scientific reports

OPEN

Lesion identification and malignancy prediction from clinical dermatological images

Meng Xia[1]✉, Meenal K. Kheterpal[2], Samantha C. Wong[3], Christine Park[3], William Ratliff[4], Lawrence Carin[1] & Ricardo Henao[1]

We consider machine-learning-based lesion identification and malignancy prediction from clinical dermatological images, which can be indistinctly acquired via smartphone or dermoscopy capture. Additionally, we do not assume that images contain single lesions, thus the framework supports both focal or wide-field images. Specifically, we propose a two-stage approach in which we first identify all lesions present in the image regardless of sub-type or likelihood of malignancy, then it estimates their likelihood of malignancy, and through aggregation, it also generates an image-level likelihood of malignancy that can be used for high-level screening processes. Further, we consider augmenting the proposed approach with clinical covariates (from electronic health records) and publicly available data (the ISIC dataset). Comprehensive experiments validated on an independent test dataset demonstrate that (1) the proposed approach outperforms alternative model architectures; (2) the model based on images outperforms a pure clinical model by a large margin, and the combination of images and clinical data does not significantly improves over the image-only model; and (3) the proposed framework offers comparable performance in terms of malignancy classification relative to three board certified dermatologists with different levels of experience.

Prior to the COVID-19 pandemic, access to dermatology care was challenging due to limited supply and increasing demand. According to a survey study of dermatologists[1], the mean ± standard deviation (SD) waiting time was 33±32 days, 64% of the appointments exceeded the criterion cutoff of 3 weeks and 63% of the appointments exceeded 2-week criterion cutoff for established patients. During the COVID-19 pandemic, the number of dermatology consultations were reduced by 80-90% to urgent issues only, leading to delay in care of dermatologic concerns. Moreover, the issue of access is very significant for the growing Medicare population, expected to account for 1 in 5 patients by 2030[2], due to a higher incidence of skin cancer.

Access issues in dermatology are concerning as there has been an increasing incidence of skin cancers, particularly a threefold increase in melanoma over the last 40 years[3]. Many of the skin lesions of concern are screened by primary care physicians (PCPs). In fact, up to one third of primary care visits contend with at least one skin problem, and skin tumors are the most common reason for referral to dermatology[4]. High volume of referrals places a strain on specialty care, delaying visits for high-risk cases. Given the expected rise in baby boomers, with significantly increased risk of skin cancer, there is an urgent need to equip primary care providers to help screen and risk stratify patients in real time, high quality and cost-conscious fashion. PCPs have variable experience and training in dermatology, causing often low concordance between their evaluation and dermatology[4]. A consistent clinical decision support (CDS) system has the potential to mitigate this variability, and to create a powerful risk stratification tool, leveraging the frontline network of providers to enhance access to quality and valuable care. In addition, such a tool can aid tele-dermatology workflows that have emerged during the global pandemic.

Over the last decade, several studies in the field of dermatology have demonstrated the promise of deep learning models such as convolutional neural networks (CNN) in terms of classification of skin lesions[5,6], with dermoscopy-based machine learning (ML) algorithms reaching sensitivities and specificities for melanoma diagnosis at 87.6% (95% CI 72.72–100.0) and 83.5% (95% CI 60.92–100.0), respectively, by meta-analysis[7]. Several authors have reported superior performance of ML algorithms for classification of squamous cell carcinoma (SCC) and basal cell carcinomas (BCC), with larger datasets improving performance[5,8].

From a machine-learning methods perspective, a common approach for classification with dermoscopy images consists on refining pre-trained CNN architectures such as VGG16 as in[9] or AlexNet after image pre-processing,

[1]Department of Electrical and Computer Engineering, Duke University, Durham, USA. [2]Department of Dermatology, Duke University, Durham, USA. [3]School of Medicine, Duke University, Durham, NC, USA. [4]Duke Institute for Health Innovation, Duke University, Durham, NC, USA. ✉email: meng.xia@duke.edu

https://doi.org/10.1038/s41598-022-20168-w

e.g., background removal[10]. Alternatively, some approaches consider lesion sub-types independently[11], sonified images[12], or by combining clinical data with images to increase the information available to the model for prediction[13]. However, dermoscopy images are generally of good quality, high resolution and minimal background noise, making them less challenging to recognize compared to clinical, wide-field, images.

Beyond dermoscopy images, similar refinement approaches have been proposed based on architectures such as ResNet152[8,14], with additional pre-processing (illumination correction)[15], by using detection models to account for the non-informative background[16,17], or by first extracting features with CNN-based models, e.g., Inception v2, to then perform feature classification with other machine learning methods[12]. Moreover, comparative studies[6,18] have shown that models based on deep learning architectures can perform similarly to dermatologists on various classification tasks.

However, these ML algorithms are often developed with curated image datasets containing high quality clinical and dermoscopy photographs with limited skin variability, i.e., majority Caucasian or Asian sets in the ISIC dataset (dermoscopy), Asan dataset, Hallym dataset, MED-NODE, Edinburgh dataset[8]. The use of such algorithms trained on images often acquired from high quality cameras and/or dermatoscopes may be limited to specialty healthcare facilities and research settings, with questionable transmissibility in resource-limited settings and the primary care, thus creating a gap between healthcare providers and patients. Smartphone-based imaging is a promising image capture platform for bridging this gap and offering several advantages including portability, cost-effectiveness and connectivity to electronic medical records for secure image transfer and storage. To democratize screening and triage in primary care setting, an ideal ML-based CDS tool should be trained, validated and tested on smartphone-acquired clinical and dermoscopy images, representative of the clinical setting and patient populations for the greatest usability and validity.

While there are challenges to consumer grade smartphone image quality such as variability in angles, lighting, distance from lesion of interest and blurriness, they show promise to improve clinical workflows. Herein, we propose a two-stage approach to detect skin lesions of interest in wide-field images taken from consumer grade smartphone devices, followed by binary lesion classification into two groups: Malignant vs. Benign, for all skin cancers (melanoma, basal cell carcinoma and squamous cell carcinoma) and most common benign tumors. Ground truth malignancy was ascertained via biopsy, as apposed to consensus adjudication. As a result, the proposed approach can be integrated and generalized into primary care and dermatology clinical workflows. Importantly, our work also differs from existing approaches in that our framework can detect lesions from both wide-field clinical and dermoscopy images acquired with smartphones.

Method

In this section, we will first explain how we formulate the problem and introduce the model details. Then we describe the datasets we used in this study. Finally, we present the evaluation metrics used for each task.

Problem formulation.

We represent a set of annotated images as $\mathscr{D} = \{X_n, Z_n, U_n, y_n\}_{n=1}^{N}$, where N is the number of instances in the dataset, $X_n \in \mathbb{R}^{h\times w\times 3}$ denotes a color (RBG) image of size $w \times h$ (width × height) pixels, Z_n is a non-empty set of annotations $Z_n = \{z_{n1}, \ldots, z_{nm_n}\}$, with elements z_{ni} corresponding to the ith region of interest (ROI) represented as a bounding box with coordinates $(x_{ni}, y_{ni}, w_{ni}, h_{ni})$ (horizontal center, vertical center, width, height) and ROI labels $U_n = \{u_{n1}, \ldots, u_{nm_n}\}$, where m_n is the number of ROIs in image X_n. Further, $y_n \in \{0, 1\}$ is used to indicate the global image label.

