

The role of artificial intelligence and convolutional neural networks in the management of melanoma: a clinical, pathological, and radiological perspective

Joshua Yee^a, Cliff Rosendahl^b and Lauren G. Aoude^c

Clinical dermatoscopy and pathological slide assessment are essential in the diagnosis and management of patients with cutaneous melanoma. For those presenting with stage IIC disease and beyond, radiological investigations are often considered. The dermatoscopic, whole slide and radiological images used during clinical care are often stored digitally, enabling artificial intelligence (AI) and convolutional neural networks (CNN) to learn, analyse and contribute to the clinical decision-making. A keyword search of the Medline database was performed to assess the progression, capabilities and limitations of AI and CNN and its use in diagnosis and management of cutaneous melanoma. Full-text articles were reviewed if they related to dermatoscopy, pathological slide assessment or radiology. Through analysis of 95 studies, we demonstrate that diagnostic accuracy of AI/CNN can be superior (or at least equal) to clinicians. However, variability in image acquisition, pre-processing, segmentation, and feature extraction remains challenging. With current technological abilities, AI/CNN and clinicians synergistically working together are better than one another in all subspecialty domains relating to cutaneous melanoma. AI has the

potential to enhance the diagnostic capabilities of junior dermatology trainees, primary care skin cancer clinicians and general practitioners. For experienced clinicians, AI provides a cost-efficient second opinion. From a pathological and radiological perspective, CNN has the potential to improve workflow efficiency, allowing clinicians to achieve more in a finite amount of time. Until the challenges of AI/CNN are reliably met, however, they can only remain an adjunct to clinical decision-making. *Melanoma Res* 34: 96–104 Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc.

Melanoma Research 2024, 34:96–104

Keywords: artificial intelligence, convolutional neural network, dermatoscopy, melanoma, radiomics, whole slide imaging

^aFaculty of Medicine, University of Queensland, St Lucia, ^bPrimary Care Clinical Unit, Medical School, The University of Queensland, Herston and ^cFrazer Institute, The University of Queensland, Woolloongabba, QLD, Australia

Correspondence to Lauren G. Aoude, PhD, Frazer Institute, The University of Queensland, 37 Kent Street, Woolloongabba, QLD, 4102, Australia
Tel: +61 7 3443 8030; e-mail: l.aoude@uq.edu.au

Received 16 August 2023 Accepted 29 November 2023.

Introduction

Over the last 30 years, dermatoscopy has become a widely accepted non-invasive clinical tool for assessment of skin lesions. If melanoma is suspected, the patient is offered an excisional biopsy for histopathological analysis. Despite the use of dermatoscopy, dermatologists rarely achieve greater than 80% accuracy in clinical test sensitivity [1]. To improve diagnostic accuracy for melanoma patients, additional tools such as artificial intelligence (AI) are warranted.

Historically, the excision biopsy specimen is assessed by a dermatopathologist through microscopic analysis using glass slides. Through technological advances, digital pathology whole slide imaging (WSI) has allowed these specimens to be assessed using digital monitors, scaling magnification as necessary and interpreting these specimens from remote locations [2,3].

Radiological investigations are used to examine and define the extent of disease for patients presenting with regional or metastatic disease [4,5]. These modalities include positive emission tomography (PET), computer tomography (CT) and MRI. These images are digitally stored for future interpretation by a radiologist and also allow for supplementary analyses.

Exponential increases in computing power have dramatically increased the ability of AI subfields of machine learning, neural networks, and deep learning to process and identify traits in imaging datasets [6–8]. Artificial neural networks are based on a set of algorithms that can mimic the human decision-making processes and can be trained in pattern recognition with greater consistency and accuracy than the human brain [7].

Convolutional neural networks (CNN) are a form of supervised deep learning, an extension of artificial neural networks. CNN are predominantly used in image-based pattern recognition [9]. Studies utilising CNN to interpret clinical dermatoscopy, pathological whole slide, and radiological images, have gained increasing popularity. More specifically for cutaneous melanoma, CNN use

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

in clinical dermatoscopy has been studied to determine accuracy and reliability in comparison to dermatologists (summarised in Table 1). Similarly, this has been performed with pathological whole slide images. AI or CNN have also been utilised in radiology to correlate image features with biomarkers, genetics, immunotherapy treatment effectiveness [20–26], and diagnostic refinement in metastatic disease of unknown primary [27].

We have reviewed the literature on the capabilities, limitations and progression of AI and CNN in the clinical management of cutaneous melanoma. In this context, we holistically examined the use of dermatoscopic digital photographs, pathological whole-slide images, and radiology, to analyse whether common benefits, limitations and challenges exist.

Methods

A keyword search of the Medline (PubMed) database was performed. The inclusion criteria were articles on cutaneous melanoma published in English. For dermatoscopy, the keywords ‘Convolutional neural network melanoma dermatoscopy’ were used resulting in 73 articles for review. Within these articles, 55 discussed diagnostic, algorithmic and differentiative accuracy and image segmentation processes, 12 articles compared CNN to clinicians, 2 articles discussed CNN working with clinicians, 2 articles used CNN in serial digital dermatoscopic monitoring, and 1 article compared CNN to traditional AI.

The keywords ‘Convolutional neural network whole slide image melanoma’ revealed 12 articles for review. Of these, 8 were related to dataset and segmentation, and 4 compared the accuracy of CNN to pathologist diagnosis.

From a radiological perspective, the keywords ‘artificial intelligence radiomics melanoma’ resulted in 10 articles for review.

The exclusion criteria were articles not relevant to cutaneous melanoma or the subject matter, article replies and articles lacking an abstract. Articles relating to uveal and acral melanomas were also excluded. In total, 36 articles were excluded.

