SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

A patient-centric dataset of images and metadata for identifying melanomas using clinical context

Veronica Rotemberg[1,15 ✉], Nicholas Kurtansky[1,15], Brigid Betz-Stablein[2], Liam Caffery[2], Emmanouil Chousakos[1,3], Noel Codella[4], Marc Combalia[5], Stephen Dusza[1], Pascale Guitera[6], David Gutman[7], Allan Halpern[1], Brian Helba[8], Harald Kittler[9], Kivanc Kose[1], Steve Langer[10], Konstantinos Lioprys[3], Josep Malvehy[5], Shenara Musthaq[1,11], Jabpani Nanda[1,12], Ofer Reiter[1,13], George Shih[14], Alexander Stratigos[3], Philipp Tschandl[9], Jochen Weber[1] & H. Peter Soyer[2]

Prior skin image datasets have not addressed patient-level information obtained from multiple skin lesions from the same patient. Though artificial intelligence classification algorithms have achieved expert-level performance in controlled studies examining single images, in practice dermatologists base their judgment holistically from multiple lesions on the same patient. The 2020 SIIM-ISIC Melanoma Classification challenge dataset described herein was constructed to address this discrepancy between prior challenges and clinical practice, providing for each image in the dataset an identifier allowing lesions from the same patient to be mapped to one another. This patient-level contextual information is frequently used by clinicians to diagnose melanoma and is especially useful in ruling out false positives in patients with many atypical nevi. The dataset represents 2,056 patients (20.8% with at least one melanoma, 79.2% with zero melanomas) from three continents with an average of 16 lesions per patient, consisting of 33,126 dermoscopic images and 584 (1.8%) histopathologically confirmed melanomas compared with benign melanoma mimickers.

Background & Summary

Artificial intelligence (AI) use in medical imaging is rapidly progressing and has the potential to reduce melanoma-associated mortality, morbidity, and healthcare costs by improving access to expertise, diagnostic accuracy, and screening efficiency[1–3]. Here we present a dermatology image dataset that includes patient- and lesion-related clinical context, which can be used in studies to examine whether this additional information further improves recognition performance.

Recent studies have demonstrated the ability of AI algorithms to match, if not outperform, clinicians in the diagnosis of individual skin lesion images in controlled reader studies. Algorithms derived from the 2018 ISIC Grand Challenge have been shown to outperform over 500 clinical readers and experts in such a reader study[1]. However, the reader study did not accurately reflect clinical scenarios where clinicians have access to examine all lesions on a patient.

[1]Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]The University of Queensland Diamantina Institute, The University of Queensland, Dermatology Research Centre, Brisbane, Australia. [3]University of Athens Medical School, Athens, Greece. [4]Microsoft, Redmond, WA, USA. [5]Melanoma Unit, Dermatology Department, Hospital Clínic Barcelona, Universitat de Barcelona, IDIBAPS, Barcelona, Spain. [6]Melanoma Institute Australia and Sydney Melanoma Diagnostic Center, Sydney, Australia. [7]Emory University School of Medicine, Department of Biomedical Informatics, Atlanta, GA, USA. [8]Kitware, Inc., Clifton Park, NY, USA. [9]Medical University of Vienna, Department of Dermatology, Vienna, Austria. [10]Division of Radiology Informatics, Department of Radiology, Mayo Clinic, Rochester, MN, USA. [11]SUNY Downstate Medical School, New York, NY, USA. [12]Stony Brook Medical School, Stony Brook, NY, USA. [13]Rabin Medical Center, Tel Aviv, Israel. [14]Department of Radiology, Weill Cornell Medical College, New York, NY, USA. [15]These authors contributed equally: Veronica Rotemberg, Nicholas Kurtansky. ✉e-mail: rotembev@mskcc.org
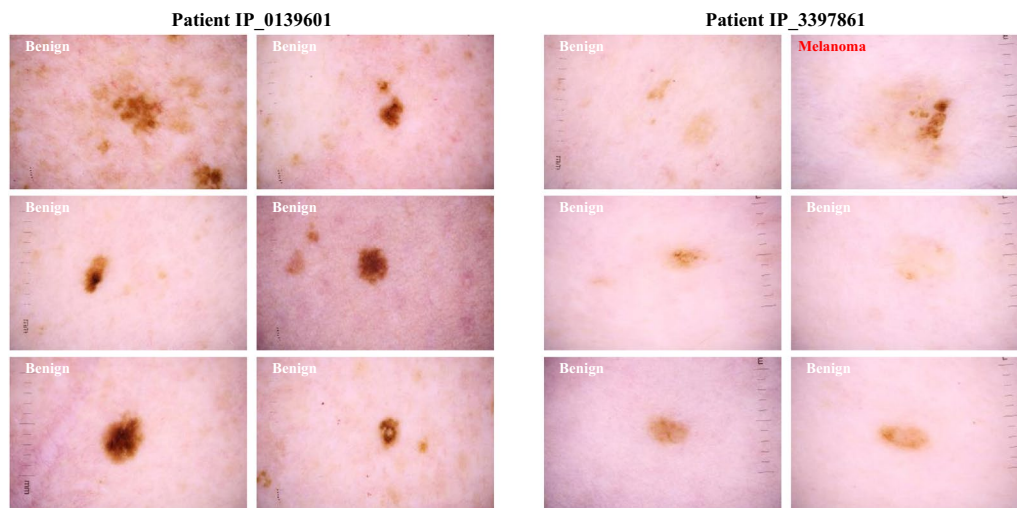
Fig. 1 Example of clinical context. An atypical lesion found on a patient with many atypical lesions is less suspicious for malignancy as opposed to an atypical lesion that is an outlier on the patient.

