

## Datasets en dermatologie : catalyseurs de l'intelligence artificielle



C. SKAYEM<sup>1</sup>, T. A. DUONG<sup>2</sup>

<sup>1</sup> UFR Simone Veil, UVSQ, MONTIGNY-LE-BRETONNEUX

<sup>2</sup> Service de Dermatologie générale et oncologique

Hôpital Ambroise-Paré, GHU Paris-Saclay AP-HP,  
BOULOGNE-BILLANCOURT

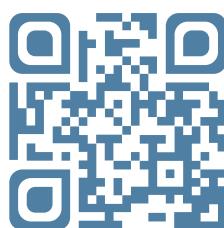
**RÉSUMÉ :** Spécialité visuelle, la dermatologie se prête particulièrement bien à l'enseignement et au diagnostic par l'image. La constitution de *datasets* de qualité, diversifiés et volumineux, est essentielle au développement d'algorithmes d'intelligence artificielle performants. Ces bases de données sont également indispensables pour valider les nouveaux modèles sur des cas connus, ce qui permet d'évaluer leur fiabilité et de faciliter les comparaisons. Enfin, elles jouent un rôle clé en recherche épidémiologique et en formation, contribuant ainsi à l'évolution de la discipline et à l'apprentissage des non-dermatologues.



### POINTS FORTS

- Les *datasets* en dermatologie permettent d'atteindre des performances diagnostiques comparables à celles des experts.
- Ils constituent une ressource clé pour la formation continue et l'apprentissage assisté par intelligence artificielle (IA).
- Leur diversité et leur représentativité conditionnent la fiabilité et l'équité des algorithmes.
- De nouveaux modèles multimodaux, comme PanDerm, ouvrent la voie à une dermatologie numérique plus intégrée.
- Les limites actuelles des *datasets* – hétérogénéité, biais de représentation et contraintes éthiques – rappellent l'importance d'une standardisation rigoureuse.

Retrouvez cette fiche en flashant  
le QR code ci-dessous



Le développement rapide de l'intelligence artificielle (IA) et de l'apprentissage automatique (*machine learning*) au cours de la dernière décennie a fondamentalement transformé l'approche de nombreuses disciplines médicales [1]. Ces technologies reposent sur une composante essentielle : les *datasets*. Par définition, un *dataset* est une collection organisée d'éléments, utilisée pour entraîner, valider et tester des modèles algorithmiques [2].

La dermatologie est une spécialité fortement visuelle, où la reconnaissance des patterns morphologiques et des structures superficielles est centrale pour établir un diagnostic. Cette dépendance à l'imagerie fait de la discipline un terrain privilégié pour le développement d'algorithmes de vision [2]. Toutefois, l'efficacité et la généralisabilité de ces modèles dépendent directement de la qualité, de la diversité et de la taille des *datasets* utilisés lors de leur conception [3]. En effet, des bases de données riches et représentatives permettent non seulement d'entraîner des réseaux neuronaux profonds (DNN) capables de rivaliser avec des dermatologues experts, mais aussi de valider leur performance dans des contextes cliniques variés [4].

## ■ Intérêts des datasets d'images

L'intérêt des *datasets* en dermatologie ne se limite pas au développement d'outils diagnostiques. Ils constituent également une ressource précieuse pour :

- l'enseignement et la formation assistée par ordinateur ;
- la recherche épidémiologique pour mieux comprendre la prévalence des affections cutanées selon les groupes démographiques ;
- la création de modèles prédictifs pour anticiper l'évolution des maladies chroniques comme le psoriasis ou la dermatite atopique.

En premier lieu, les *datasets* permettent de former des modèles de classification et de détection des lésions cutanées. Les réseaux neuronaux convolutifs (CNN), par exemple, nécessitent des milliers, voire des millions d'images annotées pour apprendre à différencier des entités

cliniques telles que les nævus bénins, les carcinomes cutanés, et les mélanomes [4]. Ce paradigme d'apprentissage supervisé, illustré par les travaux d'Esteva *et al.* [1] qui ont démontré une performance équivalente à celle de dermatologues experts, repose sur l'accès à des bases de données d'images dermatoscopiques et cliniques massives et correctement étiquetées.

