The American Journal of Pathology, Vol. 193, No. 12, December 2023

ELSEVIER

The American Journal of PATHOLOGY

ajp.amjpathol.org

MACHINE LEARNING, COMPUTATIONAL PATHOLOGY, AND BIOPHYSICAL IMAGING

Tumor-Infiltrating Lymphocyte Recognition in Primary Melanoma by Deep Learning Convolutional Neural Network

Check for updates

Filippo Ugolini,* Francesco De Logu,† Luigi Francesco Iannone,† Francesca Brutti,‡ Sara Simi,* Vincenza Maio,* Vincenzo de Giorgi,§ Anna Maria di Giacomo,¶ Clelia Miracco,¶ Francesco Federico,‖ Ketty Peris,** Giuseppe Palmieri,†† Antonio Cossu,‡‡ Mario Mandalà,§§ Daniela Massi,* and Marco Laurino‡

From the Sections of Pathological Anatomy, Clinical Pharmacology and Oncology,† and Dermatology,§ Department of Health Sciences, University of Florence, Florence; the Institute of Clinical Physiology,‡ National Research Council, Pisa; the Center for Immuno-Oncology, Medical Oncology and Immunotherapy,¶ University of Siena, Siena; the Institutes of Pathology‖ and Dermatology,** Sacred Heart Catholic University, Rome; the Unit of Cancer Genetics,†† Institute of Genetic and Biomedical Research, National Research Council, Sassari; the Department of Medical, Surgical and Experimental Sciences,‡‡ University of Sassari, Sassari; and the Oncology Unit,§§ Department of Medicine and Surgery, University of Perugia, Perugia, Italy*

Accepted for publication August 2, 2023.

Address correspondence to Marco Laurino, Ph.D., Institute of Clinical Physiology, National Research Council, Via Moruzzi 1, 56124 Pisa, Italy.
E-mail: marco.laurino@cnr.it.

The presence of tumor-infiltrating lymphocytes (TILs) is associated with a favorable prognosis of primary melanoma (PM). Recently, artificial intelligence (AI)-based approach in digital pathology was proposed for the standardized assessment of TILs on hematoxylin and eosin—stained whole slide images (WSIs). Herein, the study applied a new convolution neural network (CNN) analysis of PM WSIs to automatically assess the infiltration of TILs and extract a TIL score. A CNN was trained and validated in a retrospective cohort of 307 PMs including a training set (237 WSIs, 57,758 patches) and an independent testing set (70 WSIs, 29,533 patches). An AI-based TIL density index (AI-TIL) was identified after the classification of tumor patches by the presence or absence of TILs. The proposed CNN showed high performance in recognizing TILs in PM WSIs, showing 100% specificity and sensitivity on the testing set. The AI-based TIL index correlated with conventional TIL evaluation and clinical outcome. The AI-TIL index was an independent prognostic marker associated directly with a favorable prognosis. A fully automated and standardized AI-TIL appeared to be superior to conventional methods at differentiating the PM clinical outcome. Further studies are required to develop an easy-to-use tool to assist pathologists to assess TILs in the clinical evaluation of solid tumors. *(Am J Pathol 2023, 193: 2099—2110; https://doi.org/10.1016/j.ajpath.2023.08.013)*

Primary melanoma (PM) is the most dangerous form of skin tumor and causes 90% of skin cancer mortality.[1] Patients with PM are staged primarily according to the American Joint Committee on Cancer[2] criteria, including Breslow thickness, ulceration, sentinel lymph node (SLN) status, and presence of distant metastases.[3] Numerous studies have identified additional factors beyond American Joint Committee on Cancer staging with varying degrees of prognostic impact. Despite the heterogeneity of the studies, there is robust evidence that in PM the presence of tumor-infiltrating lymphocytes (TILs) predicts SLN status[4–7] and improves survival.[7–11]

In routine diagnostic practice, TILs in PM are classified according to their distribution and intensity as "brisk," "non-brisk," or "absent," according to criteria originally set forth by Clark et al.[12] Alternative methods of TIL classification based on the density (absent/mild/moderate/marked; score, 0 to 3) and distribution (absent/focal/multifocal/diffuse;

Supported by the Fondazione Associazione Italiana per la Ricerca sul Cancro Programma di Ricerca 5 per Mille 2018 grant 21073, Tuscany Region—Bando Ricerca Salute 2018 Telemo project, and THE—Tuscany Health Ecosystem grant ECS_00000017.

F.U. and F.D.L. contributed equally to this work.

D.M. and M.L. contributed equally to this work.

Copyright © 2023 American Society for Investigative Pathology. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).
https://doi.org/10.1016/j.ajpath.2023.08.013

score, 0 to 3) of the immune infiltrate[5] remain to be validated. In The Cancer Genome Atlas project, a modified TIL seven-tier scoring system was used (mostly in metastatic melanomas) for correlation with mRNA expression profiling and prognosis.[13] However, further studies are required to adapt the TIL scoring system to the metastatic setting and, especially for metastatic lymph nodes, the interpretation and interobserver agreement may be challenged by the presence of pre-existing lymphoid stroma.[14]

To date, standardized protocols for TIL immunophenotyping and quantification are lacking.[15–17] Despite the limited availability of specific large studies, the immunophenotype seems to play a role in the efficacy of the local immune response against melanoma. Previous studies, focused mostly on T-lymphocyte populations, showed that the prevalence of CD8+ T cells is related to a better clinical outcome as compared to T-regulatory infiltration.[18–20]

The development of a convolution neural network (CNN) for image-based automated assessment of TILs on hematoxylin and eosin–stained sections based on an artificial intelligence (AI) approach promises to be a useful tool. The implementation of an easy-to-use and standardized digital analysis solution able to offer fast and accurate information to pathologists is one of the great challenges of recent years.[21,22] Abousamra et al[23] recently implemented a deep learning–based mapping of TILs in whole slide images (WSIs) of multiple cancer types, including PMs. A nonoverlapping patch-wise classification approach was used to train a CNN model. The performance was satisfactory, and the model achieved a polyserial correlation coefficient of 0.82 over PM slides. Acs et al[24] developed an algorithm (NN192) to automatically extract a TIL score from a WSI showing that higher TIL scores are associated with a favorable prognosis. They collected different features and used a neural network with eight hidden layers for cell classification. Recently, the NN192 algorithm was used by Aung et al[25] to validate the association of an automated TIL score with survival in patients with melanoma. This study highlighted the importance of automated TIL scores as a robust prognostic marker in patients with melanoma.

Moore et al[26] proposed a study using an automated digital TIL analysis to evaluate TILs in early stage melanoma to predict disease-specific survival. In a prior study,[27] they implemented a CNN able to detect TILs in 13 tumor types, including melanoma, with a global area under the receiver operating characteristic curve of 0.954 on a test cohort of 2480 patches; and evaluated the extracted automated digital TIL analysis score as a prognostic factor. The validation of more than 145 patients provided by two different hospitals confirmed that the automated digital TIL analysis score correlated with the Clark et al[12] criteria and significantly improved multivariable Cox analysis considering ulceration and depth. Their method requires the manual extraction of the tumor area to compute the automated digital TIL analysis score.

The current study developed a CNN to recognize lymphocytic density within the tumor area from the hematoxylin and eosin WSI to automatically extract a TIL score. It also used preliminary multivariate survival analyses to evaluate the prognostic potential of the proposed AI-based TIL score. The final aim was to obtain a standardized AI-based TIL score useful for pathologists to assess PM prognosis.

Materials and Methods

Sample Population and Data Set

The study included a retrospective collection of formalin-fixed, paraffin-embedded stage II to III invasive PMs (N = 307), comprising a training set (N = 237 WSIs) from the University of Florence (N = 115; data set 1), University of Sassari (N = 43; data set 2), University of Siena (N = 15; data set 3), and Papa Giovanni XXIII Cancer Center Hospital Bergamo (N = 64; data set 4) from 2000 to 2015, and a disjoint testing set (N = 70; data set 5) from Catholic University of the Sacred Heart, Fondazione Policlinico Universitario A. Gemelli IRCCS Rome from 2016 to 2020 (Figure 1). The clinical and pathologic parameters extracted from the database included sex, age (continuous variable), date of the primary tumor diagnosis, ulceration status (absent/present), Breslow thickness (in millimeters), histotype, Clark et al[12] level, mitotic rate, TILs, sentinel lymph node biopsy (SLNB) status, and follow-up evaluation.

