European Journal of Cancer 127 (2020) 21–29

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Melanoma recognition by a deep learning convolutional neural network—Performance in different melanoma subtypes and localisations

Julia K. Winkler [a], Katharina Sies [a], Christine Fink [a], Ferdinand Toberer [a], Alexander Enk [a], Teresa Deinlein [b], Rainer Hofmann-Wellenhof [b], Luc Thomas [c], Aimilios Lallas [d], Andreas Blum [e], Wilhelm Stolz [f], Mohamed S. Abassi [g], Tobias Fuchs [h], Albert Rosenberger [i], Holger A. Haenssle [a,*]

[a] *Department of Dermatology, University of Heidelberg, Heidelberg, Germany*
[b] *Department of Dermatology and Venerology, Medical University of Graz, Graz, Austria*
[c] *Hospices Civils de Lyon, Department of Dermatology, Lyon Sud University Hospital, Pierre Bénite, France*
[d] *First Department of Dermatology, Aristotle University, Thessaloniki, Greece*
[e] *Public, Private and Teaching Practice, Konstanz, Germany*
[f] *Department of Dermatology, Allergology and Environmental Medicine II, Hospital Thalkirchner Street, Munich, Germany*
[g] *Faculty of Computer Science and Mathematics, University of Passau, Passau, Germany*
[h] *Department of Research and Development, FotoFinder Systems GmbH, Bad Birnbach, Germany*
[i] *Institute of Genetic Epidemiology at the Center of Statistics, University of Goettingen, Goettingen, Germany*

Received 22 August 2019; received in revised form 21 October 2019; accepted 16 November 2019

KEYWORDS
Melanoma;
Nevi;
Dermoscopy;
Deep learning;
Convolutional neural network

Abstract ***Background:*** Deep learning convolutional neural networks (CNNs) show great potential for melanoma diagnosis. Melanoma thickness at diagnosis among others depends on melanoma localisation and subtype (e.g. advanced thickness in acrolentiginous or nodular melanomas). The question whether CNN may counterbalance physicians' diagnostic difficulties in these melanomas has not been addressed. We aimed to investigate the diagnostic performance of a CNN with approval for the European market across different melanoma localisations and subtypes.

Methods: The current market version of a CNN (Moleanalyzer-Pro®, FotoFinder Systems GmbH, Bad Birnbach, Germany) was used for classifications (malignant/benign) in six dermoscopic image sets. Each set included 30 melanomas and 100 benign lesions of related

* *Corresponding author*: Department of Dermatology, University of Heidelberg, Im Neuenheimer Feld 440, 69120, Heidelberg, Germany. Fax: +49 6221 56 8510.
E-mail address: holger.haenssle@med.uni-heidelberg.de (H.A. Haenssle).

https://doi.org/10.1016/j.ejca.2019.11.020

localisations and morphology (set-SSM: superficial spreading melanomas and macular nevi; set-LMM: lentigo maligna melanomas and facial solar lentigines/seborrhoeic keratoses/nevi; set-NM: nodular melanomas and papillomatous/dermal/blue nevi; set-Mucosa: mucosal melanomas and mucosal melanoses/macules/nevi; set-AM_{skin}: acrolentiginous melanomas and acral (congenital) nevi; set-AM_{nail}: subungual melanomas and subungual (congenital) nevi/lentigines/ethnical type pigmentations).

Results: The CNN showed a high-level performance in set-SSM, set-NM and set-LMM (sensitivities >93.3%, specificities >65%, receiver operating characteristics-area under the curve [ROC-AUC] >0.926). In set-AM_{skin}, the sensitivity was lower (83.3%) at a high specificity (91.0%) and ROC-AUC (0.928). A limited performance was found in set-mucosa (sensitivity 93.3%, specificity 38.0%, ROC-AUC 0.754) and set-AM_{nail} (sensitivity 53.3%, specificity 68.0%, ROC-AUC 0.621).

Conclusions: The CNN may help to partly counterbalance reduced human accuracies. However, physicians need to be aware of the CNN's limited diagnostic performance in mucosal and subungual lesions. Improvements may be expected from additional training images of mucosal and subungual sites.

1. Introduction

Malignant melanoma accounts for the highest mortality rate among all skin cancers, and incidence rates are still increasing in many countries around the world [1]. In advanced disease stages, the prognosis is limited, making early diagnosis vital [2]. The reasons for a delayed melanoma diagnosis may be separated into patient-related and physician-related factors [3]. While an improvement of patient-related factors may be achieved by public campaigns increasing melanoma awareness, physician-related factors are mostly modelled by the individual diagnostic expertise [4]. Compared with the naked eye examination, dermoscopy may increase the diagnostic accuracy of physicians [5–7], and a number of validated algorithms have been designed to improve early diagnosis [8–10]. However, while many dermoscopic features and algorithms target the diagnosis of the most frequent melanoma subtypes, namely superficial spreading melanoma (SSM) and lentigo maligna melanoma (LMM), the morphological traits of melanomas of other specific anatomic sites (e.g. mucosa, acral skin and nail unit) are rarely covered. In addition, nodular melanomas may show a symmetrical growth pattern with only one colour, thus evading detection [11]. The aforementioned details may explain why dramatic improvements have been observed in the early detection of SMM (nearly 50% decrease in median thickness over two decades), while the thickness of nodular or acral melanomas at diagnosis remained largely unchanged [12,13].