In our specific use case, the images in $\mathscr{D}$ are a combination of smartphone-acquired wide-field and dermoscopy images with ROIs of 8 different biopsy-confirmed lesion types (ROI labels): Melanoma, Melanocytic Nevus, Basal Cell Carcinoma, Actinic Keratosis/Bowen's Disease, Benign Keratosis, Dermatofibroma, Vascular Lesions and Other Benign lesions. The location of different lesions was obtained by manual annotation as described below in the "Dataset" section. For malignancy prediction, the set of malignant lesions denoted as $\mathscr{M}$ is defined as Melanoma, Basal Cell Carcinoma, and Actinic Keratosis/Bowen's Disease/Squamous cell carcinoma while the set of benign lesions contains all the other lesion types. For the global image label y_n, a whole image (smartphone or dermoscopy) is deemed as malignant if at least one of its ROI labels are in the malignant set, $\mathscr{M}$.

Below, we introduce deep-learning-based models for lesion identification, malignancy prediction and image-level classification for end-to-end processing. An illustration of the two-step lesion identification and malignancy prediction framework is presented in Fig. 1.

Malignancy prediction. Assuming we know the position of the ROIs, i.e., $\{X_n, Z_n\}_{n=1}^{N}$ are always available, the problem of predicting whether a lesion is malignant can be formulated as a binary classification task. Specifically, we specify a function $f_\theta(\cdot)$ parameterized by θ whose output is the probability that a single lesion is consistent with a malignancy pathohistological finding in the area, i.e.,

$$p(u_{ni} \in \mathscr{M} | X_n, z_{ni}) = f_\theta(X_n, z_{ni}) \tag{1}$$

where $f_\theta(\cdot)$ is a convolutional neural network that takes the region of X_n defined in z_{ni} as input. In practice, we use a ResNet-50 architecture[19] with additional details described in the "Model details" section.

Lesion identification. Above we assume that the location (ROI) of the lesions is known, which may be the case in dermoscopy images as illustrated in Fig. 1. However, in general, wide-field dermatology images are likely to contain multiple lesions, while their locations are not known or recorded as part of clinical practice. Fortunately, if lesion locations are available for a set of images (via manual annotation), the task can be formulated as a super-

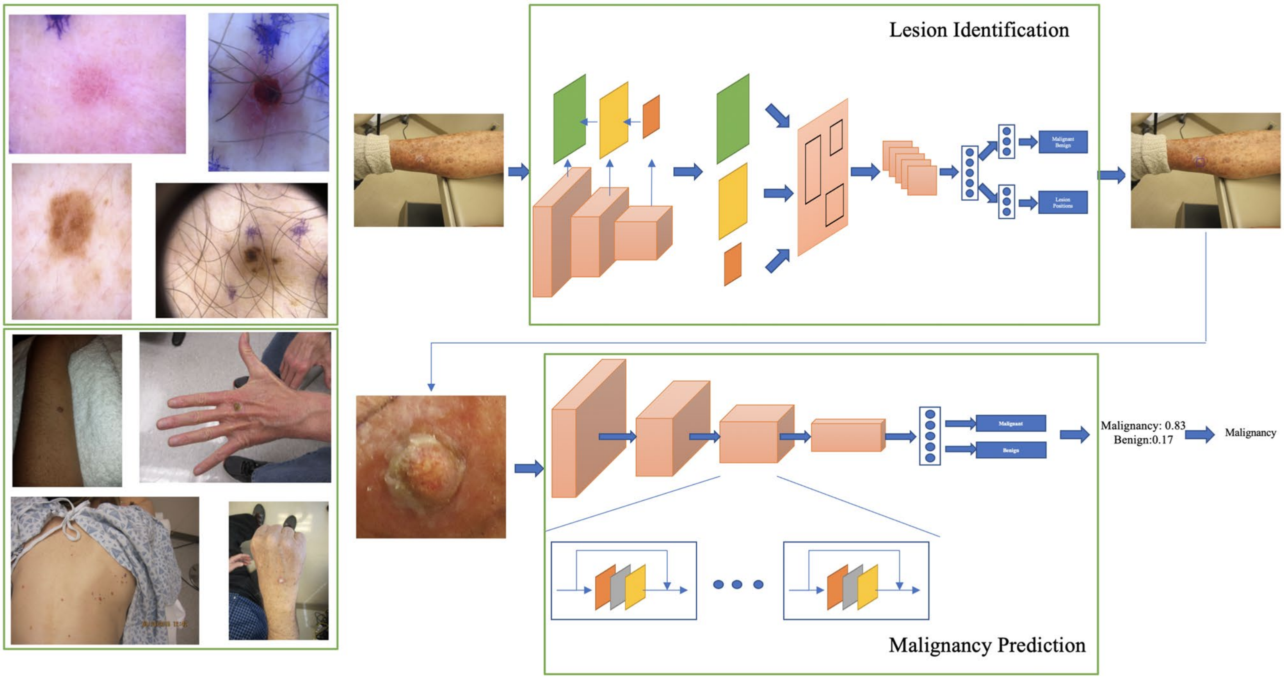

Figure 1. Two-stage lesion identification and malignancy prediction Framework. Top left: Examples of dermoscopy images. Bottom left: Examples of wide-field images. Top right: The lesion identification model estimates lesion locations (bounding boxes) from whole images (dermoscopy or wide-field) via a faster-RCNN architecture (see "Lesion identification" section). Bottom right: The malignancy prediction model specified via a ResNet-50 architecture predicts the likelihood that a lesion is malignant (see "Malignancy prediction" section). The lesions identified by the lesion identification model are fed into the malignancy prediction model for end-to-end processing.

vised object detection problem, in which the model takes the whole image as input and outputs a collection of predicted ROIs along with their likelihood of belonging to a specific group. Formally,

$$\{\hat{z}_{ni}, \hat{p}_{ni},\}_{i=1}^{\hat{m}_n} = g_\psi(X_n), \tag{2}$$

where $\hat{p}_{ni} = [\hat{p}_{ni1}, \ldots, \hat{p}_{niC}] \in (0,1)^C$ is the likelihood that the predicted region $\hat{z}_{ni} = \{\hat{x}_{ni}, \hat{y}_{ni}, \hat{w}_{ni}, \hat{h}_{ni}\}$ belongs to one of C groups of interest, i.e., $p(\hat{z}_{ni} \in c) = \hat{p}_{nic}$. In our case, we consider three possible choices for C, namely, (1) $C = 1$ denoted as *one-class* where the model seeks to identify any lesion regardless of type; (2) $C = 2$ denoted as *malignancy* in which the model seeks to separately identify malignant and benign lesions; and (3) $C = 8$ denoted as *sub-type*, thus the model is aware of all lesion types of interest.

Note that we are mainly interested in finding malignant lesions among all lesions present in an image as opposed to identifying the type of all lesions in the image. Nevertheless, it may be beneficial for the model to be aware that different types of lesions may have common characteristics which may be leveraged for improved detection. Alternatively, provided that some lesion types are substantially rarer than others (e.g., dermatofibroma and vascular lesions only constitutes 1% each of all the lesions in the dataset described in the "Dataset" section), seeking to identity all lesion types may be detrimental for the overall detection performance. This label granularity trade-off will be explored in the experiments. In practice, we use a Faster-RCNN (region-based convolutional neural network)[20] with a feature pyramid network (FPN)[21] and a ResNet-50[19] backbone as object detection architecture. Implementation details can be found in the "Model details" section.