Representative hematoxylin and eosin slides were reviewed, and the histopathologic features were re-assessed by dedicated experienced dermatopathologists (V.M. and D.M.), manually labeling melanoma regions of each WSI by detecting TILs and no TIL regions. The regions labeled by dermatopathologic experts were selected, avoiding borderline and uncertain visually recognizable as TIL or no TIL areas. The tumor stage was assessed according to the American Joint Committee on Cancer.[2] TILs were assessed as brisk, non-brisk, and absent according to conventional criteria formulated by Clark et al.[12] Specifically, lymphocytes had to surround and disrupt tumor cells in the vertical growth phase to be defined as TILs. These lymphocytes were termed “brisk” if they infiltrated the entire invasive component diffusely or across the base of the vertical growth phase. TILs were termed “absent” if no lymphocytes were present or if they were present but did not infiltrate the tumor. When lymphocytes only infiltrated the melanoma focally with one or scattered foci, the term “non-brisk” was used.

The use of formalin-fixed, paraffin-embedded sections of human samples was approved by the local Ethics Committee (13676_bio; 17033_bio; Comitato Etico Regione Toscana-Area Vasta Centro) according to the Declaration of Helsinki.

Immunohistochemistry

Representative 3-μm–thick, formalin-fixed, paraffin-embedded tissue sections of PMs were selected for

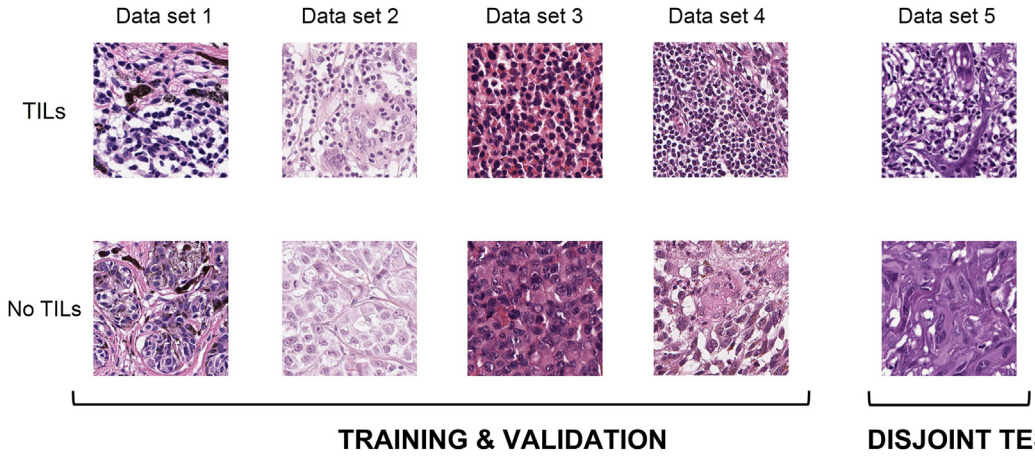

Figure 1 Examples of patches extracted from the four data sets used for the training and validation of the network, and from the disjoint data set used for testing the model in tumor-infiltrating lymphocyte (TIL) recognition. Original magnification: ×400.

immunohistochemical analysis. Sample processing was performed with the Ventana Discovery Ultra immunostainer (Ventana Medical Systems, Tucson, AZ). The sections were deparaffinized in EZ prep (950-102; Ventana Medical Systems), and antigen retrieval was achieved by incubation with cell-conditioning solution 1 (950-124; Ventana Medical Systems), pH 8.2, for 32 minutes at 100°C. Sections were incubated with anti-CD3 antibody (790-4341, rabbit monoclonal, clone 2GV6 ready to use; Ventana Medical Systems). The signal was developed with the UltraMap Red anti-rabbit Detection Kit (Ventana Medical Systems). Sections were counterstained with hematoxylin. Immunostained slides were evaluated semiquantitatively by two pathologists (V.M. and D.M.) and a CD3 score was assigned as follows: score of 0, lack of intratumoral inflammatory cells; score of 1, scattered inflammatory cells; score of 2, conspicuous inflammatory cell infiltration under low magnification; and score of 3, diffuse/clustering infiltration.

Digital Image Analysis

Representative histopathologic slides of PMs stained with hematoxylin and eosin and immunostained for CD3 were anonymized and digitalized at an original magnification of ×400 using an Aperio AT2 on a WSI (Leica Biosystems, Wetzlar, Germany). Individual SVS format files were imported into HALO digital imaging analysis software version 3.6.4134 (Indica Labs, Albuquerque, NM). Two expert pathologists (V.M. and D.M.) drew the image annotations of the whole surface of PMs. Using the Multiplex immunohistochemistry (IHC) module v3.1.4 (Indica Labs), TIL detection was performed based on cytonuclear features such as stain intensity, size, and roundness for CD3-positive cells. The software automatically excludes tissue gaps from analysis and the settings were set to include the full range of staining intensity (from weak to strong). Data were expressed as cellular density (ie, the number of positive cells divided by the square millimeter of the annotation layer area).

AI Methodology

The proposed AI-based methodology aims to extract a TIL density score automatically from WSIs of a PM. Figure 2 shows the workflow of our approach once a WSI is provided.

Three different steps were performed to estimate the lymphocyte density within the tumor area from each WSI. First, the automatic recognition of melanoma and non-melanoma regions of WSIs was performed using the CNN (tumor-CNN; Figure 1) developed by De Logu et al[28] with a model based on a pretrained Inception-ResNet-v2.[29]

Then, a specific CNN (TIL-CNN; Figure 1) was trained to recognize areas in the tumor region containing TILs and areas without TILs. Finally, the TIL density was assessed based on the distribution of TIL and no TIL areas and a global score was assigned to each WSI. The TIL-CNN for TIL area recognition and the TIL density score are addressed specifically in this work.

Melanoma Tissue Detection

Healthy tissue versus melanoma tissue detection was performed as previously described by De Logu et al.[28] Briefly, the training model was based on a pretrained Inception-ResNet-v2,[29] the data set was formed by 100 WSIs: the regions of interests of the WSIs were labeled in the two classes and each regions of interest was tiled in nonoverlapping patches of 299 × 299 pixels, to fit the input dimension of the CNN.

CNN-Based TIL Region Recognition

The original and the labeled WSIs, both cropped on the melanoma region, form the data set used to implement the TIL-CNN for TIL recognition.

To adapt the data set to the input size of the TIL-CNN, each image was tiled in 299 × 299 pixel square patches

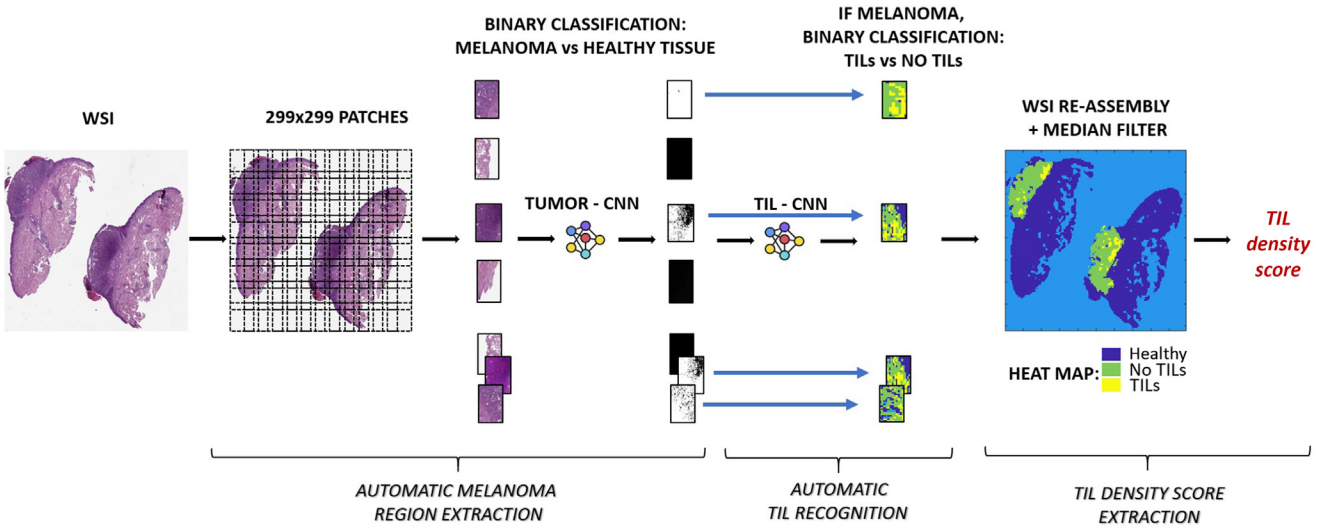

Figure 2 Pipeline of the proposed model application. Whole slide image (WSI) example. Original magnification: ×200. CNN, convolution neural network; TIL, tumor-infiltrating lymphocyte.