The increasing application of artificial intelligence and machine learning in areas of healthcare and medicine has attracted a great deal of research interest in recent decades [14]. Recently, deep learning convolutional neural networks (CNNs) have entered the arena of image-based melanoma diagnosis and proven a dermatologist-level diagnostic performance [15,16]. Diagnostic support for physicians with regard to difficult-to-diagnose melanomas of special localisations and rare subtypes would be desirable. Yet, the diagnostic performance of a CNN across such a variety of melanomas has not been evaluated.

We provide data by comparing the sensitivity, specificity, and receiver operating characteristics-area under the curve (ROC-AUC) of a CNN approved as a medical device in Europe across six image sets including melanomas of different localizations and subtypes admixed with benign lesions matched for localisation and morphology.

2. Material and methods

The study was approved by the local ethics committee and performed in accordance with the Declaration of Helsinki principles. In this study, we used the current version of a CNN that recently gained regulatory approval as a medical device for the European market (Moleanalyzer-Pro®; FotoFinder Systems GmbH, Bad Birnbach, Germany). Details on the CNN's architecture and training procedures are given in the supplement methods.

For the study, six dermoscopic image sets were compiled (set-SSM, set-LMM, set-NM, set-Mucosa, set-AM_{skin} and set-AM_{nail}), each containing 30 melanomas (Table 1) and 100 benign lesions (Table 2) that were matched for localisation and morphology. To this end, J.K.W. and H.A.H. manually screened image libraries

Table 1
Characteristics of patients and melanomas within image sets.

Characteristics	set-SSM (Melanomas)		set-LMM (Melanomas)		set-NM (Melanomas)		set-Mucosa (Melanomas)		set-AM_{skin} (Melanomas)		set-AM_{nail} (Melanomas)	
Age (mean, ± SD)	58.6	±14.8	67.2	±13.3	61.0	±15.9	63.0	±17.9	69.6	±13.7	49.2	±15.6
Sex (n, %)												
Female	16	53.3%	11	36.7%	12	40%	21	70%	22	73.3%	14	46.7%
Male	14	46.6%	19	63.3%	18	60%	9	30%	8	26.7%	16	53.3%
Localisation (n, %)					–	–					–	–
Scalp/Face	–	–	30	100%	3	10%	–	–	–	–	–	–
Trunk/Extremities	30	100%	–	–	25	83.3%	–	–	–	–	–	–
Palmoplantar skin	–	–	–	–	2	6.7%	–	–	30	100%	–	–
Mucosa	–	–	–	–	–	–	30	100%	–	–	–	–
Nail unit	–	–	–	–	–	–	–	–	–	–	30	100%
Invasiveness												
In situ (n, %)	5	16.7%	5	16.7%	0	0%	5	16.7%	5	16.7%	5	16.7%
Breslow thickness (mean, ± SD)	0.7	0.3	0.7	0.8	4.0	2.8	4.1	3.3	1.5	1.7	1.2	1.8

SSM, superficial spreading melanoma; LMM, lentigo maligna melanoma (including lentigo maligna, i.e. *in situ* lesions); Mucosa, mucosal melanoma; NM, nodular melanoma, AM_{skin}, acral melanoma of palmoplantar skin, $AM_{nail,}$ acral melanoma of the nail unit; SD, standard deviation.

(Departments of Dermatology, Universities of Heidelberg, Munich and Lyon) and corresponding metadata comprising more than 50.000 dermoscopic cases. Requirements for matching by localisation and morphology were as follows: set-SSM: superficial spreading melanomas admixed with flat, macular, benign melanocytic lesions of trunk and extremities including junctional, compound, Reed and Spitz nevi; set-LMM: lentigo maligna melanomas admixed with flat, macular solar lentigines, seborrheic keratosis and nevi located on face and scalp; set-NM: nodular melanomas admixed with raised, nodular, papillomatous, dermal and blue nevi; set-Mucosa: mucosal melanomas admixed with flat, mucosal melanoses, melanotic macules and nevi; set-AM_{skin}: acrolentiginous melanomas admixed with acral nevi (including congenital acral nevi); set-AM_{nail}: subungual melanomas admixed with subungual nevi (including congenital subungual nevi), lentigines as well as ethnical type pigmentations. The ground truth in all melanoma cases ($n = 180$) was based on the histopathological diagnosis, while the ground truth in benign lesions ($n = 600$) was either based on histopathology ($n = 363$, 60.5%), on an unremarkable follow-up by sequential digital dermoscopy over at least 2 years ($n = 210$, 35.0%) or on expert opinion ($n = 27$, 4.5%). Various camera/dermoscope combinations were used for image acquisition. No overlap between data sets for training, validation and testing was allowed. All images complied with quality standards for skin imaging [17].