Image classification. For screening purposes, one may be interested in estimating whether an image is likely to contain a malignant lesion so the case can be directed to the appropriate dermatology specialist. In such case, the task can be formulated as a whole-image classification problem

$$p(y_n = 1|X_n) = h_\phi(X_n), \tag{3}$$

where $p(y_n = 1|X_n) \in (0,1)$ is the likelihood that image X_n contains a malignant lesion.

The model in Eq. (3) can be implemented in a variety of different ways. Here we consider three options, two of which leverage the lesion identification and malignancy prediction models described above.

Direct image-level classification $h_\phi(\cdot)$ is specified as a convolutional neural network, e.g., ResNet-50[19] in our experiments, to which the whole image X_n is fed as input. Though this is a very simple model that has advantages from an implementation perspective, it lacks the context provided by (likely) ROIs that will make it less susceptible to interference from background non-informative variation, thus negatively impacting classification performance.

Two-stage approach $h_\phi(\cdot)$ is specified as the combination of the *one-class* lesion identification and the malignancy prediction models, in which detected lesions are assigned a likelihood of malignancy using Eq. (1). This is illustrated in Fig. 1(Right). Then we obtain

$$p(y_n = 1|X_n) = a\left(\{p(u_{ni} \in \mathscr{M}|X_n, \hat{z}_{ni})\}_{i=1}^{\hat{m}_n}\right), \tag{4}$$

where we have replaced the ground truth location z_{ni} in Eq. (1) with the $\hat{m}_n$ predicted locations from Eq. (2), and $a(\cdot)$ is a permutation-invariant aggregation function. In the experiments we consider two simple parameter-free options:

$$a(\cdot) = \frac{1}{\hat{m}_n}\sum_{i=1}^{\hat{m}_n} p(u_{ni} \in \mathscr{M}|X_n, \hat{z}_{ni}), \quad \text{(Average)} \tag{5}$$

$$a(\cdot) = \max(\{p(u_{ni} \in \mathscr{M}|X_n, \hat{z}_{ni})\}_{i=1}^{\hat{m}_n}) \quad \text{(Maximum)} \tag{6}$$

$$a(\cdot) = 1 - \prod_{i=1}^{\hat{m}_n} p(u_{ni} \in \mathscr{M}|X_n, \hat{z}_{ni}) \quad \text{(Noisy OR)} \tag{7}$$

Other more sophisticated (parametric) options such as noisy AND[22], and attention mechanisms[23], may further improve performance but are left as interesting future work.

One-step approach $h_\phi(\cdot)$ is specified directly from the *sub-types* lesion identification model in Eq. (2) as

$$p(y_n = 1|X_n) = a\left(\{\hat{p}_{ni}\}_{i=1}^{\hat{m}_n}\right), \tag{8}$$

where $a(\cdot)$ is either Eqs. (5), (6) or Eq. (7).

From the options described above, the direct image-level classification approach is conceptually simpler and easier to implement but it does not provide explanation (lesion locations) to its predictions. The one-step approach is a more principled end-to-end system that directly estimates lesion locations, lesion sub-type likelihood, and overall likelihood of malignancy, however, it may not be suitable in situations where the availability of labeled sub-type lesions may be limited, in which case, one may also consider replacing the *sub-type* detection model with the simpler *malignancy* detection model. Akin to this simplified one-step approach, the two-stage approach provides a balanced trade-off between the ability of estimating the location of the lesions and the need to identify lesion sub-types. All these options will be quantitatively compared in the experiments below.

Model details.

Malignancy classification. For malignancy classification we use a ResNet-50 architecture[19] as shown in Fig. 1(Bottom right). The feature maps obtained from the last convolutional block are aggregated via average pooling and then fed through a fully connected layer with sigmoid activation that produces the likelihood of malignancy. The model was initialized from a ResNet-50 pre-trained on ImageNet and then trained (refined) using a stochastic gradient descent (SGD) optimizer for 120 epochs, with batch size 64 initial learning rate 0.01, momentum 0.9 and weight decay 1e−4. The learning rate was decayed using a half-period cosine function, i.e., $\eta(t) = 0.01 \times [0.5 + 0.5\cos(t\pi/T_{\max})]$, where t and $T_{\max}$ are the current step and the max step, respectively. We augment the data by randomly resizing and rotating images and note that other augmentation techniques such as flips and random crops have non-substantial impact on model performance.

Lesion identification. The lesion identification model is specified as a Faster-RCNN[20] with a FPN[21] and a ResNet-50[19] backbone. The feature extraction module is a ResNet-50 truncated to the 4th block. The FPN then reconstructs the features to higher resolutions for better multi-scale detection[21]. Higher resolution feature maps are built as a combination of the same-resolution ResNet-50 feature map and the next lower-resolution feature map from the FPN, as illustrated in Fig. 1(Top right). The combination of feature maps from the last layer of the feature extraction module and all feature maps from the FPN are then used for region proposal and ROI pooling. See the original FPN work for further details[21]. The model was trained using an SGD optimizer for 25 epochs, with batch size of 512 per image, initial learning rate 0.001, momentum 0.9 and weight decay 1e-4. Learning rate was decayed 10x at 60,000-th and 80,000-th step, respectively. We augment the data by random resizing.

Direct image-level classification model. The direct image-level classification model in the "Image classification" section has the same architecture and optimization parameters as the malignancy classification model described above.

Clinical model. The clinical model was built using logistic regression with standardized input covariates and discrete (categorical) covariates encoded as one-hot-vectors.

Combined model. In order to combine the clinical covariates with the images into a single model, we use the malignancy classification model as the backbone while freezing all convolutional layers during training. Then, we concatenate the standardized input covariates and the global average-pooled convolutional feature maps, and feed them through a fully connected layer with sigmoid activation that produces the likelihood of malignancy.

Binary classes	Lesion type	Discovery	ISIC2018	Test
Malignant	MEL	510 (7%)	1,113 (11%)	50 (10%)
Benign	NV	1,170 (16%)	6,705 (67%)	139 (28%)
Malignant	BCC	1,481 (21%)	514 (5%)	76 (15%)
Malignant	AKIEC	2,122 (29%)	327 (3%)	121 (24%)
Benign	BKL	897 (13%)	1,099 (11%)	83 (17%)
Benign	DF	88 (1%)	115 (1%)	11 (2%)
Benign	VASC	102 (1%)	142 (2%)	5 (1%)
Benign	OB	826 (12%)	–	17 (3%)

Table 1. Lesion type counts by dataset. The 8 lesion types considered are: Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis/Bowen's Disease (AKIEC), Benign Keratosis (BKL), Dermatofibroma (DF), Vascular Lesions (VASC) and Other Benign (OB) lesions. Note that in our dataset we included Squamous Cell Carcinoma lesion in the AKIEC category.

The combined model was trained using an SGD optimizer for 30 epochs, with batch size 64, initial learning rate 0.001, momentum 0.9 and weight decay 1e-4. The learning rate was decayed using a half-period cosine function as in the malignancy classification model.

Implementation. We used Detectron[24] for the lesion identification model. All other models were coded in Pyhton 3.6.3 using the PyTorch 1.3.0 framework except for the clinical model that was implemented using scikit-learn 0.19.1. The source code for all the models used in the experiments is available (upon publication) at https://github.com/mx41-m/Lesion-Identification-and-Malignant-Prediction.git.

Dataset.

