Artificial Intelligence in Melanoma Dermatopathology: A Review of Literature

Hannah Neimy, BS, John Elia Helmy, BS,* Alan Snyder, MD, MSCR,† and Manuel Valdebran, MD†*

Abstract: Pathology serves as a promising field to integrate artificial intelligence into clinical practice as a powerful screening tool. Melanoma is a common skin cancer with high mortality and morbidity, requiring timely and accurate histopathologic diagnosis. This review explores applications of artificial intelligence in melanoma dermatopathology, including differential diagnostics, prognosis prediction, and personalized medicine decision-making.

Key Words: artificial intelligence, melanoma, dermatopathology, computer-assisted diagnosis, machine learning

(*Am J Dermatopathol* 2024;46:83–94)

INTRODUCTION

The societal shift toward integration of artificial intelligence (AI) into daily life has been aptly dubbed the Fourth Industrial Revolution. Between 2015 and 2020, 222 AI medical devices were approved by the Food and Drug Administration.[1] Machine learning is 1 example of AI algorithm technology that is trained to make predictions based on pattern detection. Deep learning is a subset of machine learning capable of processing massive datasets and supplying decision-making outputs. Deep learning is typically operated using convolutional neural networks (CNNs), wherein image data processing achieves a desired outcome.[2,3] Pathology serves as a promising field to integrate AI into medicine. CNNs can be trained with pixels from digitized whole-slide images of hematoxylin and eosin (H&E)-stained histopathologic slides.

In the United States, melanoma is the fifth most common cancer[4] and the current gold standard for melanoma diagnostics is histopathologic examination.[5] Given its high mortality and morbidity, a timely and accurate diagnosis is essential for early intervention and outcome optimization. In the research setting, AI has shown utility as a tool for dermatopathologists diagnosing melanoma, highlighting a potential adjunct for improving patient outcomes.[6–13] As such, this review aims to explore AI-dermatopathology applications in melanoma differential diagnostics, prognosis prediction, and related personalized medicine decision-making. Frequently used AI terminology is defined in Table 1.

From the *College of Medicine, Medical University of South Carolina, Charleston, SC; and †Department of Dermatology & Dermatologic Surgery, Medical University of South Carolina, Charleston, SC.

The authors declare no conflicts of interest.

Correspondence: Hannah Neimy, BS, College of Medicine, Medical University of South Carolina, 96 Jonathan Lucas St. Ste. 601, MSC 617, Charleston, SC 29425 (e-mail: neimy@musc.edu).

METHODS

This review investigated the available literature on melanoma dermatopathology and AI tools. The following Boolean search string was input to PubMed: ("melanoma" OR "skin cancer") AND ("dermatopathology" OR "histopathology" OR "pathology") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "computer vision" OR "image analysis" OR "pattern recognition"). Inclusion criteria included studies published between 2013 and 2023, and methods involving AI technologies in melanoma dermatopathology. Articles were excluded if they were reviews, published before 2013, or did not investigate melanoma dermatopathology AI tools. 599 studies were selected for screening and 90 full texts were eligible for full-text assessment. Ultimately, 34 studies were included in this review (Figs. 1,2,3).

DISCUSSION

Differential Diagnosis

Deep learning networks have successfully identified melanoma in digitized biopsies with accuracy similar to expert dermatopathologists.[14] Earlier methods of CNN-based melanoma detection from H&E whole-slide images (WSIs) worked using pixel patches from the digitized tissue samples to train and evaluate the system. Using this method, CNNs can find regions of interest to flag for potential melanoma diagnoses. One example of this is Dika et al,[6] wherein they compared melanoma region-of-interest (ROI) detection by CNNs versus expert dermatopathologists; the network and pathologists agreed 94% of the time. When compared with nevi identification, pathologists and deep learning networks agreed 100% of the time. They concluded AI could be a helpful screening tool for flagging suspicious regions in WSIs.

De Logu et al[7] later evaluated the deep learning recognition of melanoma from WSIs using small image patches from H&E slides. The deep learning system achieved an overall image patch classification accuracy of 96.5%, a sensitivity of 95.7%, and a specificity of 97.7% and the system achieved a promising F1 score of 97%. Using a CNN-based algorithm, the authors developed a method for defining healthy and pathologic tissues from scanned lesioned tissues. Although the algorithm produced promising results, only melanomas with a Breslow thickness of more than 2 mm were included.

Alheejawi et al[8] used a different approach to deep learning for melanoma detection. The investigators trained a CNN using 100 digitized H&E WSI. Their NSNet adapted-CNN

TABLE 1. AI Technical Terminology Definition Table

Term	Definition
CNN	A type of deep learning neural network commonly used for image recognition and computer vision tasks. CNNs are designed to automatically learn and extract relevant features from input images through a series of convolutional and pooling layers
Convolution layers	Layers within a CNN that perform the convolution operation. Convolution involves applying a set of filters to input data (eg, images) to extract relevant features, such as edges, textures, or shapes
eTIL	Electronic assessment of the presence of lymphocytes (immune cells) within tumor tissue
Heat map	A visualization technique commonly used in deep learning and computer vision to highlight regions of interest in an image or indicate the importance or activation of specific areas. Heat maps typically use a color gradient to represent the intensity or concentration of a feature
Patch	A patch refers to a small rectangular or square region extracted from an image. It is a subset of the original image used for analysis, such as feature extraction, classification, or segmentation
Pixel level	The level of individual pixels in an image or video. Analyzing or manipulating images at the pixel-level involves considering and processing each pixel individually
Random forest model	A machine learning algorithm that builds an ensemble of decision trees. Each tree is trained on a random subset of the training data, and predictions are made by aggregating the predictions of individual trees
RNN	A type of neural network architecture designed to process sequential data by maintaining internal memory. RNNs are widely used in NLP and speech recognition tasks, where the order and context of the data are crucial
Segmentation mask	A binary image or pixel-level annotation that identifies and labels specific regions or objects of interest within an image. It assigns a unique value or color to each pixel or region belonging to a particular object
SSL	A model learns from the data itself without any explicit human supervision
WSI	Also known as virtual microscopy, it involves scanning entire glass slides containing specimens (eg, tissue samples) into high-resolution digital images. WSI enables the viewing, analysis, and sharing of pathology slides in a digital format

NLP, natural language processing; SSL, self-supervised learning.

