewer studies on immunohistochemical marker assessment for dermatological diseases. Accordingly, we developed a method for scoring CD30 based on convolutional neural networks for a few primary cutaneous CD30$^+$ lymphoproliferative disorders and used this method to evaluate other biomarkers.

Methods A multipatch spatial attention mechanism and conditional random field algorithm were used to fully fuse tumour tissue characteristics on immunohistochemical slides and alleviate the few sample feature deficits. We trained and tested 28 CD30$^+$ immunohistochemical whole slide images (WSIs), evaluated them with a performance index, and compared them with the diagnoses of senior dermatologists. Finally, the model's performance was further demonstrated on the publicly available Yale HER2 cohort.

Results Compared with the diagnoses by senior dermatologists, this method can better locate the tumour area and reduce the misdiagnosis rate. The prediction of CD3 and Ki-67 validated the model's ability to identify other biomarkers.

Conclusions In this study, using a few immunohistochemical WSIs, our model can accurately identify CD30, CD3 and Ki-67 markers. In addition, the model could be applied to additional tumour identification tasks to aid pathologists in diagnosis and benefit clinical evaluation.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Biomarker evaluation depends heavily on the experience of pathologists and can be subjective and can experience staining problems. Artificial intelligence automatically evaluates biomarkers based on big data, which is almost impossible to obtain for rare diseases.

WHAT THIS STUDY ADDS

⇒ Multipatch spatial attention mechanism network framework (MPSANet) was trained with a few samples and achieved accurate positioning and quantitative evaluation of CD30, CD3 and Ki-67.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ MPSANet can be applied to more biomarkers assessments, reducing the workload of pathologists and helping to realise precision medicine.

INTRODUCTION

Primary cutaneous CD30-positive lymphoproliferative diseases (CD30$^+$LPDs), include lymphomatoid papilloma (LyP) and primary cutaneous anaplastic large cell lymphoma (pcALCL). It is a heterogeneous group of malignant diseases originating from the lymphatic system and is the second most common cutaneous T-cell lymphoma.[1 2] CD30 is a member of the tumour necrosis factor receptor superfamily. It is an essential differentiator between LyP and pcALCL and is a target for disease treatment (table 1).[3–5] Some overlap exists in the clinical, histopathological and molecular features between LyP and pcALCL. In particular, improper treatment of LyP may increase the risk of transformation into pcALCL or mycosis fungoides, leading to the occurrence of a second lymphoma.[5 6]

Immunohistochemistry (IHC) is the gold standard for the diagnosis of cutaneous lymphomas.[7] Dermatologists quantitatively and qualitatively evaluate abnormal immunohistochemical cells through optical microscopy, which is clinically significant for diagnosing specific disease types and for the developmental stages of cutaneous lymphomas.[8 9] However, this process is highly dependent on the expertise of dermatologists and is easily interfered with by staining and isomerism.

Digital pathology assists pathologists in making diagnoses, simplifying complex and time-consuming tasks, and reducing the risks and biases caused by various intraobserver and interobserver factors in the pathological diagnosis process.[10–12] Several methods for the automated evaluation of immunohistochemical markers by machine learning and deep learning have been widely used[13–22] (online supplemental table 1). However, deep learning typically requires large amounts of data to achieve optimal performance. This runs counter to the difficulty of obtaining medical data. In addition to the fact that few deep-learning methods have been applied to CD30 scoring, the lack of publicly available high-quality IHC databases has hindered the development of deep learning for IHC. Therefore, there is an urgent need to develop a deep-learning method that addresses the problem of automated

Table 1 Primary cutaneous CD30$^+$ T-cell lymphoproliferative disorders: histopathological and immunohistochemical characteristics

Type		Histological characteristics	Immunophenotype
Lymphoma papulosis	A	Atypical lymphocytes and mixed wedge*shaped skin infiltration	CD4$^+$ CD8$^+$ CD30$^+$
	B	Medium*sized lymphocytes with brain-like nuclei and epidermal infiltration	CD4$^+$ CD8* CD30$^+$
	C	Atypical lymphocytes, skin infiltration, no obvious mixed inflammatory background	CD4$^+$ CD8* CD30$^+$
	D	Small to medium pleomorphic lymphocytes usually expressed as cytotoxic markers	CD4* CD8$^+$ CD30$^+$
	E	Atypical lymphocyte vascular infiltration and skin infiltration under the background of mixed inflammation	CD4* CD8$^+$ CD30$^+$
Primary cutaneous anaplastic large*cell lymphoma		Single or local nodule or tumour	CD4$^+$ CD8* CD30$^+$ CD3$^+$ or CD3*

$^+$The biomarker is present in the disease type.
*The biomarker is not present in the disease type.

quantitative CD30 scoring using small samples of LyP and pcALCL diseases.