Clinicians frequently assess skin lesions for biopsy by assessing them in context with the rest of the lesions on a given patient's body, taking into consideration the individual "biologic skin ecosystem". As demonstrated in Fig. 1, a lesion with malignancy-predictive features among many similar lesions is thought not to be as dangerous as an odd lesion on a patient whose other lesions are more benign looking. The latter is known in dermatology as the "ugly duckling sign" and is frequently used to diagnose melanoma, especially in patients with multiple melanocytic lesions[4,5]. Until now, the ugly duckling concept has not been explored with machine learning due to the lack of large datasets with multiple labeled images per patient. Here, we present the first dataset of melanoma and comparative lesions from the same patient to support new machine learning challenges. This dataset is composed of 33126 images collected from 2056 patients at multiple centers around the world such as Memorial Sloan Kettering Cancer Center, New York; the Melanoma Institute Australia and the Melanoma Diagnosis Centre, Sydney; the University of Queensland, Brisbane; the Medical University of Vienna, Vienna; and Hospital Clínic de Barcelona, Barcelona. In this article, we present the methods by which we created this multicenter dataset with clinical contextual information.

Methods

General. We queried clinical imaging databases across the six centers to generate a multicenter imaging dataset. Among patients with dermoscopy imaging from 1998 to 2020, those with multiple skin lesions were identified. Histopathology reports corresponding to internal biopsied lesions were reviewed for diagnosis labelling. Non-biopsied lesions that were monitored for at least six months were considered benign without further granularity[6]. Patients with appropriate qualifying diagnoses: melanoma or benign lesions that could be considered melanoma mimickers including nevi, atypical melanocytic proliferation, café-au-lait macule, lentigo NOS, lentigo simplex, solar lentigo, lichenoid keratosis, and seborrheic keratosis were included[7–9]. Lesions satisfying the described criteria were represented in the dataset with a single dermoscopic image[8,10,11]. These include images captured with or without polarized light using a contact or noncontact dermatoscope. When multiple image types were available, the selected image was either the one of highest resolution or if multiple images at the same resolution were available, one was chosen randomly. Images containing any potentially identifying features, such as jewelry or tattoos, or from patients without at least three qualifying images were excluded during quality assurance review.

In order to test algorithm generalizability, a subset of images were allocated for the testing dataset of the 2020 ISIC Grand Challenge[12]. These images included contributions from all sites included in the training set and a held out set of cases from the Andreas Syngros Hospital of Cutaneous & Venereal Diseases, Athens, Greece. These test images are available for download, but the test labels are not yet public due to planned future challenges and experiments.

Quality assurance. A software annotation tool, called 'Tagger,' was developed internally to review diagnostic labeling of grouped images (https://github.com/dgutman/webix_image_organizer). Using this tool, dermoscopy expert reviewers (EC, OR) were presented sets of 30 images with a shared diagnosis in order to identify the ones with erroneous labeling. Reviewers invested 22 hours over three weeks of quality assurance in 'Tagger' and spent an average of 4 seconds per set when flagging a single image, and 11 seconds per set when flagging several images. Out of all images reviewed in Tagger, 2.7% were removed, out of concern for erroneous labels.

Memorial Sloan Kettering Cancer Center. The MSK Dermatology Service is a high-risk clinic that relies heavily on imaging for high risk individuals with or without a history of melanoma[13]. Images were acquired using a dermoscopic attachment to either a digital single reflex lens (SLR) camera or to a smartphone. Each lesion was imaged

with polarized and/or nonpolarized dermoscopy. For each lesion, 3–5 images are collected during each patient visit and stored in a specialized image database called Vectra™ (Canfield Scientific Inc., Parsippany, NJ, USA).