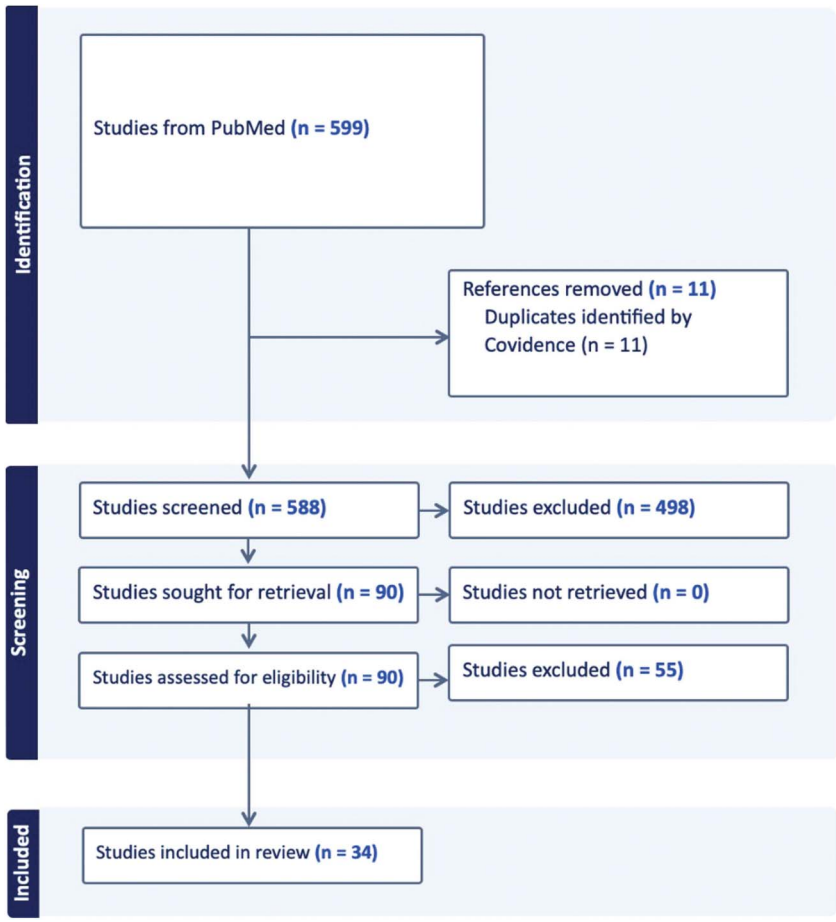

FIGURE 1. Prisma diagram.

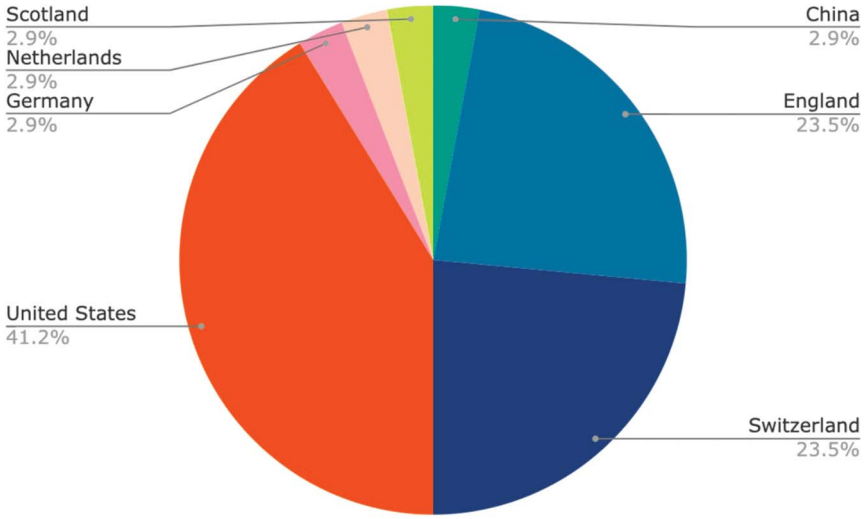

FIGURE 2. Pie chart depicting the country of publication of articles included in this review.

first annotated nuclei as belonging to melanoma cells, non-melanoma cells, or background.[15] This allowed the researchers to detect coarse and fine features of cells and extracellular backgrounds to generate a segmentation mask with more granularity. A segmentation mask refers to a binary image or pixel-level annotation that identifies and labels specific regions or objects of interest within an image. It assigns a unique value or color to each pixel or region belonging to a particular object. Once the segmentation mask was applied, a secondary melanoma region detection module used the CNN-generated mask to predict melanoma ROI. The authors proposed that this method could be more accurate than pre-existing CNN patch-fed architectures because it first detects the nuclei and then predicts the melanoma cells instead of relying on patch input. This method segmented nuclei with 94% accuracy and the Dice coefficient score was 85% for segmenting melanoma regions. From start to finish, each CNN-generated diagnosis took around 7 minutes to complete.[8]

In 2021, another group proposed a CNN model that differentiated and annotated benign lesions versus melanoma. The investigators collected 701 WSI of H&E-stained malignant skin lesions from multiple centers. The resulting model produced an area under the curve (AUC) of 0.971.[9] In 2022, Snyder et al[10] also used CNN to differentiate H&E images of melanocytic nevi, Spitz nevi, and malignant melanoma. Their ultimate tile-classification model correctly predicted patches within H&E images as belonging to melanoma cells at a sensitivity of 93%, nevi tiles with a sensitivity of 94%, and Spitz nevi tiles at a sensitivity of 73%. When evaluated with unseen cases, the architecture was capable of accurately predicting diagnosis 85.7% of the time. The authors suggested that this model could be expanded and improved using more tissue samples from multiple centers.

One group tested Mechanomind software, a CNN capable of diagnosing 40 skin disorders based on H&E samples. The investigators used 300 samples from the United States and Africa. The algorithm performed with high sensitivity and specificity. For melanoma identification, Mechanomind achieved a sensitivity of 97.8% and a specificity of 97.6%.[11] Another group compared the performance of 4 neural network models in classifying melanoma versus non-melanoma in H&E WSIs and tiles of varying pixel sizes. The most effective model (×20, 512 × 512 pixels) produced an AUC of 0.821 for WSl classification and 0.936 for tile-level classification, but detected a high false-positive rate.

Screening Tool for Diagnosing Rare Lesions

Eyelid melanoma makes up <1% of all cutaneous melanomas.[16] This makes correct pathologic interpretation crucial for treatment choice and patient wellbeing. Wang et al[13] developed a deep learning system to automatically detect melanoma in the eyelid from 155 H&E-stained WSIs. The researchers used WSIs to train and evaluate a deep learning model designed to assign patch-level classifications of disease states to regions extracted from the WSIs. The annotated patches were used to generate WSI heat maps. Then, random forest models produced a WSI-level diagnosis. When evaluated for patch diagnostic capabilities, the model achieved an AUC of 0.989, an accuracy of 94.9%, a sensitivity of 94.7%, and a specificity of 95.3%. When the model was evaluated for evaluation of malignant potential based on the WSI it achieved a sensitivity of 100%, a specificity of 96.5%, an accuracy of 98.2%, and an AUC of 0.998. Wang et al[17] contributed an effective diagnostic model that could be implemented as a supplemental tool for pathologists in the screening of rare skin disorders. Their more recent study aimed to address the obstacle of limited annotation by using self-supervised learning. A CNN model was trained and tested on patch-level classification and WSI-level diagnosis. The model achieved an AUC of 0.981 with an accuracy, sensitivity, and specificity of 90.9%, 85.2%, and 96.3% for the patch-level classification and an AUC, accuracy, sensitivity, and specificity of 0.974, 93.8%, 75.0%, and 100%, respectively. Self-supervised CNNs create an opportunity to apply AI screening tools to rare diseases with limited training datasets.