(original magnification, ×20) with an overlapping parameter of 100 pixels. The patch dimension of each pixel was approximately 0.2428 mm; therefore, the dimension of each patch was approximately 72.59 mm × 72.59 mm, corresponding to an area of approximately 5269.3 mm^2. Patches covered the entire melanoma-labeled region.

The sampled data set was divided into two sets: a training set used to train (70%; 40,429 patches) and validate (30%; 17,329 patches) the model for TIL recognition (data sets 1, 2, 3, and 4); and a disjoint test set (29,533 patches) provided by Catholic University of the Sacred Heart, Fondazione Policlinico Universitario A. Gemelli IRCCS Rome (data set 5) used to assess the performance of the model.

Data augmentation was performed on the training and the validation sets, using the parameters in Table 1. The TIL-CNN model is based on a pretrained Inception-ResNet-v2[29] with hyperparameters reported in Table 1. Training was performed using MATLAB software (R2021b; The MathWorks, Inc., Natick, MA) and its Deep Learning Toolbox. The training phase took approximately 9.4 hours. The trained TIL-CNN model is publicly available (*https://zenodo.org/record/7962742*, last accessed May 23, 2023).

To evaluate the performance of the model for TIL recognition, the same metrics (accuracy, specificity, sensitivity, Cohen's kappa, and F1 score) used by De Logu et al[28] were computed. Once TIL detection over the patches was performed, each WSI was re-assembled to compute the TIL density of each slide. The heatmap representing TIL, no-TIL, and healthy tissue regions (beyond peritumoral areas) was computed (Figure 2) and a 3 × 3 median filter was applied to avoid possible noise and spurious misclassified patches. The proposed automatic AI-TIL score was computed for each WSI as a rate between the following:

$$\text{AI-TIL} = \frac{\text{TILs}}{\text{TILs} + \text{no TILs}} * 100$$

where TILs are the number of patches classified by the TIL-CNN as patches containing TILs, and no TILs is the number of patches classified by the TIL-CNN as patches not containing TILs. The AI-TIL is used to assign a TIL density score to each WSI, and it is expressed as a percentage.

Statistical and Survival Analysis

To validate the accuracy of the TIL density score computed by our model, a comparison with common clinical PM classifications was performed using Kruskal-Wallis tests with Tukey honestly significant difference tests for *post hoc* comparisons.[30] A nonparametric method was used after the assessment of the non-normal distribution of AI-TIL in our data set by using the Lilliefors test. For comparison with the proposed AI-TIL index, brisk classification was considered as the gold standard clinical index of the TIL distribution in WSIs.[12] In addition, the AI-TIL was compared with an

Table 1 Summary of the Training Information for TIL Recognition CNN

Data augmentation			Hyperparameters			
Pixel range	Scale range	Rotation	Frozen layers	Learning rate	Minibatch size	Training time, hours
30—30	0.8—1.2	−90 degrees to 90 degrees	0	0.001	16	9.4

CNN, convolution neural network; TIL, tumor-infiltrating lymphocyte.

Table 2 Patient Characteristics of the Entire Data Set Subdivided into AI Training Cohort for Automatic TIL Region Recognition, AI Testing Cohort for Evaluation of the Automatic TIL Region Recognition, and a Survival Validation Subset to Assess AI-TIL Index Accuracy as a Prognostic Factor of Primary Melanoma

Clinical characteristics	AI training cohort (n = 237)	AI testing cohort (n = 70)	Survival validation subset (n = 170)
Sex, *n* (%)			
Male	132 (55.7)	37 (52.8)	100 (58.8)
Female	105 (44.3)	33 (47.2)	70 (41.2)
Age			
Median, years (range)	63, (14–89)	58 (12–93)	62 (14–89)
Anatomic site, *n* (%)			
Trunk	116 (48.9)	33 (47.1)	84 (49.4)
Limbs	88 (37.1)	29 (41.4)	60 (35.3)
Head/neck	16 (6.8)	8 (11.4)	12 (7.0)
Acral	15 (6.3)	—	13 (7.6)
Anal canal	1 (0.4)	—	1 (0.6)
Vulva	1 (0.4)	—	—
Pathologic characteristics			
Histotype *n*, (%)			
Superficial spreading melanoma	152 (64.1)	54 (77.1)	101 (59.4)
Nodular melanoma	66 (27.8)	15 (21.4)	54 (31.8)
Acral melanoma	15 (6.3)	—	13 (7.6)
Desmoplastic melanoma	1 (0.4)	—	—
Lentigo maligna melanoma	1 (0.4)	1 (1.4)	1 (0.6)
Mucosal melanoma	2 (0.8)	—	1 (0.6)
Breslow thickness, mm			
Median, mm (range)	3.17 (0.3–35.5)	1.7 (1.0–27.0)	4.0 (0.30–35.0)
Ulceration *n*, (%)			
Absent	103 (43.5)	43 (61.4)	60 (35.3)
Present	134 (56.5)	27 (38.6)	110 (64.7)
Mitotic rate/mm^2			
Median, *n* (range)	4 (0–57)	1 (0–20)	5 (0–42)
TILs, *n* (%)			
Absent	60 (25.3)	39 (55.7)	41 (24.1)
Non-brisk	138 (58.2)	18 (25.7)	107 (62.9)
Brisk	39 (16.5)	13 (18.6)	22 (12.9)
Staging characteristics			
T-stage, n (%)			
pT1	41 (17.2)	13 (18.6)	15 (8.8)
pT2	23 (9.7)	30 (42.9)	7 (4.2)
pT3	86 (36.3)	11 (15.8)	73 (42.9)
pT4	87 (36.7)	16 (22.7)	75 (44.1)
SLNB status, *n* (%)			
Positive			52 (30.6)
Negative			89 (52.3)
Not performed			29 (17.1)
Overall survival, *n* (%)			
Alive			76 (44.7)
Dead			94 (55.3)

Percentages are on column totals.
AI, artificial intelligence; SLNB, sentinel lymph node biopsy; TIL, tumor-infiltrating lymphocyte.

automatic TIL density score estimated by the HALO image analysis software. Finally, to evaluate the prognostic potential of the proposed AI-TIL, multivariate survival analyses were computed considering the AI-TIL as a potential risk factor. Multivariate survival analyses were performed by using Cox proportional hazard regression models.[31] Overall survival (OS) was defined as the time between diagnosis and death from any cause. Disease-free survival (DFS) was defined as the elapsed time from the date of the primary diagnosis of the tumor to the date of tumor

recurrence or death from any cause (ie, DFS events are the events of relapse and/or death). Patients who had not relapsed/died or had died were censored at the date of the last follow-up visit. The Cox proportional hazards regressions used in survival analysis look for independent variables related to variations in the risk function of the patients with respect to a specific event. Therefore, the OS evaluated by Cox regression assesses the relationship between the risk of death from any cause and the AI-TIL, while the DFS evaluated by Cox regression assesses the relationship between the risk of tumor recurrence or death from any cause and the AI-TIL. As well as AI-TIL, several other clinicopathologic factors were taken into consideration in both OS and DFS, according to the main prognostic factors reported in the literature. A Cox regression yields a hazard ratio with associated 95% CIs and P value, which indicates the patients who have a lower or higher risk of suffering the event according to the variable or variables included in the analysis.

