Fusion between an Algorithm Based on the Characterization of Melanocytic Lesions' Asymmetry with an Ensemble of Convolutional Neural Networks for Melanoma Detection

Jules Collenne, Jilliana Monnier, Rabah Iguernaissi, Motasem Nawaf, Marie-Aleth Richard, Jean-Jacques Grob, Caroline Gaudy-Marqueste, Séverine Dubuisson, Djamal Merad

JIDOpen

Fusion between an Algorithm Based on the Characterization of Melanocytic Lesions' Asymmetry with an Ensemble of Convolutional Neural Networks for Melanoma Detection

Jules Collenne[1,4], Jilliana Monnier[1,2,3,4], Rabah Iguernaissi[1], Motasem Nawaf[1], Marie-Aleth Richard[3], Jean-Jacques Grob[2,3], Caroline Gaudy-Marqueste[2,3], Séverine Dubuisson[1] and Djamal Merad[1]

Melanoma is still a major health problem worldwide. Early diagnosis is the first step toward reducing its mortality, but it remains a challenge even for experienced dermatologists. Although computer-aided systems have been developed to help diagnosis, the lack of insight into their predictions is still a significant limitation toward acceptance by the medical community. To tackle this issue, we designed handcrafted expert features representing color asymmetry within the lesions, which are parts of the approach used by dermatologists in their daily practice. These features are given to an artificial neural network classifying between nevi and melanoma. We compare our results with an ensemble of 7 state-of-the-art convolutional neural networks and merge the 2 approaches by computing the average prediction. Our experiments are done on a subset of the International Skin Imaging Collaboration 2019 dataset (6296 nevi, 1361 melanomas). The artificial neural network based on asymmetry achieved an area under the curve of 0.873, sensitivity of 90%, and specificity of 67%; the convolutional neural network approach achieved an area under the curve of 0.938, sensitivity of 91%, and specificity of 82%; and the fusion of both approaches achieved an area under the curve of 0.942, sensitivity of 92%, and specificity of 82%. Merging the knowledge of dermatologists with convolutional neural networks showed high performance for melanoma detection, encouraging collaboration between computer science and medical fields.

Keywords: Asymmetry, Bioinformatics, Computer-aided diagnosis, Machine learning, Melanoma

Journal of Investigative Dermatology (2024) ■, ■–■; doi:10.1016/j.jid.2023.09.289

INTRODUCTION

Melanoma is a major public health concern, with studies reporting an increasing incidence in developed countries (Ferlay et al, 2021; Forsea, 2021; Sung et al, 2021). Although systemic therapies have led to enhanced survival rates, melanoma-related fatality rates continue to remain elevated. Early diagnosis of primary cutaneous melanoma is critical in reducing melanoma mortality, but it is challenging for dermatologists because this cancer can look very similar to common nevi. These concerns are also exacerbated by the lack of dermatologists in many countries. To face these problems, computer-aided diagnoses have been developed to make the diagnosis faster and more accurate. Although they may provide accurate predictions (Pham et al, 2021), they are still underused in practice. Dermatologists may find it difficult to trust models that provide predictions without any insight into how they work. Solving these problems might increase the use of computer-aided diagnosis in clinical routine, potentially leading to more diagnosed melanomas.

Existing machine learning models are mainly trained and tested on the International Skin Imaging Collaboration dataset. This dataset, although containing >30,000 images, is also an online challenge where participants compete to have the best model to detect melanomas. Winners of the 2020 challenge, Ha et al (2020[1]), used an ensemble of fine-tuned EfficientNets, showing high performance on this dataset. To characterize the state of the art in machine learning, we used a similar method for our melanoma versus nevus classification.

However, understanding the predictions of these models is tedious, if not impossible, which is a significant limitation to

[1]*Computer Science and Systems Laboratory, CNRS UMR 7020, Aix-Marseille University, Marseille, France;* [2]*Cancer Research Center of Marseille, Inserm UMR1068, CNRS UMR7258, Aix-Marseille University, Marseille, France; and* [3]*Dermatology and Skin Cancer Department, La Timone Hospital, Assistance Publique Hôpitaux de Marseille, Aix-Marseille University, Marseille, France*

[4]*These authors contributed equally to this work.*

Correspondence: Jules Collenne, Laboratoire d'Informatique et Systèmes, CNRS UMR 7020, Aix-Marseille University, Marseille, France. E-mail: jules.collenne@lis-lab.fr

Abbreviations: ANN, artificial neural network; AUC, area under the curve; CNN, convolutional neural network; HSV, hue; saturation, value

Received 25 May 2023; revised 21 September 2023; accepted 26 September 2023; accepted manuscript published online XXX; corrected proof published online XXX

[1] Ha Q, Liu B, Liu F. Identifying melanoma images using EfficientNet ensemble: winning solution to the SIIM-ISIC melanoma classification challenge. arXiv 2020.