2.1. Statistical analysis

The primary outcome measures were the CNN's sensitivity, specificity and AUC of ROC for the diagnostic classification of lesions in the six test-sets. The CNN put out a 'malignancy score' ranging from 0 to 1, and an a priori cut-off of >0.5 set by large validation data was applied for the dichotomous classification of melanoma versus benign lesions. Descriptive statistics as frequency, mean, range and standard deviation were used. The ROC AUCs of the six test-sets were compared pairwise by controlling for multiple testing according to Tukey's method. All analyses were carried out using SPSS, version 25 (IBM, SPSS, Chicago, IL).

3. Results

3.1. Characteristics of imaged lesions and patients

Dermoscopic images of melanomas ($n = 180$) for the six different sets were randomly selected from image libraries and categorised by localisation or subtypes (Table 1). Resulting image sets (set-SSM, set-LMM, set-NM, set-Musoca, set-AM_{skin}, set-AM_{nail}) each contained 30 melanomas (25 invasive, 5 in situ). The mean (±SD) age of melanoma patients ($n = 180$) at diagnosis was youngest in patients with AM_{nail} (49.2, ±15.6 years) and oldest in patients with AM_{skin} (69.6, ±13.7 years). Invasiveness (mean Breslow thickness, ± SD) was most advanced in NM (4.0, ±2.8 mm) and MM (4.0, ±3.3 mm), while LMM (0.7, ±0.8 mm) and SSM (0.7, ±0.3 mm) were least progressed (Table 1). The ground truth in all melanoma cases was based on the histopathological diagnosis.

For each image set, 100 benign lesions of comparable localisation and morphology were included (e.g. set-SSM comprised SSM admixed with benign, flat, macular, junctional or compound nevi of the trunk and extremities). Benign lesions were either diagnosed by histopathology, by unremarkable follow-up examinations over at least 2 years or by expert opinion (Table 2).

Table 2
Characteristics of patients and benign lesions within image sets.

Characteristics	set-SSM (benign controls)		set-LMM (benign controls)		set-NM (benign controls)		set-Mucosa (benign controls)		set-AM_{skin} (benign controls)		set-AM_{nail} (benign controls)	
Age (mean, ± SD)	42.8	±11.0	54.6	±22.4	44.9	±14.4	40.7	±17.5	39.9	±19.4	29.4	±19.6
Sex (*n*, %)												
Female	34	34%	44	44%	60	60%	67	67%	57	57%	52	52%
Male	66	66%	56	56%	40	40%	33	33%	43	43%	48	48%
Localisation (*n*, %)					–	–					–	–
Scalp/Face	–	–	100	100%	–	–	–	–	–	–	–	–
Trunk/Extremities	100	100%	–	–	100	100%	–	–	–	–	–	–
Palmoplantar skin	–	–	–	–	–	–	–	–	100	100%	–	–
Mucosa	–	–	–	–	–	–	100	100%	–	–	–	–
Nail unit	–	–	–	–	–	–	–	–	–	–	100	100%
Ground truth (n, %)												
Histopathology	100	100%	87	87%	12	12%	83	83%	64	64%	17	17%
Follow-up	–	–	–	–	88	88%	17	17%	35	35%	70	70%
Expert opinion	–	–	13	13%	–	–	–	–	1	1%	13	13%

set-SSM, superficial spreading melanomas and macular nevi; set-LMM, lentigo maligna melanomas and facial solar lentigines/seborrheic keratoses/nevi; set-NM, nodular melanomas and papillomatous/dermal/blue nevi; set-Mucosa, mucosal melanomas and mucosal melanoses/macules/nevi; set-AM_{skin}, acrolentiginous melanomas and acral (congenital) nevi; set-AM_{nail}, subungual melanomas and subungual (congenital) nevi/lentigines/ethnical type pigmentations.
Ground truth 'Follow-up': uneventful dermoscopic follow-up over at least 2 years.

Representative images of all image sets are depicted in Fig. 1.

3.2. Diagnostic performance of the CNN

Boxplots in Fig. 2 show the distribution of the CNN's 'probability of malignancy' scores that resulted from the softmax output layer and ranged from 0 to 1, with higher scores indicating a higher probability of melanoma. The average (±SD) scores of the classifier for melanomas within each set were 0.98 (±0.06) in set-SSM, 0.98 (±0.06) in set-LMM, 0.90 (±0.23) in set-NM, 0.94 (±0.19) in set-Mucosa, 0.80 (±0.36) in set-AM_{skin}, and 0.51 (±0.43) in set-AM_{nail} (Fig. 2). Benign lesions were matched to melanomas of each set for localisation and morphology. Malignancy scores were 0.07 (±0.23) for macular nevi in set-SSM, 0.37 (±0.43) for facial solar lentigines, seborrhoeic keratoses and nevi in set-LMM, 0.04 (±0.18) for nodular, papillomatous, dermal, blue and combined nevi in set-NM, 0.59 (±0.44) for mucosal melanoses, lentigines and nevi in set-Mucosa, 0.10 (±0.26) for acral nevi (including congenital acral nevi) in set-AM_{skin} and 0.33 (±0.40) for subungual nevi (including congenital subungual nevi) and ethnical type pigmentations in set-AM_{nail}. Higher standard deviations reflect an increased range of scattered malignancy scores and were observed in subungual melanomas of set-AM_{nail} and in benign lesions of set-LMM, set-Mucosa and set-AM_{nail}. Yet, for all melanoma localisations and subtypes, malignancy scores were significantly higher when compared with benign lesions (Mann–Whitney U test, $p < 0.05$).

