Translational Oncology 14 (2021) 101161

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Original Research

Diagnostic assessment of deep learning for melanocytic lesions using whole-slide pathological images

Wei Ba [a], Rui Wang [a], Guang Yin [a], Zhigang Song [b], Jinyi Zou [c], Cheng Zhong [c], Jingrun Yang [a], Guanzhen Yu [d], Hongyu Yang [e], Litao Zhang [f], Chengxin Li [a,*]

[a] Department of Dermatology, Chinese PLA General Hospital & Medical School, No. 28 Fuxing Road, Beijing 100853, China
[b] Department of Pathology, Chinese PLA General Hospital & Medical School, Beijing, China
[c] Artificial Intelligence (AI) Lab, Lenovo Research, Beijing, China
[d] Hospital of Chengdu University of Traditional Chinese Medicine, Sichuan, China
[e] St Vincent Evansville medical center, Washington, United States
[f] Department of Dermatology, Tianjin Chang Zheng Hospital, Tianjin, China

ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning algorithm
Melanoma
Nevus
Whole-slide pathological images

ABSTRACT

Background: Deep learning has the potential to improve diagnostic accuracy and efficiency in medical image recognition. In the current study, we developed a deep learning algorithm and assessed its performance in discriminating melanoma from nevus using whole-slide pathological images (WSIs).
Methods: The deep learning algorithm was trained and validated using a set of 781 WSIs (86 melanomas, 695 nevi) from PLA General Hospital. The diagnostic performance of the algorithm was tested on an independent test set of 104 WSIs (29 melanomas, 75 nevi) from Tianjin Chang Zheng Hospital. The same test set was also diagnostically classified by 7 expert dermatopathologists.
Results: The deep learning algorithm receiver operating characteristic (ROC) curve achieved a sensitivity 100% at the specificity of 94.7% in the classification of melanoma and nevus on the test set. The area under ROC curve was 0.99. Dermatopathologists achieved a mean sensitivity and specificity of 95.1% (95% confidence interval [CI]: 92.0%-98.2%) and 96.0% (95% CI: 94.2%-97.8%), respectively. At the operating point of sensitivity of 95.1%, the algorithm revealed a comparable specificity with 7 dermatopathologists (97.3% vs. 96.0%, $P = 0.11$). At the operating point of specificity of 96.0%, the algorithm also achieved a comparable sensitivity with 7 dermatopathologists (96.5% vs. 95.1%, $P = 0.30$). A more transparent and interpretable diagnosis could be generated by highlighting the regions of interest recognized by the algorithm in WSIs.
Conclusion: The performance of the deep learning algorithm was on par with that of 7 expert dermatopathologists in interpreting WSIs with melanocytic lesions. By pre-screening the suspicious melanoma regions, it might serve as a supplemental diagnostic tool to improve working efficiency of pathologists.

Introduction

Deep learning is a kind of computational algorithm, which programs itself by learning a large number of examples to achieve desired behavior, without establishing rules specifically. One of the deep learning algorithms, the convolutional neural network (CNN) has been proven to have great potential to classify images and detect objects in pictures [1, 2]. Gulshan et al. [3] and Esteva et al. [4] proved that CNN achieved high sensitivity and specificity in diabetic retinopathy and skin lesion images classification, respectively. Except for traditional images, CNN also has been demonstrated application potential in whole-slide pathological images (WSIs) recognition. Ciresan et al. [5] applied a deep learning algorithm to the task of mitosis counting for primary breast cancer grading using WSIs. Ehteshami et al. [6] demonstrated CNN achieved better diagnostic performance than a panel of 11 pathologists in detection of lymph node metastasis of breast carcinoma. Song et al. [7] developed a clinically applicable histopathological diagnosis system for gastric cancer detection using CNN.

Melanoma is the fifth most common form of cancer in adults and is potentially the most deadly form of skin cancer [8]. The incidence of

* Corresponding author.
E-mail addresses: chengxinderm@163.com, 443966199@qq.com (C. Li).

https://doi.org/10.1016/j.tranon.2021.101161
Received 10 March 2021; Received in revised form 25 May 2021; Accepted 20 June 2021

melanoma continues to increase in many countries [9-11]. It was estimated that 232,100 new cases were diagnosed with cutaneous melanoma in the world annually [8]. Accurate diagnosis for melanoma is an important task for pathologists. This process requires highly experienced pathologists and is time-consuming and error-prone [12, 13]. And inter and intra-observer variability exists in the pathological diagnosis of melanocytic lesions [13]. In addition, there is a critical shortage of pathologists both nationally and globally, which has created overloaded workforces, thus effecting diagnostic accuracy. The aim of our research was to investigate the potential of CNN to discriminate melanoma from nevus using WSIs and compare its performance with dermatopathologists.

