scientific data

OPEN

DATA DESCRIPTOR

A fusocelular skin dataset with whole slide images for deep learning models

Rocío del Amor[1,2,6] ✉, Miguel López-Pérez[1,6], Pablo Meseguer[1,6], Sandra Morales[1], Liria Terradez[3], Jose Aneiros-Fernandez[4], Javier Mateos[5], Rafael Molina[5] & Valery Naranjo[1,2] ✉

Cutaneous spindle cell (CSC) lesions encompass a spectrum from benign to malignant neoplasms, often posing significant diagnostic challenges. Computer-aided diagnosis systems offer a promising solution to make pathologists' decisions objective and faster. These systems usually require large-scale datasets with curated labels for effective training; however, manual annotation is time-consuming and expensive. To overcome this challenge, crowdsourcing has emerged as a popular and valuable strategy to scale up the labeling process by distributing the effort among different non-expert annotators. This work introduces AI4SkIN, the first public dataset Whole Slide Images (WSIs) for CSC neoplasms, annotated using an innovative crowdsourcing protocol. AI4SkIN dataset contains 641 Hematoxylin and Eosin stained WSIs with multiclass labels from both expert and trainee pathologists. The dataset improves CSC neoplasm diagnosis using advanced machine learning and crowdsourcing based on Gaussian Processes, showing that models trained on non-expert labels perform comparably to those using expert labels. In conclusion, we illustrate that AI4SkIN provides a good resource for developing and validating methods for multiclass CSC neoplasm classification.

Background & Summary

Cutaneous spindle cell (CSC) lesions represent a diagnostically challenging group of tumors composed of spindle-shaped cells, often arranged in fascicles. Despite this common morphology, they include a wide range of entities, from reactive to malignant, with overlapping features that complicate the accurate classification[1]. A reliable diagnosis is necessary to predict their behavior, the outcome and survival rates[2]. The most frequent CSC neoplasms are: leiomyomas (lm), leiomyosarcomas (lms), dermatofibromas (df), dermatofibrosarcomas (dfs), spindle cell melanomas (scm), atypical fibroxanthomas (afx), and squamous cell carcinoma (scc). Visual examples of those neoplasms are provided in Fig. 1.

Spindle cell tumor diagnosis depends on both morphological and immunohistochemical criteria, but subtle variations between entities can make interpretation difficult, especially for pathologists in training[1]. A structured, pattern-oriented diagnostic framework can support more accurate identification and differentiation of these lesions. Table 1 outlines the key features that differentiate various spindle cell neoplasms. Accurate classification relies on the pathologist's ability to evaluate specific histological criteria, including: (1) growth pattern, (2) cellular density, (3) cytological features, (4) nature and extent of extracellular matrix, (5) tumor-stroma interface, (6) vascular characteristics, (7) presence of necrosis, and (8) mitotic count. At low magnification, features such as preserved tissue architecture, zonation, symmetry, and cellularity assist in distinguishing benign from malignant lesions. In contrast, high-power views highlighting nuclear atypia and atypical mitoses are more indicative of malignancy.

[1]Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, HUMAN-tech Universitat Politècnica de València, Valencia, Spain. [2]Artikode Intelligence S.L, Valencia, Spain. [3]Pathology Department. Hospital Clínico Universitario de Valencia, Universidad de Valencia, Valencia, Spain. [4]Pathology Department. Hospital Universitario San Cecilio de Granada, Granada, Spain. [5]Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain. [6]These authors contributed equally: Rocío del Amor, Miguel López-Pérez, Pablo Meseguer. ✉e-mail: madeam2@upv.es; vnaranjo@upv.es