From the entire data set (data sets 1, 2, 3, 4, and 5), the study detected a subset of 170 (55.4% of the entire data set) patients with a complete clinical history and prognostic information (ie, elapsed time from the date of the primary diagnosis of the tumor to the date of death and/or relapse). The AI-TIL scores of these patients were estimated, and the continuous AI-TIL scores was transformed into a dichotomous factor to discriminate between the PM with high and low TIL density. The dichotomous AI-TIL index was considered and tested as a risk factor in the multivariate survival analyses. An AI-TIL density threshold of 5%, which corresponds to approximately the 30th percentile of the AI-TIL density distribution, was set as the cut-off value to assign to each WSI the class of low TILs (AI-TIL score $<5\%$) or high TILs (AI-TIL score $>5\%$). The threshold value was chosen because it maximized the prognostic significance of the dichotomous AI-TIL factor (ie, minimum P value for associated hazard ratio) in multivariate survival analyses.

The following clinicopathologic prognostic factors were included in the multivariate analyses: age, Breslow thickness, and ulceration. According to Gershenwald et al,[32] these factors are considered main clinically significant factors to predict OS or DFS probability in melanoma.

To evaluate the prognostic value of the AI-TIL index, several multivariate survival models for OS (O models) and DFS (D models) were tested. For the O1 model, OS was estimated with the following predictors: dichotomous AI-TIL, age, Breslow thickness, and ulceration. For the D1 model, DFS was estimated with the following predictors: dichotomous AI-TIL, age, Breslow thickness, and ulceration.

According to Keung and Gershenwald,[3] SLN status is considered a prognostic factor to predict survival probability in patients with a PM. Therefore, the SLN was included in the multivariate survival models, considering a subset of 141 patients (45.9% of the entire data set) with SLNB performed. We tested two other survival models including the SLN predictor. For the O11 model, OS was estimated with the following predictors: dichotomous AI-TIL, SLNB, age, Breslow thickness, and ulceration. For the D11 model, DFS was estimated with the following predictors: dichotomous AI-TIL, SLNB, age, Breslow thickness, and ulceration.

To validate our data set, the same multivariate analyses were performed considering gold standard brisk classification[12] instead of the AI-TIL index. The following multivariate survival models were tested.

- O2 model: OS was estimated with the following predictors: brisk classification, age, Breslow thickness, and ulceration;
- D2 model: DFS was estimated with the following predictors: brisk classification, age, Breslow thickness, and ulceration;
- O22 model: OS was estimated with the following predictors: brisk classification, SLNB, age, Breslow thickness, and ulceration; and
- D22 model: DFS was estimated with the following predictors: brisk classification, SLNB, age, Breslow thickness, and ulceration.

Log-likelihood ratio tests, based on Wilks's[33] theorem, were performed to compare the Cox proportional hazard regression models with the AI-TIL index, with respect to those including brisk classification as a predictor. For the univariate descriptive visualization, the crude Kaplan-Meier survival curves segregated by the dichotomous AI-TIL and brisk classification were reported. All analyses were performed with MATLAB (R2021b; The MathWorks, Inc.). In all statistical tests, the level of significance was set at $P < 0.05$.

Results

TIL-CNN Validation

Demographic and clinical characteristics at diagnosis are reported in Table 2. For patients with clinical information available ($n = 170$) at a median follow-up period of 42 months, a relapse was reported for 105 patients (61.7%), and death for 103 (60.6%) patients.

As reported in Figure 3, the global accuracy of the trained model for automatic recognition of patches with TILs (TIL-CNN) was 99.5% for the validation set, and 100.0% for the test set. For the validation set, specificity and sensitivity was 99.7% and 99.3%, respectively. Cohen's kappa returned 0.99 and the F1 score was 0.99. For the test set, specificity and sensitivity both was 100.0%. Cohen's kappa returned 1.00 and the F1 score was 1.00.

AI-TIL Index Validation

The Kruskal-Wallis test showed a statistically significant difference in AI-TIL scores between the different TIL density score classifications estimated by HALO

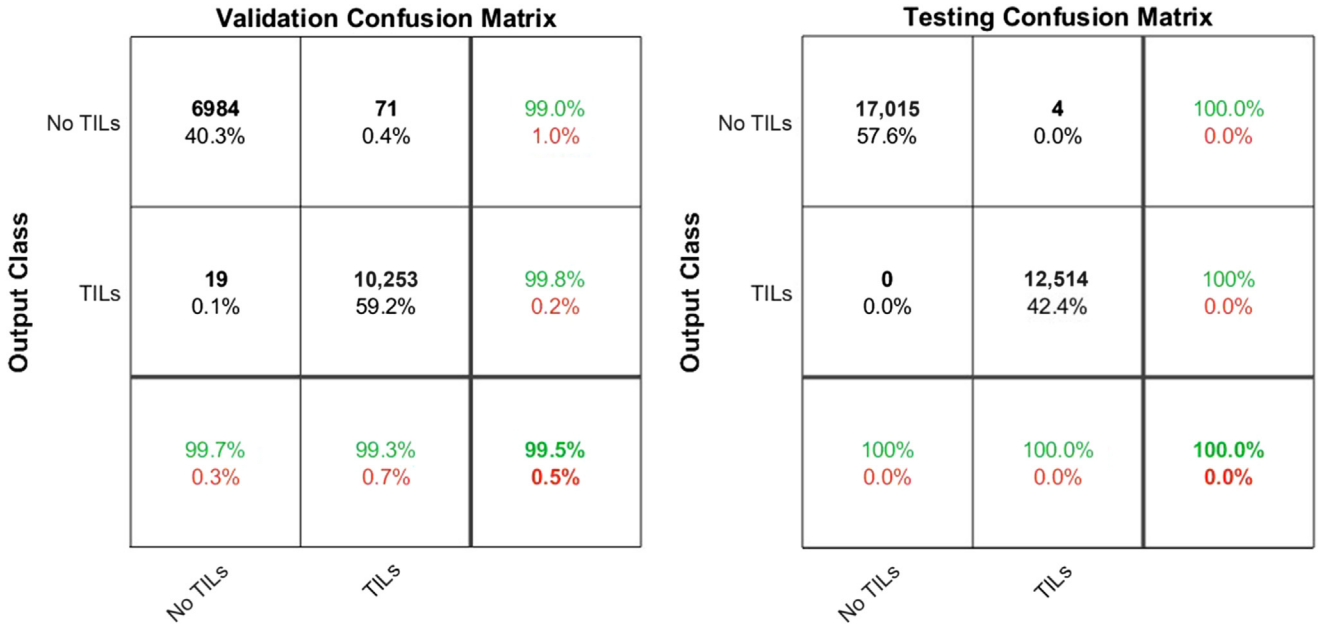

Figure 3 Validation and testing confusion matrices of the tumor-infiltrating lymphocyte (TIL)- convolution neural network (CNN) for recognition of patches with (TILs) and without TILs (No TILs). The validation cohort belongs to the training set, and the test cohort belongs to disjoint data set 5. The diagonal cells correspond to observations that are classified correctly, and the off-diagonal cells correspond to observations classified incorrectly. In each cell, both the number of observations and the percentage of the total number of observations are shown. The **far right column** in each confusion matrix shows the percentages of all items predicted to belong to each class that are classified correctly (**green text**) and incorrectly (**red text**). These metrics often are called the precision and false discovery rate, respectively. The **bottom row** in each confusion matrix shows the percentages of all the examples belonging to each class that are classified correctly (**green text**) and incorrectly (**red text**). These metrics often are called the recall and false negative rate, respectively. The **bottom right cell** in each confusion matrix shows the overall accuracy (**green text**) and error rate (**red text**).

(Figure 4A) [H(2) = 47.43; $P < 0.001$; with a median AI-TIL score of 5.4% for class 1+, a median AI-TIL score of 11.4% for class 2+, and a median AI-TIL score of 25.7% for class 3+]. As shown in Figure 4A, *post hoc* tests indicated significant differences in the AI-TIL score between class 1+ and class 2+ ($P < 0.001$), between class 1+ and class 3+ ($P < 0.001$), and between class 2+ and class 3+ ($P < 0.01$).