Patients and methods

Image data and reference standard diagnosis

A total of 885 melanocytic skin lesions (115 melanomas, 770 nevi) were retrospectively selected from patients aged 20 years and older who underwent surgery for clinical care at PLA General Hospital (86 melanomas, 695 nevi) and Tianjin Chang Zheng Hospital (29 melanomas, 75 nevi). The dataset represented the most common melanocytic lesions encountered in routine pathology practice (Table 1).

Because the complexity and observer variability existed in the pathological diagnosis of melanocytic lesions [12-15], reference gold standard diagnosis was established for each of 885 cases. Three senior specialists in pathology department independently reviewed the glass slides and given the diagnosis of each case into one of the five MPATH-Dx (Melanoma Pathology Assessment & Treatment Hierarchy for Diagnosis) categories: 1) Nevus/mild atypia; 2) Moderate atypia; 3) Severe atypia/Melanoma in situ; 4) T1a invasive melanoma; and 5) ≥T1b invasive melanoma [16]. For the cases with inconsistent diagnoses, all three dermatopathologists reviewed the case together using a multiheaded microscope. A consensus diagnosis was reached [13]. We assumed cases in MPATH-DX categories 1 and 2 as nevus and cases in MPATH-DX categories 3 to 5 as melanoma.

These 885 slides were scanned into WSIs using a digital slide scanner (PRECICE 500B) with a 40x objective lens (specimen-level pixel size, 0.25 μm × 0.25 μm). All WSIs must pass an image quality review before they were made available for inclusion in this study.

Ethics statement

Institutional review board approval was obtained for a multi-center retrospective study with the PLA General Hospital & Medical School as the coordinating center (Approval no. S2018–123–01). The informed consents were waived by the institutional review board since the slides were anonymized.

Image annotation

Our deep learning algorithm was trained using a supervised learning method, which requires manual annotations. Regions of severe atypical and melanoma presented in the WSIs (MPATH-DX categories 3 to 5) were first manually annotated by 2 dermatopathologists, and every WSI was then checked in detail by 2 senior specialists (using the labeling software wkentaro/labelme, Image Polygonal Annotation with Python, https://github.com/wkentaro/labelme). There was no annotation on WSIs of nevi (MPATH-DX category 1 to 2).

Training, validation and test datasets

In the process of deep learning algorithm development, the training dataset is a set of examples used for the algorithm to learn, which means it is used to fit the parameters of the algorithm (e.g., weights of connections between neurons in CNN). A validation set is used to optimize the algorithm [17]. A test set is independent of the training and validation set, which is used to evaluate the performance of a fully specified algorithm [17].

Table 1
Characteristics of the whole-slide images for melanocytic lesions.

Dataset	MPATH-DX categories	Terms on histology type	Training and Validation from PLA (781 WSI)	Test from CZ (104 WSI)	Total (885 WSI)
Common nevi (770 WSI)	Nevus/mild atypia (733WSI)	Junctional nevus	150	15	165
		Compound nevus	150	15	165
		Intradermal nevus	150	15	165
		Blue nevus	150	15	165
		Atypia-Atypical nevus NOS, mild	67	6	73
	Moderate atypia (37 WSI)	Atypia-Atypical nevus NOS, moderate	28	9	37
Melanoma (115 WSI)	Severe atypia/ Melanoma in situ (25 WSI)	Atypia-Atypical nevus NOS, severe	6	2	8
		Melanoma in situ	13	4	17
	T1a invasive melanoma (31 WSI)	Acral melanoma	12	4	16
		lentigo maligna melanoma	6	2	8
		Superficial spreading melanoma	5	2	7
	≥T1b invasive melanoma (59 WSI)	Acral melanoma	25	8	33
		lentigo maligna melanoma	5	2	7
		Superficial spreading melanoma	6	2	8
		Nodular melanoma	5	2	7
		Metastatic melanoma	3	1	4

Abbreviations: MPATH-DX, Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis; WSIs, whole-slide images; NOS, Not otherwise specified (including atypical nevus of special anatomic site: acral, genital, flexural, breast, scalp, ect); PLA, PLA General Hospital; CZ, Tianjin Chang Zheng Hospital.