Clinical dermatoscopy and convolutional neural networks

Basic science

Of the three forms of deep neural network (CNN, Recurrent Neural Networks and XG-Boost), CNN is the most studied and suitable for image classification, especially for cutaneous melanoma. When Computer-aided diagnosis or AI is involved with image processing, input dermatoscopic images are fed through the system undergoing five key steps: raw image acquisition, image pre-processing, region of interest segmentation, feature extraction, and classifier output [28].

Raw image acquisition is standardised, to a degree, with image pre-processing which addresses illumination, colour indifferences, artefacts, and background/foreground differentiation. Artefacts, such as hair, present a more significant challenge compared to non-essential vessels and skin lines.

Segmentation is a process where an image is isolated to a region of interest to allow for accurate classification. From a computer-assisted diagnostic perspective, this can either be semi-automated (with interactive user input) or a fully automated. Fully automated segmentation can be unsupervised (untrained) or supervised (pre-trained) [29]. Ideally, hair removal, colour standardisation, and dermatoscopic scales are managed during the pre-processing phase. Fully automated segmentation continues to be challenging and can be a rate-limiting step affecting the overall accuracy of AI.

Classification output is the final step and can either be binary (i.e. melanoma or naevus) or multi-class. While a subset of studies published original CNN algorithms, most use an established CNN algorithm that has been pre-trained using transfer learning. The details of each CNN algorithm and its respective history are beyond the scope of this review. Pre-training of CNN models is usually completed through a widely accepted database called ImageNet, which allows the model to recognise basic shapes and objects. These models are further trained in dermatoscopic image recognition and tested for accuracy either on the same or alternative dataset.

Results from CNN are usually reported based on International Skin Imaging Collaboration (ISIC) challenges with a common goal of attaining state-of-the-art performance. Performance improved by increasing number of layers or complexity of the CNN [30] and by combining predictions from multiple CNN algorithms (ensembling) and pooling the outcomes [31].

Performance comparison between CNN and clinicians

The first study comparing board-certified dermatologists and CNN was published in 2017 with promising results showing comparable performance between both groups [10]. In total, we have identified 10 studies comparing CNN and dermatologists, summarised in Table 1 [1,11,12,14–19].

Nine of the 10 studies favoured CNN’s performance over the dermatologists [1,11–17,19].

Conversely, Winkler *et al.*, showed results in favour of the dermatologist group, possibly due to a small number of cases which included rare diseases for which training opportunities are not frequent [18]. Importantly, they concluded that hive dermatologists (multiple clinicians work together) have superior diagnostic accuracy when compared to individual dermatologists or AI algorithms, specifically CNN models.

