AI in Dermatology Summit 2026

External Validation and Characterization of a Convolutional Neural Network in a Large High-Risk Melanoma Cohort

Laura Serra-García, MD[1,2,3], Nuria Ferrera, MD[2], Clément Lenoir, MD[2], Ana Claudia Rivas, MD[2], Adriana García-Herrera, MD, PhD [3,4], Llúcia Alòs, MD, PhD[3,4], Natalia Castrejón, MD, PhD[3,4], Brian M D'Alessandro, PhD[5], Míriam Potrony, MD, PhD[2,6,7], Cèlia Bàdenas, MD, PhD[2,6,7], Paula Aguilera, MD, PhD[1,2,3,7], Daniel Rizo-Potau, MD[1,3], Natalia Espinosa, MD[2], Cristina Carrera, MD, PhD[1,2,3,7], Sebastian Podlipnik, MD, PhD[1,2,3], Susana Puig, MD, PhD[1,2,3,7], Josep Malvehy, MD, PhD[1,2,3,7]

[1] Dermatology Department, Hospital Clínic de Barcelona, Barcelona, Spain; [2] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain; [3] University of Barcelona, Hospital Clínic de Barcelona, Barcelona, Spain; [4] Pathology Department, Hospital Clínic de Barcelona, Barcelona, Spain; [5] Canfield Scientific Inc., Parsippany, New Jersey, USA; [6] Biochemistry and Molecular Genetics Department, Hospital Clínic de Barcelona, Barcelona, Spain.; [7] Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain

Introduction

Convolutional neural networks (CNNs) have demonstrated high diagnostic accuracy for skin cancer detection in experimental settings. Evidence on the independent performance of artificial intelligence (AI)-based tools integrated into digital dermoscopy in real-world high-risk cohorts remains limited. We externally validated a convolutional neural network (CNN, DEXI® v2.1) integrated into dermoscopy images of patients at high risk of developing melanoma.

Materials and Methods

Patients at high risk of developing melanoma who underwent follow-up using three-dimensional total body photography and digital dermoscopy between July 2021 and December 2024 at the Melanoma Unit, Hospital Clínic de Barcelona, Spain, were identified (Table 1). Excised lesions and stable melanocytic lesions (≥ 3 follow-ups and/or 18 months) with available digital dermoscopy were included. Area under the curve (AUC), sensitivity and specificity at predefined thresholds (1.5-, 2.5- and 3.5-), mean DEXI score by histological diagnosis, and top-class diagnostic accuracy were assessed.

Results

Table 1. Baseline characteristics of included patients and excised lesions

Patient characteristics	N = 423
Sex	
Female	208 (49.2%)
Male	215 (50.8%)
Mean age at first 3D-TBP (SD)	55.6 (14.9)
Personal history	
Melanoma and atypical mole syndrome	151 (35.7%)
Melanoma	138 (32.6%)
Atypical mole syndrome	93 (22.0%)
Familial melanoma	19 (4.5%)
Xeroderma Pigmentosum	7 (1.7%)
Other:	15 (3.5%)
Multiple basal or squamous cell carcinoma	10 (66.7%)
Gorlin- Goltz Syndrome	4 (26.7%)
Spitz nevus	1 (6.7%)
Family history for melanoma	
No	290 (68.6%)
Yes	133 (31.4%)
Number of primaries (Patients with previous melanoma)	N = 296
1	145 (49.0%)
2 or more	151 (51.0%)
Lesion characteristics	**N = 685**
Histological diagnosis	
Nevus	262 (38.2%)
Basal cell carcinoma	193 (28.2%)
Melanoma	117 (17.1%)
Squamous cell carcinoma	24 (3.5%)
Seborrheic Keratosis/Solar Lentigo	21 (3.1%)
Actinic Keratosis	19 (2.8%)
Dermatofibroma	6 (0.9%)
Vascular	3 (0.4%)
Other[a]	40 (5.8%)
Anatomical Location	
Trunk	361 (52.7%)
Head and Neck	150 (21.9%)
Lower Limbs	106 (15.5%)
Upper Limbs	62 (9.1%)
Acral	3 (0.4%)
Mucosa	3 (0.4%)
Melanoma, Histological subtype	
Superficial Spreading	73 (62.4%)
Lentigo Maligna	44 (37.6%)
Melanoma, Invasiveness	
In situ	82 (70.1%)
Invasive	35 (29.9%)
Breslow index (invasive melanomas), median (IQR)	0.5 (0.3-0.6)
Ulceration	
Not present	117 (100.0%)
Melanoma, associated nevus	
Not present	93 (79.5%)
Present	24 (20.5%)

[a] Other diagnoses: atypical melanocytic proliferation (n = 18), melanoma dermal metastasis (n = 7), low grade melanocytoma (n = 2), clear cell acanthoma (n = 2), lichenoid lesion (n = 2), collision tumour of basal cell carcinoma and nevus (n = 2), hidradenoma (n = 1), cutaneous leishmaniasis (n = 1), fibrolipoma (n = 1), fibrous nasal papule (n = 1), SAMPUS (n = 1), sebaceous adenoma (n = 1), xanthogranuloma (n = 1).