Images were extracted after searching the database for patients with multiple lesions imaged and who had biopsy confirmed melanoma from 2015–2019. The clearest image per time point was selected by medical student research fellows using a selection tool designed uniquely for the task (SM, JN, OR, EC) (https://github.com/ISIC-Research/lesionimagepicker). Images were collected and shared with institutional review board approval number 16–974.

Hospital Clínic Barcelona. The Department of Dermatology of the Hospital Clínic of Barcelona is a tertiary referral center for melanoma patients, includes a high-risk melanoma patient clinic. The dermatology department is equipped with the digital dermatoscopy system MoleMax™ HD (Derma Medical Systems, Vienna, Austria) and corresponding image database to store the collected images. Each lesion was photographed using polarized dermoscopy.

Candidate images were extracted after searching the database for benign lesions with >1.5 years of digital dermoscopy follow-up and excised lesions with a histopathology report between the years 1998 and 2020. The images were examined by three expert dermatologists at the clinic for image quality assurance and label accuracy. From a series of multiple sequential images of the same nevus, we extracted the median timepoint. Images were collected and shared with institutional ethics approval number HCP/2019/0413.

The University of Queensland. The Clinical Research Facility of the Translational Research Institute in Brisbane, Queensland, Australia is the clinical trial site following both general population and high-risk individuals participating in studies carried out by the Dermatology Research Center of The University of Queensland Diamantina Institute. Contributed images came from three prospective longitudinal studies. The first study, "Changing Naevi Study", consisted of two groups of participants; advanced stage (III – IV) melanoma patients undergoing treatment with immunotherapy and/or targeted therapy and people at high risk of developing melanoma due to personal or family history but were not undergoing treatment at time of enrollment. Ethics approval was obtained from the Human Research Ethics Committees of Metro South Health (HREC/16/QPAH/37) and The University of Queensland (2016000429). The second study, "Mind Your Moles", consisted of participants from a general population cohort recruited from the Brisbane Electoral Role[14]. All nevi >5 mm were imaged in these two studies, as well as any lesions of interest/concern to the participant or clinician. This study has been approved by the Metro South Health Human Research Ethics Committee on April 21, 2016 (approval number: HREC/16/QPAH/125). Ethics approval has also been obtained from the University of Queensland Human Research Ethics Committee (approval number: 2016000554), Queensland University of Technology Human Research Ethics Committee (approval number: 1600000515), and QIMR Berghofer (approval number: P2271). The third study, "Evaluation of the Efficacy of 3D Total-Body Photography With Sequential Digital Dermoscopy in a High-Risk Melanoma Cohort", consisted of participants at high risk of melanoma[15], half of which underwent imaging intervention[16]. Lesions of interest to the participant or clinician were imaged dermoscopically. This study has received Human Research Ethics Committee (HREC) approval from both Metro South Health HREC (HREC/17/QPAH/816) and The University of Queensland HREC (2018000074).

All images used for the studies were extracted from the Vectra™ image database and were captured between the years 2016 and 2020 (Canfield Scientific Inc., Parsippany, NJ, USA).

Medical University Vienna. The Early Recognition Unit of the Department of Dermatology of Medical University of Vienna is a tertiary referral center for high-risk patients. It offers total digital dermatoscopic follow-up to patients with multiple nevi[17]. Most patients in the program are of European descent with fair skin types (usually skin type 1–3) and have a high number of nevi and a personal or family history of melanoma.

We extracted polarized dermoscopic images from 2015–2019 which were stored in the MoleMax HD System (Derma Medical Systems, Vienna, Austria). We searched the database of this system for patients with at least 3 dermoscopic images by filtering SQL-tables with a proprietary tool provided by the manufacturer. From these patients we selected all benign melanocytic lesions with >1 year follow-up and all lesions that were excised. Histopathology reports were matched manually to all excised lesions. Non-melanocytic lesions, duplicate images, images captured before 2015 with older systems, low-quality images, and images that depicted only parts of the lesion were excluded. Furthermore, we excluded images of lesions that were already included in the 2018 or 2019 ISIC challenges. Images were collected and shared with institutional ethics approval number 1804/2017.