When operating the CNN at the a priori cut-off of >0.5 for making the diagnosis of melanoma, the calculated sensitivities (95% confidence interval [CI]) and specificities (95% CI) were (i) 100% (88.7%–100%) and 94% (87.5%–97.2%) in set-SSM; (ii) 100% (88.7%–100%) and 65.0% (55.3%–73.6%) in set-LMM; (iii) 93.3% (78.7%–98.2%) and 96.0% (90.2%–98.4%) in set-NM; 83.3% (66.4%–92.7%) and 91% (83.8%–95.2%) in set-AM_{skin}; (iv) 53.3% (36.1%–69.8%) and 68.0% (58.3%–76.3%) in set-AM_{nail} and 93.3% (78.7%–98.2%) and 38.0% (29.1%–47.8%) in set-Mucosa.

The resulting diagnostic accuracy (proportion of true positive and true negative in all evaluated cases) was highest in set-SSM (95.4%) and set-NM (95.4%), followed by set-AM_{skin} (89.2%), set-LMM (73.1%), set-AM_{nail} (64.4%) and lowest in set-Mucosa (50.8%).

For a direct head-to-head comparison of the CNN's diagnostic performance across the image-sets, we plotted the corresponding ROC curves into one graph (Fig. 3). ROC-AUC [95% CI] in descending order was 0.989 [0.976–1.000] in set-SMM, 0.982 [0.964–1.000] in set-NM, 0.928 [0.867–0.990] in AM_{skin}-MM, 0.926 [0.881–0.970] in set-LMM, 0.754 [0.668–0.841] in set-Mucosa and 0.621 [0.506–0.735] in set-AM_{nail}. In a pairwise comparison of the CNN's AUCs as attained in the six image sets, we were able to identify two clusters (cluster A: set-SSM, set-LMM, set-NM, set-AM_{skin}; cluster B: set-Mucosa, set-AM_{nail}) of test-sets showing non-significant AUC differences within each cluster (Table 3).

3.3. Subgroup analysis of CNN's performance in invasive melanomas only

The differentiation of benign lesions from invasive melanomas may be considered less challenging than