The Kruskal-Wallis test linked the AI-TIL score of our model and the brisk classification (absent, non-brisk, and brisk). As shown in Figure 4B, there was a statistically significant difference in AI-TIL scores between the different brisk classifications [H(2) = 24.83; $P < 0.001$; with a median AI-TIL score of 4.5% for class absent, a median AI-TIL score of 8.0% for class non-brisk, and a median AI-TIL score of 19.7% for class brisk]. *Post hoc* tests indicated a significant difference in the AI-TIL score between class absent and class non-brisk ($P < 0.05$), between class absent and class brisk ($P < 0.001$), and between class non-brisk and class brisk ($P < 0.01$).

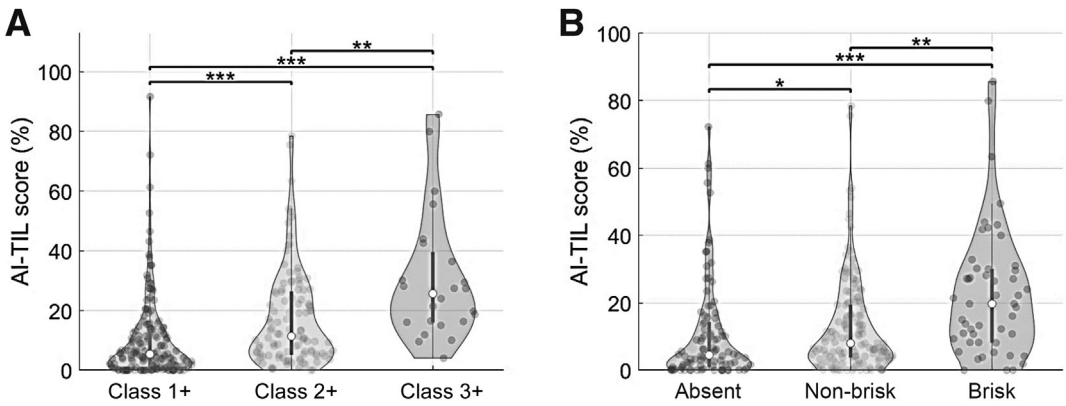

Figure 4 **A:** Violin plot of the tumor-infiltrating lymphocyte (TIL) density score of the model against HALO density score classification. The Kruskal-Wallis test was performed for the three classes (1+, 2+, and 3+). **B:** Violin plot of the TIL density score of the model against brisk classification. The Kruskal-Wallis test was performed for the three classes (absent, non-brisk, and brisk). *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$. AI, artificial intelligence.

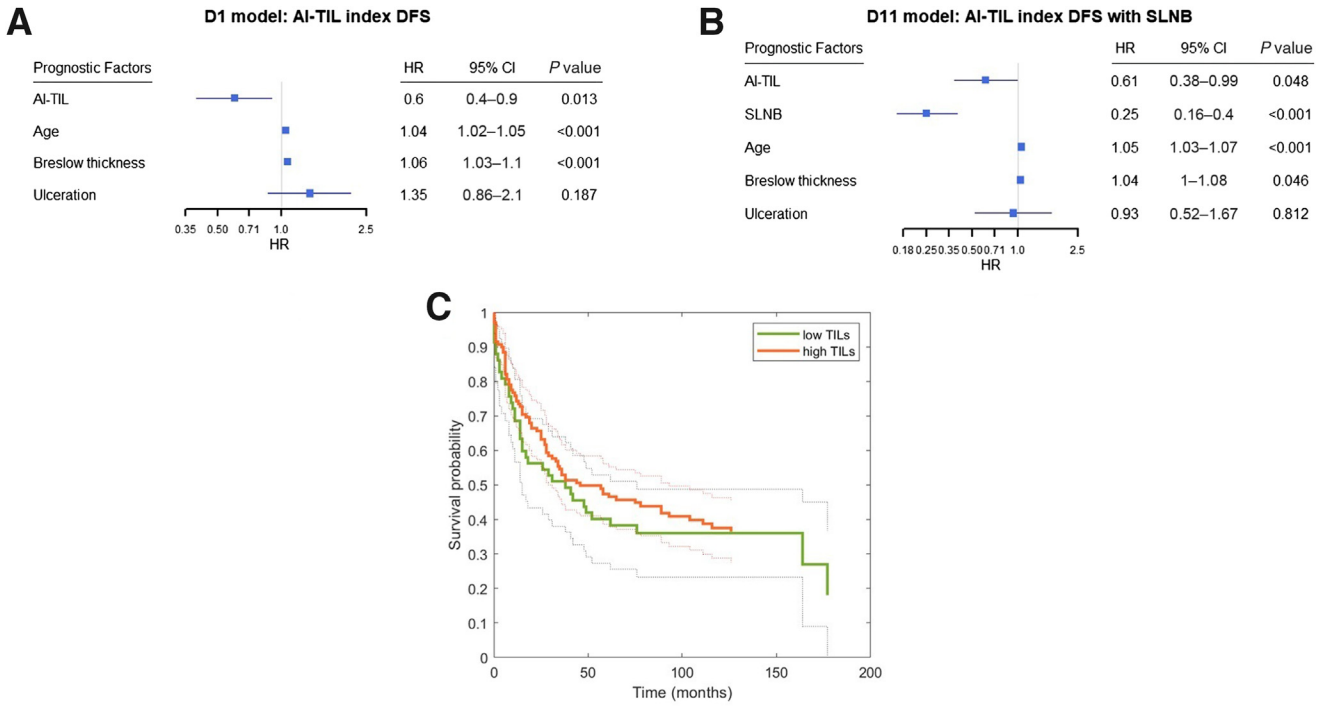

Figure 5 Forest plot of Cox proportional hazard regression ratios and the corresponding *P* values that refer to the D1 model. **A:** The artificial intelligence–based tumor-infiltrating lymphocyte density index (AI-TIL) is the dichotomous AI-TIL density index. The other regressors are the clinicopathologic prognostic factors. **B:** Forest plot of Cox proportional hazard regression ratios and the corresponding *P* values that refer to the D11 model. Sentinel lymph node biopsy (SLNB) has been added with respect to the D1 model. **C:** A crude Kaplan-Meier curve with respect to the dichotomous AI-TIL (low TILs and high TILs) of both survival models. DFS, disease-free survival; HR, hazard ratio.

As mentioned in the previous paragraph, different multivariate survival models were computed to evaluate the prognostic potential of the proposed AI-TIL.

DFS Analysis

The DFS analysis was performed on a subset of 170 patients with a complete clinical history and prognostic information. In the D1 survival model (Figure 5A), the presence of high TIL (AI-TIL score >5%) in a patient indicated a significantly longer DFS (HR, 0.6; SD, 0.2; $P = 0.013$) compared with low TIL (AI-TIL score <5%) patients. Other covariates associated with death or relapse hazard were age ($P < 0.001$) and Breslow thickness ($P < 0.01$). A crude Kaplan-Meier curve with respect to dichotomous AI-TIL (low versus high) showed a similar DFS for AI-TIL high compared with AI-TIL low (median DFS, 42 and 38 months, respectively) (Figure 5C). By adding the SLNB to the DFS model, the AI-TIL index lost significance with respect to D1, but the index remained statistically significant in the D11 model ($P < 0.05$).

In particular, in the D11 survival model (Figure 5B), the high TIL factor reduced the DFS events hazard by 39% (HR, 0.61; SD, 0.2; $P = 0.048$) with respect to low TIL. Other covariates associated with DFS were age ($P < 0.001$), Breslow thickness ($P < 0.05$), and SLNB ($P < 0.001$).

In the D2 survival model (Supplemental Figure S1A), DFS was not associated with brisk classification. Regressors associated with death hazard or relapse were age ($P < 0.001$) and Breslow thickness ($P < 0.001$). A crude Kaplan-Meier curve with respect to brisk classification showed a different DFS distribution. Pairwise analysis showed a longer DFS for brisk (median DFS not reached) compared with non-brisk (median DFS, 58 months) or absent class (median DFS, 25 months) (Supplemental Figure S1C).

In the D22 survival model (Supplemental Figure S1B), brisk class was associated with a better DFS (HR, 0.33; SD, 0.5; $P = 0.03$) with respect to non-brisk class. No difference was found between non-brisk or absent classes, as in the D2 model. Other covariates associated with DFS were age ($P < 0.001$) and SLNB ($P < 0.001$), whereas Breslow thickness lost significance with respect to the D2 model.