In this study, we trained an end-to-end neural network framework to automate the scoring for CD30$^+$ LPDs. By taking into account the clustered distribution of tumour cells in pathology and fusing features from the tumour surrounding tissue, the data scarcity problem is alleviated. The output heatmap visualisation can effectively reduce false-positives and the misdiagnosis risk. In addition, experiments with CD3 and Ki-67 markers validate that our approach can be applied to similar immunohistochemical scoring. The test results on the Yale HER2 cohort also exceeded the previous best performance. The study process is shown in figure 1.

MATERIALS AND DATASET

Data acquisition

This study conducted experiments on two datasets. The first dataset contains 28 whole slide images (WSIs) from 28 patients diagnosed with LyP or pcALCL at the First Hospital of China Medical University from 2016 to 2021. Before data collection, three dermatologists reviewed all HE-stained and immunohistochemical slides with clinical correlation to confirm the diagnoses (online supplemental table 2). The methodology used in the research process was strictly in accordance with the Declaration of Helsinki of the World Medical Congress[23] and its related guidelines. For the Yale HER2 cohort, we collected 186 invasive breast cancer slides (96 HER2$^-$ and 90 HER2$^+$) expressing HER2-positive (HER2$^+$) and negative (HER2$^-$) HE staining from the publicly available dataset created by SamanFarahmand[24] (online supplemental table 3).

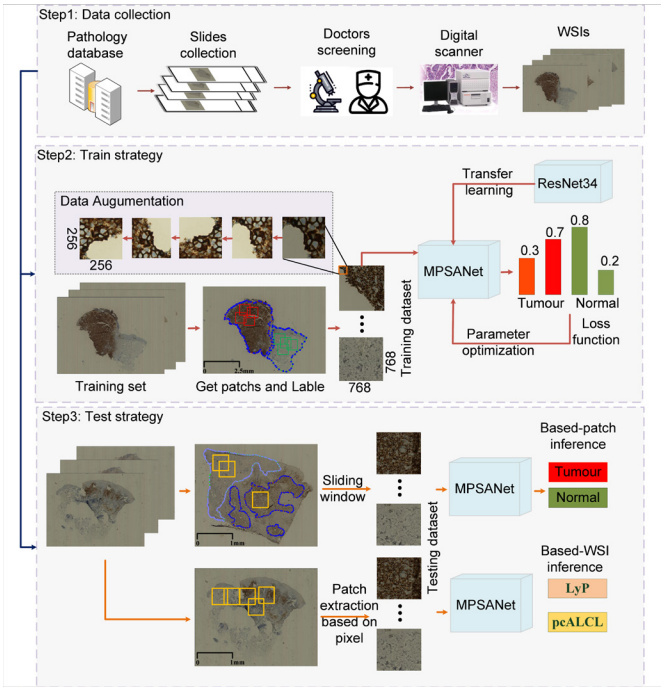

Figure 1 Flow chart for analysis of CD30$^+$ LPDs research. CD30$^+$ LPDs, CD30-positive lymphoproliferative diseases; LyP, lymphomatoid papilloma; MPSANet, multipatch spatial attention mechanism network framework; pcALCL, primary cutaneous anaplastic large cell lymphoma; WSIs, whole slide images.

Dataset preparation

In LyP and pcALCL, the distribution of CD30$^+$ cells is diffuse in histopathology. Hence there are three tissue regions in the collected patches, namely, (1) CD30$^+$ cells, (2) CD30-negative (CD30$^-$) cells and (3) CD30$^+$ cells mixed with CD30$^-$ cells. For training, the cells in (1) are positive samples, those in (2) are negative samples and the ones in (3) are discarded (online supplemental figure 1).

Dividing the dataset into training and test sets, the data processing process consists of three steps (online supplemental table 2 and figure 1). Step 1: Obtain the regions of interest (ROI). Senior dermatopathologists use the Automated Slide Analysis Platform[25] to mark out CD30$^+$ and CD30$^-$ tissue areas in slides. The red colour indicates CD30$^+$, and the green colour indicates CD30$^-$. ROI for positive and negative samples were extracted according to the different colour annotations. During the testing phase, nonbackground tissue regions in the slide were extracted using the Otsu algorithm[26] and morphological operations. Step 2: Obtain the patch dataset. Sample points were randomly selected from the ROI at ×40 magnification (0.23 µm/pixel) to obtain a patch of size 768×768. Step 3: Data enhancement: Adjustment of patch brightness (64.0/255), contrast (0.75), saturation (0.04), hue and rotation (90, 180 and 270), and normalised data.

For the Yale HER2 cohort, we downloaded 186 slides corresponding to annotation files of Her2$^+$ or Her2$^-$ regions annotated by senior pathologists. The 186 slides were randomly divided into a training set and a test set at a ratio of 7:3. When the slide magnified ×20, the training set of 12 000 patches and the test set of 1200 patches were cropped on the ROI using the same data augmentation approach as for the first dataset.