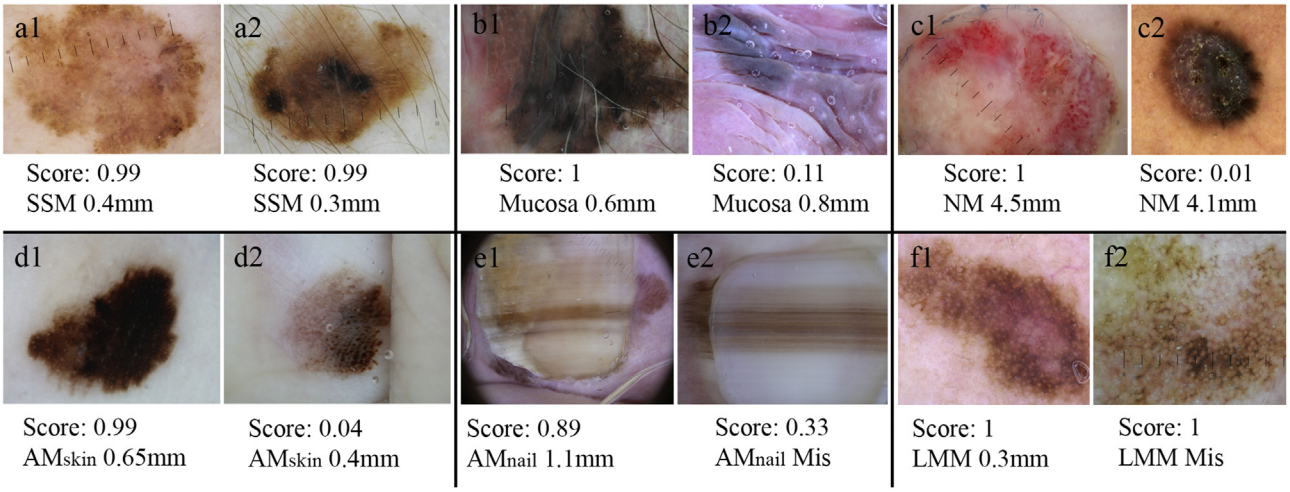

Fig. 1. The CNN's melanoma probability score ('score') of representative pairs of melanoma images from set-SSM (a1, a2), set-Mucosa (b1, b2), set-NM (c1, c2), set-AM_{skin} (d1, d2) and set-AM_{nail} (e1, e2) and set-LMM (f1, f2) with true-positive (left) versus false-negative (right) classifications are depicted. In set-SSM (a1, a2) and set-LMM (f1, f2) both images represent correctly classified lesions, as no false-negative cases were observed in the study. Breslow thickness is indicated in millimetre (mm). Mis: Melanoma in situ. ROC, receiver operating characteristics; AUC, area under the curve; CNN, convolutional neural network; set-SSM, superficial spreading melanomas and macular nevi; set-LMM, lentigo maligna melanomas and facial solar lentigines/seborrheic keratoses/nevi; set-NM, nodular melanomas and papillomatous/dermal/blue nevi; set-Mucosa, mucosal melanomas and mucosal melanoses/macules/nevi; set-AM_{skin}, acrolentiginous melanomas and acral (congenital) nevi; set-AM $_{nail,}$ subungual melanomas and subungual (congenital) nevi/lentigines/ethnical type pigmentations.

their differentiation from in situ melanomas. To rule out potential bias by situ melanomas included in the image sets (none included in set-NM), we additionally calculated the sensitivities, specificities, diagnostic accuracy, and ROC data after excluding in situ melanomas. In this setting, diagnostic accuracy was found almost unchanged in comparison to the original image sets. Best results were attained in set-SSM (95.2%), followed by set-NM (94.5%), set-AM_{skin} (89.6%), set-LMM (72%), set-AM_{nail} (66.4%), and lowest in set-Mucosa (48.8%). Similarly, the CNN's ROC-AUC (95% CI) for the six image sets of invasive melanomas and benign lesions was largely unchanged at 0.990 (0.979–1) in set-SSM, 0.982 (0.964–1) in set-NM, 0.944 (0.906–0.982) in set-LMM, 0.926 (0.854–0.998) in set-AM_{skin}, 0.737 (0.643–0.831) in set-MM and 0.668 (0.552–0.784) in set-AM_{nail}.

4. Discussion

Recent improvements in the early detection of melanoma vary with melanoma histotypes [13]. While time trends in melanoma thickness revealed a strong decrease in thickness for SSM and LMM, a constant high thickness was observed for NM and melanomas of palmoplantar skin and the nail unit [18]. Similarly, mucosal melanomas are more often diagnosed at advanced stages with a significantly worse prognosis [19]. Reports of deep learning CNN's with an expert-level performance in the diagnosis of skin cancer [15,16] have fuelled expectations that physician-related delays due to false-negative diagnoses may be avoided in the near future. In the present study, we used the current versions of a deep learning CNN approved as a medical device for the European market and systematically assessed its diagnostic performance in image sets of melanomas admixed with benign lesions from different localisations and subtypes.

As expected by us, the tested CNN showed a high-class diagnostic performance in set-SSM, which included the melanoma histotype with the highest prevalence and, therefore, the highest availability of validated training images. LMM represents the second most frequent melanoma histotype making up 5–15% of melanoma cases in population-based studies [20]. Besides the availability of sufficient amounts of validated LMM training images, LMM shows many overlapping morphological traits to SMM, so that a high level of 'transfer learning' during training of both histotypes may be assumed [21]. These reasons may contribute to the CNN's favourable diagnostic performance in set-LMM.

Surprisingly, diagnostic performance was also high-class in set-NM, accounting for approximately 5% of all melanomas [20]. Of note, our set of NM included amelanotic and hypomelanotic cases mostly recognisable (from a human perspective) by the presence of atypical vascular patterns.

The CNN's sensitivity for melanomas of acral skin was lower (84%) than in the aforementioned three image sets (sensitivities between 93.3% and 100%). However, the high specificity of 91% contributed to a reasonably high