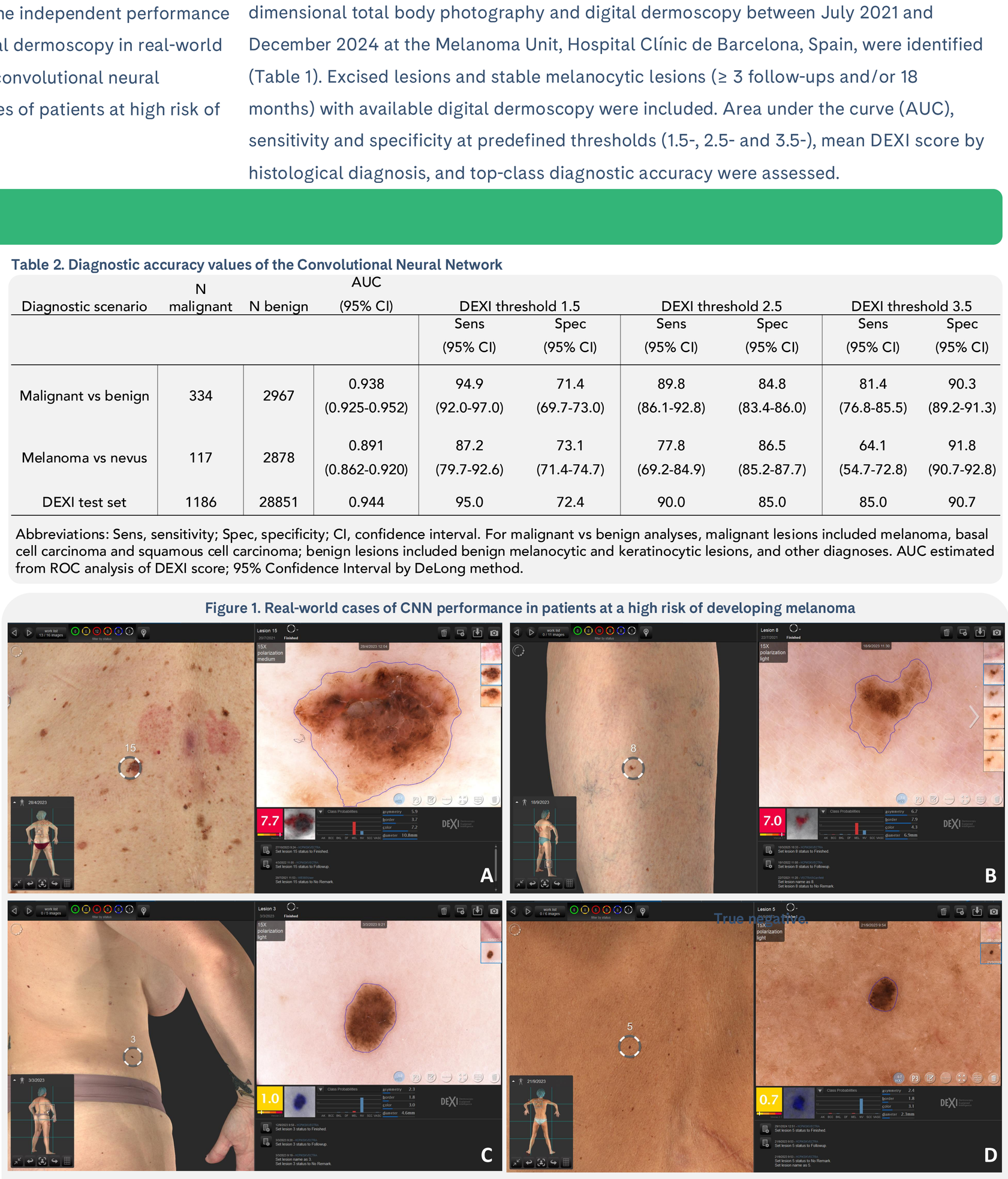
Abbreviations: TBP, total body photography; SD, standard deviation; IQR, interquartile range.