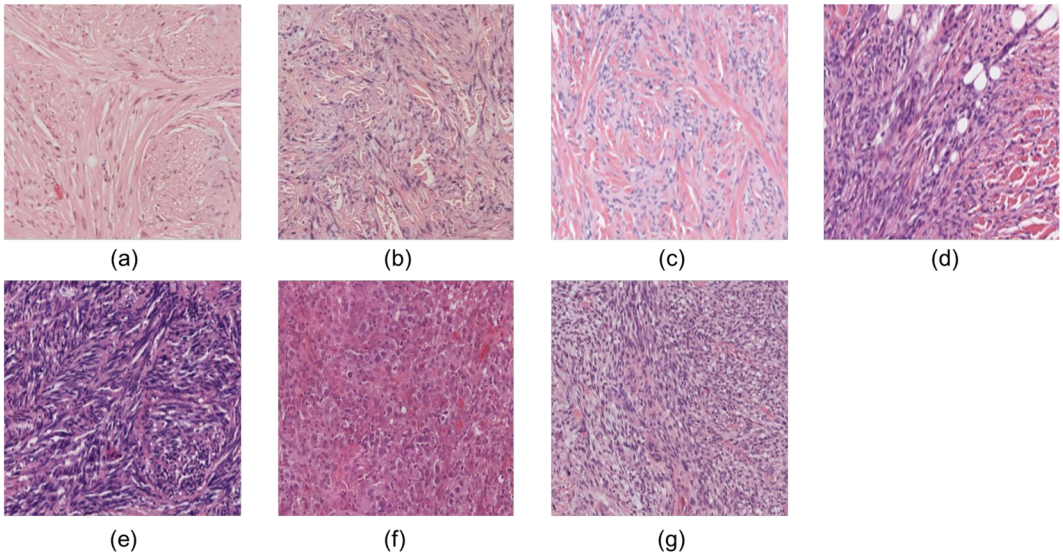

Fig. 1 Representative patches extracted from the most common CSC neoplasms[19]: (**a**) leiomyoma, (**b**) leiomyosarcoma, (**c**) dermatofibroma, (**d**) dermatofibrosarcoma, (**e**) spindle cell melanomas, (**f**) atypical fibroxanthoma and (**g**) squamous cell carcinoma.

Tumor type		**Benignity**		**Malignancy**	
Origin tumor	*Significant patterns*	*Name*	*Features*	*Name*	*Features*
Smooth muscle cells	Spindle cells with eosinophilic cytoplasm Elongated nuclei (pure form)	**Leiomyoma**	No mitosis (exceptional) No frequent atypia	**Leiomyosarcoma**	Mitoses always present Nuclear atypia
Connective tissue cells	Spindle cells with a swirling or storiform pattern	**Dermatofibroma**	May have mitosis Multinucleated cells Epidermal ridges	**Dermatofibrosarcoma**	Few but present mitoses No multinucleated cells Epidermis more flattened
Melanocytic cells	Spindle cell fascicles	—	—	**Spindle cell melanoma**	Mitoses $\geq$ 2/mm2 Significant atypia
Squamous cells	Spindle cell fascicles with variable cohesiveness	—	—	**squamous cell carcinoma**	Mitosis Nuclear atypia
Fibroblasts	Spindle-shaped, histiocytoid and multinucleated cells	—	—	**Atypical fibroxanthoma**	Solar elastosis Multinucleated cells Atypical mitoses

Table 1. Overview of the histopathological features of spindle cell neoplasms included in the dataset[19].

Machine Learning (ML) techniques have excelled as a tool to perform automatic diagnosis in histopathological images[3]. These methods are included in Computer-Aided Diagnosis (CAD) systems, aiming to provide further insights and support to pathologists[4]. For ML techniques to work effectively, they need large and diverse datasets often with labeled data, which are time-consuming and difficult to create. Crowdsourcing has become a popular strategy for labeling histopathological data by distributing the task among many annotators with different levels of expertise[5].

While crowdsourcing has yielded accurate results in simpler tasks, such as detecting nuclei or cancer cells[6], it can introduce errors in more complex tasks like subtype classification[7,8]. To mitigate this, labels from different annotators are often combined to create a more accurate dataset, which is then used for training a classifier, e.g., by label aggregation[9–11]. However, recent studies show that this approach may not be the best, as models that account for each annotator's errors during training tend to perform better[12]. New methods have been developed to learn from noisy labels provided by non-experts, achieving performance similar to that of expert-labeled datasets in histopathological tissue classification. Among them, stands out SVGPCR[13,14], which is based on Gaussian Processes (GPs), a non-parametric probabilistic classifier that takes into account the uncertainty while modeling and prediction. The main advantage of GPs is that they are not prone to overfit and can perform fairly well with a small amount of data, which is usually the case in medical imaging tasks[15–18].