A total of 781 WSIs (86 melanomas, 695 nevi) from PLA General Hospital were used for training and validation. Considering the unbalance of melanoma and nevus samples, oversampling method were adopted in deep learning algorithm development (details in supplementary method). Cross-validation was applied to compare and select the best algorithm. The 104 WSIs (29 melanomas, 75 nevi) from Tianjin Chang Zheng Hospital, were used to test the screening performance of the fully specified algorithm (Table 1).

Deep learning algorithm development

We introduced an efficient framework for WSI prediction. The proposed framework was conducted by integrating three modules: a region of interest tissue (RIT) extraction module, a CNN-based melanoma patch detection module, and a slide-level classification module. Fig. 1 showed the overall scheme of our algorithm, of which the details were illustrated in supplementary method. We also provided a movie to show the workflow of deep learning algorithm (Supplementary Movie).

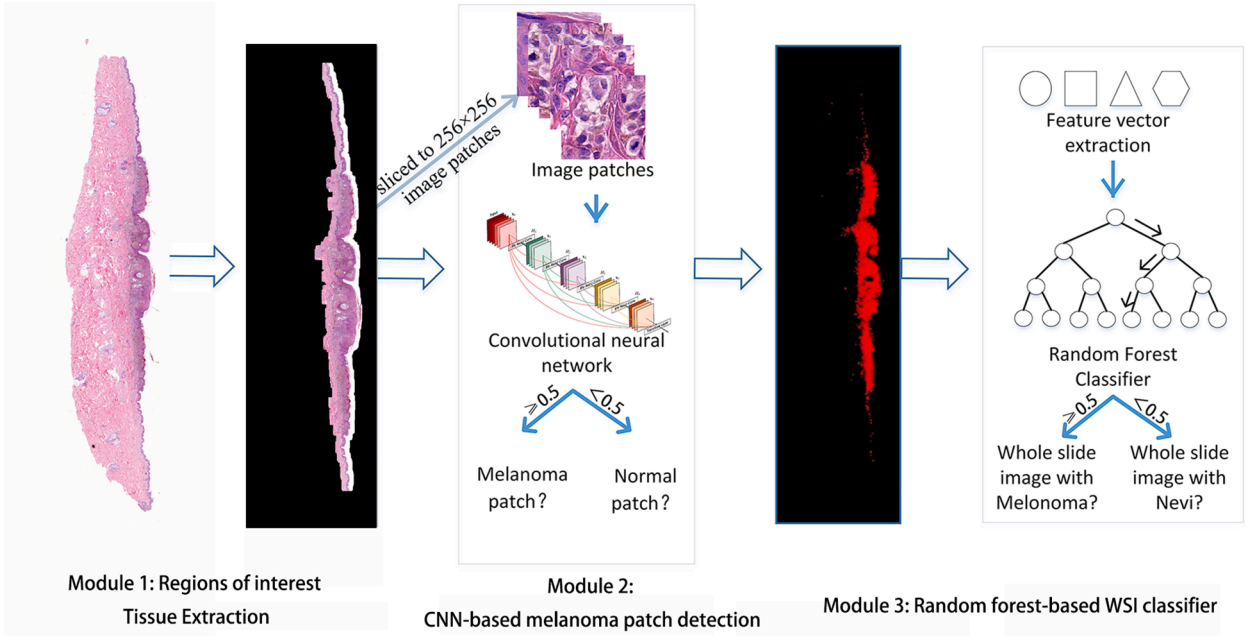

Fig. 1. Overall architecture of the prediction framework. The proposed framework was conducted by integrating three modules: a region of interest tissue (RIT) extraction module, a convolutional neural network (CNN)-based melanoma patch detection module, and a slide-level classification module. First, RIT module proposed candidate tissue regions from WSI. Second, CNN-based melanoma patch detection module predicted melanoma and non-melanoma patches within RIT. If the WSI prediction only relied on CNN-based patch classifier, a single suspicious melanoma patch can change the WSI prediction, possibly resulting in a large number of false positives. So, we introduced the random forest to determine each WSI result. Third, the predicted patches were converted to a heat map. Based on the morphological and geometrical information of the heat map, a random forest-based classifier was built to determine a slide-level prediction.