Les *datasets* jouent un rôle crucial dans la **segmentation des images médicales** [5]. Cette tâche consiste à délimiter précisément les contours des lésions afin d'en extraire des mesures morphologiques pertinentes, telles que l'asymétrie, la régularité des bords ou la répartition des pigments. Ces paramètres, historiquement évalués de manière subjective, peuvent être quantifiés automatiquement grâce à des modèles entraînés sur des données segmentées manuellement par des experts [1, 3, 4].

Les jeux de données contribuent également à l'**enseignement et à la formation continue en dermatologie**. En rassemblant des images représentatives d'un large spectre de maladies cutanées, ces ressources peuvent être intégrées à des plateformes d'apprentissage en ligne qui permettent aux praticiens de se familiariser avec des présentations cliniques rares ou atypiques [5]. L'intégration de modules d'IA utilisant ces bases de données favorise un apprentissage adaptatif, où les erreurs de l'utilisateur peuvent être détectées et corrigées en temps réel [6].

Un autre intérêt majeur réside dans l'évaluation et la validation des algorithmes d'IA. Les *datasets* permettent de mesurer la robustesse, la sensibilité et la spécificité des modèles dans des conditions simulant la pratique clinique réelle [7]. La disponibilité de bases publiques et standardisées facilite en outre la comparaison entre différents algorithmes et favorise une science reproductible [8, 9].

Enfin, les *datasets* peuvent être exploités pour **des études épidémiologiques**, notamment dans la quantification de la prévalence des affections dermatologiques à l'échelle mondiale [3]. L'agrégation de

données issues de multiples populations permet d'identifier des disparités liées au sexe, à l'âge ou au phototype cutané, et d'orienter les politiques de santé publique [4, 7].

## Classification du cancer de la peau par réseaux de neurones au niveau d'un dermatologue

Les travaux d'Esteva *et al.* comptent parmi les premiers à utiliser un *dataset* extrêmement vaste, **composé de plus de 1,28 millions d'images dermatologiques annotées**, pour entraîner un réseau de neurones profond capable de classifier les cancers de la peau. Esteva *et al.* décrivent un algorithme qui atteint une performance diagnostique remarquable, avec un taux de précision comparable à celui de dermatologues experts : leur modèle a obtenu une précision globale de 72,1 à 77,0 % selon les tests, avec une aire sous la courbe ROC (AUC) allant jusqu'à 0,96 pour la détection du mélanome, démontrant ainsi une **capacité élevée à différencier lésions bénignes et malignes**. Ce travail pionnier a ainsi posé les bases d'une dermatologie assistée par IA, capable d'atteindre un niveau d'expertise humaine sur des tâches complexes de diagnostic d'imagerie cutanée.

## Datasets d'imagerie majeurs en dermatologie

PanDerm [10] est un modèle d'IA multimodal en dermatologie, pré-entraîné par apprentissage autosupervisé **sur plus de 2 millions d'images réelles de maladies de la peau provenant de 11 institutions cliniques et couvrant 4 modalités d'imagerie** : l'image clinique, la dermatoscopie, la vidéodermoscopie totale et l'histopathologie. PanDerm a été évalué sur 28 benchmarks diversifiés, incluant le dépistage du cancer de la peau, la stratification du risque, le diagnostic différentiel de pathologies cutanées courantes et rares, la segmentation des lésions, le suivi longitudinal ainsi que la prédiction de métastases et le pronostic. Le modèle PanDerm a atteint des performances à la pointe de la technologie pour l'ensemble des tâches évaluées, surpassant