Melanoma Institute Australia and the Sydney Melanoma Diagnosis Centre. Both services are high-risk, tertiary referral dermatology clinics that rely heavily on imaging of individuals with or without a history of melanoma[13]. Lesions imaged for short term monitoring are selected at the discretion of the clinician or which are of concern to the patient. Additionally, all lesions are imaged prior to surgical removal. Images are acquired using a dermoscopic attachment to either a digital single reflex lens (SLR) camera or to a smartphone and stored in DermEngine™ (Metaoptima, Vancouver, British Columbia, Canada). Histopathology reports between 2008 and 2019 were reviewed and lesions followed for six months or more without malignant changes were considered benign. All images were manually reviewed to assure de-identification and image detail quality after database extraction. Images were collected and shared with institutional ethics approval number X20-0241 & 2020/ETH01411: Melanoma Image Annotation and Analysis Collaboration project.

Dataset compilation. The quality assurance and collection steps we performed for curating the images from various sources are detailed in Fig. 2.

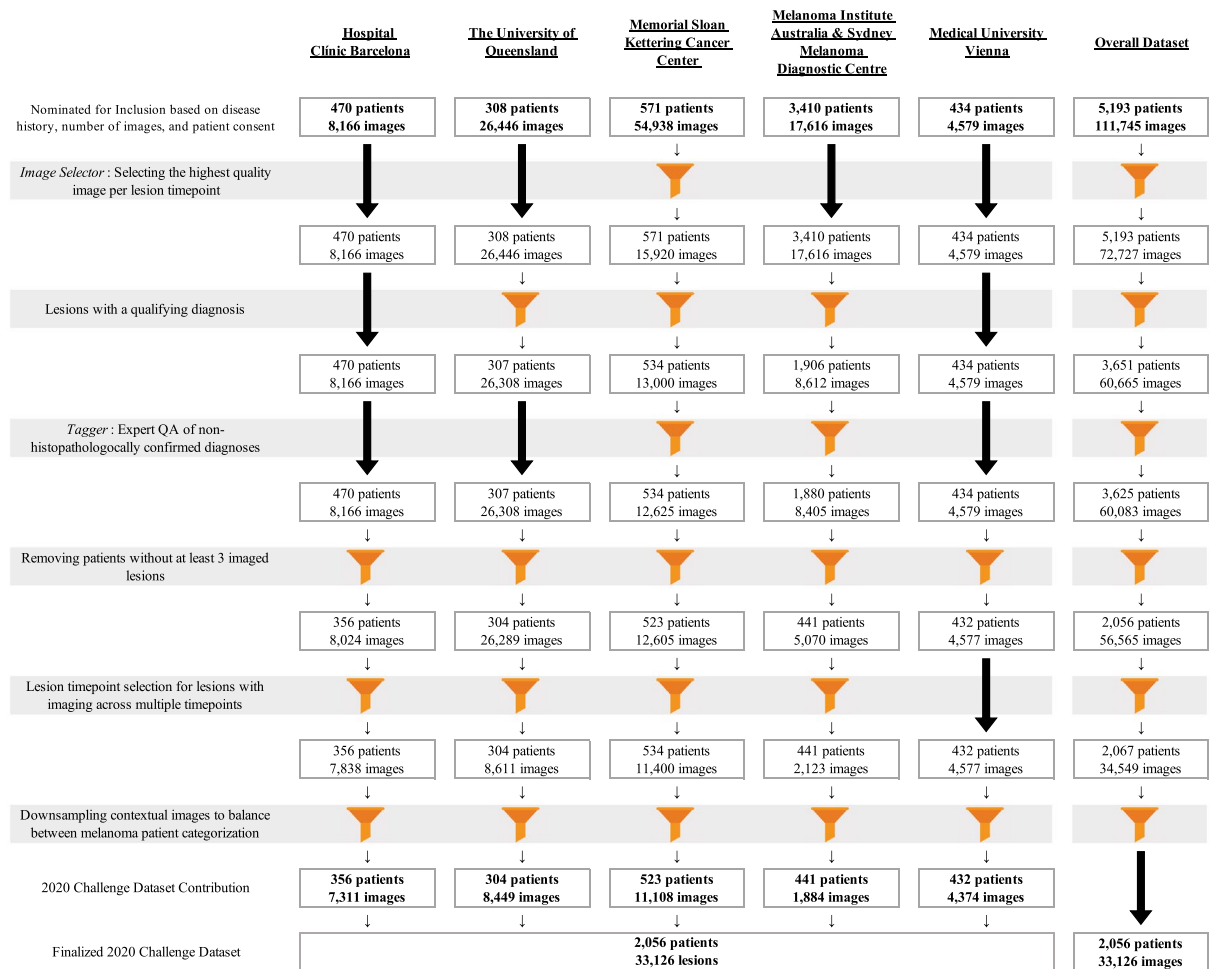| Process step	Hospital Clínic Barcelona	The University of Queensland	Memorial Sloan Kettering Cancer Center	Melanoma Institute Australia & Sydney Melanoma Diagnostic Centre	Medical University Vienna	Overall Dataset
Nominated for Inclusion based on disease history, number of images, and patient consent	470 patients / 8,166 images	308 patients / 26,446 images	571 patients / 54,938 images	3,410 patients / 17,616 images	434 patients / 4,579 images	5,193 patients / 111,745 images
Image Selector: Selecting the highest quality image per lesion timepoint	470 patients / 8,166 images	308 patients / 26,446 images	571 patients / 15,920 images	3,410 patients / 17,616 images	434 patients / 4,579 images	5,193 patients / 72,727 images
Lesions with a qualifying diagnosis	470 patients / 8,166 images	307 patients / 26,308 images	534 patients / 13,000 images	1,906 patients / 8,612 images	434 patients / 4,579 images	3,651 patients / 60,665 images
Tagger: Expert QA of non-histopathologocally confirmed diagnoses	470 patients / 8,166 images	307 patients / 26,308 images	534 patients / 12,625 images	1,880 patients / 8,405 images	434 patients / 4,579 images	3,625 patients / 60,083 images
Removing patients without at least 3 imaged lesions	356 patients / 8,024 images	304 patients / 26,289 images	523 patients / 12,605 images	441 patients / 5,070 images	432 patients / 4,577 images	2,056 patients / 56,565 images
Lesion timepoint selection for lesions with imaging across multiple timepoints	356 patients / 7,838 images	304 patients / 8,611 images	534 patients / 11,400 images	441 patients / 2,123 images	432 patients / 4,577 images	2,067 patients / 34,549 images
Downsampling contextual images to balance between melanoma patient categorization / 2020 Challenge Dataset Contribution	356 patients / 7,311 images	304 patients / 8,449 images	523 patients / 11,108 images	441 patients / 1,884 images	432 patients / 4,374 images	
Finalized 2020 Challenge Dataset			2,056 patients / 33,126 lesions			2,056 patients / 33,126 images