Discovery dataset. To develop the model we consider a single institution, retrospective collection of skin lesion images taken with smartphones with and without dermoscopy from Duke University Medical Center patients of age 18 and older from 2013 to 2018. These data are collected under the approval of the Duke Institute for Health Innovation and each participant has provided written informed consent. The *discovery* dataset consists of 6819 images from 3853 patients with 7196 manually annotated lesions. Malignancy was ascertained by separate Histopathological diagnosis. There are 4113 (57%) lesions in 3894 images diagnosed as malignant. For images with multiple lesions, the image is deemed malignant if it contains at least one malignant lesion. In terms of skin tone, the Fitzpatrick scale was used to group skin tones: Light (Fitzpatrick type 1 &2), Medium (Fitzpatrick type 3 &4), Dark (Fitzpatrick type 5 &6). Specifically, 6022 lesions (5721 images) are light, 1073 lesions (1020 images) are medium and 101 lesions (96 images) are dark tone (lesions from one image may be different skin tone.) Lesions were manually annotated as bounding boxes (ROIs) by a dermatology trained medical doctor (Dr. Kheterpal, MK) using a in-house annotation application. Diagnoses taken from the biopsy reports associated with the lesion images were designated as the ground truth (Malignant vs. Benign). Further, there are 589 (9%) dermoscopy images and 6230 (91%) wide-field images. Based on the guidelines of the International Skin Imaging Collaboration (ISIC) dataset, the lesions in our dataset can be further divided into several specific lesion types, whose counts and proportions are shown as Table 1. Additional details concerning lesion types, including proportions of malign and benign lesions or proportions of wide-field and dermoscopy images in discovery dataset are presented as supplementary Figure s1. The average area of the lesion is 307,699 (Q1–Q3: 9192–184,900) pixels^2 (roughly 554 × 554 pixels in size) while the average area of the images is 8,036,107 (3,145,728-12,000,000) pixels^2 (roughly 2834 × 2834 pixels in size). We split the dataset, at the patient level, into 6115 lesions (5781 images) for training and 1081 lesions (1038 images) for validation. The validation set was used to optimize the model parameters, architecture and optimization parameters.

Clinical dataset. We also consider a subset of 4130 images from 2270 patients for which we also have demographic (age at encounter, sex and self-reported race, as found in the medical record, from which 104 patients are self-reported as non-white), lesion characteristics (location and number of previous dermatology visits), comorbidities (history of chronic ulcer of skin, diseases of white blood cells, human immunodeficiency virus infection, Hodgkin's disease, non-Hodgkin's lymphoma, infective arthritis and osteomyelitis, leukemias, Parkinson's disease, rheumatologic diseases, skin and subcutaneous tissue infections, inflammatory condition of skin, systemic lupus erythematosus, other connective tissue disease, other sexually transmitted diseases, other hematologic diseases, and other skin disorders) and skin-cancer-related medications (immunosuppressants, corticosteroids, antihypertensives, antifungals, diuretics, antibiotics, antiarrhythmics, antithrombotics, chemotherapy, targeted therapy, immunotherapy, and other), their risk (Low vs. High), and frequency of administration. Among these patients, 1411 (2537 images) are diagnosed as malignant and 859 (1593 images) as benign. Similar to the discovery dataset, we split these data into 85% for training and the remaining 15% for validation.

ISIC2018. Provided that we have a smaller number of dermoscopy images, we also consider augmenting our discovery dataset with the ISIC2018 training dataset[25,26] consisting of 10,015 dermoscopy images, from which 1954 correspond to malignant lesions and 8061 benign lesions. Detailed lesion type counts are presented in

Table 1. In the experiments, we also consider the ISIC2018 validation dataset to test the model with and without ISIC2018 augmentation.

Independent test set. In order to evaluate the performance of the model relative to human diagnosis, we consider an independent set of 488 images also from Duke University Medical Center patients. In terms of skin tone, 369 lesions (359 images) are light, 122 lesions (118 images) are medium and 11 lesions (11 images) are dark. From these images, 242 are malignant and 246 are benign. Consistent with the Discovery Dataset, we use the same malignancy and skin tone definitions. To compare the proposed model with human experts, we had three dermatology trained medical doctors with different levels of experience label each of the images without access to the biopsy report or context from the medical record. In terms of experience, MJ has 3 years dermoscopy experience, AS has 6 years of dermoscopy experience and MK has 10 year dermoscopy experience. Provided that MK also participated in lesion annotation with access to biopsy report information, we allowed 12 months separation between the lesion annotation and malignancy adjudication sessions. Detailed lesion type counts are presented in Table 1. The average area of the lesion is 458,619 (17,161–395,483) pixels^2 (roughly 677 × 677 pixels in size) while the average area of the images is 7,755,934 (3,145,728–12,000,000) unbelievable it is same as train but it is true) pixels^2 (roughly 2785 × 2785 pixels in size).

Performance metrics. For malignancy prediction, two threshold-free metrics of performance are reported, namely, area under the curve (AUC) of the receiving operating characteristic (ROC) and the average precision (AP) of the precision recall curve, both described below. AUC is calculated as:

$$\begin{aligned}
\mathrm{AUC} &= \frac{1}{2}\sum_i \left[\mathrm{FPR}_{\hat{p}_{i+1}} - \mathrm{FPR}_{\hat{p}_i}\right]\left[\mathrm{TPR}_{\hat{p}_{i+1}} + \mathrm{TPR}_{\hat{p}_i}\right] \\
\mathrm{TPR}_t &= p(\hat{p} > t|y = 1) \\
\mathrm{FPR}_t &= p(\hat{p} > t|y = 0),
\end{aligned}$$

where $t \in [\hat{p}_1, \ldots, \hat{p}_i, \hat{p}_{i+1}, \ldots]$ is a threshold that takes values in the set of sorted test predictions $\{\hat{p}_i\}_{i=1}^{N}$ from the model, and the true positive rate, TPR_t, and false positive rate, FPR_t, are estimated as sample averages for a given threshold t.

Similarly, the AP is calculated as:

$$\begin{aligned}
\mathrm{AP} &= \frac{1}{2}\sum_i \left[\mathit{TPR}_{\hat{p}_{i+1}} - \mathit{TPR}_{\hat{p}_i}\right]\left[\mathrm{PPV}_{\hat{p}_{i+1}} + \mathrm{PPV}_{\hat{p}_i}\right] \\
\mathrm{PPV}_t &= p(y = 1|\hat{p} > t),
\end{aligned}$$

where PPV_t is the positive predictive value or *precision* for threshold t. The calculation for the AUC and AP areas follow the trapezoid rule.

The intersection over union (IoU) is defined as the ratio between the overlap or ground truth and estimated ROIs, $\{z_{ni}\}_{i=1}^{m_n}$ and $\{\hat{z}_{ni}\}_{i=1}^{\hat{m}_n}$, respectively, and the union of their areas. For a given ROI, IoU=1 indicates complete overlap between prediction and ground truth. Alternatively, IoU=0 indicates no overlap. In the experiments, we report the median and interquartile range IoU for all predictions in the test set.

The mean average precision (mAP) is the AP calculated on the binarized predictions from the detection model such that predictions with an IoU$\geq t$ are counted as correct predictions or incorrect otherwise, if IoU$< t$, for a given IoU threshold t set to 0.5, 0.75 and (0.5, 0.95) in the experiments. These values are standard in object detection benchmarks, see for instance[27].

We also report the recall with IoU> 0 as a general, easy to interpret, metric of the ability of the model to correctly identify lesions in the dataset. Specifically, we calculate it as the proportion of lesions (of any type) in the dataset for which predictions overlap with the ground truth.