# Dossier – Dermatologie et IA

souvent les modèles existants même lorsqu'il opérait avec seulement 10 % des données annotées. Trois études de lecteurs ont été réalisées afin d'évaluer le potentiel clinique de PanDerm : le modèle a surpassé les cliniciens de 10,2 % dans la détection précoce du mélanome à travers une analyse longitudinale, amélioré la précision diagnostique des cancers de la peau des cliniciens de 11 % sur des images de dermoscopie, et augmenté la capacité des professionnels de santé non-dermatologues à réaliser un diagnostic différentiel de 16,5 % sur 128 affections cutanées à partir de photographies cliniques. Ces résultats démontrent le potentiel de PanDerm pour améliorer la prise en charge des patients dans des scénarios cliniques variés et pour servir de modèle dans le développement de modèles d'analyses multimodaux dans d'autres spécialités médicales, ce qui pourrait accélérer l'intégration de l'IA dans le domaine de la santé.

Derm1M [11] est un *dataset* vision-langage de 1 029 761 paires image-texte, couvrant

plus de 390 affections et permettant l'entraînement de DermLIP, qui surpassé les modèles précédents. ISIC 2019 [12] regroupe environ 19 400 images dermoscopiques réparties en 8 à 11 catégories diagnostiques, incluant une classe "hors distribution" pour les cas atypiques. HAM10000 [13] contient 10 015 images multisource couvrant 7 types de lésions pigmentées, dont plus de 50 % confirmées histologiquement. SD-198 [14] compile environ 6 500 images cliniques (réduites à 4 467 pour 136 classes, chaque classe ayant entre 10 et 60 images). ISIC 2018 [12] propose environ 12 500 images pour la segmentation, la détection d'attributs et la classification. DermNet [15] fournit environ **20 000 images cliniques** réparties en **23 maladies** (eczéma, psoriasis, mélanome, etc.). Derm7pt [16] contient environ 20 000 images cliniques/dermoscopiques annotées selon le système des 7 points et sert de benchmark. DermIS [17] est un atlas en ligne avec des milliers d'images couvrant des centaines de dermatoses, sans structure de classes fixes. ISIC 2020 [12] atteint

33 126 images avec métadonnées (âge, sexe, localisation, type clinique). PH2 [18] contient 200 images portugaises pour la classification de nævus communs, atypiques et mélanomes, utilisées pour validation segmentée. Enfin, MED-NODE [19] comprend 170 images cliniques (70 mélanomes, 100 nævus) destinées à la détection de mélanomes via des algorithmes comme SENet et ResNeXt. Le **tableau I** reprend les principaux *datasets* d'images disponibles.

## Limitations et enjeux des datasets

L'utilisation croissante des *datasets* en dermatologie s'accompagne d'enjeux majeurs qui nécessitent une attention particulière afin de garantir leur utilisation sûre, équitable et efficace. Ces enjeux se situent à l'interface entre **la qualité scientifique des données, les aspects éthiques et réglementaires, ainsi que les considérations sociétales relatives à l'équité et à la représentativité**.