Statistical analysis

The main outcome measures were the sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) for the diagnostic classification of melanoma and nevus using WSIs by the algorithm versus a panel of 7 pathologists.

To compare the algorithm with the 7 pathologists, a two-sided, one-sample *t*-test was applied to compare the sensitivity (specificity) of the deep learning algorithm with the average of that on 7 pathologists. The results were considered statistically significant at the $P<0.05$ level. All the data were analyzed using SPSS Version 22.

Results

Comparison of the deep learning algorithm to pathologists

To evaluate the deep learning algorithm in the context of 7 expert pathologists, a test set of 104 WSIs, independent of the training and validation set, was used to compare our algorithm predicted diagnosis with the diagnosis made by pathologists.

We asked 7 dermatopathologists (all with over 20 years of cutaneous pathology experience) to independently assess 104 test WSIs with their diagnoses mapped into one of five MPATH-DX categories (categories 1 to 2 assumed as nevus and categories 3 to 5 as melanoma), without knowledge of others' diagnoses.

All dermatopathologists voluntarily participated in this study, understood and agreed with the basic principles and objectives of this study. The 7 dermatopathologists evaluated the corresponding glass slides for the WSIs, because they used to making diagnosis using a microscope. There was no time restriction, and participants could complete evaluations over multiple sittings.

The sensitivity and specificity of the 7 dermatopathologists for the classification of melanoma and nevus were 95.1% (95% confidence interval [CI]: 92.0%−98.2%) and 96.0% (95% CI: 94.2%−97.8%), respectively (Fig. 2a). Using the dermatopathologists' mean sensitivity (95.1%) as the operating point on the algorithm's ROC curve, there was no statistical difference between the specificity of algorithm (97.3%) and the mean specificity of the dermatopathologists (96.0%, $P = 0.11$). Using the pathologists' mean specificity (96.0%) as the operating point on the algorithm's ROC curve, there was no statistical difference between the algorithm's sensitivity (96.5%) and the mean sensitivity of dermatopathologists (95.1%, $P = 0.30$).

Weighted error scoring

In the pathological diagnosis of melanocytic lesions, failing to diagnose (a false-negative result) is more harmful than making a melanoma diagnosis when it was not (a false-positive result) [18] . We defined the weighted error scoring to reflect that fact (Supplemental Method). In the defined formula, the lower score presented the better diagnostic performance. The deep learning algorithm yielded a score of 1.82%, and the weighted error of the 7 dermatopathologists ranged from 1.36% to 7.27%, with a mean weighted error of 4.61%. As shown in Fig. 2b, the deep learning algorithm outperformed almost all except one dermatopathologist based on this weighted error scale.

Malignancy identification

A more transparent and explanatory diagnosis were provided by highlighting the predicted melanoma regions by deep learning algorithm in WSIs (Supplementary Movie). Fig. 3 showed examples of regions of melanoma annotated by dermatopathologists and predicted by