METHODS

Overview of network architecture

The lack of high-quality big data has always hindered the application of deep-learning technology in the analysis of digital pathological images. In this paper, we propose a neural network

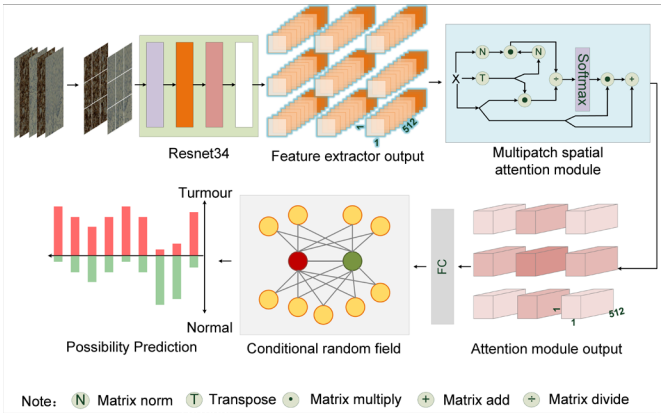

Figure 2 Framework of our MPSANet. MPSANet, multipatch spatial attention mechanism network framework.

framework (multipatch spatial attention mechanism network framework, MPSANet) based on a 34-layer residual convolution neural network (ResNet34),[27] which combines a MPSA[28] and a conditional random field (CRF) algorithm.[29] It is suitable for working with a few samples[30] (figure 2).

MPSANet consists of three parts. The first is the feature extractor. ResNet34 is very easy to optimise, overcomes the gradient disappearance problem in the training process, and solves the problem of rapid degradation associated with the deepening of the network. It has been widely used and recognised in processing digital pathological images and computer-aided diagnosis (CAD). The second part is MPSA. Inspired by the human visual system, the attention mechanism can quickly focus on ROI. The convolutional neural network fused with the attention mechanism significantly improves the accuracy of model prediction in medical image classification[31] and survival prediction.[32] Using the attention mechanism obtained by metric learning,[33] the similarity and interchangeability of features between images belonging to the same category are fully considered. In the case of data scarcity, the use of this attention mechanism can obtain richer features. The third part is CRF which uses the label information of adjacent patches to suppress noise and improve classification accuracy (Detailed descriptions and formulas of the ResNet34 feature extractor, the MPSA and CRF algorithms are shown in online supplemental documents).

Training and implementation details

The model and framework used in this experiment were implemented with the PyTorch library and trained and tested on a server configured with two TITAN_XP GPUs. The ImageNet dataset[34] was used to train the model to initialise the network weights. The optimisation of the network parameters was completed using stochastic gradient descent momentum. The initial learning rate, momentum and batch size were 0.001, 0.9 and 32, respectively. Cross-validation was carried out five times, and training was executed for 50 continuous epochs. All network models used the cross-entropy loss function and the backpropagation method for end-to-end training.

Performance testing

To verify the accuracy and reliability of the model, two testing methods were used on an internal independent test set. (1) Patch-level diagnostics: The $CD30^+/CD30^-$ patch obtained from the ROI on the test set is input into the network and the output is the benign and malignant predictions for each patch. (2) Slide-level diagnosis: The LyP and pcALCL diseases are distinguished based on the proportion of $CD30^+$ cells, with less than 30% $CD30^+$ denoting LyP and more than 75% $CD30^+$ denoting pcALCL. The patch input model is extracted by traversing the nonbackground tissue regions in the WSIs in 32 steps. Mixing patches of $CD30^+$ cells with normal cells can cause an overpositive rate. To avoid this, this study used a threshold majority vote[35] for the calculation. This is done as follows: first, it records the number of patches with a positive probability greater than 0.8 from the model output. Second, the number of positive patches is divided by the total number of patches entered into the model to obtain the percentage of CD30 +cells (Detailed description of the threshold majority voting in online supplemental method 4).

Statistical analysis

Recall, precision, specificity, F1-score, receiver operating characteristic curve and area under the receiver operating characteristic curve (AUC) were used to evaluate model performance (online supplemental table 4).

EXPERIMENTS AND RESULTS

Cross-validation and testing analyses

Fivefold cross validation was conducted on the training set, and the AUC was 0.99. The accuracy, F1-score, precision, recall and specificity index achieved on the verification set were 0.994 (95% CI 0.989 to 0.998), 0.993 (95% CI 0.988 to 0.998), 0.993 (95% CI 0.986 to 1.000), 0.994 (95% CI 0.989 to 0.998) and 0.993 (95% CI 0.986 to 1.000), respectively (online supplemental table 5).