Table 1 Published studies comparing CNN versus dermatologists

Reference	CNN algorithm	Clinician characteristics	Dataset	Summary
Esteva <i>et al.</i> , 2017 [10]	GoogleNet Inception v3 Pre-trained on 2014 ImageNet	21 board-certified dermatologists	Two test series (epidermal and melanocytic) derived from ISIC, Edinburgh Dermofit Library and Stanford Hospital	CNN matched dermatoscopic interpretation performance to board-certified dermatologists on keratinocyte carcinoma versus seborrheic keratosis and malignant melanoma versus benign naevi classification.
Haenssle <i>et al.</i> , 2018 [11]	GoogleNet Inception v4	58 dermatologists from 17 countries (17 beginners, 11 skilled, 30 expert)	100 images extracted from private dataset collated by the University of Heidelberg. 20% invasive or in-situ melanoma, 80% non-melanoma. Not all benign lesions confirmed histologically.	Binary comparison between melanoma and melanocytic naevi. Mean specificity, sensitivity, ROC were statistically superior in CNN. 13 dermatologists outperformed CNN.
Brinker <i>et al.</i> , 2019 [12]	ResNet 50 Pre-trained using same dataset	144 dermatology clinicians from 9 university hospitals in Germany (92 junior, 52 board-certified)	804 histological confirmed images extracted from the International ISIC dataset. Equal ratio of melanoma and naevi.	Binary comparison between melanoma and melanocytic naevi. Significant superiority by CNN when compared to both junior dermatology clinicians and board-certified dermatologists. Sensitivity by junior clinicians higher than board-certified dermatologists.
Brinker <i>et al.</i> , 2019 [1]	ResNet 50	157 dermatologists from 12 university hospitals in Germany	100 images extracted from ISIC public dataset. 80% atypical naevi, 20% melanoma.	Binary comparison between melanoma and melanocytic naevi. Only 7/157 dermatologists outperformed CNN in both sensitivity and specificity. Junior physicians showed highest sensitivity, but low specificity.
Tschanl <i>et al.</i> , 2019 [13]	MetaOptima Technology, DAISYLab, Sun Yat-sen University (Ensemble CNN)	511 clinicians (283 board-certified dermatologists, 118 dermatology residents, 83 general practitioners/primary care physicians)	1195 images from private and custom datasets from ViDIR in Vienna, Austria, Australia and external images from Turkey, New Zealand, Sweden, Argentina. (a subset of the HAM10000 dataset)	Multiclass comparison of 7 disease categories: intraepithelial carcinoma, BCC, benign keratinocytic lesion, dermatofibroma, melanoma, melanocytic naevus, and vascular lesions. Mean of 6.7–6.8% superiority in favour of all ensemble CNN in correct diagnoses. Mean sensitivity for melanoma in all clinicians, expert clinicians (>10 years experience) and combination of top three ensemble CNNs were 73.1%, 67.8, and 81.9% respectively. Equally, mean specificity were 92.8%, 94.0% and 96.2% respectively.
Fink <i>et al.</i> , 2020 [14]	Google Inception v4 (Moleanalyzer Pro, FotoFinder Systems)	11 dermatologists (3 beginner, 5 skilled, 3 expert)	72 histologically confirmed images from the University of Heidelberg private dataset. 36 melanomas and 36 combined naevi 'melanoma stimulators'	Binary comparison between melanoma and combined naevus. Specificity increased with experience. Significant increase in specificity of beginners when using CNN in as an adjunct.
Hanssle <i>et al.</i> , 2021 [15]	Google Inception v4 (Moleanalyzer Pro, FotoFinder Systems, Bad Birnbach, Germany)	64 dermatologists	100 randomised dermatoscopy images of face and scalp lesions from private university and hospital-based datasets from Germany, Australia and Greece. Additional Australian data from primary care setting– ISIC 2018, MSK-1, prospective series.	Binary comparison of malignant and benign lesions. CNN outperformed dermatologists mean sensitivity by 12%. Australian dataset showed significantly lower specificity, potentially related to CNN segmentation issues, and co-located pre-cancerous/benign lesions. More experienced dermatologists achieved better results.
Maron <i>et al.</i> , 2019 [16]	ResNet50	112 dermatologists from 13 German university-based hospitals	300 biopsy proven images from ISIC/ HAM10000 dataset. Comprises of images from different skin colours, collected by different camera systems.	Primary end-point was binary comparison of malignant versus benign lesions. Secondary end-point was correct multi-class. CNN significantly outperformed dermatologists in primary and secondary end-points.
Minagawa <i>et al.</i> , 2021 [17]	ResNet Inception v2	30 Japanese dermatologists (6 beginners, 8 skilled, 16 expert) with mean age of 32.5 years	50 biopsy proven images extracted from ISIC 2017/HAM10000/BCN20000 datasets. 50 images (not all biopsy proven) extracted from private Japanese Shinshu University hospital dataset.	Binary comparison of malignant versus non-malignant lesions. Compared multi-class diagnostic accuracy between dermatologists and trained CNN. Japanese dermatologist had lower sensitivity and specificity for 50 non-Japanese dermatoscopic images. Performance hindered on non-familiar images. CNN was statistically superior and may help close the gap for unfamiliar images.
Winkler <i>et al.</i> , 2021 [18].	GoogleNet Inception v4 (FotoFinder Systems, Bad Birnbach)	120 dermatologists (no data on epidemiology) who attended a conference (30 years of dermoscopy, Germany)	30 cases presented to a group of 120 dermatologists, including clinical, dermatoscopic and images with clinical information. CNN analysed dermatoscopic images.	Binary and multi-class comparison between collective group, individual dermatologist, and CNN. CNN was inferior to both collective and individual dermatologists.
Barata <i>et al.</i> , 2023 [19]	Deep Q-Learning – supervised learning (SL) and reinforcement learning (RL) models	89 dermatologists	10,5015 images from the HAM10000 dataset used to train the models. Included melanoma, BCC, pre-cancerous lesions and benign lesions. Reader study with dermatologist input was performed using 1511 retrospective images from Austria, Australia, New Zealand, Sweden and Argentina.	Comparison of diagnosis with and without AI support. The rate of correct diagnosis using the SL model increased accuracy from 68.0% to 75.5%. Using the RL model, this further increased to 79.9%. Using the SL model, sensitivity increased from 62.4% to 69.4%. this further increased to 83.9% using RL.

BCC, basal cell carcinoma; CNN, convolutional neural network; ISIC, International Skin Imaging Collaboration; ROC, receiver operating characteristic; ViDIR, Vienna Dermatologic Imaging Research Group.

Notably, CNN models incorporating ensembled algorithms, achieved the top three places in the ISIC 2018 challenge [32]. Ensembled algorithms (considered as hives) are superior to individual algorithms in multiple studies [31,33,34]. Moreover, in a clinician versus ensemble CNN, multi-class reader study, Tschandl *et al.* showed that there was a 6.7–6.8% advantage in favour of ensemble CNN [13].

Multiple studies examined the role of clinical experience in melanoma diagnostics [1,12,15]. Studies indicated that junior dermatology trainees showed higher sensitivity but lower specificity when compared to more experienced dermatologists. This suggests that for less experienced dermatoscopists, CNN would undoubtedly be a powerful adjunct in clinical decision-making.

Current challenges and limitations

While CNN has been shown to be superior, there are challenges and limitations associated with this approach. Maron *et al.*, describe a study that replicated real-world application of CNN [35]. Significant shortcomings included brittleness, adversarial attacks and difficulty in discernment. Brittleness occurs when small changes in an image such as zoom, rotation and dermatoscopic have significant effect on CNN classification. Adversarial attacks are the result of intentionally deceptive images created and designed to fool CNN. To offset this challenge, including the CNN output confidence level will allow greater trust by the treating physician.

CNNs have difficulty in discerning non-biological and insignificant artefacts. For example, if all melanomas in a dataset had a dermatoscopic scale, the CNN would falsely learn that the dermatoscopic scale is a ‘feature’ of melanoma, which is inaccurate [35,36]. Actual CNN performance is based on prior training or transfer learning. Depending on training quality, training bias can be a significant problem that may overestimate the robustness and usefulness of CNN in clinical decision-making [36]. Lastly, the exact learning process of CNN is still difficult to grasp, and if significant erroneous outputs are seen, they may be challenging to explain and troubleshoot [37].

Current utility

It has been shown that dermatologists working in conjunction with AI significantly increases diagnostic accuracy [19,38,39]. A recent study investigated AI-based decision support in skin cancer diagnostics [19]. Using a reinforcement learning (RL) model, dermatologists created a reward/penalty system based on the type of skin lesion. The rate of correct diagnoses for this group of 89 dermatologists with AI RL model support increased by 12%.