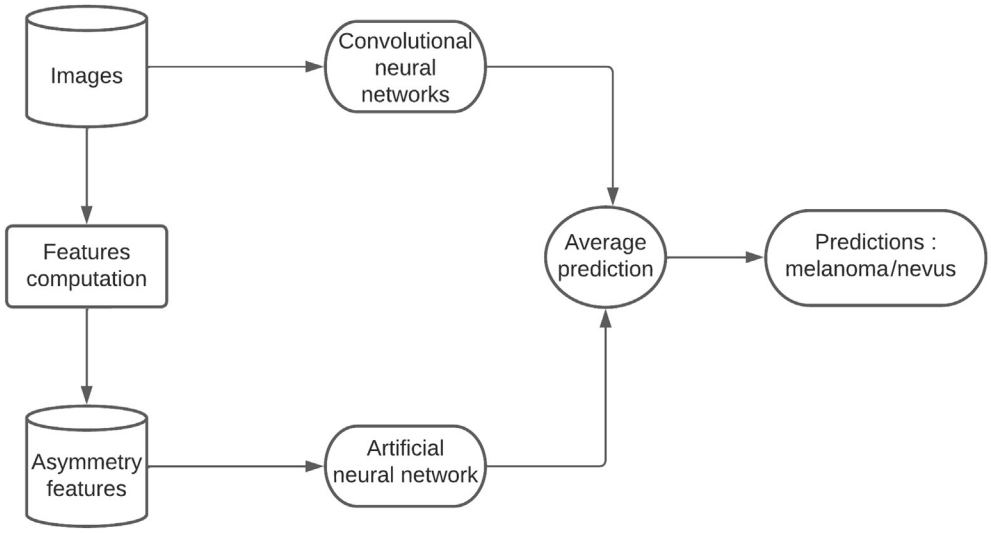


Figure 1. Architecture of our framework. One branch directly uses images, whereas the other one focuses on asymmetry.

their use in clinical practice. A possible solution is the design of features corresponding to visual characteristics used by dermatologists to detect melanomas. Several research works have proposed different features to make the diagnosis, including the mean color, luminescence, and texture analysis. Nezhadian and Rashidi (2017) explored color and texture characteristics, and Chandra et al (2019) proposed a model using the borders of the lesions. Yet, current works use preprocessing techniques without taking into account the axes of symmetry. Asymmetry was studied by Srivastava et al (2021), but they used only the 2 halves of the lesions. Milczarski and Stawska (2020) solely analyzed the lesions' shape, not using the important asymmetrical color distributions. Our research explores the analysis of asymmetrical color distribution within lesions, shedding light on this aspect. We use multiple axes of symmetry as well as a method comparing pairs of opposed pixels according to the center. Indeed, lesions with an anarchic structure led to asymmetrical color distribution and outlines, whereas other lesions tend to be more symmetrical in terms of borders and colors. In this study, we sought to develop a handcrafted model that can effectively extract innovative features representing global color asymmetry, consistent with the approach of dermatologists.

Once extracted, our features are fed into a classifier trained to make the diagnosis. There are various commonly used classifiers such as support vector machines (Rastgoo et al, 2015; Zghal and Kallel, 2020), Random Forest (Rastgoo et al, 2015), Gradient Boosting methods (Rastgoo et al, 2015), and k-nearest neighbors (Oukil et al, 2019), but we chose artificial neural network (ANN) owing to its superior performance for the classification task. To gain deeper insights into the model's predictions, various analytical tools are available. Among these, SHAP (Shapley Additive exPlanations) (Lundberg and Lee, 2017) values stand out as particularly noteworthy. SHAP values play a crucial role in identifying and quantifying the significance of features in the prediction process. Moreover, several works (Hagerty et al, 2019; Shekar and Hailu, 2023; Nancy Jane et al, 2022) showed that the fusion of handcrafted models and convolutional neural network (CNN) can lead to better performance than using models individually. Toward this path of research, we also merge our ANN with a CNN ensemble, as described in Figure 1, to analyze the final performance.

Finally, our results are composed of the ANN model on the basis of the asymmetry features of the CNN ensemble and of the fusion of the 2 approaches.