In the test set (online supplemental figure 2), the AUC, accuracy and F1-score were 0.985 (95% CI 0.982 to 0.988), 0.957 (95% CI 0.948 to 0.966) and 0.953 (95% CI 0.950 to 0.956), respectively, which indicated that the model could effectively distinguish between normal and abnormal tissues. The specificity and recall rates were 0.950 (95% CI 0.943 to 0.957) and 0.965 (95% CI 0.945 to 0.984), respectively, indicating that the model was sensitive to the characteristics of $CD30^+$ cells, which reduced the misdiagnosis rate of the model.

Ablation study and $CD30^+$ characterisation analysis

To verify the effectiveness of the MPSA and CRF algorithm proposed in this work combined with different models, we conducted ablation studies on the classical networks ResNet18, ResNet34, VGG16 and VGG19. The experimental results show that our proposed method achieves optimal performance and that the performance of the underlying networks can be significantly improved using the MPSA and the CRF algorithm. In the training set, fivefold cross-validation was carried out, and the mean value and SD of each index in the validation set were obtained (online supplemental table 6). First, adding the MPSA to the basic feature extractor can significantly improve the accuracy of the network. This may be related to the clumpy distribution of $CD30^+$ cells. Second, the addition of the CRF model fully integrates the information extracted from ResNet34 and MPSA. Although the result did not improve by leaps and bounds, it made the result more stable and made the model have good performance on an independent test set. Considering the characteristics of adjacent patches, it was helpful for the current patch to learn more characteristics of normal or abnormal tissues.

Visualisation of features map

To better assist pathologists in diagnosis, the results of model prediction were clearly described by visualisation. The tissue

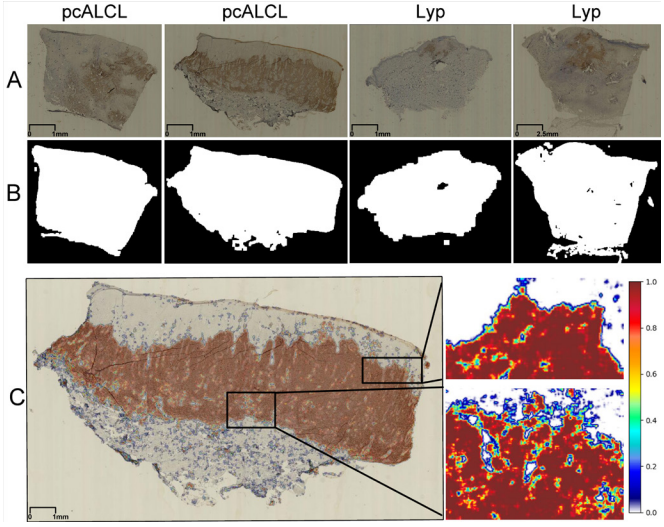

Figure 3 Heatmap visualization of CD30$^+$ regions were identified through the network. (A) The four low-resolution views in the test set, (B) Extraction of binarised maps of non-background tissue regions, (C) The model output a probabilistic heatmap of pcALCL, with yellow to red indicating CD30+ tissue areas. LyP, lymphomatoid papilloma; pcALCL, primary cutaneous anaplastic large cell lymphoma.

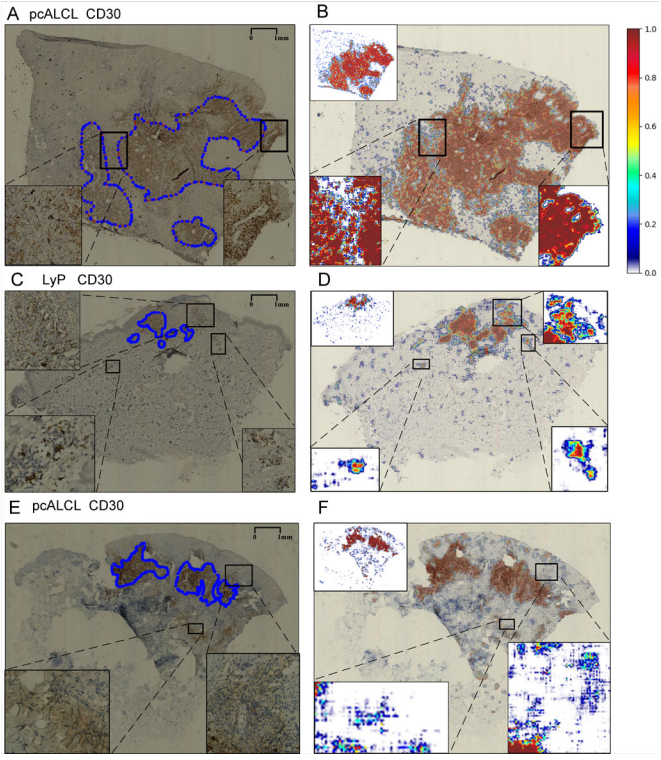

Figure 4 Comparison between CD30$^+$ regions annotated by pathologists and predicted results of models. (A), (C), and (E) Low-resolution views marked by dermatologists. (B), (D), and (F) are the combined pictures of the probability graphs of the model output and the original graphs, where the upper left corner of each image is the probability heatmap of the model output. The areas marked with black boxes are the areas that show the comparison between the pathologist's annotations and the model result. The model can accurately identify CD30+ tissue regions (A–D) and ignore false positives due to staining problems (E–F). LyP, lymphomatoid papilloma; pcALCL, primary cutaneous anaplastic large cell lymphoma.