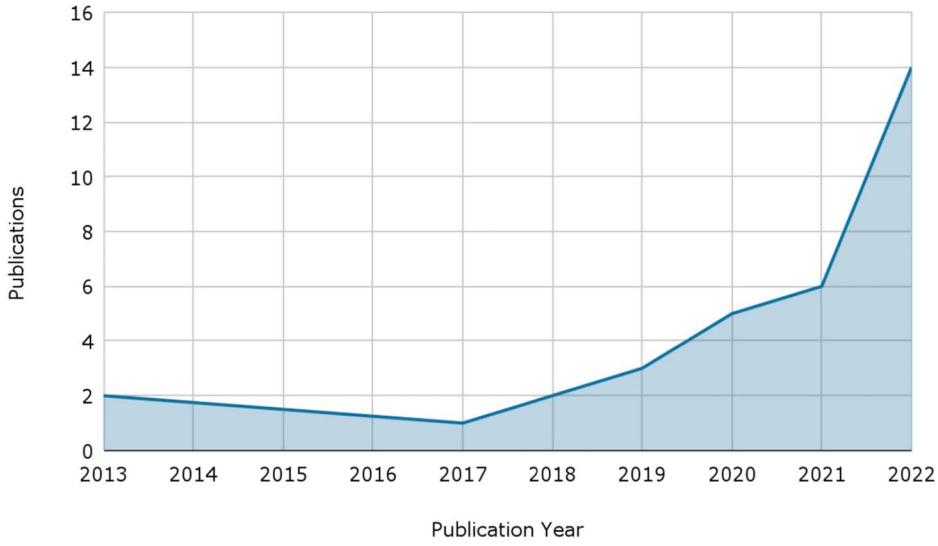

FIGURE 3. AI and melanoma dermatopathology article publication by year depicting an increase in publications in recent years. Includes only articles discussed in this review.

AI-Pathologist Differential Diagnosis Discordance

Before AI can be adopted as a casual screening tool, it is worthwhile to consider how it performs compared with expert pathologists. Earlier studies have explored the discordance between AI diagnosis and pathologist diagnosis. In 2019, a CNN was trained to differentiate nevi from melanoma on H&E. Misclassification rates of the neural network were compared with board-certified pathologists. The study reported a discordance of 18% for melanoma and 20% for nevi.[18] Binary diagnostics (melanoma vs. nevi) are not reflective of a pathologist's daily workload. Ba et al[19] designed a program to differentiate WSIs of melanoma and nevi; the program was tested against pathologists for the identification of melanoma and nevi. The program predicted ROI, CNN annotation of probable melanoma patches, and the implementation of a forest-based slide-level module to produce a final diagnosis. The program and dermatopathologists were limited to WSIs to make diagnostic decisions. The pathologists diagnosed lesions with a sensitivity of 95.1% and a specificity of 96.0%. When comparing the algorithm's performance to the pathologists', they found that the algorithm had similar sensitivity and specificity. This held true at multiple operating points. The investigators then applied a scoring system to consider the harm caused by false-negatives versus false-positives. The computer program produced a weighted error score of 1.82%, whereas the 7 dermatopathologists ranged between 1.36% and 7.27%.

Cazzato et al[20] trained a fast random forest algorithm to classify clusters of pixels from digitized H&E slides as belonging to malignant melanoma or dysplastic nevi. The fast random forest algorithm was discordant with the dermatopathologist 17% of the time. This was an improvement from the observed interpathologist variability of around 25%. However, these results may be limited by the lack of available clinical information that practicing pathologists would normally have access to.

Immunohistochemistry-Based Diagnosis

Immunohistochemistry is vital to the diagnosis of lesions. In some instances, however, it may be prohibitively expensive, take a long time, and consume finite amounts of tissue resources. Research has been conducted to evaluate the possibility of creating virtual immunohistochemistry assays. Jackson et al[21] trained CNNs to predict SOX10 immunophenotypes based on H&E slides of melanocytic neoplasms. Upon evaluation of this CNN architecture in sorting individual nuclei, an AUC of 0.942 was achieved.

Nielson et al[22] evaluated the utility of virtual immunohistochemistry assaying of melanoma images by training CNN to calculate tumor burden from H&E/SOX10 stains. The network annotated primary melanoma tumor cells and normal cells with an accuracy of 100% and 99%, respectively. The CNN calculated the tumor burden at 6% (95% CI, −1 to 13, $P = 0.10$), which was discordant with the pathologists' tumor burden of 16% (95% CI, 4–28, $P = 0.02$). Tumor burden is a critical part of molecular tumor marker interpretation, and thus treatment choice. Overestimation of tumor burden may hinder proper treatment protocols. As a result, patients may not be prescribed targeted therapy, resulting in a decreased likelihood of success.[22,23] Future studies may contribute relevant findings by calculating tumor burden with CNNs trained with different stains. In addition, they may expand generalizability to other cancer types.

Rexhepaj et al[24] developed a supervised melanoma pattern recognition tool from digital tissue microarray sections. Tissue sections were immunoassayed for Melan-A, a marker for melanoma cells, to provide researchers with ground truth. The resulting images were deconvoluted to separate color channels from immunostaining and hematoxylin counterstaining. Three classification methods were compared: Naïve Bayes, random forest, and a support vector machine (SVM). SVM was found to be the most effective at predicting melanoma.

The aforementioned studies are examples of AI applying virtual immunohistochemistry as a diagnostic tool. However, digital image analysis has also been used to find markers not right for use as a diagnostic tool. Koh et al[25] used a digital image analysis algorithm to compare cyclin D1 expression in immunohistochemically stained tissues in superficial spreading melanoma and compound nevi. Results showed no difference in cyclin D1 levels between lesion groups. Digital immunohistochemistry is a promising diagnostic tool that could be further evaluated for differentiating melanoma and nevi through different markers. However, it would be necessary to further compare the virtual immunohistochemistry marker to the conventional immunohistochemistry marker. This is to ensure the continuity of the traditional mechanisms. Future studies could try this method using CNNs and PReferentially expressed Antigen in MElanoma to produce virtual immunohistochemistry assays capable of diagnosing melanoma.[26]

Prognostic Potential

AI can learn how to correlate image patterns with prognostic outcomes. Earlier studies have used transfer learning to research AI's ability to predict breast cancer recurrence and response to neoadjuvant chemotherapy based on magnetic resonance imagings.[27,28] AI could be used to predict melanoma disease-free survival based on whole-slide images resulting in further optimized and individualized treatment protocols. Comes et al[29] investigated a deep learning model trained to learn prognostic biomarkers from WSIs. The goal of this experiment was to predict 1-year disease-free survival in cutaneous melanoma patients. When evaluated for prognostic power, the model had an AUC of 69.5% and median accuracy of 72.7%.