RESULTS

Dataset

For our experiments, we used the public International Skin Imaging Collaboration 2019 dataset (Codella et al, 2018a[2]; Tschandl et al, 2018). We included only dermoscopic images of melanocytic lesions melanoma and nevus classes because our work focused on the detection of melanoma beyond the nevus. These images are from real clinical practice from different dermatology departments and contain many artifacts such as hairs and bubbles. Examples of these artifacts are visually demonstrated in Supplementary Figure S2. Because the handcrafted features approach requires the extraction of descriptors within the lesions without further interference, a dermatologist removed images containing many artifacts within the lesion (eg, excessive hair, the ruler from the dermoscope). The resulting dataset still contains >7600 dermoscopies, including 6371 nevi and 1301 melanomas. Owing to the class imbalance (there are 4.59 times as many nevi as melanomas), these models have a weight in favor of melanomas implemented as loss multiplied by 4.59 when dealing with melanomas. For our experiments on handcrafted features, we segmented lesions using a U-Net (Ronneberger et al, 2015) trained on the International Skin Imaging Collaboration dataset of 2016 (Gutman et al, 2016[3]) and 2018 (Codella et al, 2018b[4]; Tschandl et al, 2018) to

[2] Codella N, Rotemberg V, Tschandl P, Celebi E, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). arXiv 2018a

[3] Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, et al. Skin lesion analysis toward melanoma detection: a challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). arXiv 2016.

[4] Codella NCF, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the international skin imaging collaboration (ISIC) arXiv 2018b.

remove perilesional skin and ease our computations of handcrafted features.

We did not use the segmentation for CNN, but we used multiple preprocessing functions for data augmentation, such as random rotation, translation, mirror image, and contrast change. Finally, we randomly divided the dataset into 3 sets corresponding to the train, validation, and test set, each containing respectively 70, 10, and 20% of the total data.

Handcrafted model characterizing global asymmetry

To characterize the global asymmetry of melanocytic lesions, we used a first method based on the axial symmetry and a second method based on the central symmetry. For the first method, we divided the lesion into 4 equal parts, named quadrants, and computed multiple features on each one. This step produces a vector per quadrant that we compare with other quadrants by computing the Euclidean distance between each of them. Thus, the final vector for each image contains the distance of features of each quadrant, representing the asymmetry within lesions. More details are given in the Materials and Methods section. This model achieved good performance, with an area under the curve (AUC) of 0.80, a sensitivity of 0.61, a specificity of 0.80, and a balanced accuracy of 0.71. We then added the color histograms per quadrant, and the performance improved: AUC = 0.82, sensitivity = 0.87, specificity = 0.66, and balanced accuracy = 0.77 (Table 1).

The second method uses the asymmetry according to the center of the lesion by computing the difference of color between pairs of pixels that are symmetrically opposed. This model showed good performance, with AUC = 0.81, sensitivity = 0.73, specificity = 0.74, and balanced accuracy = 0.74. Finally, we concatenated all the features together into a single vector that we gave as input to a final ANN, which provided 2 outputs representing probabilities for each image being a melanoma (first output) or a nevus (second output). Details about our ANN architecture, training setup, as well as computation of the features are explained in the Materials and Methods section.

The fusion of all features yielded the highest performance with a balanced accuracy of 0.79, an AUC of 0.87, a sensitivity of 0.90, and a specificity of 0.67. On the final ANN, 82% of melanomas were correctly detected, revealing a good sensitivity and keeping a good specificity. Moreover, the increase in performance during the fusion of our different features means that they are diverse enough and represent different discriminating aspects of the melanocytic lesions. Overall, our developed representation of symmetry in skin lesions proved that it was a discriminant characteristic for detecting melanomas.

CNNs ensemble

For the ensemble of CNN, we used the EfficientNets, which are currently among the most efficient models for image classification (Chandra et al, 2019). We trained and tested all models from B0 to B7 and made an ensemble model consisting of the average of all the predictions. Each model takes as input the dermoscopic images from our dataset and outputs 2 scalars representing the probability of the image being a melanoma or a nevus.

All the CNNs reached high levels of performance (Table 1). The ensemble method based on the average of predictions outperformed every single model by a large margin. As for the majority vote, it achieved a slightly lower balanced accuracy than the mean predictions method (0.85 vs 0.87). The combination of multiple models has indeed led to a considerable increase in performance, albeit with a higher computational demand. Interestingly enough, we can note that the handcrafted features of ANN obtain results relatively similar to those of these models alone, albeit utilizing significantly fewer computing resources.