OS Analysis

The OS analysis was performed on a subset of 170 patients with complete clinical history and prognostic information. In the O1 survival model (Figure 6A), a high TIL factor was associated with longer OS (reduced death hazard, 41%; HR, 0.59; SD, 0.2; $P = 0.017$) with respect to a low TIL factor. Other covariates associated with death hazard were age ($P < 0.001$), Breslow thickness ($P < 0.01$), and ulceration ($P < 0.05$). A crude Kaplan-Meier curve with respect to dichotomous AI-TIL (low versus high) showed a longer OS for AI-TIL high compared with AI-TIL low (median OS, 75 and 68 months, respectively) (Figure 6C).

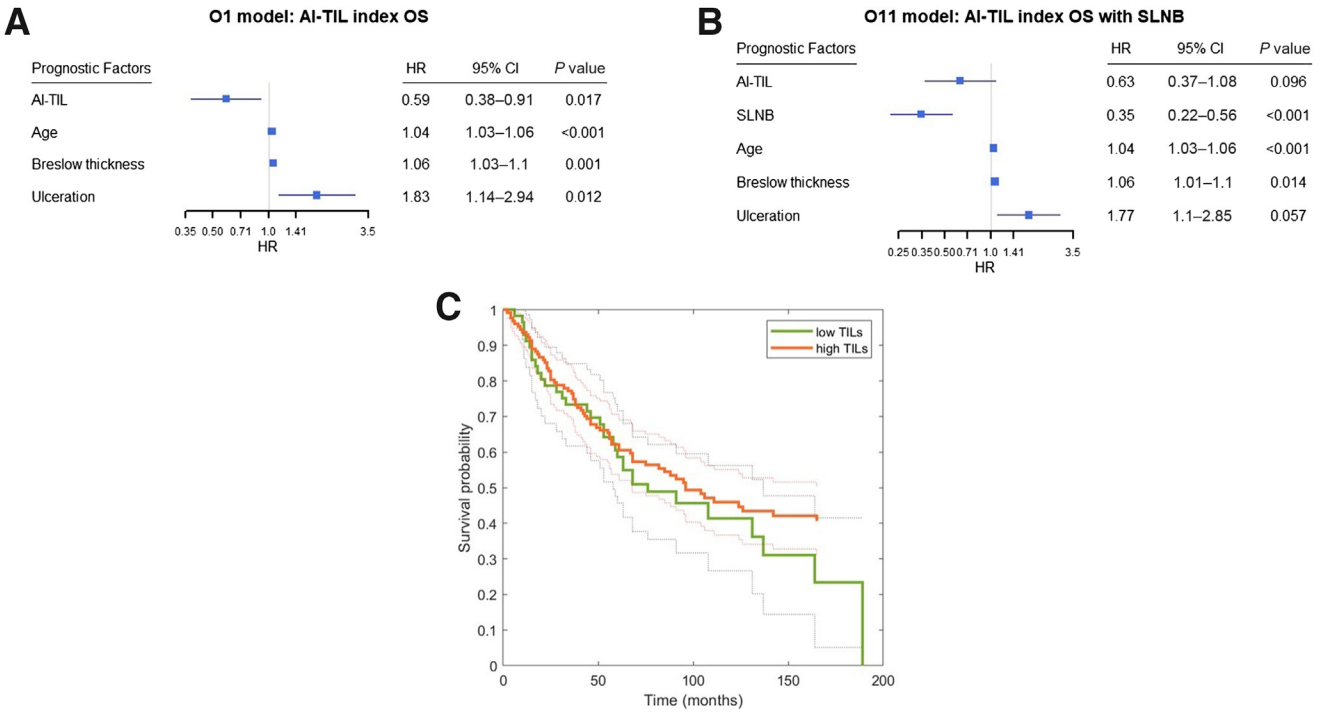

Figure 6 Forest plot of Cox proportional hazard regression ratios and the corresponding *P* values that refer to the O1 model. **A:** The artificial intelligence–based tumor-infiltrating lymphocyte density index (AI-TIL) is the dichotomous AI-TIL density index. The other regressors are the clinicopathologic prognostic factors. **B:** Forest plot of the Cox proportional hazard regression ratios and corresponding *P* values that refer to the O11 model. Sentinel lymph node biopsy (SLNB) has been added with respect to the O1 model. **C:** A crude Kaplan-Meier curve with respect to the dichotomous AI-TIL (low TILs and high TILs) of both survival models. HR, hazard ratio; OS, overall survival.

The addition of SLNB returned a loss of significance of the AI-TIL index for the O11 model, although its trend was consistent with O1. In fact, in the O11 survival model (Figure 6B), the death hazard was not related significantly to the AI-TIL index (HR, 0.63; SD, 0.3; $P = 0.096$). Age ($P < 0.001$), Breslow thickness ($P < 0.05$), and SLNB ($P < 0.001$) were associated with death hazard, whereas the ulceration lost significance with respect to the O1 model.

In the O2 survival model (Supplemental Figure S2A), brisk class was associated with a better OS (HR, 0.42; SD, 0.4; $P = 0.031$) compared with non-brisk class; no difference was found between non-brisk and absent classes. Other covariates associated with death hazards were age ($P < 0.001$), Breslow thickness ($P < 0.01$), and ulceration ($P < 0.05$). A crude Kaplan-Meier curve with respect to brisk classification showed a different survival distribution. Pairwise analysis showed a longer OS for brisk (median OS not reached) compared with non-brisk (median OS, 70 months) or absent class (median OS, 63 months) (Supplemental Figure S2C).

In the O22 survival model (Supplemental Figure S2B), brisk class conferred a longer OS (HR, 0.26; SD, 0.6; $P = 0.019$) with respect to non-brisk class; no difference was found between non-brisk or absent classes similar to the O2 model. Other covariates associated with OS were age ($P < 0.001$), Breslow thickness ($P < 0.05$), and SLNB ($P < 0.001$), whereas ulceration lost significance with respect to the O2 model.

Log-likelihood ratio tests indicated that the models considering AI-TIL as a prognostic factor (O1, O11, D1, and D11) fit the data significantly better than the models considering the brisk classification (O2, O22, D2, and D22). The comparison is reported in Supplemental Table S1.

Discussion

TILs have been investigated extensively as potential prognostic markers in PMs, with several studies suggesting that high densities of TILs correlate with favorable clinical outcome, including longer DFS, OS, and negative SLN,[11,14,16,19] whereas other studies failed to show such a correlation.[34,35] Such discrepancies could be explained in part by the subjective nature of conventional TIL assessment, performed by different pathologists with potential intraoperator and interinstitutional variability.[36] In fact, melanoma is a morphologically heterogeneous tumor that shows diverse cytomorphologic and architectural patterns of the host response, thereby presenting several diagnostic challenges.[37]

A standardized TIL assessment by digital pathology and AI-based approaches that minimizes observer variability is essential to expand our understanding of the tumor-immune microenvironment and its role in disease progression, recurrence, and treatment response. Recently, the central role of immunotherapy generated an increasing need to

characterize the density and distribution of TILs in cancer tissue samples effectively, including melanoma.[38] In clinical practice, a rapid and automated identification of TIL infiltrate might be helpful in determining if options for immunotherapy should be explored.[24,39]

Previous studies[23,24,26,27] have attempted to automate the analysis of TILs in cancer patients quantitatively, including those with melanoma, but the studies lacked sufficient substantial improvement in the precision of conventional pathology evaluation. In this study, a specifically trained CNN analyzed digital pathology slides of melanoma to provide a standardized TIL score that could predict clinical pathology evaluations. This has the potential to significantly impact clinical care. To the best of our knowledge, this is the first study in which a completely automated CNN-based digital analysis of TILs not only correlated with OS and DFS in melanoma, but also improved upon the standard brisk classification of TILs.