Fig. 2 Ingestion pipeline. The process of selecting images from each center and incorporating into a cohesive dataset emphasizing the importance of contextual lesions. Patients without three or more contextual lesions of appropriate diagnosis and image quality were removed.

Lesion timepoints. Each lesion in the dataset is represented by a single image. The image of non-biopsied benign lesions with imaging at multiple time points were selected to minimize the difference in patient imaging date variability and date range between patients with and without an imaged melanoma. This was performed to reduce potential bias in image lighting, camera type, or other factors between the benign and melanoma patient class.

Lesion context images. Due to the retrospective nature of image acquisition and potential surveillance bias in different patient populations, the number of lesions per patient was not distributed identically between the class of patients with a melanoma image and the class without a melanoma image. Because the lesions in this dataset do not represent all lesions that exist on this set of patients, it is possible the imbalance is related to selection bias of imaged lesions. Lesions in both classes were subsampled through patient matching, which led to a loss of 4.1% of images. Ultimately, 50% of the patients have more than 10 contextual lesions. The matched number of images per patient ID before and after subsampling is shown in Fig. 3.

Duplicates. Due to a clerical error during the data ingestion process to the ISIC Archive, 425 pixelwise identical duplicate images were ingested and included in the dataset. The duplicates are included in the data to mirror the dataset used for the 2020 SIIM-ISIC Melanoma Classification competition. In order to preserve fidelity with the dataset that was used in the competition, the dataset itself has not been modified; however, lesion identifiers were made available as a metadata field and a comma-separated value file is available at the dataset landing page which highlights the image identifiers of the duplicates.
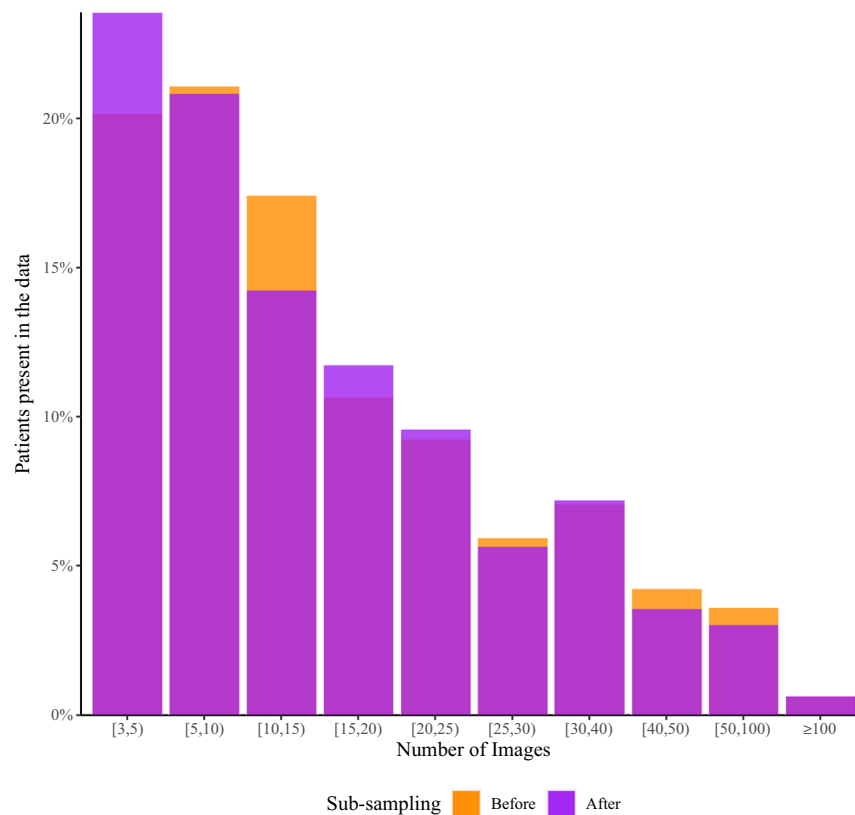
Fig. 3 Distribution of image count per patient. Orange bars represent the original number of lesions present in the dataset per patient before sub-sampling to balance on patient disease class. Transparent purple bars represent the final distribution of contextual lesions present in the dataset. While it may appear that some patients acquired contextual images, multi-colored bars represent shifts between bins (patients moving leftward as contextual images were removed).