This work presents AI4SkIN, the first public dataset of complete WSIs from CSC neoplasms. The dataset is created using a new crowdsourcing protocol, which is an extension of the one described for the binary classification of CSC neoplasms[19]. AI4SkIN includes multiclass labels and 641 WSIs with Hematoxylin and Eosin (H&E) stains, along with labels from 2 expert pathologists and 10 pathologists in training. This resource aims to help researchers improve the diagnosis and prediction of challenging CSC neoplasms by developing methods to learn from crowd labels. Additionally, the dataset is validated with state-of-the-art crowdsourcing methods based on GPs and implemented in Python. To the best of our knowledge, there is no publicly available dataset that allows the multiclass classification of CSC neoplasms.

	lm	lms	df	dfs	scm	afx	scc	Total
HCUV	31	23	102	21	48	44	15	284
HUSC	73	23	93	36	74	58	—	357
Total	104	46	195	57	122	102	15	641

Table 2. AI4SkIN dataset distribution. CSC neoplasms contained in the dataset (lm:leiomyomas; lms: leiomyosarcomas; df:dermatofibromas; dfs: dermatofibrosarcomas; scm: spindle cell melanomas; afx: atypical fibroxanthomas; scc: squamous cell carcinoma.).

Methods

Study approval. All participants in the study provided written informed consent for the use of their data in this research, in accordance with approvals granted by the Ethics Committees of the University Clinic Hospital of Valencia (*Hospital Clínico Universitario de Valencia - HCUV*) and the San Cecilio University Hospital (*Hospital Universitario San Cecilio - HUSC*). Both committees issued favorable opinions regarding the study and its consent procedures, corresponding to approvals no. 2020/245 and no. 2/22, respectively. The study was conducted within the framework of the AI4SkIN project, funded by the Spanish Ministry of Science and Innovation (grants PID2019-105142RB-C21 and PID2019-105142RB-C22), and adhered to the ethical principles outlined in the Declaration of Helsinki. Additionally, the Research Ethics Committee of the Universitat Politécnica de Valéncia (UPV), acting as the coordinating institution, issued a favorable report for the project (reference P07-08-07-2020). The authors ensured that all data in the study were appropriately de-identified and handled securely and confidentially to safeguard participants' rights and welfare, in full compliance with informed consent and ethical research standards.

Selection and preparation of the slides. The AI4SkIN dataset is composed of two different datasets obtained from the Departments of Anatomical Pathology at HCUV (Valencia, Spain) and HUSC (Granada, Spain). Each dataset comprises, respectively, 284 and 357 H&E slides from skin tissue samples taken from 588 patients collected from the hospitals' archives according to the pathology reports.

Digitization and Pre-processing. The formalin-fixed paraffin-embedded (FFPE) tissue blocks and slides from all selected cases were collected from the institutions' archives. For the AI4SkIN database, the HCUV and the HUSC selected cases from the last 10 years, ensuring recent and relevant data. However, for spindle cell carcinoma, the search was extended to 20 years, given its infrequency in comparison to the other groups.

Slides from HCUV related to the selected cases were retrieved from the hospital archives. A qualified pathologist (LT) chose the optimal slide for each case, after which all slides were digitized to be specifically used for this study. In the case of the HUSC, images already digitized from the HUSC system were used, which cover a period of up to 5 years, corresponding to the scope of the hospital's digital storage. To complement this, additional slides were selected and digitized from the physical archive, to expand the database with images that were not previously digitized. After digitization, WSIs were examined by qualified pathologists (LT at HUCV and JA at HUSC) for quality problems such as blurring, artifacts, or inadequate coloration. Images that did not meet quality standards were either rejected or re-scanned.

The digitization process at HCUV was carried out using Roche's scanner, Ventana iScan HT, equipped with a 40× objective lens (0.227 M/pixel) and directly saved in .tif file format. At HUSC, the slides were scanned using Philips Ultra Fast Scanner, which automatically provides images at 40× magnification, resulting in a resolution of 0.25 microns per pixel. Images were generated in Philips-specific .isyntax format, ensuring high quality and fidelity. For its use and analysis, these WSIs were converted to .tiff format using the Philips Image Management System (IMS). The digitization process at both centers encompassed a maximum magnification of 40×, including all levels down to 5×.

Expert labels. Two expert pathologists (LT, JA) re-evaluated the slides to confirm the diagnosis for each case and to label each image. Specifically, one pathologist (LT) examined the 284 images from HCUV, while the other pathologist (JA) reviewed the 357 images associated with HUSC. Each whole-slide image (WSI) corresponds to one of the seven types of CSC neoplasms under study. Table 2 details the content of the AI4SkIN dataset.