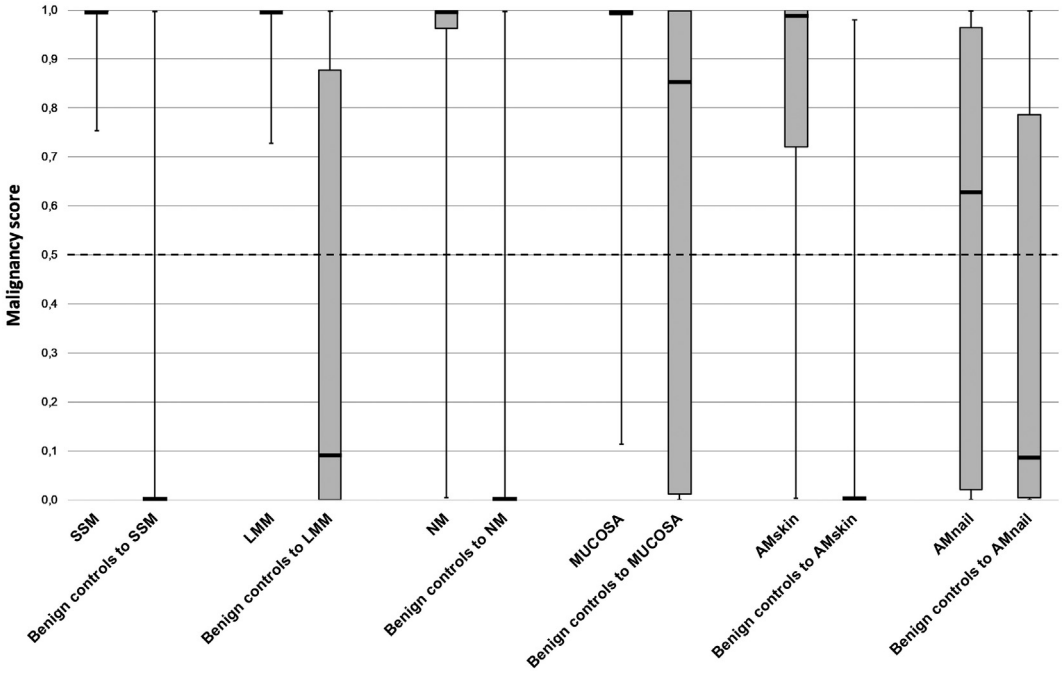

Fig. 2. Boxplots present the distribution of the CNN's malignancy scores (range 0–1) for the image sets (set-SSM, set-LMM, set-NM, set-Mucosa, set-AM_{skin}, set-AM_{nail}) showing malignant and benign lesions separately. Scores closer to 1 indicated a higher probability of malignancy. The upper and lower bounds of boxes indicate the 25th and 75th percentiles while the median is indicated by the thick line intersecting the upper and lower box. Whiskers indicate the full range of probability scores. Statistical analyses revealed significantly different malignancy scores in melanomas versus benign controls in all image sets ($p < 0.05$). The dotted line indicates the CNN's a priori threshold for classifying a lesion as being malignant (malignancy score >0.5). CNN, convolutional neural network; set-SSM, superficial spreading melanomas and macular nevi; set-LMM, lentigo maligna melanomas and facial solar lentigines/seborrheic keratoses/nevi; set-NM, nodular melanomas and papillomatous/dermal/blue nevi; set-Mucosa, mucosal melanomas and mucosal melanoses/macules/nevi; set-AM_{skin}, acrolentiginous melanomas and acral (congenital) nevi; set-AM $_{nail,}$ subungual melanomas and subungual (congenital) nevi/lentigines/ethnical type pigmentations.

accuracy of 89.6% and ROC-AUC of 0.928. As a result, the CNN's accuracy and ROC-AUC in set-AM_{skin} ranked third place directly after set-SSM (first place) and set-NM (second place). The measured diagnostic accuracy of the CNN in AM_{skin} was slightly higher than results attained by a previous report (accuracy ranging from 80.2% to 83.5%) [22]. However, it needs to be acknowledged that results may not be directly compared across different image sets and neural networks.

Of note, the CNN's AUCs of set-SSM, set-LMM, set-NM and set-AM_{skin} showed non-significant differences in a pairwise statistical comparison (forming cluster A). In contrast to the aforementioned results, we found a limited diagnostic performance of the CNN in set-Mucosa and set-AM_{nail} (forming cluster B).

Mucosal melanomas are rare and make up only 1.3% of melanoma case [19]. Besides their low frequency, mucosal melanomas are mostly localised in the oral cavity, genital area or perianal area, which often discourages dermoscopic imaging because of technical difficulties. As a result, images of mucosal melanomas are among the least frequent melanoma histotypes found in validated training images.

Melanomas of the nail unit are also rare [20]. Dermoscopic patterns of subungual lesion are unique and markedly distinct from skin lesions of any other localisation (e.g. parallel band of pigmentation in longitudinal melanonychia striata). This morphological difference is caused by the underlying anatomy of the nail unit, and conceivably, 'transfer learning' by the CNN will not help to compensate the lack of validated training images, thus leading to a limited performance.

To exclude a relevant impact of in situ melanomas, each melanoma set included a fixed number of five in-situ lesions, which were excluded for an additional subgroup analysis comparing only invasive melanomas to benign lesions. These analyses showed comparable results with regard to diagnostic accuracies and ROC-AUCs to those attained in the original overall image sets.

Our study reveals some limitations. First, dermoscopic images of test-sets were randomly selected from local image libraries of different institutions, which does not guarantee a representative sample as found in population-based studies. Second, the quality of images in this study was high. In a clinical routine setting, a lower quality of images, non-compliant with imaging