Data Records

The dataset was made available for download through the Kaggle platform as part of a live competition from May 27, 2020 through August 20, 2020. It is released under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license, and is permanently accessible to the public through the ISIC Archive[12] at this https://doi.org/10.34970/2020-ds01. Currently no modifications have been made to the dataset, however, any metadata or image modifications will be noted at that DOI landing page.

Training images consisted of 12,743,090 pixels on average but ranged from 307,200 to 24,000,000. Metadata for each image included approximate patient age at time of image capture, biological sex, general anatomic site of the lesion, anonymized patient identification number, benign/malignant category, and the specific diagnosis if one was available based on an acceptable ground truth confirmation method. A summary of the characteristics of the dataset at patient- and lesion-level is shown in Table 1.

This dataset is adjoined by a test which determined the 2020 ISIC Grand Challenge leaderboard scoring. Test images and associated metadata are available for download through the ISIC Archive at the above listed DOI, though diagnostic labels remain undisclosed at this time until further notice to serve as the basis for scoring future competitions.

Dataset format. The dataset is available in two formats.

The first is the file format described in Part 10 of Digital Imaging and Communication in Medicine (DICOM) standard[18,19], which is currently being developed for dermatology. The DICOM standard is a comprehensive, international medical image standard that was originally developed for radiology, where it has become ubiquitous as the core standard. It has since been adopted by many other medical imaging specialties including ophthalmology, dentistry, cardiology, nuclear medicine, pathology, surgical specialties who perform image-guided surgery (e.g., neurosurgery, ENT, orthopedics), and specialties that acquire endoscopic or laparoscopic imaging[20]. The DICOM file format is an amalgamation of the metadata and pixel data in a single file. The pixel data is encoded in Joint Photographic Expert Group (JPEG) format. While imaging technology changed over the time period from which images were selected and continue to change over time, the devices were not recorded upon image capture, and perhaps in the future when the DICOM standard is more widely adopted this may be metadata that could be collected and provided.

	Patient Class		
	≥1 Melanoma*	0 Melanoma*	Total**
Patients (n)	428 (21%)	1628 (79%)	2056 (100%)
Male	260 (24%)	817 (76%)	1077 (52%)
Female	168 (17%)	809 (83%)	977 (48%)
Unknown	0 (0%)	2 (100%)	2 (0%)
Average Age (years)	57.7	49.6	51.3
Lesions (n)	6927 (21%)	26199 (79%)	33126 (100%)
Benign (Not Biopsied)	5862 (19%)	25256 (81%)	31118 (94%)
Benign (Biopsied)	481 (34%)	943 (66%)	1424 (4%)
Melanoma	584 (100%)	0 (0%)	584 (2%)
Lesions per Patient			
Mean	16.2	16.1	16.1
Q1	5	5	5
Median	12	12	12
Q3	22	22	22
Timepoints per Patient			
Mean	5.9	3.4	3.9
Q1	3	1	1
Median	4	2	3
Q3	7	4	5
	Lesion Diagnosis		
	Melanoma*	Benign*	Total**
Lesions per Patient Class (n)			
≥1 Melanoma	584 (8%)	6343 (92%)	6927 (21%)
0 Melanoma	0 (0%)	26199 (100%)	26199 (79%)
Lesions per Anatomic Site (n)			
Head/Neck	74 (4%)	1781 (96%)	1855 (6%)
Torso	257 (2%)	16588 (98%)	16845 (51%)
Upper Extremity	111 (2%)	4872 (98%)	4983 (15%)
Lower Extremity	124 (1%)	8293 (99%)	8417 (25%)
Palms/Soles	5 (1%)	370 (99%)	375 (1%)
Oral Genital	4 (3%)	120 (97%)	124 (0%)
Unknown	9 (2%)	518 (98%)	527 (2%)