Pathologist-in-training labels. Ten pathologists-in-training participated in the study. In this case, four resident pathologists belonged to the HCUV and six to the HUSC. Four pathologists-in-training were in their fourth year of residency, three in their third year, one in their second year and two in their first year. An annotation protocol was designed to ensure that the 75% of WSIs were annotated by at least one pathologist-in-training. In concrete, 82 images were annotated by all pathologists-in-training (dense set). In contrast, the rest were only annotated by some pathologists (non-dense set).

Each pathologist-in-training assigned a global label (image level) to each WSI corresponding to one of the seven considered types of neoplasms. In Fig. 2, the agreement between annotators is shown by means of the Kappa score in the dense set for each pair of raters[20] and the MV and expert labels. The Kappa score ranges between −1 (total disagreement) and 1 (total agreement). In our dataset, we observe that the agreement with the expert label is low. Furthermore, the agreement between annotators is neither strong nor substantial, indicating the wide range of non-expert opinions present in the dataset.

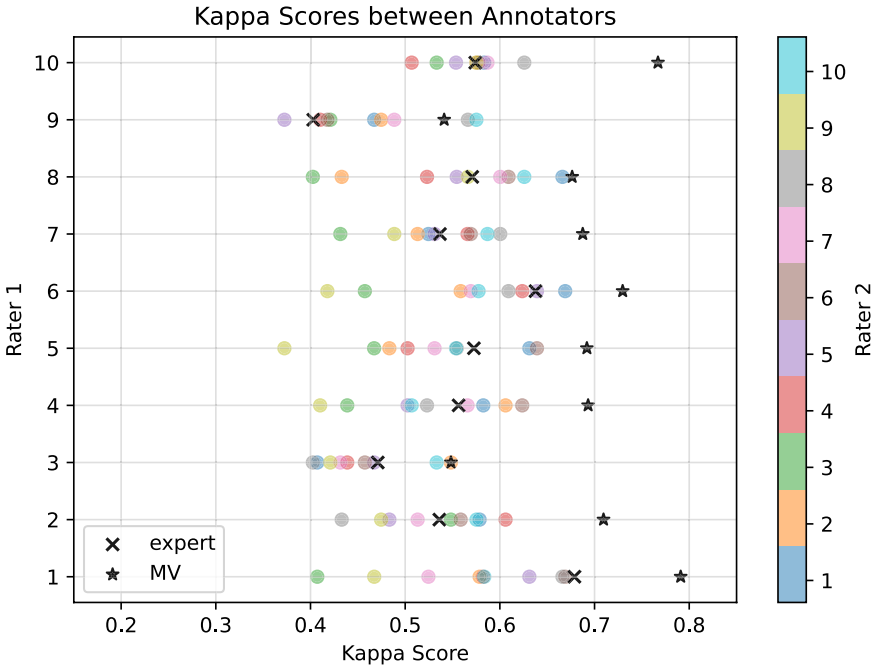

Fig. 2 Agreement between annotators: Kappa score.

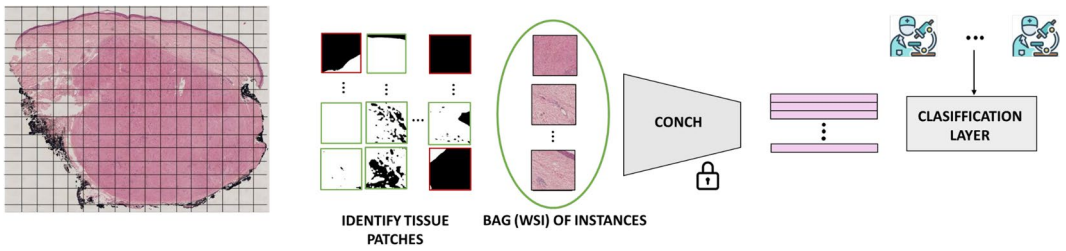

Fig. 3 Technical validation for the AI4SkIN dataset.

Data Record

The complete AI4SkIN dataset is available in Figshare[21]. The dataset consists of two components. There is a zipped file named "WSI.7z" containing the digitalized slides grouped in folders according to the subtype. Additionally, a spreadsheet referred to as "AI4SkIN_Database.xlsx", which includes the histopathological diagnosis for each case (expert and pathologist-in-training labels) is provided.