Dataset	Nombre d'images	Caractéristiques principales	Différenciateurs clés
PanDerm [10]	> 2 000 000	Multimodal, auto-apprentissage, images cliniques issues de 11 institutions, 4 modalités d'imagerie	Très grande échelle, multimodal, autosupervisé, multitâche
Derm1M [11]	~ 1 000 000	Dataset vision-langage aligné avec ontologie clinique, multimodal, grande échelle	Alignement texte-image, multitâche clinique
ISIC 2019 [12]	25 331	8 catégories diagnostiques (tumeurs), images dermoscopiques variées	Large volume, multiples classes, qualité contrôlée
HAM10000 [13]	10 015	7 classes (tumeurs, lésions vasculaires), images dermoscopiques annotées, diversité de lésions pigmentées	Référence pour mélanomes, public et bien annoté
SD-198 [14]	6 584	198 classes de maladies (y compris inflammatoires), photos cliniques et dermoscopiques	Très grand nombre de classes, diversité importante
ISIC 2018 [12]	13 715+	7 classes, données multisites, confirmations clinique et histopathologique	Classes cliniques standardisées, données multi-institutionnelles
DermNet [15]	> 20 000	Plus de 200 maladies différentes, images évaluées et annotées par des experts	Grande diversité de maladies, annotation clinique
Derm7pt [16]	> 20 000	Photos cliniques et non cliniques, données synthétiques, large variété de pathologies	Mix réel/synthétique, diversité large, multitâche
DermIS [17]	> 3 000	Environ 20 types de maladies, <i>dataset</i> universitaire allemand	Variété de pathologies, origine universitaire
ISIC 2020 [12]	2 000	Images dermoscopiques uniques, diagnostic confirmé par histopathologie, patients uniques	Liens patients uniques, confirmation histopathologique précise
PH2 [18]	~ 200	Images dermoscopiques annotées de mélanomes, nævus et dermatofibromes	Annotation au niveau pixel, utilisé pour segmentation et classification
MED-NODE [19]	170	Images macroscopiques de mélanomes et nævus	<i>Dataset</i> spécifique pour mélanomes et nævus

**Tableau I :** Listes non exhaustives de *datasets* en dermatologie.

Sur le plan scientifique, un premier défi réside dans la qualité et l'hétérogénéité des données [12-14]. Les images dermatologiques sont issues de multiples sources : appareils dermatoscopiques, caméras de smartphones, bases hospitalières, ce qui entraîne une variabilité importante en termes de résolution, d'éclairage et de mise au point [13, 15]. Cette hétérogénéité peut affecter les performances des modèles d'IA qui y sont entraînés, notamment lorsqu'ils sont appliqués à des contextes cliniques différents de ceux des données d'origine [10-18]. De plus, les erreurs d'annotation ou le manque de validation par des dermatologues experts peuvent introduire des biais qui compromettent la fiabilité des résultats.

Un enjeu central concerne la **représentativité des populations** incluses dans ces *datasets*. De nombreuses bases de données largement utilisées, comme ISIC [12], sont composées principalement d'images de patients à phototypes clairs (Fitzpatrick I-III). Cette sous-représentation des peaux foncées (phototypes IV-VI) a été mise en évidence comme un facteur de biais majeur, susceptible de réduire la précision diagnostique des algorithmes chez ces populations. Ainsi, le développement de *datasets* inclusifs et diversifiés est impératif pour éviter l'exacerbation des inégalités de santé déjà présentes en dermatologie [2, 3, 20, 21].

Les annotations sont parfois basées sur des diagnostics cliniques sans confirmation histologique systématique, introduisant un bruit d'étiquetage. Dans certains *datasets* collaboratifs comme ISIC, des erreurs d'annotation ont été identifiées après coup [12].

Les images elles-mêmes présentent une variabilité technique importante (qualité, appareil photo, conditions de prise) et les bases sont souvent monomodales, ne contenant que des images dermoscopiques (ISIC, HAM10000) ou cliniques (DermNet, DermIS), très rarement les 2, ce qui ne reflète pas la réalité multimodale de la pratique [10-18]. Par ailleurs, plusieurs jeux souffrent de tailles limitées (PH<sup>2</sup>: 200 images, MED-NODE: 170 images) [18, 19] et de métadonnées incomplètes, et

certains posent des contraintes d'accès (licence restrictive), ce qui amoindrit leur utilité pour le *deep learning*. Pour les constructeurs d'appareils d'imagerie, il est plus facile de développer des algorithmes d'aide d'interprétation des images sur des images annotées et directement acquises par le dispositif d'imagerie ; cela limite les biais liés à l'acquisition des images, l'éclairage, la distance d'acquisition et la mise au point.