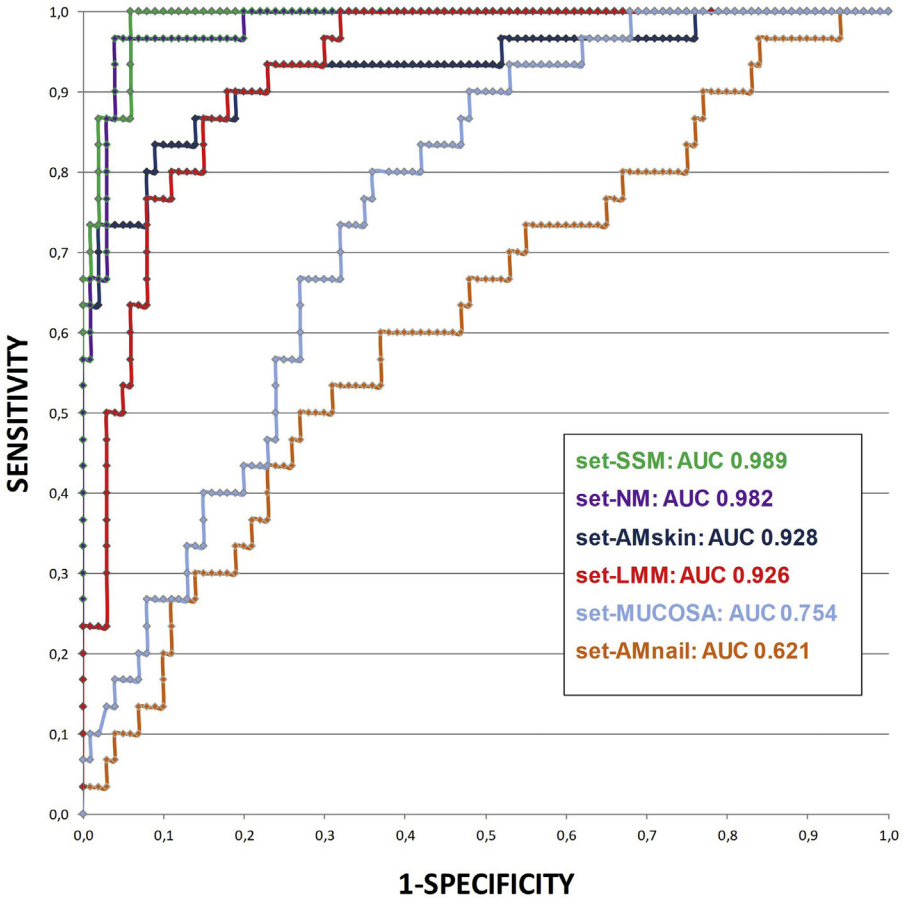

Fig. 3. Head-to-head comparison of ROC curves across the different test-sets. The CNN's ROC AUC for each individual test-set is indicated. ROC, receiver operating characteristics; AUC, area under the curve; CNN, convolutional neural network; set-SSM, superficial spreading melanomas and macular nevi; set-LMM, lentigo maligna melanomas and facial solar lentigines/seborrheic keratoses/nevi; set-NM, nodular melanomas and papillomatous/dermal/blue nevi; set-Mucosa, mucosal melanomas and mucosal melanoses/macules/nevi; set-AM_{skin}, acrolentiginous melanomas and acral (congenital) nevi; set-AM_{nail}, subungual melanomas and subungual (congenital) nevi/lentigines/ethnical type pigmentations.

Table 3
Statistical comparison of the CNN's AUCs in image sets.

		Pair-wise difference of AUCs					
Clusters	Test-sets	SSM	LMM	NM	AMskin	Mucosa	AMnail
A	**SSM**	–	0.7893	1.0000	0.7930	**<.0001**[a]	**<.0001**[a]
A	**LMM**	0.7893	–	0.8545	1.0000	**0.0061**[a]	**<.0001**[a]
A	**NM**	1.0000	0.8545	–	0.8576	**<.0001**[a]	**<.0001**[a]
A	**AMskin**	0.7930	1.0000	0.8576	–	**0.0060**[a]	**<.0001**[a]
B	**Mucosa**	**<.0001**[a]	**0.0061**[a]	**<.0001**[a]	**0.0060**[a]	–	0.0690
B	**AMnail**	**<.0001**[a]	**<.0001**[a]	**<.0001**[a]	**<.0001**[a]	0.0690	–

AUC, area under the curve; SSM, superficial spreading melanoma; LMM, lentigo maligna melanoma (including lentigo maligna, i.e. *in situ* lesions); Mucosa, mucosal melanoma; NM, nodular melanoma; AM_{skin}, acral melanoma of palmoplantar skin; AM_{nail}, acral melanoma of the nail unit; SD, standard deviation.
The AUC's of the different image sets have been compared pairwise by controlling for multiple testing according to Tukey's method. There is not enough evidence to distinguish the CNN's AUCs of set-SSM, set-LMM, set-NM and set-AM_{skin} among each other, but the diagnostic performance of the CNN in these sets (cluster A) is significantly better than for set-Mucosa and set-AM_{nail} (cluster B).
[a] The CNN's AUCs showed a significant difference by pairwise comparison ($p < 0.05$).

standards [17], may be used for analysis by a CNN. This may potentially result in a reduced diagnostic performance. Third, most images were derived from fair skinned patients, so that no conclusion on lesions of individuals of other ethnic backgrounds may be drawn. Finally, the data of our study are of retrospective nature. We strongly agree that there is a high need for performing prospective studies to investigate the potential benefits of CNNs when applied in clinical routine.