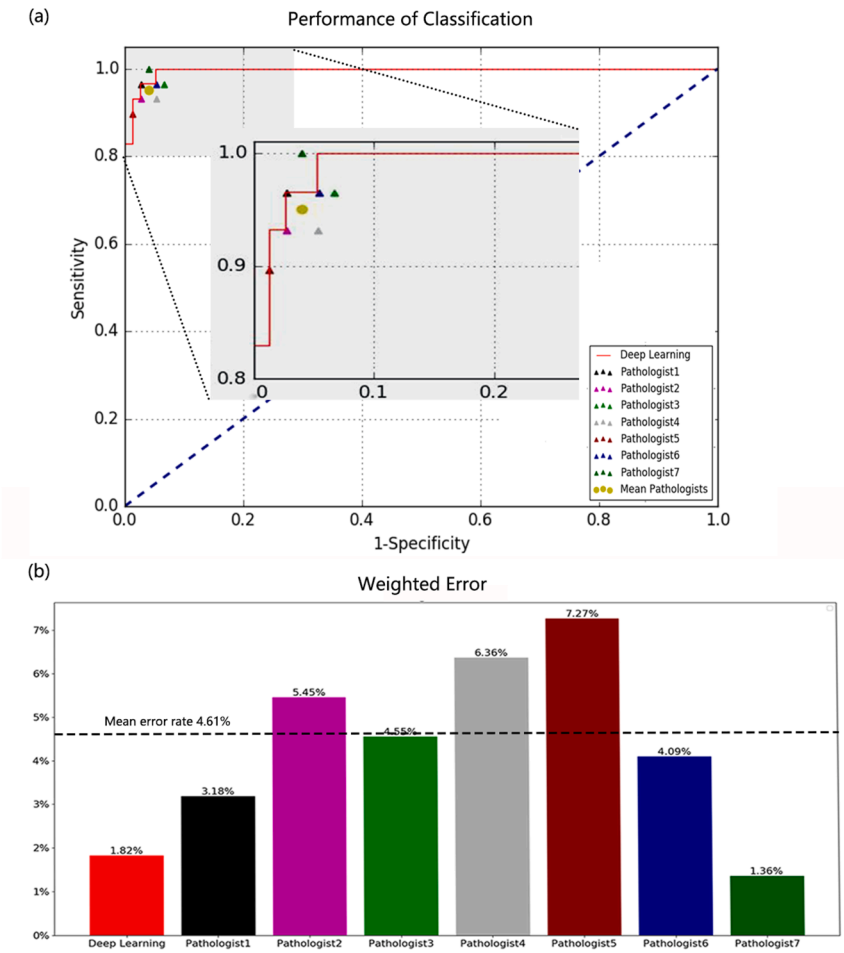

Fig. 2. Comparison of the deep learning algorithm with 7 dermatopathologists in the classification of melanoma and nevus. (a). ROC curve for the deep learning algorithm compared with the performance of human pathologists. The enlarged area was included for a detailed illustration of the comparison results. (b). Weighted error was used to reflect the fact that a false-negative result (failing to diagnose) was more detrimental than a false-positive result (making a melanoma diagnosis when it was not). The deep learning algorithm outperformed all except one dermatopathologist based on the weighted error scale. And the dotted line represented mean weighted error of 7 dermatopathologists.

the deep learning algorithm.

To evaluate the identification accuracy of melanoma regions by the algorithm, the *Jaccard index* was adopted to compute the overlap between the predicted melanoma regions and manually annotated regions (Supplemental Method). The deep learning algorithm achieved a *Jaccard index* of 77.78%, indicating that most of the melanoma regions were identified correctly by the algorithm.

Discussion

Deep learning shows promise for the application in image recognition and is approaching human performance in many tasks, including pathological image classification [19-21]. Ehteshami et al. demonstrated that a deep learning algorithm outperformed a group of 11 pathologists in the detection of lymph node metastases of breast cancer [6]. Hekler et al. demonstrated that CNN achieved better performance than 11 histopathologists in the classification of pathological images of melanocytic lesions [22, 23]. Different from Hekler's research using randomly cropped images from WSI, our study directly used WSI to compare CNN to pathologists. The randomly cropped images may lead to information lost. In a clinical setting, pathologists scanned the WSI rather than cropped images. Our results demonstrated that the performance of the deep learning algorithm was comparable to that of a panel of 7 dermatopathologists. Due to the increased cancer incidence and personalized treatment options, pathologists currently have to review a large number of slides, including immunohistochemistry, to make a final diagnosis. Therefore, this technique might be employed as a supplemental screening tool to demonstrate a biopsy as severe atypia or melanoma.

The sensitivity, specificity and AUC did not accurately reflect the impact of incorrect diagnosis on an individual patient. Compared to a false-positive result (nevus was inaccurately diagnosed with melanoma), a false-negative result (melanoma was diagnosed with nevus) could have more serious consequences. To solve this problem, weighted error scoring was incorporated during the algorithm and the pathologists' evaluation. Our study demonstrated that the deep learning algorithm outperformed almost all except one pathologist based on the weighted error scoring.

The greatest benefit of regions of interest identification was that it revealed insights into the decisions of the network. The majority regions of melanoma (*Jaccard index=77.78%*) were recognized correctly, which confirmed that the neural network used accurate distinguishing features to made its decisions. CNN could highlight abnormal areas, prompting pathologists to perform a scrutinized reassessment and may facilitate their efficiency and accuracy.

Due to the fact that a false-negative result could lead to more serious consequences than false-positive in melanoma diagnosis. Moreover, the sensitivity and specificity of neural network are adjustable. In practice, we can set the sensitivity of deep learning algorithm to 100%. In this setting, the deep learning algorithm was capable of avoiding false-negative result as much as possible. Then the pathologists could check all the suspected melanoma regions, which might improve the accuracy without significantly decrease the efficiency.