Souvent, des informations cliniques cruciales (âge, sexe, localisation anatomique, symptômes associés) sont absentes ou incomplètes, alors qu'elles sont essentielles en dermatologie [2, 3, 20].

Les classes sont souvent déséquilibrées : les nævus bénins sont massivement représentés, tandis que les cancers rares (ex : carcinome de Merkel) ou les infections dermatologiques (ex : mycoses) sont très peu présents. Exemple : dans HAM10000, les nævus représentent 67 % des images [13].

**Les questions éthiques et légales constituent également un domaine d'enjeu critique.** L'utilisation d'images médicales, même anonymisées, soulève des préoccupations relatives à la confidentialité des patients et à leur consentement éclairé [13, 22, 23]. La réglementation en vigueur, notamment le règlement général sur la protection des données (RGPD) en Europe, impose des exigences strictes en matière de protection des données personnelles [23]. Par ailleurs, l'absence de normes uniformes concernant l'anonymisation des images dermatologiques peut poser problème lorsque des caractéristiques identifiantes, telles que des tatouages ou des grains de beauté distinctifs, apparaissent dans les clichés.

Un autre enjeu est celui de la **transparence et de la reproductibilité**. L'accès restreint à certains *datasets* propriétaires limite la capacité des chercheurs indépendants à valider et comparer les performances des algorithmes. La mise à disposition de bases de données ouvertes et standardisées est donc essentielle pour favoriser une science ouverte et reproductible [23]. Cependant, cela doit être

mis en balance avec le risque potentiel de mésusage des données par des acteurs non médicaux, notamment dans des contextes commerciaux ou assurantiels.

Enfin, l'utilisation des *datasets* soulève des questions relatives à la durabilité et à l'impact environnemental. L'entraînement de modèles d'apprentissage profond sur des bases de données massives nécessite une puissance de calcul importante, avec des coûts énergétiques et environnementaux non négligeables [23]. Ces considérations doivent être intégrées dans les stratégies de développement pour garantir une innovation technologique responsable [24].

## ■ Conclusion

Les *datasets* en dermatologie représentent une ressource fondamentale pour le développement d'outils d'intelligence artificielle performants. Une combinaison de *datasets* contenant d'un côté l'image et de l'autre le contexte clinique – ou la description du dermatologue – constitue aussi une piste pour l'augmentation de la performance des algorithmes. Cependant, leur construction et leur utilisation soulèvent des défis majeurs liés à l'hétérogénéité des sources, à la standardisation des annotations et à la reproductibilité des modèles. La complexité des images dermatologiques, la variabilité inter-observateurs et la rareté de certaines pathologies limitent également la généralisation des algorithmes. Ces contraintes techniques rappellent l'importance d'une approche méthodologique rigoureuse et d'une validation continue pour exploiter pleinement le potentiel des *datasets* dans la dermatologie numérique. Dans la perspective d'une validation de certains algorithmes d'IA en tant qu'outils d'aide à la décision, il est probable que la définition par les experts de seuils de détection et de *datasets* de validation sera nécessaire pour aider les utilisateurs à mieux distinguer qualitativement les logiciels. Ces efforts collaboratifs permettront de faire des *datasets* non pas de simples outils, mais de véritables catalyseurs de la dermatologie numérique de demain.