Kulkarni et al[30] presented a deep neural network for predicting melanoma metastasis and recurrence based on digital H&E images. The network was trained by pooling samples from 108 patients from 4 institutions. The model used CNN combined with a recurrent neural network (RNN). The CNN was trained to annotate features in H&E digital images. Then, the RNN analyzed the CNN feature sequence. Distant metastatic recurrence was selected as the binary classifier to predict disease-specific survival. The pipeline was then evaluated with patients from Yale School of Medicine (YSM) and Geisinger Health Systems (GHS). This method produced an AUC of 0.905 ($P < 0.0001$) for the YSM cohort and 0.880 ($P < 0.0001$) for the GHS cohort.

Brinker et al[31] investigated the concept of a digital biomarker to predict lymph node metastasis using deep learning analysis of H&E slides. The CNN was trained using variables associated with lymph node positivity status: patient age, tumor thickness, and ulceration. These features were matched to H&E slides from melanoma patients with lymph node involvement or without lymph node involvement. Results showed that it was possible to predict lymph node involvement from H&E slides. However, the researchers concluded that this method requires additional research before it becomes clinically relevant.

Automated electronic tumor-infiltrating lymphocytes (TILs) assessment is another area of deep learning prognostic investigation. In 1 study, electronic TILs (eTILs) quantification was evaluated for its prognostic value as a way to define the risk of melanoma relapse after surgery. The investigators analyzed samples from 785 patients diagnosed with melanoma between 1994 and 2019 from 5 different institutions. The software used to analyze the images, QuPath, applied a machine learning algorithm to name distinct types of cells on H&E digitized slides. The machine-defined TILs variables included the proportion of TILs to tumor cells, the ratio of TILs over total cells, the proportion of TILs to stromal cells, and the density of TILs over the tumor region. The researchers also used 11 different antibodies for immunofluorescence staining to investigate TIL immunophenotypes ($CD34^+$, $CD68^+$, $CD56^+$, CD66b+, FOXP3+, $CD8^+$, $CD14^+$, $CD3^+$, $CD45^+$, $CD20^+$, and $CD4^+$). The tumor, stroma, and background components were annotated with cell segmentation deep learning. Results showed that the assessment of TILs through the use of eTILs quantification deep learning method can be used to stratify the risk of melanoma relapse after surgical treatment. They found that eTIL scores can be prognostic markers in primary melanoma patients. Increased eTIL proportion within the tumor environment was associated with improved prognosis for stage I and II diseases. The study also identified molecular TIL subtypes. These were predominantly $CD3^+$, $CD8^+$, or $CD4^+$ T cells.[32] Before eTILs scores can be clinically implemented to guide treatment decisions more research is needed to confirm these findings in a clinical trial. Future studies could also evaluate eTILs' predictive value for immunotherapy response. This approach could potentially enhance machine learning's prognostic abilities.

Zormpas-Petridis et al[33] designed an algorithm called SuperCRF. Their model used H&E images of melanoma samples. The investigators used a deep learning network to classify cells. Using SuperCRF, the researchers found that certain ratios of immune cells to other cells in the tumor microenvironment were associated with poor survival rates in melanoma patients. SuperCRF showed that an elevated ratio of lymphocytes to all lymphocytes within the stromal compartment correlated with a worse prognosis. In addition, SuperCRF also showed that a high ratio of stromal cells to all cells was associated with poor survival in melanoma patients. Overall, SuperCRF is a valuable tool that could help doctors and researchers enhance their understanding of the tumor microenvironment, find reliable predictors of survival, and predict patient response to therapy.

Mitosis Detection

Mitotic rate has been shown to have utility as a prognostic tool in melanoma.[34,35] Although the American Joint Committee on Cancer no longer includes mitotic rate as a component for staging, computer-assisted models could enhance pathologists' accuracy and consistency in mitotic figure detection, thus ensuring more efficient and accurate treatment plans and prognostic predictions.[36]

In 2017, Andres et al[37] designed and evaluated iDermatoPath, a tool trained to detect mitoses in H&E stains of melanoma. The system detected the tumor region, annotated mitoses, and ranked them into risk categories. When

compared with tumor region ground truth, iDermatoPath achieved a dice coefficient of 0.72. In addition, the tool achieved an accuracy of 83% for mitosis detection. Although this tool may accurately identify mitotic bodies, it is pertinent to continue training the tool with pathologist-annotated slides to improve the dice coefficient and improve agreement between the computer-generated image and the pathologist ground truth image.

Nofallah et al[38] compared 2 pre-existing CNN models, ESPNet and DenseNet, for the detection of mitotic nuclei versus non-mitotic nuclei on skin biopsy images. The ESPNet and DenseNet models achieved high accuracy, sensitivity, and specificity in classifying mitoses. DenseNet performed slightly better than ESPNet, but not at a level that was statistically significant. DenseNet and ESPNet were then compared with 2 other state-of-the-art models, ResNet and ShuffleNet, using the publicly accessible MITOS breast biopsy dataset. DenseNet achieved the highest F score of the 4 architectures tested, whereas ResNet was the most efficient architecture. The paper shows that CNNs can accurately detect mitotic figures in WSIs from multiple cancer types. This could be a diagnostic and prognostic tool for pathologists and researchers. This paper also shows that architecture designed for mitotic identification can be evaluated and tested on multiple cancer types.

Hale et al[39] used a model to detect immunopositivity in mitotically active cells using 2 mitosis-specific antibodies. The study compared manual and computer-assisted methods to detect and measure mitotic rates in melanoma cells. Tissue sections were stained with antibodies antimitotic protein monoclonal-2 (MPM-2) and antiphosphohistone-H3 (PHH3) to visualize mitotic figures. PHH3 mitotic rate was more closely correlated with progression-free survival after computer-assisted image analysis ($P = 0.02$). No advantage was clear over conventional mitotic rate determination by MPM-2 or PHH3 regardless of the detection method.

Couetil et al[40] used multiple machine learning techniques to analyze melanoma H&E images and identify predictors of metastasis and survival. Tissue samples came from multicenter cohorts with stage I-III melanoma. Features associated with improved survival included major axis length distribution entropy and major axis length SD. Major axis length distribution corresponds to increased cell size variation. This can be interpreted as an increase in cellular heterogeneity (inflammatory cells, tumor cells, and stromal cells). Major axis length bin 4 represents the length of the major axis of tumor cells. The pipeline achieved F1 scores of 0.72 and 0.73 for predicting melanoma survival and metastasis risk, respectively. The results show that machine learning models can accurately and sensitively predict 5-year survival and metastasis risks using interpretable features of cellular and nuclear morphology. Image feature extraction could 1 day help clinicians find patients at elevated risk of metastasis for increased surveillance and improved precision treatment.

Personalized Medicine

AI may help physicians predict patient response to individualized treatments. Hu et al[41] used deep learning to predict immune-checkpoint blockade response in melanoma based on H&E images. The CNN was able to differentiate anti-PD-1 responders versus nonresponders with an AUC of 0.778 (95% CI, 63.8–90.5).