Late fusion between the CNN ensemble and the handcrafted model characterizing asymmetry

Although the ANN is alone among 7 CNNs, its contribution increases the results of the CNN ensemble as shown in Table 1. When used alone, the ANN performs similarly to those of single CNNs but with very few parameters in comparison (only about 10,000 against several million), and its

Table 1. Results on the Test Set for Each Model

Architecture	Model	Sensitivity	Specificity	Balanced Accuracy	AUC
ANN	Features per quadrants without histograms	0.61	0.80	0.71	0.80
	Features per quadrant with histograms per quadrant	0.87	0.66	0.77	0.82
	Circles on superpixels	0.73	0.74	0.74	0.81
	All features	**0.90**	**0.67**	**0.79**	**0.873**
CNN	EfficientNetB0	0.65	0.90	0.77	0.897
	EfficientNetB1	0.81	0.81	0.81	0.905
	EfficientNetB2	0.81	0.84	0.82	0.905
	EfficientNetB3	0.77	0.86	0.81	0.908
	EfficientNetB4	0.85	0.80	0.83	0.914
	EfficientNetB5	0.88	0.76	0.82	0.906
	EfficientNetB6	0.93	0.67	0.80	0.889
	EfficientNetB7	0.94	0.64	0.79	0.892
	Ensemble (mean predictions)	**0.91**	**0.82**	**0.87**	**0.938**
	Ensemble (vote)	0.92	0.79	0.85	—
Fusion	**ANN + CNNs ensemble**	**0.92**	**0.82**	**0.87**	**0.942**

The best model of each architecture is presented in bold.
Abbreviations: ANN, artificial neural network; AUC, area under the curve; CNN, convolutional neural network.

Table 2. Percentage of Lesions from the Test Set Depending on the Predictions of the CNN Ensemble and ANN, Normalized by Type of Lesion

Type of Lesion	ANN Predictions	CNN Ensemble Correct Prediction	CNN Ensemble Incorrect Prediction
Melanomas	ANN — correct prediction	83%	7%
	ANN — incorrect prediction	8%	2%
Nevus	ANN — correct prediction	62%	5%
	ANN — incorrect prediction	20%	13%

Abbreviations: ANN, artificial neural network; CNN, convolutional neural network.

choices are entirely based on color asymmetry, which makes the predictions understandable for dermatologists. The increase when using the ANN is visible through the receiver operating characteristic AUC as more melanomas are detected, and fewer nevi are misidentified. A likely hypothesis of this increase is that CNNs make similar errors, whereas the ANN brings a different and original way to see the image, which is beneficial. In Table 2, both models correctly predicted 83% of melanomas, and 15% were correctly predicted by 1 model and not the other. These 15% of lesions can theoretically be corrected during the fusion of both methods.

SHAP values for model explainability

SHAP values developed by Lundberg and Lee (2017) highlight more impactful features during prediction, which can provide more insight into why a lesion might be a melanoma or a nevus. In Figure 2, SHAP values for each instance and each feature are represented, sorted by amplitude, and grouped by feature. Lower values (lower difference between quadrants) result in a nevus prediction (the label is 1), whereas higher values (representing an important asymmetry) tend to represent melanomas. We can see in Figure 2 that the most impactful feature is the centroid asymmetry on superpixels using the hue, saturation, value (HSV) color type. Centroid asymmetry seems to be the most impactful feature, followed by histograms of color. Centroid asymmetry might highlight criteria used by dermatologists such as the regression area of melanoma. Although axial asymmetry is a good feature, centroid asymmetry shows more impact, probably owing to the multiple axes of symmetry studied. Although our work only focuses on asymmetry, adding other characteristics, using the ABCD (asymmetrical, border, color, diameter) rule, would probably provide interesting insights into each lesion. Some lesions might be considered melanoma owing to their shape, whereas others might be considered because of their texture, size, or colors. Using multiple space colors, we make the detection of an asymmetric pattern easier because different space colors focus on different characteristics of the image. We can see that HSV indeed allowed better performance, which is probably because it naturally detects more easily changes in HSV, which are more representative of the way we see the world, instead of the classical RGB (red, green, and blue values), which focuses on red, green, and blue values.

DISCUSSION

This study reveals the ability of our handcrafted model based on asymmetry to detect melanoma using machine learning models, and we compared its performance with those of state-of-the-art CNN. The ANN model based on the characterization of color asymmetry within lesions allowed better performance than a conventional image analyzer (AUC = 0.738, sensitivity = 53.4%, specificity = 86.6%, and balanced accuracy = 70%) (Sies et al, 2020), with results competing with individual CNNs.