# Dossier – Dermatologie et IA

## Bibliographie

1. ESTEVA A, KUPREL B, NOVOA RA *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017;542:115-118.
2. WEHNER MR, LEVANDOSKI KA, KULLDORFF M *et al.* Research Techniques Made Simple: An Introduction to Use and Analysis of Big Data in Dermatology. *J Invest Dermatol*, 2017;137:e153-e158.
3. JEONG HK, PARK C, HENAO R *et al.* Deep learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations. *JID Innov*, 2022;3:100150.
4. WARD A, LI J, WANG J *et al.* Creating an Empirical Dermatology Dataset Through Crowdsourcing With Web Search Advertisements. *JAMA Netw Open*, 2024;7:e2446615.
5. COMBALIA M, CODELLA N, ROTEMBERG V *et al.* Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge. *Lancet Digit Health*, 2022;4:e330-e339.
6. CODELLA N, GUTMAN D, CELEBI ME *et al.* Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv*, 2018. Disponible sur: <https://arxiv.org/abs/1710.05006v3>
7. MARON RC, WEICHENTHAL M, UTIKAL JS *et al.* Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer*, 2019;119: 57-65.
8. MAR VJ, SOYER HP. Artificial intelligence for melanoma diagnosis: how can we deliver on the promise? *Ann Oncol*, 2019;30:e1-e3.
9. SUN MD, KENTLEY J, MEHTA P *et al.* Accuracy of commercially available smartphone applications for the detection of melanoma. *Br J Dermatol*, 2022;186:744-746.
10. YAN S, YU Z, PRIMERO C *et al.* A multimodal vision foundation model for clinical dermatology. *Nat Med*, 2025;31:2691-2702.
11. YAN S, HU M, JIANG Y *et al.* Derm1M: A Million-scale Vision-Language Dataset Aligned with Clinical Ontology Knowledge for Dermatology. *arXiv*, 2025. Disponible sur: <https://arxiv.org/abs/2503.14911>
12. CASSIDY B, KENDRICK C, BRODZICKI A *et al.* Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Med Image Anal*, 2022;75:102305.
13. TSCHANDL P, ROSENDAHL C, KITTNER H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*, 2018;5:180161.
14. YANG J, WU X, LIANG J *et al.* Self-Paced Balance Learning for Clinical Skin Disease Recognition. *IEEE Trans Neural Netw Learn Syst*, 2020;31:2832-2846.
15. Dermnet. Sur le site [dermnet.com](http://dermnet.com).
16. KAWAHARA J, DANESVAR S, ARGENZIANO G *et al.* 7-Point Checklist and Skin Lesion Classification using Multi-Task Multi-Modal Neural Nets. *IEEE J Biomed Health Inform*, 2018.
17. FARHAT U. "Skin Lesion Dermis Dataset". Sur le site [kaggle.com](https://www.kaggle.com/datasets/farhatullah8398/skin-lesion-dermis-dataset). <https://www.kaggle.com/datasets/farhatullah8398/skin-lesion-dermis-dataset>
18. MENDONCA T, FERREIRA PM, MARQUES JS *et al.* PH<sup>2</sup> - a dermoscopic image database for research and benchmarking. *Annu Int Conf IEEE Eng Med Biol Soc*, 2013;2013:5437-5440.
19. MEDNODE Dataset University Medical Center Groningen. Disponible sur: [https://www.cs.rug.nl/~imaging/databases/melanoma\\_naevi/index.html](https://www.cs.rug.nl/~imaging/databases/melanoma_naevi/index.html)
20. DANESHJOU R, VODRAHALLI K, NOVOA RA *et al.* Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci Adv*, 2022;8:eabq6147.
21. BARBIERI JS. Strengths and Limitations of Study Designs Using Administrative Data to Assess Incidence and Prevalence of Disease. *JAMA Dermatol*, 2023;159:991.
22. YADAV N, PANDEY S, GUPTA A *et al.* Data Privacy in Healthcare: In the Era of Artificial Intelligence. *Indian Dermatol Online J*, 2023;14:788-792.
23. VANARIA RJ, BHUPALAM V, MARRERO-PEREZ A *et al.* Environmental Impact and Sustainability Associated with the Practice of Dermatology. *J Clin Aesthet Dermatol*, 2025;18:50-55.
24. ALLWRIGHT E, ABBOTT RA. Environmentally sustainable dermatology. *Clin Exp Dermatol*, 2021;46:807-813.

Les auteurs ont déclaré ne pas avoir de liens d'intérêts concernant les données publiées dans cet article.