Studies reviewed here predominantly use laboratory simulated methods which do not account for other

potential patient factors that may be important in a clinician’s final decision. The use of CNN models would be more reliable as ‘second reader’ and adjunct for clinical decision-making as opposed to a fully reliant process [13].

Future outlook

As outlined above, the multiple challenges and limitations associated with the use of unsupervised CNNs will likely prevent a fully automated process in the near future. However, combining CNNs with additional inputs, has the potential to significantly improve diagnostic outcomes for melanoma patients. This is particularly relevant when used in combination with pathology and radiology.

Whole slide imaging and convolutional neural networks

Like clinical dermatoscopy, digital pathology WSI is a pixel-based representation of the lesion which is examined on digital monitors. Compared to dermatoscopy, the image processing steps for WSI are perhaps more simplified as there can be less variation in image acquisition and fewer artefacts [40]. Image pre-processing may not be a necessary step towards the final output classification accuracy. Therefore, a stepwise approach may be simplified and focussed on segmentation, feature extraction, and classification.

Current challenges and limitations

Digital WSI presents unique challenges. The images are huge and usually kept at 20X to 40X magnification with multiple regions of interest (ROI) within the one slide [40–42]. The average WSI image is approximately 2000 times larger than the average dermatoscopic image. With current computational capabilities, processing end-to-end WSI efficiently is not entirely realistic. These large gigapixel images can be divided into non-overlapping instances, tiles or patches, allowing for more efficient segmentation, feature extraction and classification processing [40]. Alternatively, if WSI is analysed without tiling or patching with current computing power, a lower magnification (e.g. 5X) could increase computational speed, but at the theoretical risk of increased misclassification [42]. Robust comparison studies are needed to compare classifier accuracy at different magnifications.

WSI datasets are more heterogenous than dermatoscopic imaging datasets. Studies are predominantly based on retrospective slides derived from a local pathology or university department, unlike dermatoscopic datasets where there is some degree of standardisation. Without standardisation of datasets, there may be ambiguity in training, allowing CNNs to extract features that are unrelated to biology, causing erroneous learning and final outputs. These unintended but learnable variables are known as batch effects [43]. Finally, these small datasets are based solely on H&E staining [40,41]. To significantly improve

diagnostic outcomes, other melanoma-specific staining using WSI and CNN should be studied.

Current utility

Despite the challenges and limitations, there are many possible clinical applications for WSI (Summarised in Table 2). CNN could assist with the intra- and inter-observer variability between dermatopathologists in differentiating naevus and cutaneous melanoma by providing a 'second reader' [40–42]. Where there is a lack of an expert dermatopathologist, CNN could assist with workflow efficiency improving diagnostic turnaround times [41]. Segmentation of ROIs for further pathologist verification can also accelerate the workflow and volume of work. On average, it takes a pathologist several minutes to interpret, whereas CNN can perform the task in a matter of seconds [42,44]. Using a pathomics (pattern recognition of WSI based on genomics) approach, WSI has been used to examine differences between BRAF-positive and BRAF wild-type tumours and found statistically significant differences in nuclei between the groups [45].

Finally, CNN has been used to detect prognostic features such as stroma-to-tumour ratio and immune infiltrate, which have both been shown to be prognostic and predictive biomarkers [42]. Several studies have shown that electronic quantification of tumour infiltrating lymphocytes (TILs) using CNN has prognostic significance and can be an effective tool for the identification of patients at high risk of disease recurrence [49–52]. This has been shown in both primary and metastatic disease as a complementary assessment for staging [53,54]. Furthermore, machine learning in this setting has used TILs assessment to predict patient response to immune checkpoint inhibitors [51,53].

Future outlook

As computational capability and storage increases, end-to-end WSI gigapixel images may be processed at a speed acceptable for future clinical integration. Increased computational capabilities will also allow more features to be extracted and trained, but will still require a large, collaborative, and standardised international dataset to be available. While studies have reported CNN to be similar, if not superior to pathologists [40,44], the future direction is similar to that of clinical dermatoscopy, where it is unlikely to replace the role of a pathologist entirely.

Radiomics and artificial intelligence

In melanoma care, radiology has a key role in disease staging and management of patients [4]. The various modalities (PET, CT, MRI) not only provide information for diagnosis, but also provide evidence of recurrence and disease progression as well as response to treatment. The radiology images are digitally stored, allowing for additional analyses using AI. Radiomics methods are used

to extract numerous image features from medical scans using algorithms such as CNN. While the AI workflow in radiomics is similar to the dermatoscopic and WSI workflows, there are some significant features that are unique to radiology [21].

Radiology images are sequenced in a three-dimensional perspective and then re-formatted to be interpreted on a two-dimensional computer screen. Therefore, segmentation accounts for a volume of interest (VOI) instead of an ROI. Pre-processing and VOI segmentation is crucial in radiomics, and prior to feature extraction, artefacts such as air, streak, and calcific deposits should be removed [22].

Feature extraction is synonymously the most critical step, but in radiomics, hand-crafted feature extraction is more common. Hand-crafted features of shapes and textures are defined manually but could be partially assisted with mathematical functions. Fully automated CNN feature extraction requires training on large datasets to uphold accuracy, which is lacking in radiomics for cutaneous melanoma. Current studies are based on small sample sizes [21,22,27], and similar to WSI, there is a lack of standardised datasets for AI training [27]. As seen in dermatoscopic and WSI studies, radiomics and AI studies are performed retrospectively [20].