The CNN proposed in this work demonstrated high performance in recognizing TILs in histopathology slides of PMs. On a validation data set, it could identify a fraction of tumor tissue with TILs with a specificity and sensitivity of 99.7% and 99.3%, respectively, an F1 score of 0.99, and a Cohen's kappa of 0.99. On independent and disjoint testing data sets, specificity and sensitivity achieved 100%, Cohen's kappa and the F1 score were both 1.00. These values indicate an almost perfect agreement (Cohen's kappa), high precision, and recall (F1 score) in both the validation and test cohorts. Two methodologic aspects of these studies should be underlined: first, the training and validation data sets were composed of histopathologic images collected from several Italian clinical sites; second, the testing data set was completely disjointed and independent from the training/validation cohort. Our approach has ensured high robustness to the interoperator and intersite variability and accurate testing of the generalization capability of developed CNN. Previous studies[23,24,26,27] have proposed AI-based methods to extract a TIL density index from histopathology slides and compare it with other prognostic factors in survival analyses. As in our study, the data sets from the previous studies were provided by several clinical sites, but they included fewer slides or patches.

Concerning the use of CNNs in TIL detection, Acs et al[24] trained an eight-layer neural network as a machine learning method to detect TILs in PMs. They collected features of different types of cells to classify tumor, TILs, stroma, and other cells; no validation metric was reported. On the other hand, both Abousamra et al[23] and Saltz et al[27] trained a CNN able to extract TILs from different types of tumors, PMs included, returning a polyserial correlation coefficient of 0.82 and an area under the receiver operating characteristic curve of 0.954, respectively.

In contrast to our model, all of the previously cited works required a preliminary manual extraction of the tumor area of the melanoma. The current approach is completely automated: it provides a first CNN able to detect the tumor area by extracting the patches on which the second CNN performs the TIL estimation.

As previously demonstrated,[24–26] preliminary multivariate survival analyses were performed to evaluate the potential prognostic value of the AI-TIL score. The TIL density index was a significant prognostic factor in models including age, ulceration, and thickness or depth. Because only 28% of the test data set had the SLNB performed, the SLN was considered only in multivariate Cox regression,[26] which included an unknown biopsy specimen. Automated digital indices and brisk classification were performed to validate the extracted TIL index.[26] The TIL density score correlated with the Clark et al[12] criteria with a P value less than 0.001 and was significant in differentiating between high and low TIL density in PMs.

Furthermore, our AI-TIL score is comparable with HALO classification. In particular, as reported in Figure 4, the Kruskal-Wallis test returned a significant difference between the median AI-TIL score in the three categories of brisk and HALO classifications. The weakest difference was found between absent and non-brisk classes, although it still was significant with respect to the set level of significance ($P < 0.05$). Interestingly, the proposed AI-TIL score was associated more closely with the outcome than pathologist-assessed TILs or HALO classification. Moreover, it should be noted that, although the conventional TIL assessment was based on the evaluation of selected tumor areas of vertical growth phase, the AI-TIL score evaluated the total melanoma area.

Overall, models considering AI-TIL as a prognostic factor fit the data better than models using brisk classification for both DFS and OS (Supplemental Table S1). Of note, the AI-TIL score significantly differentiated DFS in both multivariate regressions, with and without SLNB, compared with the Clark et al[12] TIL grades (Figures 5 and Supplemental Figure S1), which failed to demonstrate differences in DFS among brisk, non-brisk, and absent classes (in model D2) and between non-brisk and absent classes in model D22. However, in the latter, brisk class was associated with a significantly better DFS compared with the non-brisk class only, as reported in previous studies.[26,37] Even though the AI-TIL index was statistically verified to be highly consistent with the universally accepted Clark et al[12] TIL classification, its concordance at a statistical level may not necessarily reflect identical results in OS and DFS analyses. The AI-TIL is a quantitative index used in our OS and DFS analyses after a 2-level dichotomization; meanwhile, the Clark et al[12] TIL classification is a semiquantitative evaluation with three levels of categorization. This precludes obtaining identical results in the OS and DFS analyses.

Similarly, in the OS regression models using the Clark et al[12] TIL grades (Supplemental Figure S2), TIL did not associate with OS between absent—non-brisk patients, although a significantly better OS was reported for brisk—non-brisk patients in both multivariate models (O2

and O22). On the other hand, the AI-TIL score significantly differentiated OS in model O1, but did not uphold significance adjusting for SLNB status in model O22 (Figure 6). This discordance could be attributed to the limited number of patients with known SLNB status included in the second model.

These results suggest that the AI-TIL score is overall superior at differentiating DFS and OS, particularly for absent-graded patients, and that a standardized and fully automated AI-TIL score may improve the prognostic impact of TIL. Our AI-TIL score was shown to be a consistent independent prognostic marker adjusting for age, Breslow thickness, and ulceration. Nevertheless, further studies are needed to achieve definitive conclusions, specifically improving multivariate models, including SLNB status.

Although the proposed model achieves a very good performance level, in future studies it might be necessary to increase the data set to provide a larger number of examples, and allow generalization. A major limitation of this study is the use of a subset of patients in the survival multivariate analysis. Any future studies will benefit from avoiding this and using the complete clinical information of the entire data set. Moreover, because the AI-TIL density threshold is based on our population, an extension of the data set would be useful to confirm or adjust the chosen threshold. Finally, the training of the CNN depends strongly on the manual labeling performed.

Conclusions

The proposed pipeline for AI-TIL score extraction has the potential to increase uniformity across cohorts and clinical sites, reduce the need for monitoring by highly experienced pathologists, and prevent operator variability. Further prospective investigation is needed to determine whether AI-TIL scores should be included routinely in PM pathology reports and future American Joint Committee on Cancer staging revisions, and whether they should be applied to metastatic samples to shape the design of novel immunotherapeutic protocols.

Author Contributions

D.M. and M.L. conceived and designed the study, interpreted the results, and reviewed and revised the paper; F.U., F.D.L., F.B., and M.L. wrote the first draft of the manuscript; S.S. and V.M. collected and annotated data; F.U. and F.D.L. performed the experiments under supervision by D.M.; M.L. and F.B. developed the computational framework; L.F.I., F.B., and M.L. performed statistical analysis; V.d.G., A.M.d.G., C.M., F.F., K.P., G.P., A.C., and M.M. provided clinical data; and all authors read and approved the final paper.

Disclosure Statement

None declared.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.ajpath.2023.08.013*.

References

1. Garbe C, Amaral T, Peris K, Hauschild A, Arenberger P, Basset-Seguin N, Bastholt L, Bataille V, del Marmol V, Dréno B, Fargnoli MC, Forsea A-M, Grob J-J, Hoeller C, Kaufmann R, Kelleners-Smeets N, Lallas A, Lebbé C, Lytvynenko B, Malvehy J, Moreno-Ramirez D, Nathan P, Pellacani G, Saiag P, Stratigos AJ, Van Akkooi ACJ, Vieira R, Zalaudek I, Lorigan P: European consensus-based interdisciplinary guideline for melanoma. Part 2: treatment - update 2022. Eur J Cancer 2022, 170:256–284
2. Amin MB; American Joint Committee on Cancer, American Cancer Society. AJCC Cancer Staging Manual. Edited by Amin MB, Edge SB, Gress DM, Meyer LR. ed 8. Chicago, IL: Springer, 2017
3. Keung EZ, Gershenwald JE: The eighth edition American Joint Committee on Cancer (AJCC) melanoma staging system: implications for melanoma treatment and care. Expert Rev Anticancer Ther 2018, 18:775–784
4. Mandalà M, Imberti GL, Piazzalunga D, Belfiglio M, Labianca R, Barberis M, Marchesi L, Poletti P, Bonomi L, Novellino L, Di Biagio K, Milesi A, Guerra U, Tondini C: Clinical and histopathological risk factors to predict sentinel lymph node positivity, disease-free and overall survival in clinical stages I-II AJCC skin melanoma: outcome analysis from a single-institution prospectively collected database. Eur J Cancer 2009, 45:2537–2545
5. Azimi F, Scolyer RA, Rumcheva P, Moncrieff M, Murali R, McCarthy SW, Saw RP, Thompson JF: Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. J Clin Oncol 2012, 30:2678–2683
6. Santos FDM, Silva FCD, Pedron J, Furian RD, Fortes C, Bonamigo RR: Association between tumor-infiltrating lymphocytes and sentinel lymph node positivity in thin melanoma. An Bras Dermatol 2019, 94:47–51
7. Fortes C, Mastroeni S, Caggiati A, Passarelli F, Ricci F, Michelozzi P: High level of TILs is an independent predictor of negative sentinel lymph node in women but not in men. Arch Dermatol Res 2021, 313:57–61
8. Clemente CG, Mihm MC, Bufalino R, Zurrida S, Collini P, Cascinelli N: Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. Cancer 1996, 77:1303–1310
9. Taylor RC, Patel A, Panageas KS, Busam KJ, Brady MS: Tumor-infiltrating lymphocytes predict sentinel lymph node positivity in patients with cutaneous melanoma. J Clin Oncol 2007, 25:869–875
10. Cintolo JA, Gimotty P, Blair A, Guerry D, Elder DE, Hammond R, Elenitsas R, Xu X, Fraker D, Schuchter LM, Czerniecki BJ, Karakousis G: Local immune response predicts survival in patients with thick (t4) melanomas. Ann Surg Oncol 2013, 20:3610–3617
11. Thomas NE, Busam KJ, From L, Kricker A, Armstrong BK, Anton-Culver H, Gruber SB, Gallagher RP, Zanetti R, Rosso S, Dwyer T, Venn A, Kanetsky PA, Groben PA, Hao H, Orlow I, Reiner AS, Luo L, Paine S, Ollila DW, Wilcox H, Begg CB, Berwick M: Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based