Results

Comprehensive experiments to analyze the performance of the proposed approach were performed. First, we evaluate and compare various design choices based on the evaluation metrics described in "Method" section. Then, we study the effects of adding clinical covariates and using an auxiliary publicly available dataset for data augmentation. Lastly, we present some visualization of the proposed model predictions for qualitative analysis.

Quantitative results. *Malignancy prediction.* First, we present results for the malignancy prediction task, for which we assume that lesions in the form of bounding boxes (ROIs) have been pre-identified from smartphone (wide-field) or dermoscopy images. Specifically, we use ground truth lesions extracted from larger images using manual annotations as previously described. Table 2 shows AUCs and APs on the independent test dataset for the malignancy prediction model described in the "Malignancy prediction" section. We observe that the model performs slightly better on dermoscopy images presumably due to their higher quality and resolution.

Malignancy detection. Provided that in practice lesions are not likely to be pre-identified by clinicians, we present automatic detection (localization) results using the models presented in the "Lesion identification" section. Specifically, we consider three scenarios: (1) *one-class*: for all types of lesions combined; (2) *malignancy*: for all types of lesions combined into malignant and benign; and (3) *sub-type*: for all types of lesions separately. Table 3 shows mean Average Precision (mAP) at different thresholds, Recall (sensitivity) and IoU summaries (median

	AUC	AP
All lesions	0.790	0.775
Lesions from smartphone images	0.789	0.775
Lesions from dermoscopy images	0.791	0.790

Table 2. Malignancy prediction performance where ground truth lesions are manually drawn and annotated by a dermatology trained medical doctor (MK).

	One-class	Malignancy	Sub-type
$\text{mAP}_{@0.5}$	**0.793**	0.755	0.749
$\text{mAP}_{@0.75}$	**0.281**	0.232	0.250
$\text{mAP}_{@0.5,0.95}$	**0.370**	0.335	0.337
Recall	0.954	**0.956**	0.948
IoU	$\mathbf{0.729_{(0.59,0.82)}}$	$0.726_{(0.58,0.80)}$	$0.716_{(0.59,0.80)}$

Table 3. Lesion detection from smartphone (wide-field) and dermoscopy images. Performance is evaluated as the mean Average Precision (mAP) at three different thresholds: 0.5, 0.75 and [0.5, 0.95], recall (sensitivity) and intersection over union (IoU) summarized as median (interquartile range). Significant values are in [bold].

and interquartile range), all on the independent test set (detailed results for the sub-type lesions are shown as supplementary Table s1 and Table s2). In order to make mAP comparable across different scenarios, we calculate it for all lesions regardless of type, i.e., mAP is not calculated for each lesion type and then averaged but rather by treating all predictions as lesions. We observe that in general terms, the *one-class* lesion identification model outperforms the more granular *malignancy* and *sub-type* approaches. These observation is also consistent in terms of Recall and IoU.

For the *one-class* model specifically, 79.3% regions predicted are true lesions at at IoU$\geq$ 0.5 (at least 50% overlap with ground truth lesions), whereas the precision drops to 28.1% with a more stringent IOU$\geq$ 0.75. Interestingly, the 95.6% Recall indicates that the *one-class* model is able to capture most of the true lesions at IoU$>$ 0 and at least 50% of the predicted regions have a IoU$>$ 0.729 or IoU$>$ 0.599 for 75% of the lesions in the independent test set.

Image classification. The image-level prediction results of malignancy are reported in Fig. 2. Predictions on the independent test set were obtained from the average-pooled image classification model in the "Image classification" section with the *one-class* detection model in the "Lesion identification" section and the malignancy prediction model in the "Malignancy prediction" section. From the performance metrics reported we note that the proposed approach is comparable with manual classification by three expert dermatologists (AS, MK and MJ). Interestingly, in dermoscopy images, the model slightly outperforms two of the three dermatologists and the difference in their performance is consistent with their years of experience; MK being the most experienced and better performing dermatologist.

Additional results comparing the different image-level malignancy prediction strategies described in the "Image classification" section, namely, (1) direct image-level classification, (2) two-stage with *one-class* lesion identification, and one-step with (3) *malignancy* or (4) *sub-type* identification models with max pooling aggregation are presented in Table 4. In terms of AUC, the one-class approach consistently outperforms the others, while in terms of AP, sub-type is slightly better. Interestingly, the direct image-level classification which takes the whole image as input, without attempting to identify the lesions, performs reasonably well and may be considered in scenarios where computational resources are limited, e.g., mobile and edge devices.

Further, we also compare different lesion identification models (*one-class*, *malignancy* and *sub-types*) described in the "Lesion identification" section and aggregation strategies (average, max and noisy OR pooling) described in the "Image classification" section, and the results are presented in Table 5, from which we see that the combination of max pooling and one-class lesion detection slightly outperforms the alternatives.

Accounting for clinical data. Next, we explore the predictive value of clinical features and their combination with image-based models. Specifically, we consider three models: (1) the logistic regression model using only clinical covariates; (2) the malignancy classification model; and (3) the combined model described in the "Model details" section. Note that since we have a reduced set of images for which both clinical covariates and images are available as described in the "Dataset" section, all models have been re-trained accordingly. Figure 3 shows ROC and PR curves for the three models and the TPR and FPR values for three dermatology trained MDs on the independent test set. Results indicate a minimal improvement in classification metrics by combining clinical covariates and images, and a significant improvement of the image-based models relative to the pure clinical model, which underscores the importance and predictive value of the images for the purpose of malignancy