area of the slide was extracted and input into the model, and the model directly output the visualisation of the prediction results of the WSIs. The probability values were normalised to 0–1, with white to green representing nonmalignant tissue areas and yellow to red representing CD30$^+$ tissue areas. Our proposed method successfully and accurately located CD30$^-$ tissue regions and CD30$^+$ cell tissue regions (figure 3). In addition, to minimise the interference of false-positives (mostly normal cells and a few tumour cells) in the visualised heatmap, patches with a predicted probability of less than 0.8 were considered negative (see online supplemental material 4 for an introduction to thresholds).

Minimising the risk of misdiagnosis

The heterogeneity and accuracy of immunohistochemical staining have serious impacts on clinicopathological diagnoses and on the development of deep learning in histopathology.[36] This is especially the case when CD30$^+$ cells are lightly stained and scattered, which causes them to be easily ignored by dermatologists. According to the comparison in figure 4A–D, it can be clearly seen that our model could effectively solve the areas that are difficult for dermatologists to observe during the process of diagnosis due to the shallow CD30 staining and the presence of nonclustered tumour cells, thus reducing the risk of misdiagnosis. Figure 4E,F shows that our model can effectively identify CD30$^+$ tissue areas and abnormal staining 'noise' due to immunohistochemical staining. Thus, our method has a strong anti-interference ability and it effectively reduces the risk of misdiagnosis. In addition, the visualised heatmaps produced by the model are more accurate. Hence they can assist dermatologists in identifying tumour areas and reduce the risk of misdiagnosis.

Small-scale training set model evaluation

With the rise of rare diseases worldwide, our goal is to develop a fully automated, accurate quantification tool for pathologists that requires only a small number of slices to train. To achieve this, 10%, 20%, 40%, 60% and 80% of normal and CD30$^+$ tumour patches from the training set were randomly selected to train MPSANet. Fivefold cross-validation was performed on each subsampled training set, and testing was performed on the test set (online supplemental figure 3). As the amount of data decreased, the AUC index of the model gradually decreased but remained above 0.91. This shows that our method has good generalisation performance on small datasets. This provides the basis for the automated assessment of biomarkers for other rare or noncommon skin diseases with our method.

Automatic scoring of CD3 and Ki-67

To promote MPSANet for automatic scoring of other biomarkers, CD3 and Ki-67 immunohistochemical slides of pcALCL and LyP were tested. Visualised heatmaps of CD3 and Ki-67 were compared with areas manually annotated by three dermatologists (figure 5). Although our network had not studied the immunohistochemical pathologies of CD3 and Ki-67, it was still able to identify positive areas and conduct quantitative analyses. However, there were some false-positives, perhaps because the immunohistochemical molecular characteristics of CD30 are fundamentally different from those of CD3 and Ki-67 despite their similar immunohistochemical slides in terms of colour. It is reasonable to believe that the model could be applied to additional immunohistochemical biomarkers via simple transfer learning with other small immunohistochemical datasets.[37] This