B-raf proto-oncogene (BRAF) mutation is a relevant prognostic component of melanoma diagnosis.[42–44] Targeted BRAF inhibitor therapies have improved outcomes in patients with advanced late-stage melanoma.[45] Testing for BRAF is becoming an increasingly important component of treatment selection. Currently, BRAF mutations are detected via Sanger sequencing, polymerase chain reaction, mass spectrometry, immunohistochemistry, and next-generation sequencing. Figueroa-Silva et al[46] tested 3 machine learning platforms (XGBoost, LightGBM, and ANN) to predict the presence of BRAF V600E mutations based on histopathologic and clinical features including Breslow thickness, age at diagnosis, mitoses, morphology, ulceration, pigmentation, epidermal hyperplasia, and melanoma nest formation rate. BRAF-positive melanomas had greater Breslow thickness and mitoses. The researcher's highest-performing machine learning algorithm (XGBoost) produced an AUC of 0.704. Interestingly, the machine learning models incorrectly predicted that 5 patients had the BRAF V600E mutation. Further genomic analysis revealed that of those 5 patients, 4 had different BRAF mutations including BRAF V600E/E2/D and BRAF K601E. Although less common, these mutations are also relevant clinical variables. This study suggests AI could be a sensitive screening tool for BRAF mutations, thus it is another example of AI aiding in enhanced individualized therapy and improved prognostics.

Mund et al[47] introduced an innovative technique called deep visual proteomics (DVP), which combines submicron resolution imaging, AI interpretation and phenotyping of WSI, and evaluation with ultrasensitive proteomics. The researchers designed this model using tissue from malignant tissue biopsies including melanoma. Based on AI cell type interpretation, investigators extracted cells and/or nuclei via lasers. They evaluated intracellular proteins through mass spectrometry. Using this method, researchers could link certain proteins to specific cells or cellular components while tracking their location within the tissue. Researchers also used DVP to find changes in proteins as the melanoma progressed. DVP data can be applied to explore proteome variation at the phenotypic level. This will contribute to the development of highly personalized medicine and improved prognosis. A complete summary of AI models explored in this review is provided in Table 2.

CONCLUSIONS

The argument for AI inclusion in clinical practice should not be to replace the role of the pathologist. The limited number of dermatologic conditions used to train models emphasizes the fact that AI is not a replacement for expert pathologists. Instead, AI could be developed in such a way that it becomes a highly beneficial tool for screening, reducing workloads, and decreasing interobserver differences, particularly in analyzing the more complex and time-consuming diagnoses pathologists face.[48] This is especially helpful in cases where tissue availability is limited, making

TABLE 2. Characteristics of AI Models Used in Review Studies

Article Title	Lead Author	Year Published	AI Model	Training and Validation Dataset/Patches/WSIs/No. of Patients (n)	Pixel Level	Internal Validation Performance Metrics	External Validation Dataset/WSIs/No. of Patients (n)	External Validation Performance Metrics
Detection of malignant melanoma in H&E-stained images using deep learning techniques	Alheejawi	2021	INS-Net CNN	70 H&E-images/4 WSIs	960 × 960 px	N/A	15 H&E images/4 WSIs	Dice Coefficient: nuclei segmentation (89.17), nuclei classification (86.48), melanoma segmentation (85.10), melanoma detection (85.10)
iDermatoPath: a novel software tool for mitosis detection in H&E-stained tissue sections of malignant melanoma	Andres	2017	iDermatoPath	7 WSIs	105 × 105 px	Mitosis detection accuracy (81.4%), non-mitosis detection accuracy (84.1%)	78 WSIs	Dice coefficient (0.72), accuracy (83%)
Objective assessment of tumor infiltrating lymphocytes as a prognostic marker in melanoma using machine learning algorithms	Aung	2022	NN192	n = 139	0.4986 μm × 0.4986 μm	N/A	n = 764	AUC: 0.793
Diagnostic assessment of deep learning for melanocytic lesions using whole-slide pathologic images	Ba	2021	CNN	781 WSIs	0.25 μm × 0.25 μm	Weighted Error Scoring: deep learning algorithm (1.82%), mean weighted error (4.61%)	104 WSIs	Specificity (97.3%), sensitivity (96.5%)
Deep learning approach to predict sentinel lymph node status directly from routine histology of primary melanoma tumours	Brinker	2021	AutoPrognosis	200 WSIs	256 × 256 px	N/A	88 WSIs	AUC: Unmatched cases (61.8%), matched cases (55.0%)
Performance of automated classification of diagnostic entities in dermatopathology validated on multisite data representing the real-world variability of pathology workload	Brodsky	2022	CNN	N/A	0.26 lm/px	N/A	300 H&E images	Sensitivity 97.8%, 100%, 99%; specificity of 97.6%, 97.9%, 100% for melanoma, nevi, basal cell carcinoma respectively
Dermatopathology of malignant melanoma in the era of artificial intelligence: a single institutional experience	Cazzato	2022	Fast random forest algorithm	125 WSI/n = 63	1280 × 1080 px	Accuracy (92%), sensitivity (85%), specificity (99%)	N/A	Discrepancy (17%)
A deep learning model based on whole-slide images to predict disease-free survival in cutaneous melanoma patients	Comes	2022	ResSVM, DenseSVM, InceptionSVM	CPTAC-CM/n = 43	224 × 224 px	AUC: 57.3%	n = 11	AUC: ResSVM (64.7%), DenseSVM (64.8%), InceptionSVM (64.9%)
Predicting melanoma survival and metastasis with interpretable histopathologic features and machine learning models	Couetil	2022	CNN	90 WSIs	512 × 512 px	N/A	126 WSI/n = 90	F1 scores: 0.72, 0.73

(*continued on next page*)