Current challenges and limitations

Imaging acquisition is a common challenge in radiomics as differences in imaging modalities, scanner models, and scanning protocols are specific to radiology [20,21]. In addition, more significant imaging acquisition challenges may exist as staging melanoma can incorporate ultrasound, CT, MRI, and PET with variable sensitivity and specificity. For example, F-FDG PET CT has shown superior sensitivity for in-transit melanomas compared to other imaging modalities [21]. With such dynamic imaging modality variation, radiomic features extracted and analysed will need to account for this. Furthermore, small metastatic lesions (<5 mm) may be challenging to analyse and segment due to scan acquisition limits [22].

Specifically, to PET imaging and immunotherapy research, various criteria have been established to define solid tumour response to immunotherapy. Depending on the study design, this could alter results, and risk misclassification of response. As such, agreement to the optimal criteria is still debated [21,23].

Radiomic features are highly dependent on image acquisition (machine variability) and reconstruction (software variability) [20]. Current studies show that while there are statistically significant radiomic features associated with overall survival and progression free survival, these features are based on small sample sizes and it would be difficult to have them universally accepted [23,24,55].

Table 2 Key studies comparing outcomes between pathologists and WSI assessment using CNN

Reference	Segmentation classifier protocol	Dataset	Proposed segmentation OR training method to improve segmentation	Summary
De Logu <i>et al.</i> , 2020 [41]	ResNet V2	University of Florence Department of Pathology. H&E of primary invasive cutaneous melanoma, n = 100, Breslow >2 mm)	ROI patches extracted from WSI. ROI were defined and labelled by two dermatopathologists, then trained and tested with CNN to assess performance	CNN had potential to give more detailed information on pathological cases, defining heat maps that distinguish healthy and pathological areas. High concordance between pathologist and CNN. Misclassification seen in patients with dermal solar elastosis and epidermal atrophy (chronic UVR exposed sites).
Wu <i>et al.</i> , 2021 [40]	Scale-Aware Transformer Network	240 H&E	WSI comparison between pathologist, CNN and ground truth	While accuracy between pathologist and CNN was comparable, the subset of dysplastic naevus seemed inferior with CNN.
Xie <i>et al.</i> , 2021 [44]	Grad-CAM	841 H&E of melanoma and naevus. Central South University Xiangya Hospital	WSI comparison between proposed CNN model and 20 pathologists for specificity, sensitivity and accuracy.	CNN superior when compared to pathologist. Model identified salient features through heat maps. Additional clinical data was helpful for pathologist and may also aide CNN
Kim <i>et al.</i> , 2022 [45]	Inception	256 H&E New York University	Predicting BRAF mutation through WSI analysis (Pathonomics)	When compared to BRAF-wild type, BRAF mutated nuclei were shown to be statistically larger (in radii), and rounder (in form factor, solidity, extent, and eccentricity)
Klein <i>et al.</i> , 2021 [46]	U-Net	H&E from 90 patients with metastatic melanoma. University Hospital Cologne	Used WSI to determine association with TILs and CPI treatment response in metastatic melanoma	TIL clusters reveal a predictive response/resistance to CPI. Elevated TIL clusters showed higher response to CPI in BRAF-positive tumours. High TIL counts were associated with increased survival.
Hohn <i>et al.</i> , 2021 [47]	ResNeXt50	431 images (430 patients)	Used WSI to examine CNN accuracy when patient data was incorporated (age, sex, location).	Patient data did not improve CNN accuracy unless the confidence level without patient data was low
Li <i>et al.</i> , 2021 [48]	ResNet50	701 images (583 patients). Multi-centre database. Chinese University Hospitals	Assessing AUROC including both melanoma and naevus (intra-dermal, compound, junctional)	Very high AUROC 0.971, showing promising results for full automation ability for CNN and WSI
Schmitt <i>et al.</i> , 2021 [43]	ResNet50, DenseNet21, VGG16	427 H&E Slides from 5 different institutions	Batch effects that are learned by CNN can cause significant misclassification and accuracy issues. Batch effect variables that were studied included patient age, slide preparation date, slide origin, scanner type	Hidden variables can cause significant accuracy variability. Preparation date of the slide and patient's age were the biggest factor that caused significant accuracy variability
Zormpass-Petridis <i>et al.</i> , 2020 [42]	SuperHistopath/Xception	127 melanoma H&E	Introduction of a novel SuperHistopath framework and modified Xception CNN to analyse 5X magnified WSI and segment ROI breast cancer, melanoma and neuroblastoma	Accurate segmentation and ability to determine prognostic histological features (some are seen to have high intra and inter-observer variability within pathologists)

AUC ROC, area under the curve of the receiver operating characteristic; CNN, convolutional neural network; CPI, check point inhibitor; H&E, haematoxylin and eosin staining; ROI, region of interest; TILs, tumour-infiltrating lymphocytes; UVR, ultraviolet radiation; WSI, whole-slide imaging.

Challenging radiomics is Immuno-PET, another frontier that may prove promising. Immuno-PET is based on injecting targeted radionuclear antibodies to determine potential immune response and can prove more valuable than extracting radiomics features [56]. Refer to Table 3 for a summary of the literature.

Current utility

Despite its challenges and limitations, there are potential uses for radiomics and AI. These methods could assist with a more precise melanoma diagnosis and individualised therapy. They can be used to predict immunotherapy response in metastatic melanoma prior to commencing treatment [20,22,24–26]. Furthermore, radiomics can aide in-vivo classification of disease, reducing invasive diagnostic testing such as biopsy [21]. This can improve a radiologist's workflow efficiency, where AI may segment ROI for further interpretation [27]. More

specifically and relating to metastatic melanoma to the brain, genetic BRAF p.V600E mutation prediction in intracranial metastasis on MRI radiomics has been shown to be highly accurate [56].

Future outlook

Expanding on its current clinical utility could include more reliable prognostic biomarkers and immunotherapy prediction by combining WSI and radiomic features. Furthermore, extracting features across multiple imaging modalities and cross-referencing them could address the current imaging and staging protocols.