Table 1. Summary of combined dataset with row (*) and column (**) percentages. Patient- and lesion-level characteristics of the dataset.

The second format is where the images are in JPEG format and the metadata is included in a linked comma-separated values (CSV) file.

Technical Validation

The ground truth labels for all malignant lesions in the dataset were confirmed via retrospective review of histopathology reports, and diagnosis plausibility was visually confirmed by visual confirmation of a dermoscopy expert. Histopathology reports were double checked if the label was suspicious. Melanoma *in situ* and invasive melanoma were both coded as melanoma. All other qualifying images were coded as benign, including those diagnosed as severely dysplastic nevi[21,22].

Non-biopsied lesions with expert consensus agreement and lesions followed for six months or more without malignant changes were labelled benign without a more specific diagnosis by most contributors. Dermatofibromas, seborrheic keratosis, or vascular lesions were not monitored, as that would not reflect clinical practice, but labels were verified visually by an expert in dermoscopy. Images of lesions were attributed to patients based on the clinical imaging database identification codes which are stored at the time of capture during each clinical photography session.

Usage Notes

This dataset mimics clinical practice by labeling images from the same patient (mean = 16, median = 12, standard deviation = 16) as such and allows algorithms to assess multiple images from the same patient for malignancy. It addresses a particularly challenging area of clinical practice, those patients with multiple atypical nevi suspicious for malignancy. The dataset is designed to improve translational potential of algorithms, especially to help clinicians without access to tertiary referral centers assess high risk patients with multiple atypical nevi. Additionally, algorithms developed using this dataset may be better candidates for incorporating into dermatology imaging systems, as they can evaluate all images for a given patient in context, and perhaps even be used during clinic

visits in which multiple lesions are imaged. Given the translational potential of algorithms developed using this dataset, we hope that generating a public, well annotated dataset that mimics clinical practice will lead to prospective studies of promising automated approaches for diagnosing melanoma.

Various forms of dermoscopy imaging are included in the dataset: contact non-polarized light, contact polarized light, and non-contact polarized light. Deeper skin structures are more often visible under polarized light than non-polarized light, even without direct skin contact with the interface or the use of a liquid interface[23]. Various colors, structures, and patterns are more pronounced, or accessible, under specific forms of dermoscopy[24]. Imaging modalities are not equivalent in identifying certain morphologies but complement one another in a holistic clinical assessment. A limitation to this dataset is that each lesion is represented by a single image and type of dermoscopy, which may not reflect the full spectrum of information that would be used by a clinician[20].

Generalization of AI-assisted skin lesion classification to broad clinical use depends on the demographic agreement of the training dataset to the clinical population. Due to low population prevalence and challenges with access to care in different populations, the images gathered for large datasets such as this for AI classification have a strong tendency to under-represent darker skin types. This may lead to either overdiagnosis or underdiagnosis of melanomas in darker skin types, both of which would have significant clinical implications and will require prospective study. The ISIC Archive is actively pursuing methods by which to increase the diversity of images obtained, but at this point caution should be used when attempting to generalize algorithms trained on images from specialized referral centers (such as the dataset described herein) to the global population at large. The dataset is also enriched for melanoma in general and does not represent true incidence of melanoma.

Code availability

Custom generated code for the described methods is available at https://github.com/ISIC-Research/2020-Challenge-Curation.