This study had several limitations. These analyses are performed as a simulation process rather than a real pathological workflow. The test dataset on which the deep learning algorithm and the pathologists were evaluated was enriched with the cases of melanoma, which was not directly comparable to the mixed cases encountered during clinical practice. In our study, each pathologist was given a single H&E slide for per case to determine the diagnosis. In a real clinical setting,

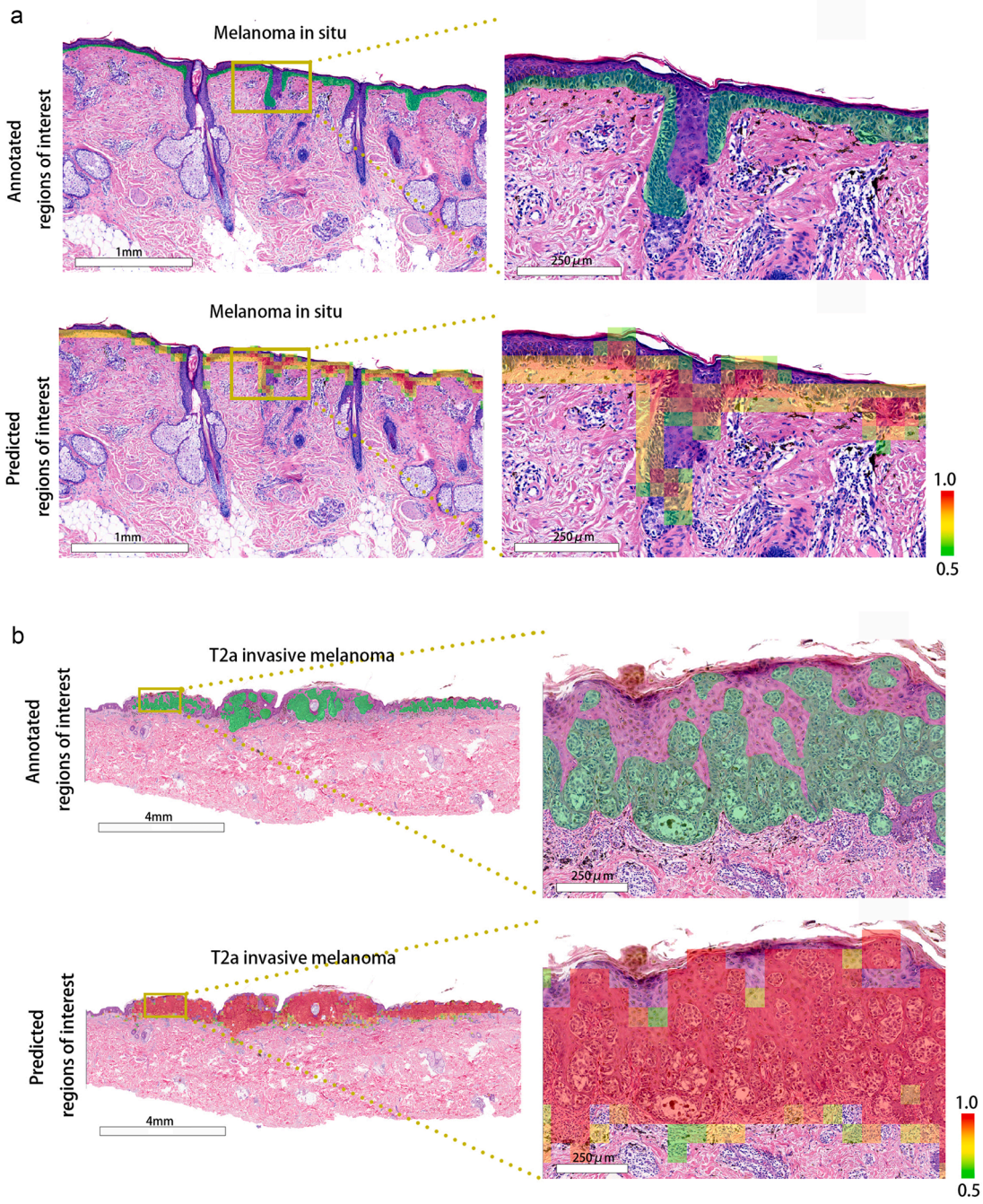

Fig. 3. Examples of regions of melanoma annotated by the dermatopathologists (Top) and predicted by the deep learning algorithm (Bottom). (a)-(c). Three annotated melanoma WSIs and local magnification taken from the test dataset compared with predicted regions of melanoma by deep learning algorithm. The colored scale bar (lower right) indicated the probability for each patch to be malignancy.

pathologists request additional IHC staining and clinical data, which might have improved diagnostic performance [24].