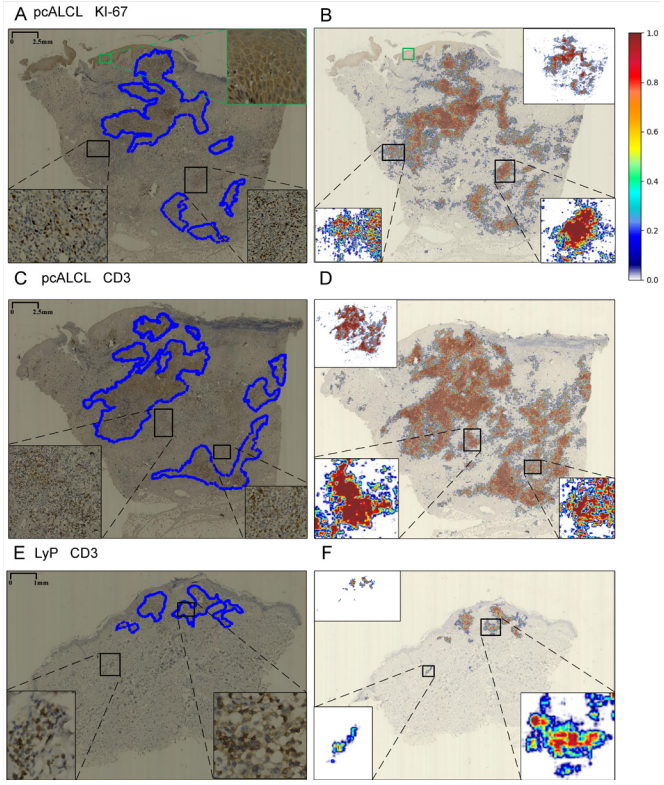

Figure 5 Comparison between CD3/Ki-67 positive regions labeled by pathologists and prediction results of models. (A), (C) and (E) Low-resolution views marked by dermatologists. (B), (D) and (F) are the combined pictures of the probability graphs of the model output and the original graphs. LyP, lymphomatoid papilloma; pcALCL, primary cutaneous anaplastic large cell lymphoma.

could reduce the workloads of dermatologists and speed up clinical applications.

HER2 status classification in the Yale HER2 cohort

To verify the efficacy and generalisation of the proposed method, we also conducted experiments on publicly available HE invasive breast cancers expressing $HER2^+$ and $HER2^-$. We employ AUC as the evaluation criterion for this binary classification problem. The AUC of the fivefold cross-validation was 0.97 on the training set and 0.82 on the test set (online supplemental figure 4). In comparison with the method proposed by Farahmand *et al*,[24] our approach has 5% and 7% higher AUCs than InceptionV3-based migration learning (AUC=0.77) and the fully trained method (AUC=0.75).

DISCUSSION

Primary cutaneous lymphomas are a group of heterogeneous tumours originating from skin lymphocytes with complex clinicopathologic manifestations. Clinical diagnosis usually requires differentiation between benign LPDs of the skin and haematopoietic malignancies.[1] IHC is necessary for disease diagnosis and differentiation, especially in cases that are easy to confuse and difficult to distinguish.[8 9] Efficient and accurate qualitative and quantitative evaluations of $CD30^+$ cell proportions play vital roles in distinguishing between $CD30^+$LPD subtypes, making treatment plans and prognoses feasible for patients. In daily routine diagnoses of $CD30^+$LPDs, dermatologists evaluate immunohistochemical markers under an optical microscope. This time-consuming method, which lacks objectivity, accuracy and reproducibility, hinders the timely diagnosis and treatment of patients. With the increase in biomarkers for clinical decision-making and the shortage of experienced dermatologists, an objective and accurate method is urgently needed to assist dermatologists.

In this paper, a reliable method based on ResNet34 was proposed to evaluate $CD30^+$ cells and identify different subtypes of $CD30^+$LPDs. According to a fivefold cross-validation on collected datasets, the results showed that our proposed deep-learning method can accurately identify $CD30^+$ cells and it exhibits favourable performance and robustness (figure 4). By comparing the visual model prediction results with the diagnosis results of experienced dermatologists, our model was proven to be highly effective in assisting dermatologists with diagnoses. Finally, the use of trained models directly applied to CD3 and Ki-67 biomarkers suggests that our approach has great potential for the automated application for other immunohistochemical biomarkers.

The application of artificial intelligence algorithms in IHC and clinical biomarker scoring has always been a research hotspot.[12 13] However, this approach is still challenging in clinical practice. At present, the mainstream research method is to score markers in IHC through a segmentation algorithm.[14 15] Although such algorithms based on traditional and deep-learning algorithms have achieved satisfactory results, cell segmentation-based counting is easily affected by staining standards and image brightness, especially when a large number of kernels overlap. This seriously affects the accuracy of the obtained scoring results.[14]

On the basis of previous studies, combined with the histopathological characteristics of $CD30^+$ cell aggregation, we proposed MPSA and considered the potential characteristics of neighbourhood patches. The accuracy of the model on the test set was improved from 93.4% to 95.7%, which verified the similarity of the structures of adjacent organisations in histopathological diagnosis. To prevent deviations in small datasets, the CRF algorithm was used to fuse context features to improve the robustness and accuracy of the developed model. In addition, the model could effectively identify false-positive areas caused by uneven staining. The heatmap output by the model showed tumour areas in WSIs in a highly detailed manner (figure 4). We demonstrated the effectiveness of CAD for $CD30^+$LPDs, thereby reducing diagnostic variability among dermatologists and facilitating clinical decision-making.

To adapt to the lack of rare disease data, this article also discusses the performance of using MPSANet for small samples. After conducting training on a randomly selected downsampled training set, a test was carried out on the independent internal test set. For 10% of the training set, the AUC of the test set still reached 0.965 (online supplemental figure 3). The use of an attention mechanism based on metric learning can effectively solve the problem of data scarcity for noncommon diseases in CAD. On 20% of the training set, the AUC of the test set was quite different from that achieved on 10%. After analysis, this was determined to be caused by a large difference between the randomly selected patch and the patch in the test set. The results suggested that the immunohistochemical staining problem not only interferes with the pathologist's judgement regarding the immunohistochemical results but also affects the performance of the CAD model.