References

1. Tschandl, P., Argenziano, G., Razmara, M. & Yap, J. Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features. *Br J Dermatol.* **181**, 155–165 (2019).
2. Tschandl, P. *et al*. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol.* **20**, 938–947 (2019).
3. Tschandl, P. *et al*. Human–computer collaboration for skin cancer recognition. *Nat Med.* **26**, 1229–1234 (2020).
4. Gaudy-Marqueste, C. *et al*. Ugly Duckling Sign as a Major Factor of Efficiency in Melanoma Detection. *JAMA Dermatol.* **153**, 279–284 (2017).
5. Scope, A. *et al*. The "Ugly Duckling" Sign: Agreement Between Observers. *Arch Dermatol.* **144**, 58–64 (2008).
6. Moscarella, E. *et al*. Both short-term and long-term dermoscopy monitoring is useful in detecting melanoma in patients with multiple atypical nevi. *J Eur Acad Dermatol Venereol.* **31**, 247–251 (2017).
7. Carrera, C. *et al*. Dermoscopic Clues for Diagnosing Melanomas That Resemble Seborrheic Keratosis. *JAMA Dermatol.* **153**, 544–551 (2017).
8. Argenziano, G. *et al*. Early diagnosis of melanoma: what is the impact of dermoscopy? *Dermatol Ther.* **25**, 403–409 (2012).
9. Kaminska-Winciorek, G. & Wydmański, J. Benign simulators of melanoma on dermoscopy – black colour does not always indicate melanoma. *J Pre-Clin Clin Res.* **7**, 6–12 (2013).
10. Argenziano, G. *et al*. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. *J Am Acad Dermatol.* **48**, 679–693 (2003).
11. 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.* **137**, 1343–1350 (2001).
12. International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset. *International Skin Imaging Collaboration* https://doi.org/10.34970/2020-ds01 (2020).
13. Halpern, A. C., Marchetti, M. A. & Marghoob, A. A. Melanoma Surveillance in "High-Risk" Individuals. *JAMA Dermatol.* **150**, 815–816 (2014).
14. Koh, U. *et al*. 'Mind your Moles' study: protocol of a prospective cohort study of melanocytic naevi. *BMJ Open.* **8** (2018).
15. Rastrelli, M., Tropea, S., Rossi, C. R. & Alaibac, M. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo.* **28**, 1005–1011 (2014).
16. Primiero, C. A. *et al*. Evaluation of the efficacy of 3D total-body photography with sequential digital dermoscopy in a high-risk melanoma cohort: protocol for a randomised controlled trial. *BMJ Open.* **9** (2019).
17. Rinner, C., Tschandl, P., Sinz, C. & Kittler, H. Long-term evaluation of the efficacy of digital dermatoscopy monitoring at a tertiary referral center. *J Dtsch Dermatol Ges.* **15**, 517–522 (2017).
18. National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM) Supplement 221: Dermoscopy https://www.dicomstandard.org/News/current/docs/sups/sup221.pdf (2019).
19. National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM) Standard PS3.10 2018d - Media Storage and File Format for Media Interchange. http://dicom.nema.org/medical/dicom/2018d/output/html/part10.html (2018).
20. Caffery, L. J. *et al*. Transforming Dermatologic Imaging for the Digital Era: Metadata and Standards. *J Digit Imaging.* **31**, 568–577 (2018).
21. Lott, J. P. *et al*. Evaluation of the Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx) classification scheme for diagnosis of cutaneous melanocytic neoplasms: Results from the International Melanoma Pathology Study Group. *J Am Acad Dermatol.* **75**, 356–363 (2016).
22. Piepkorn, M. W. *et al*. The MPATH-Dx reporting schema for melanocytic proliferations and melanoma. *J Am Acad Dermatol.* **70**, 131–141 (2014).
23. Marghoob, A. A. *et al*. Instruments and new technologies for the *in vivo* diagnosis of melanoma. *J Am Acad Dermatol.* **49**, 777–797; quiz 798–799 (2003).
24. Benvenuto-Andrade, C. *et al*. Differences Between Polarized Light Dermoscopy and Immersion Contact Dermoscopy for the Evaluation of Skin Lesions. *Arch Dermatol.* **143**, 329–338 (2007).

Author contributions

All authors have provided critical feedback during revision process and actively participated in preparation of the manuscript. V.R. oversaw dataset design, performed quality review and co-wrote the manuscript with N.K., N.K. collected cases, collated the dataset, and performed data-handling. J.W. and D.G. developed software applications for quality review of images. M.C., K.L., A.S., P.T., B.B.S., P.G. and H.K. collected cases and extracted images from their institutional databases. L.C., N.C., S.D., A.H., K.K., S.L., J.M. and H.P.S. contributed to dataset design. B.H. supported image ingestion and database hosting. E.C., S.M., J.N. and O.R. performed quality review of images.

Competing interests

H.P.S. is a shareholder of MoleMap N.Z. Limited and e-derm consult GmbH, and undertakes regular teledermatological reporting for both companies. H.P.S. is a Medical Consultant for Canfield Scientific Inc., Revenio Research Oy and also a Medical Advisor for First Derm. AH serves as a consultant to Canfield Scientific Inc., and is on the SciBase advisory panel.

Additional information

Correspondence and requests for materials should be addressed to V.R.

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.