Moreover, the histology subtypes of melanocytic lesions were complex. Our study only included the most common melanocytic lesions encountered in routine pathological workflow. And differences in the inherent difficulty of the test set will directly affect the diagnostic performance of the deep learning algorithm and the pathologists. To generate comparability among different algorithms and pathologists, it is important to include more pathologists with various levels of experience and to use larger datasets. In our study, the data set comes from Asians. The histological subtypes of melanomas are different between Caucasians and Asians. It is possible that the algorithm is likely to perform poorer on external test set. Future studies could use WSIs from multiple centers in training and test datasets, making the system universally useful. In addition to confirming our results with larger, more diverse datasets, prospective studies are needed to address the pathologists' acceptance of this new technique [25, 26].

In conclusion, our research demonstrated that an adequately trained deep learning algorithm could have a high degree of accuracy in distinguishing melanoma from nevus using WSI. Thus, a deep learning algorithm might function as a supplemental diagnostic tool to assist pathologist in demonstrating a biopsy with severe atypia or melanoma.

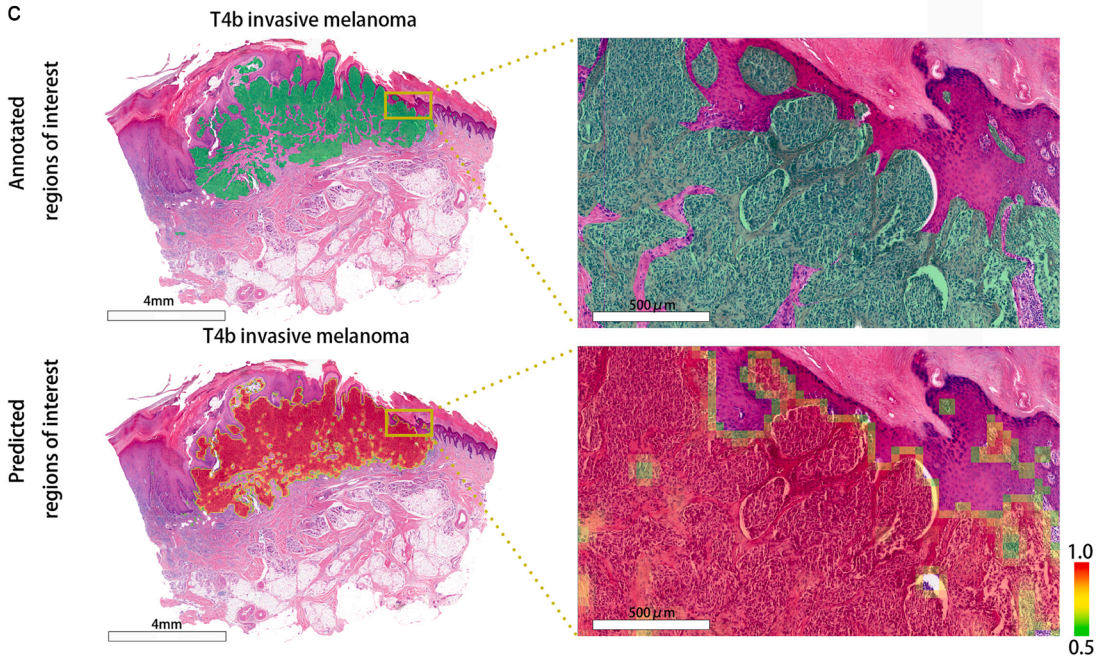

Fig. 3. (*continued*).

Author contribution

Study concepts and design: Wei Ba, Rui Wang and Chengxin Li
Data acquisition: Wei Ba, Guang Yin, Zhigang Song, Jinyi Zou, Litao Zhang, Cheng Zhong and Jingrun Yang
Data analysis and interpretation: Wei Ba, Hongyu (Henry) Yang, Guanzhen Yu and Chengxin Li
Manuscript preparation,editing and review: Wei Ba and Chengxin Li

Declaration of Competing Interest

None

Acknowledgments

This study was supported by translational medicine program of PLA General Hospital 2018TM-10.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101161.