In addition, to promoting MPSANet to other tumour marker types, we randomly tested immunohistochemical slides of CD3

and Ki-67 in LyP on a trained model to output a probabilistic visual heatmap (figure 5). Because the model had not learnt the immunohistochemical characteristics of CD3 and Ki-67, some false-positives were obtained. In general, the model can still accurately locate and identify the positive areas in WSIs. We assume that after conducting simple transfer learning on small datasets, the model can be applied to more automatic biomarker identification tasks. We have also evaluated our method on the publicly available Yale HER2 cohort, and its performance is highly superior, indicating a broad application potential for this method.

In this study, we strive to expand the applicability of MPSANet. However, our model may have some limitations because the used data volume was small and came from only one hospital. The model cannot adapt to all unseen data. In addition, it is difficult to obtain immunohistochemical data and the complexity, uneven staining and artefacts of the staining process also hindered the development of CAD in IHC. The above problems have not been fundamentally solved in this study, but we will devote ourselves to exploring the standardisation of immunohistochemical staining in the future, creating new methods to address these limitations and promoting the use of CAD for the evaluation of immunohistochemical markers.

Handling editor Runjan Chetty.

Acknowledgements Thanks to the Department of Dermatology, The First Affiliated Hospital of China Medical University and NHC Key Laboratory of Immunodermatology (China Medical University) for Providing histopathological services slides.

Contributors XC is the guarantor of this study. TZ wrote the manuscript and developed the network together with the assistance of KW; SZ annotated all the WSIs and evaluated the model results; HQ, QB and SL provided additional guidance on the network design part of the project; RQ and XG assisted in the annotation of all the WSIs data and guided on clinical and pathology related matters; RQ, XG, YZ and XC provided financial and laboratory equipment support.

Funding Supported by the National Nature Science Foundation of China (U1908206), 'Double hundred project' major scientific and technological achievements transformation project (Z19-4-010), Key research and development program of Liaoning Province (2019JH8/10300003),The Fundamental Research Funds for the Central Universities (N2219001), Ningbo Science and Technology Major Project (2021Z027).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by ID: 2020229. This study protected the privacy of the patients by receiving their consent and by obtaining the approval of the ethics committee of the First Hospital of China Medical University. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Our Code is available at https://github.com/titizheng/MPSANet_BiomarkerScores. The Yale HER2 cohort dataset was obtained from SamanFarahmand's project (https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=119702524). As the use of this dataset is subject to critical review by the hospital, if other researchers require this dataset, they can contact the corresponding author of this paper.