Conclusion

The cost of treating in situ and invasive melanoma in Australia is >\$200 million per annum [57]. With monetary policy tightening, any effort to reduce the costs of diagnosing and managing cutaneous melanoma should

Table 3 Summary of the literature regarding artificial intelligence and radiomics

Reference	Study details	Patient demographics	Segmentation protocol imaging modality	Summary
Brendlin <i>et al.</i> , 2021 [20]	Retrospective study assessing additive value on baseline DECT to predict immunotherapy response	140 consecutive patients with Stage IV melanoma lesions receiving immunotherapy	Initial baseline DECT CT VOI segmentation analysed with eXamine (Siemens Healthineers)	Patients with higher minimum lesion brightness were more likely to be non-responders. Structural heterogeneity was a good prognostic feature.
Guerrisi <i>et al.</i> , 2021 [22]	Pilot study using contrast-enhanced CT prior to treatment and after first dose to assess features associated with survival outcomes	78 metastatic melanoma patients enrolled. 32 patients involved in final study of patients receiving PD-1 inhibitor, Nivolumab.	Contrast-enhanced MDCT (Phillips) with manual segmentation and feature extraction	Kurtosis and percentage change in entropy without filtration are best predictors of survival and potential radiological biomarkers. Sample size was a study limitation.
Kniep <i>et al.</i> , 2019 [27]	Feasibility of AI feature extraction and classification when compared to radiologist.	658 brain metastases in 189 patients (NSCLC, BC, malignant melanoma). Included 89 melanoma cases.	Brain MRI with semi-automatic segmentation. Random forest algorithm for multiclass classification tumours.	AI outperformed clinician by 17% for melanoma diagnosis. AI did not perform as well with BC and NSCLC. Only high-resolution images were used for comparison (higher feature extraction)
Meissner <i>et al.</i> , 2022 [56]	Retrospective study of MRIs with melanoma brain metastases from 2010–2020 to determine whether radiomics predict BRAF status	59 patients from two German University hospitals	T2-contrast enhanced MRI used and feed through segmentation and classification with feature extraction. BRAF status determined using DNA analysis	Very good AUC 0.92 results. Potential use for radiomics to determine genetic basis of melanoma brain metastasis without biopsy
Peisen <i>et al.</i> , 2022 [24]	Comparison of whether machine learning analysis of clinical data, with or without CT radiomics, would improve prediction of therapy response and survival	262 Stage IV melanoma patients treated with PD-1 or CTLA-4 checkpoint inhibitors	Whole body CT with all lesions segmented and fed through random forest model algorithm	Marginal and non-statistically significant benefit when using radiomics in conjunction with clinical parameters compared to clinical parameters alone
Trebeschi <i>et al.</i> , 2019 [25]	Identification of a radiomic biomarker to predict immunotherapy response.	1055 lesions (203 patients) from patients with metastatic NSCLC and melanoma	Whole-body CT	Radiomics biomarker was better for NSCLC compared to melanoma due to possible diversity of therapeutic backgrounds.

AI, artificial intelligence; AUC, area under the curve; BC, breast cancer; CT, computed tomography; DECT, dual energy CT; MDCT, multi detector CT; NSCLC, non-small cell lung cancer; VOI, volume of interest.

be explored. Early intervention is critical to saving lives, increasing productivity, and reducing the healthcare burden. While the cost of establishing an AI-based system is substantial, the use of AI/CNN as an adjunct could prove to be a valuable long-term investment.

Numerous articles have concluded that the dermatoscopic and WSI diagnostic accuracy and differentiation were superior or at least equal to clinicians. CNN can calculate data much faster than a clinician and is much more able to identify patterns. To date, publications have tested AI/ CNN accuracy retrospectively. These studies were performed mostly on established datasets which have not been externally validated and this is a limitation. Conclusively however, the clinician is inventive and instinctive and can think laterally. As such, there is evidence that, CNN and clinicians collaborating are superior to either alone, certainly with respect to dermatoscopy.

AI has the potential to enhance the diagnostic capabilities of junior dermatology trainees and primary care skin cancer clinicians as well as general practitioners that do not see skin cancer regularly. For more experienced clinicians, AI provides a cost-efficient second opinion on whether to offer excision biopsy. From a pathological and radiological perspective, CNN will potentially improve workflow efficiency, allowing clinicians to achieve more in a finite amount of time. Until the challenges of AI/

CNN are reliably met however, they can only remain a powerful adjunct to clinical decision-making.

Acknowledgements

L.G.A. is supported by funding from a Cancer Australia Priority-driven Collaborative Cancer Research Scheme Project (APPID_2019962) and a University of Queensland Amplify fellowship.

This research was carried out at the Translational Research Institute, Woolloongabba, QLD 4102, Australia. The Translational Research Institute is supported by a grant from the Australian Government. L.G.A. received support from Civic Solutions.

J.Y. performed the literature review and wrote the manuscript. L.G.A. supervised the project. C.R. and L.G.A. provided intellectual input and edited the manuscript.

Conflicts of interest

There are no conflicts of interest.