Image Data. Each TIF file is named according to the following format `AI4SkIN_HOSPITAL_NUMBER_SAMPLE.tif`.

- `HOSPITAL`: This will be either `HUSC` or `HCUV`, depending on the hospital.
- `NUMBER`: A unique random number assigned to each file.
- `SAMPLE`: Indicates the biopsy sequence. Start with `A` for the initial biopsy. Additional biopsies for the same patient will be labeled from `B` to `F`.

Technical Validation

To validate the dataset proposed in this paper, we present a GP-based approach to classify WSIs of CSC neoplasms, see Fig. 3. We discarded the 'scc' class because is underrepresented, and conducted the experiments with the remaining 626 WSIs. For this validation, we compare two types of approaches: those using pathologist-in-training labels and those using expert labels. Regarding approaches using non-expert labels, we studied four label aggregation models (i.e., MV, DS[9], MACE[10] and GLAD[11]). These models perform the classification in two stages. In the first stage, they aggregate the labels into a unique 'curated' label. MV uses majority voting, while DS, GLAD, and MACE perform smarter aggregation by estimating annotator biases. In the second stage, a GP classifier is applied with the aggregated labels. Apart from these label aggregation approaches, we also explore the SVGPCR model[13], which jointly estimates annotator biases, true labels, and the GP classifier during the training process. Regarding approaches using expert labels, we present a GP classifier with these labels. Since expert labels are considered the ground truth for this task, this model would be the upper bound on performance.

Class	Train		Val		Test	
	HCUV	HUSC	HCUV	HUSC	HCUV	HUSC
lm	19	36	8	11	4	26
lms	16	9	3	8	4	6
df	75	50	23	10	4	33
dfs	9	17	1	8	11	11
scm	32	52	7	12	9	10
afx	36	33	6	4	2	21

Table 3. Number of WSIs in each class and hospital in the train, validation, and test subsets.

Model	VGG16-IN	UNI	PLIP	CONCH
MV	0.5424 ± 0.0251	0.5891 ± 0.0134	0.6101 ± 0.0055	0.6005 ± 0.0046
GLAD[11]	0.5477 ± 0.0118	0.5317 ± 0.1573	0.6142 ± 0.0094	0.6655 ± 0.0096
MACE[10]	0.5007 ± 0.0207	0.6156 ± 0.0245	0.6295 ± 0.0048	0.6589 ± 0.0096
DS[9]	0.5566 ± 0.0168	0.6541 ± 0.0216	0.6497 ± 0.0113	0.6910 ± 0.0211
SVGPCR[13]	$\mathbf{0.6760 \pm 0.0127}$	0.7540 ± 0.0180	0.7248 ± 0.0205	0.8045 ± 0.0103
Expert	0.6379 ± 0.0173	$\mathbf{0.7825 \pm 0.0225}$	$\mathbf{0.7523 \pm 0.0145}$	$\mathbf{0.8822 \pm 0.0078}$

Table 4. Macro-averaged F1 score through the six classes. We utilize several feature extractors and labeling configurations. All models were run 5 times and then the metrics were averaged. We also report the standard deviation.

The dataset is divided into train/validation/test sets (see Table 3), ensuring a homogeneous distribution across the six studied classes. We also ensure that WSIs coming from the same patient belong to the same set. To provide a learning-from-crowds setting, in the training and validation sets, we include WSIs with Pathologist-in-training labels, while the test set only contains WSIs with expert labels. For a fair comparison, each approach utilizes the same hyperparameters and follows the same optimization procedure, namely an RBF kernel, 200 inducing points, a minibatch size of 64, and the Adam optimizer with a learning rate of 10^{-2}. Five independent runs are conducted for each model and results were averaged for comparison.

Data pre-processing and feature extraction. The WSIs are divided into 512×512 patches without overlap at a magnification of $10\times$. The Otsu threshold method is applied to the magenta channel to separate tissue from the background. Patches containing less than 20% tissue are discarded. Leveraging the success of recent foundation models trained on diverse histopathological tissue, we obtain an ambedding for each patch using the following features extractors: VGG16-IN[22], UNI[23], PLIP[24], and CONCH[25]. While VGG16-IN is pre-trained on ImageNet, the rest are foundation models pretrained with histopathological images. These foundation models aim to achieve general-purpose models for any task in computational pathology. Once the embedding is obtained for each single patch within a WSI, they are averaged across all patches to obtain a global feature vector for the WSI. These vectors serve as input to the GP classifiers.