References

[1] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: a comprehensive review, Neural Comput. 29 (2017) 2352–2449.
[2] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Trans. Med. Imaging 19 (2000) 203–210.
[3] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (2016) 2402–2410.
[4] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (2017) 115–118.
[5] D.C. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, Med. Image Comput. Comput. Assist. Interv. 16 (2013) 411–418.
[6] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, JAMA 318 (2017) 2199–2210.
[7] Z. Song, S. Zou, W. Zhou, Y. Huang, L. Shao, J. Yuan, et al., Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning, Nat. Commun. 11 (2020) 4294.
[8] D. Schadendorf, A.C.J. van Akkooi, C. Berking, K.G. Griewank, R. Gutzmer, A. Hauschild, et al., Melanoma, Lancet 392 (2018) 971–984.
[9] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2018, CA Cancer J. Clin. 68 (2018) 7–30.
[10] V. Nikolaou, A.J. Stratigos, Emerging trends in the epidemiology of melanoma, Br. J. Dermatol. 170 (2014) 11–19.
[11] W.O. Jones, C.R. Harman, A.K. Ng, J.H. Shaw, Incidence of malignant melanoma in Auckland, New Zealand: highest rates in the world, J. World J. Surg. 23 (1999) 732.
[12] R.P. Braun, D. Gutkowicz-Krusin, H. Rabinovitz, A. Cognetta, R. Hofmann-Wellenhof, V. Ahlgrimm-Siess, et al., Agreement of dermatopathologists in the evaluation of clinically difficult melanocytic lesions: how golden is the 'gold standard'? Dermatology 224 (2012) 51–58.
[13] P.A. Carney, L.M. Reisch, M.W. Piepkorn, R.L. Barnhill, D.E. Elder, S. Knezevich, et al., Achieving consensus for the histopathologic diagnosis of melanocytic lesions: use of the modified Delphi method, J. Cutan. Pathol. 43 (2016) 830–837.
[14] G. Ferrara, G. Argenziano, H.P. Soyer, R. Corona, F. Sera, B. Brunetti, et al., Dermoscopic and histopathologic diagnosis of equivocal melanocytic skin lesions: an interdisciplinary study on 107 cases, Cancer 95 (2002) 1094–1100.
[15] S.L. Edwards, K. Blessing, Problematic pigmented lesions: approach to diagnosis, J. Clin. Pathol. 53 (2000) 409–418.
[16] M.W. Piepkorn, R.L. Barnhill, D.E. Elder, S.R. Knezevich, P.A. Carney, L.M. Reisch, et al., The MPATH-Dx reporting schema for melanocytic proliferations and melanoma, J. Am. Acad. Dermatol. 70 (2014) 131–141.
[17] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, 2013.
[18] D.S. Kermany, M. Goldbaum, W. Cai, C.C.S. Valentim, H. Liang, S.L. Baxter, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (2018) 1122–1131, e9.
[19] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[20] G. Litjens, C.I. Sanchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, et al., Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, Sci. Rep. 6 (2016) 26286.
[21] D. Gutman, N.C.F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, et al., Skin lesion analysis toward melanoma detection, in: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC), 2016.

[22] A. Hekler, J.S. Utikal, A.H. Enk, W. Solass, M. Schmitt, J. Klode, et al., Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images, Eur. J. Cancer 118 (2019) 91–96.
[23] A. Hekler, J.S. Utikal, A.H. Enk, C. Berking, J. Klode, D. Schadendorf, et al., Pathologist-level classification of histopathological melanoma images with deep neural networks, Eur. J. Cancer 115 (2019) 79–83.
[24] M. Binder, H. Kittler, S. Dreiseitl, H. Ganster, K. Wolff, H. Pehamberger, Computer-aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process, Melanoma Res. 10 (2000) 556–561.
[25] H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, et al., Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists, Ann. Oncol. (2018).
[26] M.A. Marchetti, N.C.F. Codella, S.W. Dusza, D.A. Gutman, B. Helba, A. Kalloo, et al., Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images, J. Am. Acad. Dermatol. 78 (2018) 270–277, e1.