References

- Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, *et al.*; Collaborators. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer* 2019; **113**:47–54.
- Chen X, Lu Q, Chen C, Jiang G. Recent developments in dermoscopy for dermatology. *J Cosmet Dermatol* 2021; **20**:1611–1617.
- Shah KK, Lehman JS, Gibson LE, Lohse CM, Comfere NI, Wieland CN. Validation of diagnostic accuracy with whole-slide imaging compared

- with glass slide review in dermatopathology. *J Am Acad Dermatol* 2016; **75**:1229–1237.
- 4 Edge SB. *AJCC cancer staging manual 8th ed American Joint Committee on Cancer*. 8 ed. 2017, Springer.
 - 5 Mar VJ, Soyer HP, Button-Sloan A, Fishburn P, Gyorki DE, Hardy M, et al. Diagnosis and management of cutaneous melanoma. *Aust J Gen Pract* 2020; **49**:733–739.
 - 6 Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc* 2020; **92**:807–812.
 - 7 Kavlakoglu E. *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?* 2020. <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>. [Accessed 06 August 2022]
 - 8 Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med* 2018; **131**:129–133.
 - 9 Madhavan S. *Introduction to convolutional neural networks*. 2021. <https://developer.ibm.com/articles/introduction-to-convolutional-neural-networks/>. [Accessed 06 August 2022]
 - 10 Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**:115–118.
 - 11 Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al.; Reader study level-I and level-II Groups. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018; **29**:1836–1842.
 - 12 Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* 2019; **119**:11–17.
 - 13 Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* 2019; **20**:938–947.
 - 14 Fink C, Blum A, Buhl T, Mitteldorf C, Hofmann-Wellenhof R, Deinlein T, et al. Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *J Eur Acad Dermatol Venereol* 2020; **34**:1355–1361.
 - 15 Haenssle HA, Winkler JK, Fink C, Toberer F, Enk A, Stolz W, et al.; Reader study level-I and level-II Groups Christina Alt. Skin lesions of face and scalp - classification by a market-approved convolutional neural network in comparison with 64 dermatologists. *Eur J Cancer* 2021; **144**:192–199.
 - 16 Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, et al.; Collaborators. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer* 2019; **119**:57–65.
 - 17 Minagawa A, Koga H, Sano T, Matsunaga K, Teshima Y, Hamada A, et al. Dermoscopic diagnostic performance of Japanese dermatologists for skin tumors differs by patient origin: A deep learning convolutional neural network closes the gap. *J Dermatol* 2021; **48**:232–236.
 - 18 Winkler JK, Sies K, Fink C, Toberer F, Enk A, Abassi MS, et al. Collective human intelligence outperforms artificial intelligence in a skin lesion classification task. *J Dtsch Dermatol Ges* 2021; **19**:1178–1185.
 - 19 Barata C, Rotemberg V, Codella NCF, Tschandl P, Rinner C, Akay BN, et al. A reinforcement learning model for AI-based decision support in skin cancer. *Nat Med* 2023; **29**:1941–1946.
 - 20 Brendlin AS, Peisen F, Almansour H, Afat S, Eigentler T, Amaral T, et al. A Machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. *J ImmunoTher Cancer* 2021; **9**:e003261.
 - 21 Filippi L, Bianconi F, Schillaci O, Spanu A, Palumbo B. The role and potential of (18)F-FDG PET/CT in malignant melanoma: prognostication, monitoring response to targeted and immunotherapy, and radiomics. *Diagnostics (Basel)* 2022; **12**:929.
 - 22 Guerrisi A, Russillo M, Loi E, Ganeshan B, Ungania S, Desiderio F, et al. Exploring CT texture parameters as predictive and response imaging biomarkers of survival in patients with metastatic melanoma treated with pd-1 inhibitor nivolumab: a pilot study using a delta-radiomics approach. *Front Oncol* 2021; **11**:704607.
 - 23 Lopci E. Immunotherapy monitoring with immune checkpoint inhibitors based on [(18)F]FDG PET/CT in metastatic melanomas and lung cancer. *J Clin Med* 2021; **10**:5160.
 - 24 Peisen F, Hänsch A, Hering A, Brendlin AS, Afat S, Nikolaou K, et al. Combination of whole-body baseline ct radiomics and clinical parameters to predict response and survival in a stage-iv melanoma cohort undergoing immunotherapy. *Cancers (Basel)* 2022; **14**:2992.
 - 25 Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* 2019; **30**:998–1004.
 - 26 Zhang C, de A F Fonseca L, Shi Z, Zhu C, Dekker A, Bermejo I, et al. Systematic review of radiomic biomarkers for predicting immune checkpoint inhibitor treatment outcomes. *Methods* 2021; **188**:61–72.
 - 27 Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. *Radiology* 2019; **290**:479–487.
 - 28 Hasan MK, Dahal L, Samarakoon PN, Tushar FI, Marti R. DSNet: automatic dermoscopic skin lesion segmentation. *Comput Biol Med* 2020; **120**:103738.
 - 29 Bi L, Kim J, Ahn E, Kumar A, Fulham M, Feng D. Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Trans Biomed Eng* 2017; **64**:2065–2074.
 - 30 Ding J, Song J, Li J, Tang J, Guo F. Two-stage deep neural network via ensemble learning for melanoma classification. *Front Bioeng Biotechnol* 2021; **9**:758495.
 - 31 Yang X, Li H, Wang Li, Yeo SY, Su Y, Zeng Z. Skin lesion analysis by multi-target deep neural networks. *Annu Int Conf IEEE Eng Med Biol Soc* 2018; **2018**:1263–1266.
 - 32 ISIC 2018 Leaderboards. 2018. <https://challenge.isic-archive.com/leaderboards/2018/>. [Accessed 25 February 2023]
 - 33 Foahom Gouabou AC, Damoiseaux J-L, Monnier J, Iguernaissi R, Moudafi A, Merad D. Ensemble method of convolutional neural networks with directed acyclic graph using dermoscopic images: melanoma detection application. *Sensors (Basel)* 2021; **21**:3999.
 - 34 Harangi B. Skin lesion classification with ensembles of deep convolutional neural networks. *J Biomed Inform* 2018; **86**:25–32.
 - 35 Maron RC, Haggemüller S, von Kalle C, Utikal JS, Meier F, Gellrich FF, et al. Robustness of convolutional neural networks in recognition of pigmented skin lesions. *Eur J Cancer* 2021; **145**:81–91.
 - 36 Tschandl P. Artificial intelligence for melanoma diagnosis. *Ital J Dermatol Venerol* 2021; **156**:289–299.
 - 37 Tognetti L, Bonechi S, Andreini P, Bianchini M, Scarselli F, Cevenini G, et al. A new deep learning approach integrated with clinical data for the dermoscopic differentiation of early melanomas from atypical nevi. *J Dermatol Sci* 2021; **101**:115–122.
 - 38 Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, et al.; Collaborators. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer* 2019; **120**:114–121.
 - 39 Maron RC, Utikal JS, Hekler A, Hauschild A, Sattler E, Sondermann W, et al. Artificial intelligence and its effect on dermatologists' accuracy in dermoscopic melanoma image classification: web-based survey study. *J Med Internet Res* 2020; **22**:e18091.
 - 40 Wu W, Mehta S, Nofallah S, Knezevich S, May CJ, Chang OH, et al. Scale-aware transformers for diagnosing melanocytic lesions. *IEEE Access* 2021; **9**:163526–163541.
 - 41 De Logu F, Ugolini F, Maio V, Simi S, Cossu A, Massi D, et al.; Italian Association for Cancer Research (AIRC) Study Group. Recognition of cutaneous melanoma on digitized histopathological slides via artificial intelligence algorithm. *Front Oncol* 2020; **10**:1559.
 - 42 Zormpas-Petridis K, Noguera R, Ivankovic DK, Roxanis I, Jamin Y, Yuan Y. SuperHistopath: a deep learning pipeline for mapping tumor heterogeneity on low-resolution whole-slide digital histopathology images. *Front Oncol* 2020; **10**:586292.
 - 43 Schmitt M, Maron RC, Hekler A, Stenzinger A, Hauschild A, Weichenthal M, et al. Hidden variables in deep learning digital pathology and their potential to cause batch effects: prediction model study. *J Med Internet Res* 2021; **23**:e23436.
 - 44 Xie P, Zuo K, Liu J, Chen M, Zhao S, Kang W, et al. Interpretable diagnosis for whole-slide melanoma histology images using convolutional neural network. *J Healthc Eng* 2021; **2021**:8396438.
 - 45 Kim RH, Nomikou S, Coudray N, Jour G, Dawood Z, Hong R, et al. Deep learning and pathomics analyses reveal cell nuclei as important features for mutation prediction of braf-mutated melanomas. *J Invest Dermatol* 2022; **142**:1650–1658.e6.
 - 46 Klein S, Mauch C, Brinker K, Noh K-W, Knez S, Büttner R, et al. Tumor infiltrating lymphocyte clusters are associated with response to immune checkpoint inhibition in BRAF V600(E/K) mutated malignant melanomas. *Sci Rep* 2021; **11**:1834.
 - 47 Hohn J, Kriehoff-Henning E, Jutzi TB, von Kalle C, Utikal JS, Meier F, et al. Combining CNN-based histologic whole slide image analysis and patient data to improve skin cancer classification. *Eur J Cancer* 2021; **149**:94–101.