5. Conclusions

In conclusion, the tested CNN revealed a high-class diagnostic performance in set-SSM, set-NM, set-AM_{skin} and set-LMM (all ROC-AUC >0.9). These results make it reasonable to assume that the CNN may help to counterbalance low human accuracies inherent to nodular melanomas and melanomas of acral skin. In contrast, we would currently like to discourage the CNN's application in mucosal and subungual lesions. The expected expansion of dermoscopic image archives and new software approaches warrant further improvements.

Ethics

Reviewed and approved by the ethic committee of the medical faculty of the University of Heidelberg (Approval number S-629/2017).

Funding sources

This research project received funding from a public, non-for-profit domain, namely the Skin Cancer Council Germany (Nationale Versorgungskonferenz Hautkrebs (NVKH) e.V.), www.nvkh.de/.

Grant/Award number

AD-LEARN DERMOSCOPY (HF01-Z02-P05).

Conflict of interest statement

FotoFinder Systems GmbH had no role in the study design or interpretation of the data. T Fuchs is an employed software developer at the research and development department of FotoFinder Systems GmbH and was responsible for technical support and for writing parts of the supplement method section covering details on the CNN architecture and training. HA Haenssle received honoraria and/or travel expenses from companies involved in the development of devices for skin cancer screening: Scibase AB, FotoFinder Systems GmbH, Heine Optotechnik GmbH, Magnosco GmbH. A Blum received honoraria and/or travel expenses from companies involved in the development of devices for skin cancer screening: Scibase AB, FotoFinder Systems GmbH, Heine Optotechnik GmbH. All other authors indicated no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2019.11.020.

References

[1] Arnold M, Holterhues C, Hollestein L, Coebergh J, Nijsten T, Pukkala E, et al. Trends in incidence and predictions of cutaneous melanoma across Europe up to 2015. J Eur Acad Dermatol Venereol 2014;28:1170–8.
[2] Geller AC, Swetter SM, Weinstock MA. Focus on early detection to reduce melanoma deaths. J Investig Dermatol 2015;135:947–9.
[3] Krige JE, Isaacs S, Hudson DA, King HS, Strover RM, Johnson CA. Delay in the diagnosis of cutaneous malignant melanoma. A prospective study in 250 patients. Cancer 1991;68:2064–8.
[4] Richard MA, Grob JJ, Avril MF, Delaunay M, Gouvernet J, Wolkenstein P, et al. Delays in diagnosis and melanoma prognosis (II): the role of doctors. Int J Cancer 2000;89:280–5.
[5] Bafounta M-L, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma?: Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Arch Dermatol 2001;137:1343–50.
[6] Vestergaard M, Macaskill P, Holt P, Menzies S. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. Br J Dermatol 2008;159:669–76.
[7] Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol 2002;3:159–65.
[8] Stolz W. ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. Eur J Dermatol 1994;4:521–7.
[9] Menzies SW, Ingvar C, Crotty KA, McCarthy WH. Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. Arch Dermatol 1996;132:1178–82.
[10] Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch Dermatol 1998;134:1563–70.
[11] Pizzichetta M, Kittler H, Stanganelli I, Bono R, Cavicchini S, De Giorgi V, et al. Pigmented nodular melanoma: the predictive value of dermoscopic features using multivariate analysis. Br J Dermatol 2015;173:106–14.
[12] Warycha MA, Christos PJ, Mazumdar M, Darvishian F, Shapiro RL, Berman RS, et al. Changes in the presentation of nodular and superficial spreading melanomas over 35 years. Cancer 2008;113:3341–8.
[13] Baumert J, Schmidt M, Giehl KA, Volkenandt M, Plewig G, Wendtner C, et al. Time trends in tumour thickness vary in subgroups: analysis of 6475 patients by age, tumour site and melanoma subtype. Melanoma Res 2009;19:24–30.
[14] Tran BX, Vu GT, Ha GH, Vuong QH, Ho MT, Vuong TT, et al. Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. J Clin Med 2019;8.
[15] 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;542:115–8.

[16] Haenssle H, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018;29:1836–42.
[17] Finnane A, Dallest K, Janda M, Soyer HP. Teledermatology for the diagnosis and management of skin cancer: a systematic review. JAMA Dermatol 2017;153:319–27.
[18] Geller AC, Elwood M, Swetter SM, Brooks DR, Aitken J, Youl PH, et al. Factors related to the presentation of thin and thick nodular melanoma from a population-based cancer registry in Queensland Australia. Cancer 2009;115:1318–27.
[19] Lerner BA, Stewart LA, Horowitz DP, Carvajal RD. Mucosal melanoma: new insights and therapeutic options for a unique and aggressive disease. Oncology 2017;31:e23–32.
[20] Markovic SN, Erickson LA, Rao RD, McWilliams RR, Kottschade LA, Creagan ET, et al. Malignant melanoma in the 21st century, part 1: epidemiology, risk factors, screening, prevention, and diagnosis. In: Mayo clinic proceedings: Elsevier; 2007. p. 364–80.
[21] Hoo-Chang S, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35:1285.
[22] Yu C, Yang S, Kim W, Jung J, Chung K-Y, Lee SW, et al. Acral melanoma detection using a convolutional neural network for dermoscopy images. PLoS One 2018;13:e0193321.