ORCID iD

Xiaoyu Cui http://orcid.org/0000-0002-0585-9813

REFERENCES

1 Vena JA, Copel LC. Cancer survivorship and quality of life outcomes of adolescents and young adults with lymphoma: an integrative review. *Eur J Oncol Nurs* 2021;52:101948.
2 Willemze R, Cerroni L, Kempf W, *et al*. The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas. *Blood* 2019;133:1703–14.
3 Duvic M, Tetzlaff MT, Gangar P, *et al*. Results of a phase II trial of Brentuximab Vedotin for CD30$^+$cutaneous T-cell lymphoma and lymphomatoid papulosis. *J Clin Oncol* 2015;33:3759–65.
4 Di Raimondo C, Parekh V, Song JY, *et al*. Primary cutaneous CD30$^+$lymphoproliferative disorders: a comprehensive review. *Curr Hematol Malig Rep* 2020;15:333–42.
5 Martinez-Cabriales SA, Walsh S, Sade S, *et al*. Lymphomatoid papulosis: an update and review. *J Eur Acad Dermatol Venereol* 2020;34:59–73.
6 Chen C, Gu YD, Geskin LJ. A Review of Primary Cutaneous CD30$^+$ Lymphoproliferative Disorders. *Hematol Oncol Clin North Am* 2019;33:121–34.
7 Ramos-Vara JA. Principles and methods of immunohistochemistry. *Methods Mol Biol* 2017;1641:115–28.
8 Rao IS. Role of immunohistochemistry in lymphoma. *Indian J Med Paediatr Oncol* 2010;31:145–7.
9 Kone D, Atimere YN, Coulibaly ZI, *et al*. Contribution of immunohistochemistry in the diagnosis of lymphomas. *Open J Blood Dis* 2018;08:17–25.
10 Tschandl P, Rosendahl C, Akay BN, *et al*. Expert-Level diagnosis of nonpigmented skin cancer by combined Convolutional neural networks. *JAMA Dermatol* 2019;155:58–65.
11 Xu Y, Jiang L, Huang S. Dual resolution deep learning network with self-attention mechanism for classification and localisation of colorectal cancer in histopathological images. *J Clin Pathol*. [Epub ahead of print: 10 Mar 2022].
12 Lee MKI, Rabindranath M, Faust K, *et al*. Compound computer vision workflow for efficient and automated immunohistochemical analysis of whole slide images. *J Clin Pathol* 2022. doi:10.1136/jclinpath-2021-208020. [Epub ahead of print: 15 Feb 2022].
13 Qaiser T, Rajpoot NM. Learning where to see: a novel attention model for automated immunohistochemical scoring. *IEEE Trans Med Imaging* 2019;38:2620–31.
14 Khameneh FD, Razavi S, Kamasak M. Automated segmentation of cell membranes to evaluate HER2 status in whole slide images using a modified deep learning network. *Comput Biol Med* 2019;110:164–74.
15 Saha M, Chakraborty C. Her2Net: a deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. *IEEE Trans Image Process* 2018;27:2189–200.
16 Qaiser T, Mukherjee A, Reddy Pb C, *et al*. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology* 2018;72:227–38.
17 Negahbani F, Sabzi R, Pakniyat Jahromi B, *et al*. PathoNet introduced as a deep neural network backend for evaluation of Ki-67 and tumor-infiltrating lymphocytes in breast cancer. *Sci Rep* 2021;11:8489.
18 S L, Sai Ritwik KV, Vijayasenan D, *et al*. Deep learning model based Ki-67 index estimation with automatically labelled data. *Annu Int Conf IEEE Eng Med Biol Soc* 2020;2020:1412–5.
19 Liu S, Zhang B, Liu Y, *et al*. Unpaired stain transfer using Pathology-Consistent constrained generative Adversarial networks. *IEEE Trans Med Imaging* 2021;40:1977–89.
20 Saha M, Chakraborty C, Arun I, *et al*. An advanced deep learning approach for Ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer. *Sci Rep* 2017;7:3213.
21 Stenman S, Bychkov D, Kucukel H, *et al*. Antibody supervised training of a deep learning based algorithm for leukocyte segmentation in papillary thyroid carcinoma. *IEEE J Biomed Health Inform* 2021;25:422–8.
22 Aprupe L, Litjens G, Brinker TJ, *et al*. Robust and accurate quantification of biomarkers of immune cells in lung cancer micro-environment using deep convolutional neural networks. *PeerJ* 2019;7:e6335.
23 World Medical Association. World Medical association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4.
24 Farahmand S, Fernandez AI, Ahmed FS, *et al*. Deep learning trained on hematoxylin and eosin tumor region of interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. *Mod Pathol* 2022;35:44–51.
25 Bándi P, Bulten W. Program for the analysis and visualization of whole-slide images in digital pathology, 2020. Available: https://github.com/computationalpathologygroup/ASAP [Accessed 26 Oct 2020].
26 Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 1979;9:62–6.
27 He K, Zhang X, Ren S, *et al*. *Deep residual learning for image recognition. 2016 IEEE conference on computer vision and pattern recognition (CVPR)*, 2016: 770–8.
28 Zagoruyko S, Komodakis N. Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer. arXiv[csCV], 2017. Available: https://arxiv.org/abs/1612.03928 [Accessed 12 Feb 2017].
29 Lafferty J, Mccallum A, Pereira F. Conditional random fields: probabilistic models for Segmenting and labeling sequence data. *proceedings of icml* 2002;2002:282–9.

30 Wang Y, Yao Q, Kwok J, *et al*. Generalizing from a few examples: a survey on few-shot learning. arXiv[csLG], 2019. Available: https://arxiv.org/abs/1904.05046v3 [Accessed 29 Mar 2020].
31 Schlemper J, Oktay O, Schaap M, *et al*. Attention gated networks: learning to leverage salient regions in medical images. *Med Image Anal* 2019;53:197–207.
32 Yao J, Zhu X, Jonnagaddala J, *et al*. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med Image Anal* 2020;65:101789.
33 Chen H, Li H, Li Y. Multi-level Metric Learning for Few-shot Image Recognition. arXiv[csCV], 2021. Available: https://arxiv.org/abs/2103.11383 [Accessed 12 Apr 2021].
34 Deng W, Dong R, Socher LL. ImageNet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009:248–55.
35 Yang H, Chen L, Cheng Z, *et al*. Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study. *BMC Med* 2021;19:80.
36 Vasiljevi J, Feuerhake F, Wemmert C. Self adversarial attack as an augmentation method for immunohistochemical stainings. arXiv[csCV], 2021. Available: https://arxiv.org/abs/2103.11362 [Accessed 21 Mar 2021].
37 Lu J, Behbood V, Hao P, *et al*. Transfer learning using computational intelligence: a survey. *Knowl Based Syst* 2015;80:14–23.