TABLE 2. (*Continued*) Characteristics of AI Models Used in Review Studies

Article Title	Lead Author	Year Published	AI Model	Training and Validation Dataset/Patches/WSIs/No. of Patients (n)	Pixel Level	Internal Validation Performance Metrics	External Validation Dataset/WSIs/No. of Patients (n)	External Validation Performance Metrics
Machine learning to predict overall short-term mortality in cutaneous melanoma	Cozzolino	2023	Logistic Regression classifier, Support-Vector Machine, Random Forest, Gradient Boosting, k-Nearest Neighbors, Deep Neural Network	n = 2646	N/A	N/A	n = 265	Accuracy: Deep Neural network (91%), random forest model (88%)
Recognition of cutaneous melanoma on digitized histopathologic slides via artificial intelligence algorithm	DeLogu	2020	CNN	3518 patches/60 WSIs	299 × 299 px	N/A	40 WSIs/791 patches of healthy tissue, 1122 patches of pathologic tissue	Accuracy (96.5%), sensitivity (95.7%), specificity (97.7%), F1 score (96.5%)
Advantages of manual and automatic computer-aided compared to traditional histopathologic diagnosis of melanoma: A pilot study	Dika	2022	CNN	N/A	0.23 μm/px	N/A	70 H&E images	Discrepancy (19%)
Machine learning techniques in predicting braf mutation status in cutaneous melanoma from clinical and histopathologic features	Figueroa-Silva	2022	Random Forest, Support Vector Machine, and Extreme Gradient Boosting	n = 106	N/A	AUC: 87.8%	N/A	N/A
Mitotic rate in melanoma: prognostic value of immunostaining and computer-assisted image analysis	Hale	2013	Computer-assisted image analysis	380 WSIs/n = 190	N/A	$P = 0.02$	N/A	N/A
Pathologist-level classification of histopathologic melanoma images with deep neural networks	Hekler	2019	CNN	595 H&E images	N/A	N/A	100 H&E images	AI-pathologist diiscordance: 19%
Deep learning outperformed 11 pathologists in the classification of histopathologic melanoma images	Hekler	2019	CNN	595 H&E images	N/A	N/A	101 H&E images	Sensitivity (76%), specificity (60%), accuracy (68%)
Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images	Hu	2021	CNN	190 H&E images	256 × 256 tiles	N/A	109 H&E images	AUC: 77.8% on 54 melanoma H&E samples., 64.5% on 55 lung cancer samples
Multiphoton microscopy of the dermoepidermal junction and automated identification of dysplastic tissues with deep learning	Huttunen	2020	GoogLeNet CNN	1.2 million WSI	75 × 25 mm	N/A	108 WSIs	Specificity (95.2%), sensitivity (95.8%), accuracy (95.4%)
A machine learning algorithm for simulating immunohistochemistry: development of SOX10 virtual IHC and	Jackson	2020	vIHC CNN	16,309 H&E images	500 × 500 px	N/A	1813 H&E images	AUC (0.9422), sensitivity (91.62%), specificity (85.66%)

TABLE 2. (*Continued*) Characteristics of AI Models Used in Review Studies

Article Title	Lead Author	Year Published	AI Model	Training and Validation Dataset/Patches/WSIs/No. of Patients (n)	Pixel Level	Internal Validation Performance Metrics	External Validation Dataset/WSIs/No. of Patients (n)	External Validation Performance Metrics
evaluation on primarily melanocytic neoplasms								
Deep learning for the detection of anatomical tissue structures and neoplasms of the skin on scanned histopathologic tissue sections	Kriegsmann	2022	EfficientNetV2	101,313 image tiles/n = 1334 patients	395 × 395 px	Matthews correlation coefficient (0.96750)	27,239 tiles/n = 347	Accuracy: 84%
Deep learning based on standard H&E images of primary melanoma tumors identifies patients at risk for visceral recurrence and death	Kulkarni	2020	CNN	n = 108	N/A	N/A	n = 104	AUC: 0.905 (YSM), 0.880 (GHS)
Application of deep learning on the prognosis of cutaneous melanoma based on full-scan pathology images	Li	2022	Deep learning (VGG-19), machine Learning (SVM)	248 WSIs/n = 42 patients	N/A	5-fold cross-validation learning rate (0.01)	64 WSIs/n = 11 patients	Accuracy: deep learning (0.769); machine learning (0.65)
Deep learning approach to classify cutaneous melanoma in a whole-slide image	Li	2023	EfficientNetB1 CNN	66 WSIs	224 × 224 px, 512 × 512 px, 768 × 768 px, 1024 × 1024 px	Adaptive moment estimation (ADAM) optimizer learning rate (0.001)	90 WSIs	AUC: 0.821 (whole-slide image level), 0.936 (tile level)
Automated diagnosis and localization of melanoma from skin histopathology slides using deep learning: a multicenter study	Li	2021	CNN	596 WSIs	224 × 224 px	Learning rate (0.1)	105 WSIs	AUROC: 0.971
Deep visual proteomics defines single-cell identity and heterogeneity	Mund	2022	Unsupervised phenotype finder model, deep learning, machine learning	nucleAIzer 3	0.22 mm/px, 5.86 μm/px	F1 scores: melanoma nucleus (0.5498), melanoma cytoplasm (0.5336)	N/A	N/A
Computer-assisted annotation of digital H&E/SOX10 dual stains generates high-performing convolutional neural network for calculating tumor burden in H&E-stained cutaneous melanoma	Nielsen	2022	CNN	19 WSIs	512 × 512 px	N/A	6 WSIs	Sensitivity (88.8%), specificity (94.4%), accuracy (92.6%)
Machine learning techniques for mitoses classification	Nofallah	2021	ESPNet, DenseNet, ResNet, and ShuffleNet CNNs	6 WSIs	101 × 101 patches	Precision, recall, and F-score: ESPNet (0.961, 0.976, 0.968), DenseNet (0.984, 0.968, 0.976)	MITOS/6 WSIs	Precision, recall, F-score: ESPNet: (0.916, 0.866, 0.890), DenseNet (0.939, 0.916, 0.927), ResNet (0.931, 0.807, 0.865), ShuffleNet (0.968, 0.753, 0.847)
A texture-based pattern recognition approach to distinguish melanoma from non-melanoma cells in histopathologic tissue microarray sections	Rexhepaj	2013	Naïve Bayes Network, Random Forest classifier, Support Vector Machine	Melan-A_DISCOVERY/ *n* = 264	N/A	N/A	*n* = 157	AUC: 0.663–0.734

(*continued on next page*)

TABLE 2. (*Continued*) Characteristics of AI Models Used in Review Studies

Article Title	Lead Author	Year Published	AI Model	Training and Validation Dataset/Patches/WSIs/No. of Patients (n)	Pixel Level	Internal Validation Performance Metrics	External Validation Dataset/WSIs/No. of Patients (n)	External Validation Performance Metrics
Histologic screening of malignant melanoma, spitz, dermal, and junctional melanocytic nevi using a deep learning model	Snyder	2022	CNN ResNet50	22 WSIs	256 × 256 px	N/A	39 WSIs	Accuracy: irrelevant (96%), melanoma (93%), melanocytic nevi (94%), Spitz nevi (73%)
Computer-aided assessment of melanocytic lesions by means of a mitosis algorithm	Sturm	2022	Computer-aided diagnosis	102 WSIs	N/A	N/A	N/A	Accuracy: 89%, excluding nevoid melanoma (n = 89), comparable with and without the use of AI (89% vs. 90%)
Automated identification of malignancy in whole-slide pathologic images: identification of eyelid malignant melanoma in gigapixel pathologic slides using deep learning	Wang	2020	Deep learning system	44 WSIs	N/A	N/A	111 WSIs	Patch Level: AUC (98.9%), accuracy (94.9%), sensitivity (94.7%), specificity (95.3%); WSI Level: AUC (99.8%), accuracy (98.2%), sensitivity (100%), specificity (96.5%)
Prediction of early-stage melanoma recurrence using clinical and histopathologic features	Wang	2022	Support Vector Machine, Gradient Boosting, Random Forest, Logistic Regression, Multilayer Perceptron	1172 WSIs	N/A	Recurrence classification AUC (0.845), time-to-event prediction AUC (0.853)	548 WSIs	AUC: recurrence classification (81.2%), time-to-event prediction (82%)
Superpixel-Based Conditional Random Fields (SuperCRF): incorporating global and local context for enhanced deep learning in melanoma histopathology	Zormpas-Petridis	2019	SuperCRF CNN	432 tiles	2000 × 2000 px	Accuracy (97.7%) at 29,997 super px, Accuracy (97.1%) at 1798 super px	290 tiles	Accuracy: 96.48%
Self-supervised learning mechanism for identification of eyelid malignant melanoma in pathologic slides with limited annotation	Wang	2022	SSL CNN	PCam dataset	N/A	AUC (0.981)	ZJU-2 dataset	AUC: WSI-level (97.4%), patch-level (98.1%)