WSI classification results. Table 4 depicts the global results using the different feature extractors. In general, we observe that the performance of VGG16-IN is inferior to that of the benchmark models, regardless of the label modeling. Overall, the best feature extractor is CONCH, which achieves an F1 score of 88.22 % with the GP using expert labels on the test set for the six-class classification problem. As expected, approaches using pathologist-in-training labels experienced a decrease in performance. In particular, models conducting label aggregation (i.e., MV, GLAD, MACE, and DS) perform worse than SVGPCR. Remind that SVPGCR benefits from learning from the annotator's experience along with the latent classifier, achieving an F1 score of 80.45% with the CONCH model.

Table 5 reports the per-class results using CONCH as feature extractor. We can observe the clear superiority of the model using expert labels. In contrast, the other models see their performance harmed by the noise introduced by pathologist-in-training labels. Among them, SVGPCR stands out, which remains stable in the different classes. The models that aggregate labels obtain a very low value in the lms class, especially MV that obtained an F1 score of 0 in this class.

Limitations. The dataset has several limitations. Firstly, the number of images is relatively small compared to other tumor types, reflecting the lower prevalence of these specific lesions. A notable limitation is the class imbalance, with some categories underrepresented due to their less frequent occurrence. Additionally, the images were acquired using two different scanners, which could affect the consistency and generalizability of the findings. Despite these challenges, the dataset offers valuable insights into spindle cell tumors and serves as a useful resource for future research and development in this area.

Model	lm	lms	df	dfs	scm	afx
MV	0.7350 ± 0.0145	0.0000 ± 0.0000	0.8230 ± 0.0123	0.6554 ± 0.0169	0.8105 ± 0.0445	0.5788 ± 0.0497
GLAD[11]	0.7873 ± 0.0131	0.1876 ± 0.0537	0.8354 ± 0.0105	0.6620 ± 0.0262	0.7922 ± 0.0107	0.7287 ± 0.0296
MACE[10]	0.7776 ± 0.0130	0.1452 ± 0.0445	0.8381 ± 0.0127	0.6630 ± 0.0287	0.8082 ± 0.0112	0.7214 ± 0.0355
DS[9]	0.8121 ± 0.0274	0.3282 ± 0.0660	0.8414 ± 0.0085	0.6691 ± 0.0485	0.8020 ± 0.0315	0.6930 ± 0.0163
SVGPCR[13]	0.8650 ± 0.0109	0.6037 ± 0.0312	0.8825 ± 0.0124	0.7829 ± 0.0146	0.8228 ± 0.0241	0.8698 ± 0.0299
Expert	**0.9237 ± 0.0123**	**0.8366 ± 0.0292**	**0.9023 ± 0.0051**	**0.8233 ± 0.0171**	**0.8570 ± 0.0137**	**0.9505 ± 0.0107**

Table 5. Per-class F1 score using CONCH[25] as the feature extractor. We compare the different labeling configurations. All models were run 5 times and then the metrics were averaged, we also reported the standard deviation.

Code availability

The code to extract processing the dataset and the technical validation is implemented in Python and is available at GitHub (https://github.com/vipgugr/AI4SkIN-technical-validation). The foundational models used to extract the WSI embeddings are implemented in Pytorch and publicly available. The crowdsourcing models and classifiers are implemented using GPflow 1.2.0, a framework that utilizes Gaussian Processes, accelerated by TensorFlow to enable GPU computation.

Received: 30 October 2024; Accepted: 28 April 2025;
Published online: 14 May 2025