- 48 Li T, Xie P, Liu J, Chen M, Zhao S, Kang W, *et al.* Automated diagnosis and localization of melanoma from skin histopathology slides using deep learning: a multicenter study. *J Healthc Eng* 2021; **2021**:5972962.
- 49 Acs B, Ahmed FS, Gupta S, Wong PF, Gartrell RD, Sarin Pradhan J, *et al.* An open source automated tumor infiltrating lymphocyte algorithm for prognosis in melanoma. *Nat Commun* 2019; **10**:5440.
- 50 Kulkarni PM, Robinson EJ, Sarin Pradhan J, Gartrell-Corrado RD, Rohr BR, Trager MH, *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.
- 51 Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y, *et al.* Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. *Clin Cancer Res* 2021; **27**:131–140.
- 52 Aung TN, Shafi S, Wilmott JS, Nourmohammadi S, Vathiotis I, Gavrielatou N, *et al.* Objective assessment of tumor infiltrating lymphocytes as a prognostic marker in melanoma using machine learning algorithms. *EBioMedicine* 2022; **82**:104143.
- 53 Chatziioannou E, Roßner J, Aung TN, Rimm DL, Niessner H, Keim U, *et al.* Deep learning-based scoring of tumour-infiltrating lymphocytes is prognostic in primary melanoma and predictive to PD-1 checkpoint inhibition in melanoma metastases. *EBioMedicine* 2023; **93**:104644.
- 54 Moore MR, Friesner ID, Rizk EM, Fullerton BT, Mondal M, Trager MH, *et al.* Automated digital TIL analysis (ADTA) adds prognostic value to standard assessment of depth and ulceration in primary melanoma. *Sci Rep* 2021; **11**:2809.
- 55 Aoude LG, Wong BZY, Bonazzi VF, Brosda S, Walters SB, Koufariotis LT, *et al.* Radiomics biomarkers correlate with cd8 expression and predict immune signatures in melanoma patients. *Mol Cancer Res* 2021; **19**:950–956.
- 56 Meissner AK, Gutsche R, Galldiks N, Kocher M, Jünger ST, Eich ML, *et al.* Radiomics for the noninvasive prediction of the BRAF mutation status in patients with melanoma brain metastases. *Neuro Oncol* 2022; **24**:1331–1340.
- 57 Elliott TM, Whiteman DC, Olsen CM, Gordon LG. Estimated healthcare costs of melanoma in australia over 3 years post-diagnosis. *Appl Health Econ Health Policy* 2017; **15**:805–816.