diagnosis potentially more difficult. In addition, deep learning networks may have valuable utility to ensure correct, cost-effective, and prompt diagnosis in cases where referring clinicians may have less experience, such as eyelid melanoma.

Dermatopathology AI differential diagnosis pipelines could promote medicine's smart technologic advancement as a smart screening tool for benign and malignant lesions. Deep learning could improve patient quality of life through the risk reduction of harm caused by false positives and false negatives. In addition, future AI differential diagnostic studies could evaluate bias in lesion detection on samples derived from patients with different Fitzpatrick tones.

Pathologists could use these models to increase their confidence in diagnosing difficult cases. Future research could examine the impact of deep learning output on clinical decision-making and patient well-being. Furthermore, incorporating AI in genetic mutation detection and screening

through the dermatopathologic route could reduce the diagnostic burden and prohibitive costs patients may experience.

There is still ample room for improvement in computer-aided diagnostic tools before they can be implemented outside of a research setting. Sturm et al[49] used WSI to study H&E-stained melanocytic skin lesions using an AI mitosis detection algorithm for melanocytic lesion diagnosis. It was concluded that the mitosis algorithm can be used to evaluate melanocytic skin lesions; however, it does not yet have utility in a clinical setting because the algorithm produces too many false-positive mitoses. In the future, AI could be regarded as a tool to observe the unobservable with the human eye. Moving forward, standardized approaches could be taken to develop a general model that could reduce intermodel variability. To do this, AI algorithms must be trained using massive amounts of data. Kriegsmann et al[50] sought to address the lack of publicly available annotated data and learning models designed to differentiate common skin cancers based on H&E images. They developed a curated dataset that could be used by other researchers to train and test future learning models. Future AI dermatopathology exploration should focus on the development of randomized clinical trials. This is especially essential when samples from independent institutions are tested on models trained on curated datasets.

AI use in clinical decision-making and research raises ethical concerns. Bias is a significant issue. A well-designed and balanced training set is crucial for accurate, precise, and helpful AI output. As of now, it is impossible to supply bias-free data sets. As such, AI outputs are currently bound to be affected by these biases.[51] These biased outputs may have profound consequences. As an example, a dataset used to train an AI algorithm may not perform as well for certain populations if it does not supply enough representative and accurate data or if it holds harmful data. The result may be an escalation in the disparities that already exist in health care and research. There could be consequences associated with this biased practice. These consequences could range from exacerbating health care disparities in historically marginalized communities to inaccurately diagnosing diseases at an early stage of development. To ensure patient well-being, it is essential to develop methods to mitigate bias in input–output before AI implementation in clinical practice.

Explainability is another concern for AI in health care and research settings. Patients and health care providers should understand how AI algorithms arrive at their decisions to assess their reliability and accuracy when making diagnoses and choosing treatments. The inability to understand how an AI system arrives at its decision has been termed the "black box problem." Without explainability, can providers and researchers truly educate patients such that they can give informed consent? Sauter et al[52] investigated the utility of automated concept-based explanation in comprehending CNN image analysis wherein results showed that automated concept-based explanation could be an effective tool for analyzing CNN decision-making and enhancing researcher/physician diagnostic and treatment decisions while improving the transparency of CNNs. When mechanisms are properly understood and data sets are used responsibly, AI could optimize pathology workflows, enhance diagnostic accuracy, help the provider engage in prognostic prediction, and find individualized treatments. AI has the potential to revolutionize medicine and help health care workers prioritize patient well-being.

REFERENCES

1. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. *Lancet Digit Health*. 2021; 3:e195–e203.
2. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60:84–90.
3. Sultan AS, Elgharib MA, Tavares T, et al. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J Oral Pathol Med*. 2020;49:849–856.
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:7–34.
5. Bolick NL, Geller AC. Epidemiology of melanoma. *Hematol Oncol Clin North Am*. 2021;35:57–72.
6. Dika E, Curti N, Giampieri E, et al. Advantages of manual and automatic computer-aided compared to traditional histopathological diagnosis of melanoma: a pilot study. *Pathol Res Pract*. 2022;237:154014.
7. De Logu F, Ugolini F, Maio V, et al. Recognition of cutaneous melanoma on digitized histopathological slides via artificial intelligence algorithm. *Front Oncol*. 2020;10:1559.
8. Alheejawi S, Berendt R, Jha N, et al. Detection of malignant melanoma in H&E-stained images using deep learning techniques. *Tissue Cell*. 2021;73:101659.
9. Li T, Xie P, Liu J, et al. Automated diagnosis and localization of melanoma from skin histopathology slides using deep learning: a multicenter study. *J Healthc Eng*. 2021;2021:5972962.
10. Snyder AN, Zhang D, Dreesen SL, et al. Histologic screening of malignant melanoma, Spitz, dermal and junctional melanocytic nevi using a deep learning model. *Am J Dermatopathol*. 2022;44:650–657.
11. Brodsky V, Levine L, Solans EP, et al. Performance of automated classification of diagnostic entities in dermatopathology validated on multisite data representing the real-world variability of pathology workload. *Arch Pathol Lab Med*. 2023;147:1093–1098.
12. Li M, Abe M, Nakano S, et al. Deep learning approach to classify cutaneous melanoma in a whole slide image. *Cancers (Basel)*. 2023; 15:1907.
13. Wang L, Ding L, Liu Z, et al. Automated identification of malignancy in whole-slide pathological images: identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning. *Br J Ophthalmol*. 2020;104:318–323.
14. Ghazvinian Zanjani F, Zinger S, Piepers B, et al. Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images. *J Med Imaging (Bellingham)*. 2019;6:027501.
15. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, et al, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*. Vol 9351.Cham: Springer; 2015:234–241.
16. Rodriguez-Sains RS, Jakobiec FA, Iwamoto T. Lentigo maligna of the lateral canthal skin. *Ophthalmology*. 1981;88:1186–1192.
17. Wang L, Jiang Z, Shao A, et al. Self-supervised learning mechanism for identification of eyelid malignant melanoma in pathologic slides with limited annotation. *Front Med (Lausanne)*. 2022;9:976467.
18. Hekler A, Utikal JS, Enk AH, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur J Cancer*. 2019;115:79–83.
19. Ba W, Wang R, Yin G, et al. Diagnostic assessment of deep learning for melanocytic lesions using whole-slide pathological images. *Transl Oncol*. 2021;14:101161.
20. Cazzato G, Massaro A, Colagrande A, et al. Dermatopathology of malignant melanoma in the era of artificial intelligence: a single institutional experience. *Diagnostics (Basel)*. 2022;12:1972.
21. Jackson CR, Sriharan A, Vaickus LJ. A machine learning algorithm for simulating immunohistochemistry: development of SOX10 virtual IHC