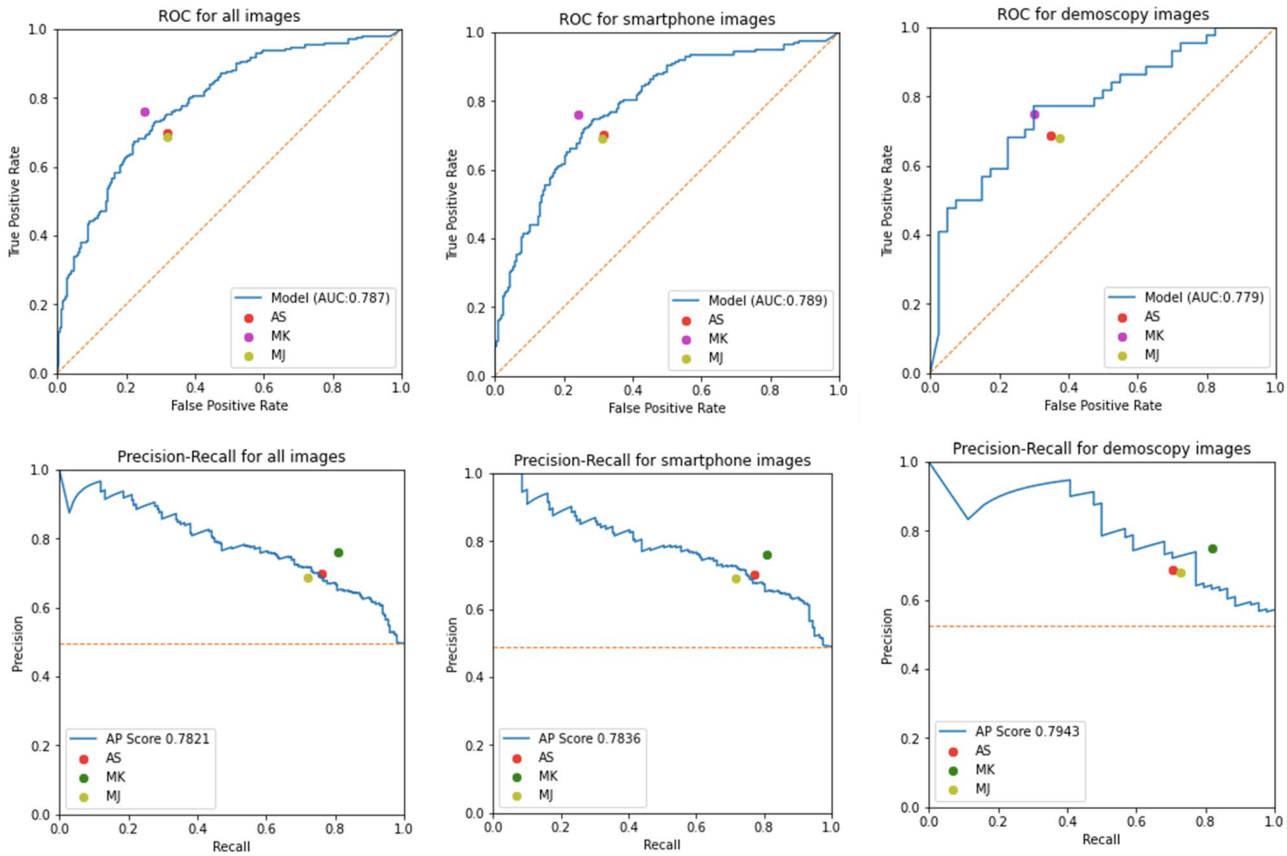

Figure 2. Performance metrics of the malignancy prediction models. ROC and PR curves, top and bottom rows, respectively, for all images (Left), smartphone (wide-field) images (Middle) and dermoscopy images (Right) on the test set. Predictions were obtained from the *one-class* model followed by the malignancy prediction model and the image classification aggregation approach. Also reported are the TPR (sensitivity) and FPR (1-specificity) for three dermatology trained MDs (AS, MK and MJ).

	All images		Smartphone only		Dermoscopy only	
	AUC	**AP**	**AUC**	**AP**	**AUC**	**AP**
Image-level	0.755	0.731	0.768	0.729	0.696	0.723
Malignancy	0.769	0.777	0.775	0.779	0.735	0.788
Sub-type	0.764	0.771	0.767	0.769	0.754	**0.805**
One-class	**0.787**	**0.782**	**0.789**	**0.784**	**0.779**	0.794

Table 4. Performance metrics with different image-level classification strategies (direct image-level, two-stage with *one-class* lesion identification and one-step with *malignancy* or *sub-type* identification) stratified into all images and smartphone or dermoscopy only subsets. Significant values are in [bold].

	Average		Noisy OR		Max	
Detection	**AUC**	**AP**	**AUC**	**AP**	**AUC**	**AP**
Malignancy	0.7698	0.7759	0.7698	0.7461	0.7690	0.7772
Sub-type	0.7784	0.7822	0.7698	0.7600	0.7643	0.7705
One-class	0.7759	0.7650	0.7783	0.7661	**0.7868**	**0.7821**

Table 5. Performance metrics of the image-level malignancy prediction model with different lesion identification models (*one-class*, *malignancy* and *sub-types*) and aggregation strategies (average, noisy OR and max pooling). The best performing combination is highlighted in boldface.

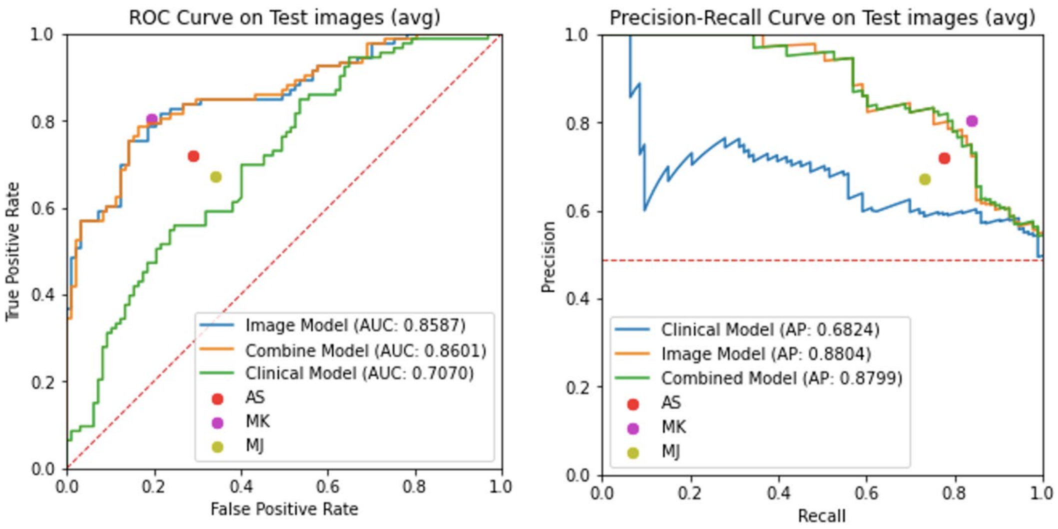

Figure 3. Performance metrics of the malignancy prediction models including clinical covariates. ROC and PR curves for three models are presented, namely, combined (clinical + images), image only and clinical covariates only. Also reported are the TPR (sensitivity) and FPR (1-specificity) for three dermatology trained MDs (AS, MK and MJ). Note that we also provide performance characteristics for a clinical model trained without self-reported race in supplementary Figure s2.

		All images		Smartphone only		Dermoscopy only	
		AUC	AP	AUC	AP	AUC	AP
Malignancy prediction	Discovery	0.783	0.769	0.786	0.769	**0.775**	**0.783**
	Discovery + ISIC2018	**0.805**	**0.783**	**0.822**	**0.787**	0.734	0.777
Direct image-level	Discovery	0.755	0.731	0.768	0.739	0.696	**0.723**
	Discovery + ISIC2018	**0.774**	**0.741**	**0.778**	**0.753**	**0.745**	0.719
Two-step approach	Discovery	0.787	**0.782**	0.789	**0.784**	**0.779**	**0.794**
	Discovery + ISIC2018	**0.793**	0.773	**0.802**	0.773	0.734	0.762
Malignancy prediction	ISIC2018	–	–	–	–	0.959	0.849
	Discovery + ISIC2018	–	–	–	–	**0.961**	**0.881**

Table 6. Performance metrics (AUC and AP) of the models with data augmentation. We consider three models with and without ISIC2018 dermoscopy image dataset augmentation. The three models considered are the malignancy prediction model described in the "Malignancy prediction" section, and the direct image-level classification and two-step approach with one-class lesion identification described in the "Image classification" section. Significant values are in [bold].

prediction. Moreover, we verified that self-reported race has minimal impact on performance in the clinical model (see Supplementary Figure s2).

Dermoscopy data augmentation. Finally, we consider whether augmenting the discovery dataset with the publicly available ISIC2018 dataset improves the performance characteristics of the proposed model. Specifically, the ISIC20128 (training) dataset which consists of only dermoscopy images is meant to compensate for the low representation of dermoscopy images in our discovery dataset, i.e., only 9% of the discovery images are dermoscopy. Results in Table 6 are stratified by image type (all images, smartphone (wide-field) only and dermoscopy only) are presented for three different models: (1) malignancy prediction (assuming the positions of the lesions are available); (2) direct image-level classification; and (3) the two-stage approach with *one-class* lesion identification. As expected, data augmentation consistently improve the performance metrics of all models considered. But, performance metrics on dermoscopy images do not improve. We think it may be caused by the domain gap between our dermoscopy images and ISIC dataset since the dermoscopy parameters, resolutions and light conditions are different.