References

1. Choi, J. H. & Ro, J. Y. Cutaneous spindle cell neoplasms: Pattern-based diagnostic approach. *Archives of Pathology & Laboratory Medicine* **142**, 958–972 (2018).
2. Winnepenninckx, V. *et al.* New phenotypical and ultrastructural findings in spindle cell (desmoplastic/neurotropic) melanoma. *Applied Immunohistochemistry and Molecular Morphology* **11**, 319–325 (2003).
3. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60–88 (2017).
4. Van der Laak, J. *et al.* Deep learning in histopathology: The path to the clinic. *Nature Medicine* **27**, 775–784 (2021).
5. Zhang, J. *et al.* Learning from crowdsourced labeled data: A survey. *Artificial Intelligence Review* **46**, 543–576 (2016).
6. Irshad, H. *et al.* Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: Evaluating experts, automated methods, and the crowd. In *Pacific Symposium on Biocomputing*, 294–305 (World Scientific, 2014).
7. Grote, A. *et al.* Crowdsourcing of histological image labeling and object delineation by medical students. *IEEE Trans. Med. Imaging* **38**, 1284–1294 (2018).
8. Amgad, M. *et al.* Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics* **35**, 3461–3467 (2019).
9. Dawid, A. P. & Skene, A. M. Maximum likelihood estimation of observer error-rates using the EM algorithm. *J. R. Stat. Soc. Series C (Appl. Stat.)* **28**, 20–28 (1979).
10. Hovy, D. *et al.* Learning Whom to Trust with MACE. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1120–1130 (Association for Computational Linguistics, 2013).
11. Whitehill, J. *et al.* Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems* **22** (2009).
12. Raykar, V. C. *et al.* Learning from crowds. *Journal of Machine Learning Research* **11**, 4 (2010).
13. Morales-Álvarez, P. *et al.* Scalable Variational Gaussian Processes for Crowdsourcing: Glitch Detection in LIGO. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 1534–1551 (2022).
14. López-Pérez, M. *et al.* Learning from crowds in digital pathology using scalable variational Gaussian processes. *Scientific Reports* **11**, 1–9 (2021).
15. Haußmann, M. *et al.* Variational Bayesian multiple instance learning with Gaussian processes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6570–6579 (IEEE, 2017).
16. Kandemir, M. Asymmetric transfer learning with deep Gaussian processes. In *International Conference on Machine Learning*, 730–738 (PMLR, 2015).
17. Esteban, Á. E. *et al.* A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes. *Computer Methods and Programs in Biomedicine* **178**, 303–317 (2019).
18. López-Pérez, M. *et al.* Deep Gaussian Processes for Classification With Multiple Noisy Annotators: Application to Breast Cancer Tissue Classification. *IEEE Access* **11**, 6922–6934 (2023).
19. del Amor, R. *et al.* Annotation protocol and crowdsourcing multiple instance learning classification of skin histological images: The CR-AI4SkIN dataset. *Artificial Intelligence in Medicine* **145**, 102686 (2023).
20. McHugh, M. L. Interrater reliability: The kappa statistic. *Biochemia Medica* **22**, 276–282 (2012).
21. del Amor, R. *et al.* A Fusocelular Skin Dataset with Whole Slide Images for Deep Learning Models. *Figshare* https://doi.org/10.6084/m9.figshare.27118035 (2024).
22. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representatations* (ICLR, 2015).
23. Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nature Medicine* **30**, 850–862 (2024).
24. Huang, Z. *et al.* A visual-language foundation model for pathology image analysis using medical twitter. *Nature Medicine* **29**, 2307–2316 (2023).
25. Lu, M. Y. *et al.* A visual-language foundation model for computational pathology. *Nature Medicine* **30**, 863–874 (2024).

Acknowledgements

This work has received funding from the Spanish Ministry of Economy and Competitiveness through project PID2019-105142RB-C21 and PID2019-105142RB-C22 (AI4SkIN) and PID2022-140189OB-C21 and PID2022-140189OB-C22 (ASSIST), and Generalitat Valenciana through project COM-TACTS2 (CIPROM/2022/20). The work of Rocío del Amor was supported by the Spanish Ministry of Universities (FPU20/05263). The work

of Pablo Meseguer was funded by valgrAI - Valencian Graduate School and Research Network of Artificial Intelligence. The work of Sandra Morales was co-funded by the Universitat Politécnica de Valéncia through the program PAID-10-20. The work of M. López-Pérez was supported by the grant JDC2022-048318-I funded by MICIU/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR.

Author contributions

Conceptualization: R.A., M.L.-P., P.M., S.M., L.T., J.A.-F., J.M., R.M. and V.N.; methodology: M.L.-P., P.M., and R.A.; investigation, formal analysis, writing, and visualization: R.A., M.L.-P., S.M. and P.M.; review and editing: S.M., J.M., R.M. and V.N.; supervision: R.M. and V.N. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to R.d.A. or V.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.