and evaluation on primarily melanocytic neoplasms. *Mod Pathol*. 2020; 33:1638–1648.
22. Nielson PS, Georgsen JB, Vinding MS, et al. Computer-assisted annotation of digital H&E/SOX10 dual stains generates high-performing convolutional neural network for calculating tumor burden in H&E-Stained cutaneous melanoma. *Int J Environ Res Public Health*. 2022;19:14327.
23. Cheng L, Lopez-Beltran A, Massari F, et al. Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine. *Mod Pathol*. 2018;31:24–38.
24. Rexhepaj E, Agnarsdóttir M, Bergman J, et al. A texture based pattern recognition approach to distinguish melanoma from non-melanoma cells in histopathological tissue microarray sections. *PLoS One*. 2013;8: e62070.
25. Koh SS, Lau SK, Cassarino DS. Absence of differential cyclin D1 immunohistochemical protein expression in nevi and melanoma evaluated by digital image analysis. *Appl Immunohistochem Mol Morphol*. 2022;30: 441–445.
26. Lezcano C, Jungbluth AA, Nehal KS, et al. PRAME expression in melanocytic tumors. *Am J Surg Pathol*. 2018;42:1456–1465.
27. Comes MC, Fanizzi A, Bove S, et al. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs. *Scientific Rep*. 2021;11:14123.
28. Comes MC, La Forgia D, Didonna V, et al. Early prediction of breast cancer recurrence for patients treated with neoadjuvant chemotherapy: a transfer learning approach on DCE-MRIs. *Cancers (Basel)*. 2021;13: 2298.
29. Comes MC, Fucci L, Mele F, et al. A deep learning model based on whole slide images to predict disease-free survival in cutaneous melanoma patients. *Sci Rep*. 2022;12:20366.
30. Kulkarni PM, Robinson EJ, Sarin Pradhan J, et al. Deep learning based on standard H&E images of primary melanoma tumors identifies patients at risk for visceral recurrence and death. *Clin Cancer Res*. 2020;26:1126–1134.
31. Brinker TJ, Kiehl L, Schmitt M, et al. Deep learning approach to predict sentinel lymph node status directly from routine histology of primary melanoma tumours. *Eur J Cancer*. 2021;154:227–234.
32. Aung TN, Shafi S, Wilmott JS, et al. Objective assessment of tumor infiltrating lymphocytes as a prognostic marker in melanoma using machine learning algorithms. *EBioMedicine*. 2022;82:104143.
33. Zormpas-Petridis K, Failmezger H, Raza SEA, et al. Superpixel-based conditional random fields (SuperCRF): incorporating global and local context for enhanced deep learning in melanoma histopathology. *Front Oncol*. 2019;9:1045.
34. Azzola MF, Shaw HM, Thompson JF, et al. Tumor mitotic rate is a more powerful prognostic indicator than ulceration in patients with primary cutaneous melanoma: an analysis of 3661 patients from a single center. *Cancer*. 2003;97:1488–1498.
35. Thompson JF, Soong SJ, Balch CM, et al. Prognostic significance of mitotic rate in localized primary cutaneous melanoma: an analysis of patients in the multi-institutional American Joint Committee on Cancer melanoma staging database. *J Clin Oncol*. 2011;29:2199–2205.
36. 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.
37. Andres C, Andres-Belloni B, Hein R, et al. iDermatoPath: a novel software tool for mitosis detection in H&E-stained tissue sections of malignant melanoma. *J Eur Acad Dermatol Venereol*. 2017;31:1137–1147.
38. Nofallah S, Mehta S, Mercan E, et al. Machine learning techniques for mitoses classification. *Comput Med Imaging Graph*. 2021;87:101832.
39. Hale CS, Qian M, Ma MW, et al. Mitotic rate in melanoma: prognostic value of immunostaining and computer-assisted image analysis. *Am J Surg Pathol*. 2013;37:882–889.
40. Couetil J, Liu Z, Huang K, et al. Predicting melanoma survival and metastasis with interpretable histopathological features and machine learning models. *Front Med (Lausanne)*. 2022;9:1029227.
41. Hu J, Cui C, Yang W, et al. Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images. *Translational Oncol*. 2021;14:100921.
42. Gershenwald JE, Scolyer RA, Hess KR, et al. Melanoma staging: evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA: A Cancer J Clinicians*. 2017;67:472–492.
43. Nagore E, Requena C, Traves V, et al. Prognostic value of BRAF mutations in localized cutaneous melanoma. *J Am Acad Dermatol*. 2014;70: 858–862.e1-2.
44. Garcia-Casado Z, Traves V, Banuls J, et al. BRAF, NRAS and MC1R status in a prospective series of primary cutaneous melanoma. *Br J Dermatol*. 2015;172:1128–1131.
45. Vanni I, Tanda ET, Spagnolo F, et al. The current state of molecular testing in the BRAF-mutated melanoma landscape. *Front Mol Biosci*. 2020;7:113.
46. Figueroa-Silva O, Pastur Romay LA, Viruez Roca RD, et al. Machine learning techniques in predicting braf mutation status in cutaneous melanoma from clinical and histopathologic features. *Appl Immunohistochem Mol Morphol*. 2022;30:674–680.
47. Mund A, Coscia F, Kriston A, et al. Deep visual proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol*. 2022;40:1231–1240.
48. Ronen S, Al-Rohil RN, Keiser E, et al. Discordance in diagnosis of melanocytic lesions and its impact on clinical management. *Arch Pathol Lab Med*. 2021;145:1505–1515.
49. Sturm B, Creytens D, Smits J, et al. Computer-aided assessment of melanocytic lesions by means of a mitosis algorithm. *Diagnostics (Basel)*. 2022;12:436.
50. Kriegsmann K, Lobers F, Zgorzelski C, et al. Deep learning for the detection of anatomical tissue structures and neoplasms of the skin on scanned histopathological tissue sections. *Front Oncol*. 2022;12: 1022967.
51. Reva S, Apostol V, Kristen KG, et al. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. Gaithersburg, MD: Special Publication (NIST SP), National Institute of Standards and Technology; 2022.
52. Sauter D, Lodde G, Nensa F, et al. Validating automatic concept-based explanations for AI-based digital histopathology. *Sensors (Basel)*. 2022; 22:5346.