Qualitative results. Figure 4 shows examples of the *one-class* lesion identification model described in the "Lesion identification" section. Note that the model is able to accurately identify lesions in images with vastly different image sizes, for which the lesion-to-image ratio varies substantially. We attribute the model ability to

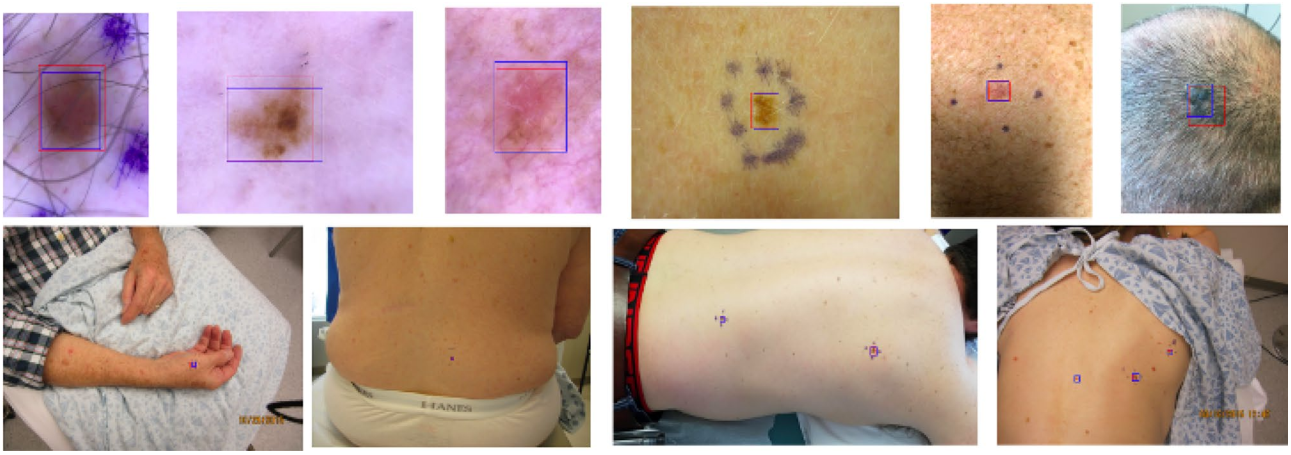

Figure 4. Lesion detection examples. Top: Dermoscopy images. Bottom: Smartphone (wide-field) images. The ground-truth, manually annotated lesion is represented by the red bounding box, while the predicted lesion is denoted by the blue bounding box.

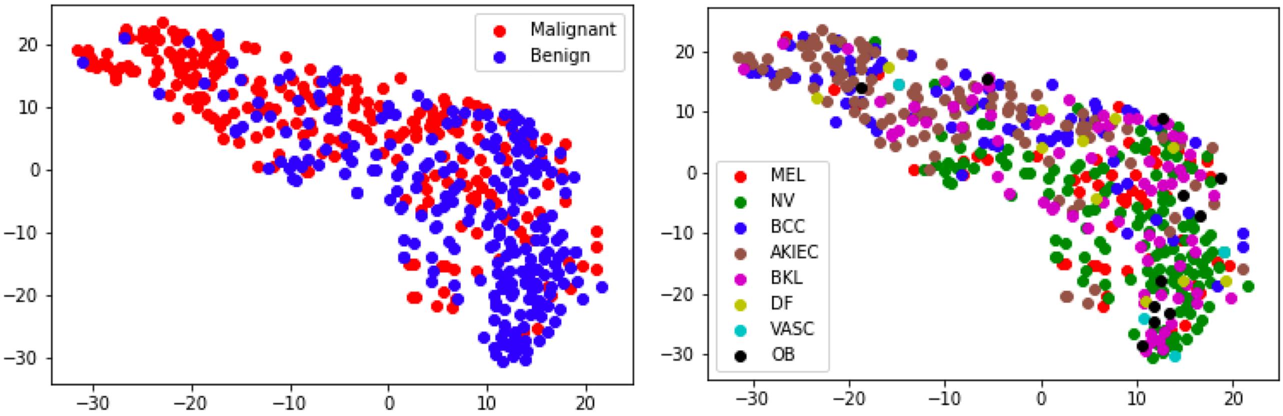

Figure 5. *t*-SNE Map. Each point in the figure represents a test-set lesion separately colored by malignancy (Top) and lesion sub-type (Bottom): Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis/Bowen's Disease (AKIEC), Benign Keratosis (BKL), Dermatofibroma (DF), Vasular Lesion (VASC) and Other Benign (OB).

do so to the FPN network that allows to obtain image representations (features) at different resolution scales. Further, in Fig. 5 we show through a two-dimensional *t*-SNE map[28] that the representations produced by the lesion detection model (combined backbone and FPN features) roughly discriminate between malignant and benign lesions, while also clustering in terms of lesion types.

Discussion

The early skin lesion classification literature used largely high-quality clinical and dermoscopy images for proof of concept[5,6,8–10]. Specifically, Esteva et al.[5] showed that when trained with over 100,000 images, Inception v3 achieved 72.1(±0.9%) accuracy in a three-level (i.e., benign, malignant and non-neoplastic) lesion classification and surpassed two dermatologists, whose accuracies were 65.6% and 66.0% respectively. Haenssle et al.[6] showed that Inception V4, an improved version of model architecture over Inception V3, obtained a higher specificity of 82.5% in a binary lesion classification (i.e., melanoma and benign) when compared to 58 international dermatologists, whose average specificity was 71.3%(±11.2%). Han *et al.*[8] applied ResNet-152 in classifying 12 types of skin diseases then tested the model performance on three different public datasets, i.e., Asan, Hallym and Edinburgh. They further analyzed the performance of the models built separately on each dataset and on each type of skin disease, and proposed several hypothesis for the observed performance differences, e.g., unequal lighting and background of images, imbalanced number of images for each class. However, usability of these algorithms in the real-world remains questionable and must be tested prospectively in clinical settings. Consumer-grade devices produce images of variable quality, however, this approach mimics the clinical work flow and provides a universally applicable image capture for any care setting. The utility of wide-field clinical images taken with smartphone was recently demonstrated by Soenksen *et. al* for detection of "ugly duckling" suspicious pigmented lesions vs. non-suspicious lesions with 90.3% sensitivity (95% CI 90.0–90.6) and 89.9% specificity (95% CI 89.6–90.2) validated against three board certified dermatologists[29]. This use case demonstrates how clinical work flow in dermatology can be replicated with ML-based CDS. However, the limitation is that the number needed to treat (NNT) for true melanoma detection from pigmented lesion biopsies by dermatologists is 9.60 (95% CI 6.97–13.41) by meta-analysis[30]. Hence, the task of detecting suspicious pigmented lesions should be compared against histological ground truth rather than concordance with dermatologists, for improved accuracy and comparability of model performance. Furthermore, pigmented lesions are a small subset of the overall task to detect skin cancer, as melanomas constitute fewer than 5% of all skin cancers. Our approach utilizing wide-field images to detect lesions of interest demonstrated encouraging mAP, IoU and Recall metrics, considering the sample size used. This primary step is critical in the clinical workflow where images are captured for lesions of interest but lesion annotation is not possible in real time. An ideal ML-based CDS would identify lesion of interest and also provide the likelihood of malignancy and the sub-type annotations as feedback to the user. Our study demonstrates malignancy classification for the three most common skin cancers (BCC, SCC and Melanoma) vs. benign tumors with smartphone images (clinical and dermoscopy) with encouraging accuracy when validated against histopathological ground truth. The usability of this algorithm is further validated by comparison with dermatologists with variable levels of dermoscopy experience, showing comparable performance to dermatologists in both clinical and dermoscopy binary classification tasks, despite low dermoscopy image data (9%) in the Discovery set. This two-stage model, with the current performance level, could be satisfactorily utilized in a PCP triage to dermatology (pending prospective validation) at scale for images concerning for malignancy as a complete end-to-end system. Interestingly, the additional ISIC high-quality dataset (predominantly dermoscopy images) improved performance across both clinical and dermoscopy image sets. This suggests that smartphone image data can be enriched by adding higher quality images. It is unclear if this benefit is due to improvement in image quality or volume, and remains an area of further study.