genes, environment and melanoma study. J Clin Oncol 2013, 31: 4252–4259
12. Clark WH, Elder DE, Guerry D, Braitman LE, Trock BJ, Schultz D, Synnestvedt M, Halpern AC: Model predicting survival in stage I melanoma based on tumor progression. J Natl Cancer Inst 1989, 81: 1893–1904
13. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD, et al: The molecular taxonomy of primary prostate cancer. Cell 2015, 163: 1011–1025
14. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al: Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: part 2: TILs in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors. Adv Anat Pathol 2017, 24: 311–335
15. Meneveau MO, Sahli ZT, Lynch KT, Mauldin IS, Slingluff CL: Immunotyping and quantification of melanoma tumor-infiltrating lymphocytes. Methods Mol Biol 2021, 2265:515–528
16. Maibach F, Sadozai H, Seyed Jafari SM, Hunger RE, Schenk M: Tumor-infiltrating lymphocytes and their prognostic value in cutaneous melanoma. Front Immunol 2020,, 11:2105
17. Nishimura T, Nakui M, Sato M, Iwakabe K, Kitamura H, Sekimoto M, Ohta A, Koda T, Nishimura S: The critical role of Th1-dominant immunity in tumor immunology. Cancer Chemother Pharmacol 2000, 46(Suppl):S52–S61
18. Viguier M, Lemaître F, Verola O, Cho M-S, Gorochov G, Dubertret L, Bachelez H, Kourilsky P, Ferradini L: Foxp3 expressing CD4+CD25(high) regulatory T cells are overrepresented in human metastatic melanoma lymph nodes and inhibit the function of infiltrating T cells. J Immunol 2004, 173:1444–1453
19. De Logu F, de Araujo DSM, Ugolini F, Iannone LF, Vannucchi M, Portelli F, Landini L, Titiz M, De Giorgi V, Geppetti P, Massi D, Nassini R: The TRPA1 channel amplifies the oxidative stress signal in melanoma. Cells 2023, 10:3131
20. Sobottka B, Nowak M, Frei AL, Haberecker M, Merki S, Aebersold R, et al: Establishing standardized immune phenotyping of metastatic melanoma by digital pathology. Lab Invest 2021, 101: 1561–1570
21. Dimitriou N, Arandjelović O, Caie PD: Deep learning for whole slide image analysis: an overview. Front Med 2019, 6:264
22. Srinidhi CL, Ciga O, Martel AL: Deep neural network models for computational histopathology: a survey. Med Image Anal 2021, 67: 101813
23. Abousamra S, Gupta R, Hou L, Batiste R, Zhao T, Shankar A, Rao A, Chen C, Samaras D, Kurc T, Saltz J: Deep learning-based mapping of tumor infiltrating lymphocytes in whole slide images of 23 types of cancer. Front Oncol 2022, 11:806603
24. Acs B, Ahmed FS, Gupta S, Fai Wong P, Gartrell RD, Sarin Pradhan J, Rizk EM, Gould Rothberg B, Saenger YM, Rimm DL: An open source automated tumor infiltrating lymphocyte algorithm for prognosis in melanoma. Nat Commun 2019, 10:5440
25. Aung TN, Shafi S, Wilmott JS, Nourmohammadi S, Vathiotis I, Gavrielatou N, Fernandez A, Yaghoobi V, Sinnberg T, Amaral T, Ikenberg K, Khosrotehrani K, Osman I, Acs B, Bai Y, Martinez-Morilla S, Moutafi M, Thompson JF, Scolyer RA, Rimm DL: Objective assessment of tumor infiltrating lymphocytes as a prognostic marker in melanoma using machine learning algorithms. EBioMedicine 2022, 82:104143
26. Moore MR, Friesner ID, Rizk EM, Fullerton BT, Mondal M, Trager MH, Mendelson K, Chikeka I, Kurc T, Gupta R, Rohr BR, Robinson EJ, Acs B, Chang R, Kluger H, Taback B, Geskin LJ, Horst B, Gardner K, Niedt G, Celebi JT, Gartrell-Corrado RD, Messina J, Ferringer T, Rimm DL, Saltz J, Wang J, Vanguri R, Saenger YM: Automated digital TIL analysis (ADTA) adds prognostic value to standard assessment of depth and ulceration in primary melanoma. Sci Rep 2021, 11:2809
27. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al: Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell Rep 2018, 23: 181–193.e7
28. De Logu F, Ugolini F, Maio V, Simi S, Cossu A, Massi D, Nassini R, Laurino M: Recognition of cutaneous melanoma on digitized histopathological slides via artificial intelligence algorithm. Front Oncol 2020, 10:1559
29. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA: Inception-v4, inception-ResNet and the impact of residual connections on learning. Edited by Proceedings of the AAAI Conference on Artificial Intelligence, 31. AAAI Press, 2017. pp. 4278–4284
30. Kruskal WH, Wallis WA: Use of ranks in one-criterion variance analysis [Erratum appeared in J Am Stat Assoc 1952, 48:907–911]. J Am Stat Assoc 1952, 47:583–621
31. Harrell FE: Cox proportional hazards regression model, Regression Modeling Strategies. Springer Series in Statistics. New York, NY, Springer, 2001. pp. 465–507
32. Gershenwald JE, Scolyer RA, Hess KR, Thompson JF, Long GV, Ross MI, et al: Melanoma of the Skin. AJCC Cancer Staging Manual. Cham, Switzerland, Springer International Publishing, 2017. pp. 563–586
33. Wilks SS: The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Statist 1938, 9:60–62
34. Schatton T, Scolyer RA, Thompson JF, Mihm MC: Tumor-infiltrating lymphocytes and their significance in melanoma prognosis. Methods Mol Biol 2014, 1102:287–324
35. Lee HJ, Park IA, Song IH, Shin SJ, Kim JY, Yu JH, Gong G: Tertiary lymphoid structures: prognostic significance and relationship with tumour-infiltrating lymphocytes in triple-negative breast cancer. J Clin Pathol 2016, 69:422–430
36. Busam KJ, Antonescu CR, Marghoob AA, Nehal KS, Sachs DL, Shia J, Berwick M: Histologic classification of tumor-infiltrating lymphocytes in primary cutaneous malignant melanoma. A study of interobserver agreement. Am J Clin Pathol 2001, 115:856–860
37. Chou M, Illa-Bochaca I, Minxi B, Darvishian F, Johannet P, Moran U, Shapiro RL, Berman RS, Osman I, Jour G, Zhong H: Optimization of an automated tumor-infiltrating lymphocyte algorithm for improved prognostication in primary melanoma. Mod Pathol 2021, 34:562–571
38. Paijens ST, Vledder A, de Bruyn M, Nijman HW: Tumor-infiltrating lymphocytes in the immunotherapy era. Cell Mol Immunol 2021, 18: 842–859
39. Acs B, Hartman J: Next generation pathology: artificial intelligence enhances histopathology practice. J Pathol 2020, 250:7–8