Table 2. Diagnostic accuracy values of the Convolutional Neural Network

Diagnostic scenario	N malignant	N benign	AUC (95% CI)	DEXI threshold 1.5		DEXI threshold 2.5		DEXI threshold 3.5	
				Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)
Malignant vs benign	334	2967	0.938 (0.925-0.952)	94.9 (92.0-97.0)	71.4 (69.7-73.0)	89.8 (86.1-92.8)	84.8 (83.4-86.0)	81.4 (76.8-85.5)	90.3 (89.2-91.3)
Melanoma vs nevus	117	2878	0.891 (0.862-0.920)	87.2 (79.7-92.6)	73.1 (71.4-74.7)	77.8 (69.2-84.9)	86.5 (85.2-87.7)	64.1 (54.7-72.8)	91.8 (90.7-92.8)
DEXI test set	1186	28851	0.944	95.0	72.4	90.0	85.0	85.0	90.7

Abbreviations: Sens, sensitivity; Spec, specificity; CI, confidence interval. For malignant vs benign analyses, malignant lesions included melanoma, basal cell carcinoma and squamous cell carcinoma; benign lesions included benign melanocytic and keratinocytic lesions, and other diagnoses. AUC estimated from ROC analysis of DEXI score; 95% Confidence Interval by DeLong method.

Figure 1. Real-world cases of CNN performance in patients at a high risk of developing melanoma



A. True positive: A 55-year-old woman with a history of multiple melanomas and familial melanoma presented with changes in a melanocytic lesion on the back. Histopathology after excision revealed a 0.5-mm Breslow thickness superficial spreading melanoma arising in a nevus, without ulceration; **B. False positive**. A 62-year-old woman with a history of three melanomas presented with a melanocytic lesion on the right leg showing progressive change on serial digital dermoscopy. Histopathology after excision revealed a junctional melanocytic nevus with mild atypia; **C. False negative.** A 49-year-old woman with a history of four melanomas presented with a melanocytic lesion on the right flank showing a subtle border change on 3D total body photography. Excision demonstrated a 0.3-mm invasive superficial spreading melanoma without ulceration; **D. True negative.** A 53-year-old woman with a prior stage IIIB melanoma presented with a new melanocytic lesion detected through 3D total–body photography comparison. Histopathology after excision revealed a compound melanocytic nevus with moderate atypia.

Table 3. Characteristics of True positive vs False negative Melanomas

Characteristic	Melanomas scoring DEXI > 1.5 (True Positive) N = 102	Melanomas scoring DEXI < 1.5 (False Negative) N = 15	p-value[1]
Melanoma subtype			0.038
Superficial Spreading	60 (58.8%)	13 (86.7%)	
Lentigo maligna	42 (41.2%)	2 (13.3%)	
Melanoma invasiveness			0.5
In situ	70 (68.6%)	12 (80.0%)	
Invasive	32 (31.4%)	3 (20.0%)	
Median Breslow, invasive melanomas (IQR)	0.5 (0.38-0.6)	0.3 (0.25-0.45)	0.3
Melanoma ulceration			-
Not present	102 (100.0%)	15 (100.0%)	
Melanoma, associated nevus			>0.9
Not present	81 (79.4%)	12 (80.0%)	
Present	21 (20.6%)	3 (20.0%)	
Anatomical Location			0.7
Trunk	56 (54.9%)	11 (73.3%)	
Upper limbs	18 (17.6%)	2 (13.3%)	
Lower limbs	17 (16.7%)	1 (6.7%)	
Head and neck	11 (10.8%)	1 (6.7%)	

[1] Fisher's exact test; Wilcoxon rank sum test

Table 4. Characteristics of True negative vs False positive Nevi

Characteristic	Nevus scoring DEXI < 1.5 (True Negative) N = 2103	Nevi scoring DEXI > 1.5 (False Positive) N = 775	p-value[1]
Gold Standard			
Follow-up	1979 (94.1%)	637 (82.2%)	
Histology	124 (5.9%)	138 (17.8%)	
Nevus, grade of dysplasia			>0.9
Mild	7 (5.6%)	8 (5.8%)	
Moderate	78 (62.9%)	82 (59.4%)	
Moderate-Severe	7 (5.6%)	6 (4.3%)	
Severe	16 (12.9%)	20 (14.5%)	
Non-dysplastic	16 (12.9%)	22 (15.9%)	
Anatomical Location			<0.001
Trunk	89 (71.8%)	89 (64.5%)	
Lower limbs	23 (18.5%)	24 (17.4%)	
Upper limbs	11 (8.9%)	6 (4.3%)	
Head and neck	0 (0.0%)	18 (13.0%)	
Acral	1 (0.8%)	1 (0.7%)	

[1] Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum test

Conclusions

This study highlights the importance of external validation of AI tools before clinical implementation, particularly within their intended-use population. Although the CNN demonstrated high diagnostic performance in a real-world high-risk melanoma cohort, clinically relevant melanomas would still have been missed and many benign lesions unnecessarily flagged for excision. Our findings support the cautious use of CNNs as adjunctive support rather than stand-alone tools for melanoma detection.

Clínic Barcelona | Universitat de Barcelona