Finally, we demonstrated that comprehensive demographic and clinical data is not critical for improving model performance in a subset of patients, as the image classification model alone performs at par with the combination model. Clinicians often make contextual diagnostic and management decisions when evaluating skin lesions to improve their accuracy. Interestingly, this clinical-context effect that improves diagnostic accuracy at least in pigmented lesions maybe dependent on years of dermoscopy experience[6]. The value of clinical context in model performance has not been studied extensively and remains an area of further study in larger datasets.

Limitations. Limitations of the study include a small discovery image dataset, predominantly including light and medium skin tones, and with less than 2% of images included with dark skin tone. However, this may represent the bias in the task itself as skin cancers are more prevalent in light- followed by medium-skin tones. Given the large range of skin types and lesions encountered in clinical practice, additional images may improve performance and generalizability. At scale, image data pipelines with associated metadata are a key resource needed to obtain inclusive ML-based CDS for dermatology. Improved image quality and/or volume improves performance as demonstrated by the ISIC dataset incorporation into the model, however, this theoretical improvement in performance needs validation in prospective clinical settings. While the pure clinical model incorporates a comprehensive list and accounts for temporal association of this metadata with detection of lesions, it is not an exhaustive list as it does not include social determinants such as sun-exposure behavior and tanning bed usage; two critical factors contributing to increasing incidence of skin cancer. In particular, metadata including lesion symptoms and evolution is missing and should be incorporated in future studies. Finally, it should be noted that lesions included in this study were evaluated and selected for biopsies in dermatology clinics. If this model was to be utilized in other clinical settings such as primary care, additional validation will be needed as pre-test probability of lesion detection is different among clinical settings[30].

Data availibility

In this study, the discovery dataset, clinical dataset and independent test dataset are collected from Duke University Medical Center and are not available for participant privacy protecting rules. Researchers interested in accessing the clinical dataset should contact Meenal Kheterpal at meenal.kheterpal@duke.edu. The ISIC dataset is a public dataset, which is available at https://challenge.isic-archive.com/data#2018.

Received: 8 June 2022; Accepted: 9 September 2022
Published online: 23 September 2022

References

1. Tsang, M. W. & Resneck, J. S. Jr. Even patients with changing moles face long dermatology appointment wait-times: A study of simulated patient calls to dermatologists. *J. Am. Acad. Dermatol.* **55**, 54–58 (2006).
2. Vincent, G. K. *The next four decades: The older population in the United States: 2010 to 2050*. 1138 (US Department of Commerce, Economics and Statistics Administration, US, 2010).
3. Cancer Stat Facts melanoma of the skin. https://seer.cancer.gov/statfacts/html/melan.html.
4. Lowell, B. A., Froelich, C. W., Federman, D. G. & Kirsner, R. S. Dermatology in primary care: Prevalence and patient disposition. *J. Am. Acad. Dermatol.* **45**, 250–255 (2001).
5. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
6. Haenssle, H. A. *et al.* Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **29**, 1836–1842 (2018).
7. Safran, T. *et al.* Machine learning and melanoma: The future of screening. *J. Am. Acad. Dermatol.* **78**, 620–621 (2018).
8. Han, S. S. *et al.* Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J. Investig. Dermatol.* **138**, 1529–1538 (2018).
9. Lopez, A. R., Giro-i Nieto, X., Burdick, J. & Marques, O. Skin lesion classification from dermoscopic images using deep learning techniques. In *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, 49–54 (IEEE, 2017).
10. Salido, J. A. A. & Ruiz, C. Using deep learning to detect melanoma in dermoscopy images. *Int. J. Mach. Learn. Comput.* **8**, 61–68 (2018).

11. Polat, K. & Koc, K. O. Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all. *J. Artif. Intell. Syst.* **2**, 80–97 (2020).
12. Dascalu, A. & David, E. Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope. *EBioMedicine* **43**, 107–113 (2019).
13. Tognetti, L. *et al.* A new deep learning approach integrated with clinical data for the dermoscopic differentiation of early melanomas from atypical nevi. *J. Dermatol. Sci.* **101**, 115–122 (2021).
14. Fujisawa, Y. *et al.* Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br. J. Dermatol.* **180**, 373–381 (2019).
15. Nasr-Esfahani, E. *et al.* Melanoma detection by analysis of clinical images using convolutional neural network. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1373–1376 (IEEE, 2016).
16. Jafari, M. H. *et al.* Skin lesion segmentation in clinical images using deep learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, 337–342 (IEEE, 2016).
17. Jinnai, S. *et al.* The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules* **10**, 1123 (2020).
18. Brinker, T. J. *et al.* A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur. J. Cancer* **111**, 148–154 (2019).
19. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern rRecognition*, 770–778 (2016).
20. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems* Vol. 28 (eds Cortes, C. *et al.*) (Curran Associates Inc, 2015).
21. Lin, T.-Y. *et al.* Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2117–2125 (2017).
22. Kraus, O. Z., Ba, J. L. & Frey, B. J. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics* **32**, i52–i59 (2016).
23. Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems* Vol. 30 (eds Guyon, I. *et al.*) (Curran Associates Inc, 2017).
24. Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P. & He, K. Detectron. https://github.com/facebookresearch/detectron (2018).
25. Codella, N. *et al.* Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019).
26. Tschandl, P., Rosendahl, C. & Kittler, H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **5**, 1–9 (2018).
27. Lin, T.-Y. *et al.* Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755 (Springer, 2014).
28. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
29. Soenksen, L. R. *et al.* Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Sci. Transl. Med.* **13**, eabb3652 (2021).
30. Petty, A. J. *et al.* Meta-analysis of number needed to treat for diagnosis of melanoma by clinical setting. *J. Am. Acad. Dermatol.* **82**, 1158–1165 (2020).

Acknowledgements

The authors would like to thank Melodi J. Whitley, Ph.D., MD (MJ) and Amanda Suggs, MD (AS) for their assistance with the manual classification of images for the test dataset; and the Duke Institute for Health Innovation (DIHI) for providing access to the clinical data.

Author contributions

R.H., M.X. formulated the problem and conceived the framework. M.X. processed the data and implemented the framework. R.H., M.X. designed the experiments and analyzed the results. M.K.K. provided professional medical analysis. M.K.K., S.C.W., C.P., W.R. collected the data and are responsible for data curation. R.H., M.X., M.K.K. drafted the final manuscript. All authors contributed to revise the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20168-w.

Correspondence and requests for materials should be addressed to M.X.

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.