Analyzing the predicted images (Figure 3) allowed us to better understand the differences between the 2 models. False negatives in the CNNs, which are correctly identified by the ANN, exhibit an asymmetrical color distribution with dark patches. This characteristic likely explains why the ANN accurately classifies them as melanomas. On the contrary, false negatives in the ANN are rare cases where melanomas are symmetrical, and it is less evident why they were correctly detected by the CNNs. As for false positives, the CNN ensemble classified symmetrical lesions with uniform colors as melanomas, and the exact reasons for these models classifying them as such remain unclear. False positives of the ANN mostly consist of highly atypical nevi, which are preferable to analyze through biopsy regardless, whereas the CNNs did not predict these lesions as malignant.

The collaboration between dermatologists and artificial intelligence models is being increasingly studied (Han et al, 2020; Tschandl et al, 2020). Although such collaboration can lead to improved overall performance, instances of misleading artificial intelligence predictions have been observed. These misleading predictions often perplex physicians because they are presented as singular numerical values. Introducing understandable features in health care has the potential to enhance the utilization of artificial intelligence in this domain. This in turn could result in improved performance and greater user confidence in the algorithms. Instead of relying solely on a solitary numerical representation of the model's confidence, furnishing healthcare professionals with a wealth of information could significantly amplify the impact of machine learning models in their daily routines. These informative features encompass attributes such as asymmetry, irregular borders, the number of colors, and texture type, all of which could be provided to experts to facilitate more informed final predictions.

Nonetheless, many limitations remain in our approach. The use of a clean dataset does not fully reflect real-world scenarios, and the development of additional features for analyzing borders, size, and texture could be a key element toward improving the results. Moreover, our current set of features is specifically designed for melanoma detection, which may be insufficient for identifying other types of skin lesions, such as seborrheic keratosis or basal cell carcinoma. Recognizing these conditions might necessitate the creation of new, tailored features, which is a current limitation of handcrafted features. The necessity of the segmentation is

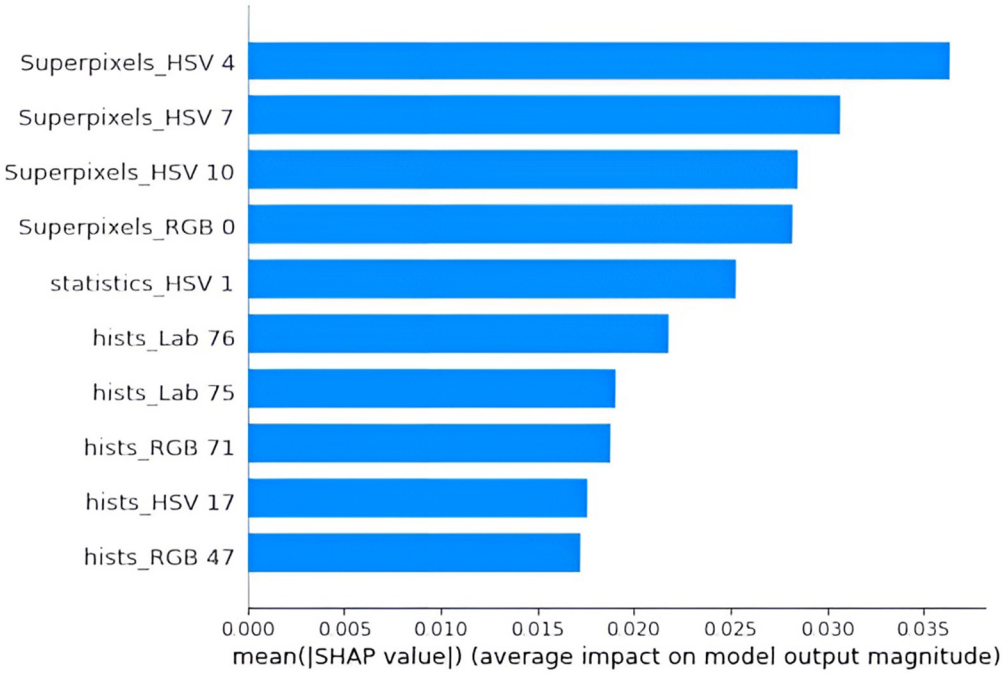


Figure 2. The 10 most impactful features according to the SHAP values. Hists are histograms of colors per quadrant, Superpixels are circles on a superpixel image, and statistics are the fusion of all other statistical features. Color spaces are also mentioned on each feature name. HSV, hue, saturation, value; RGB, red, green, green.

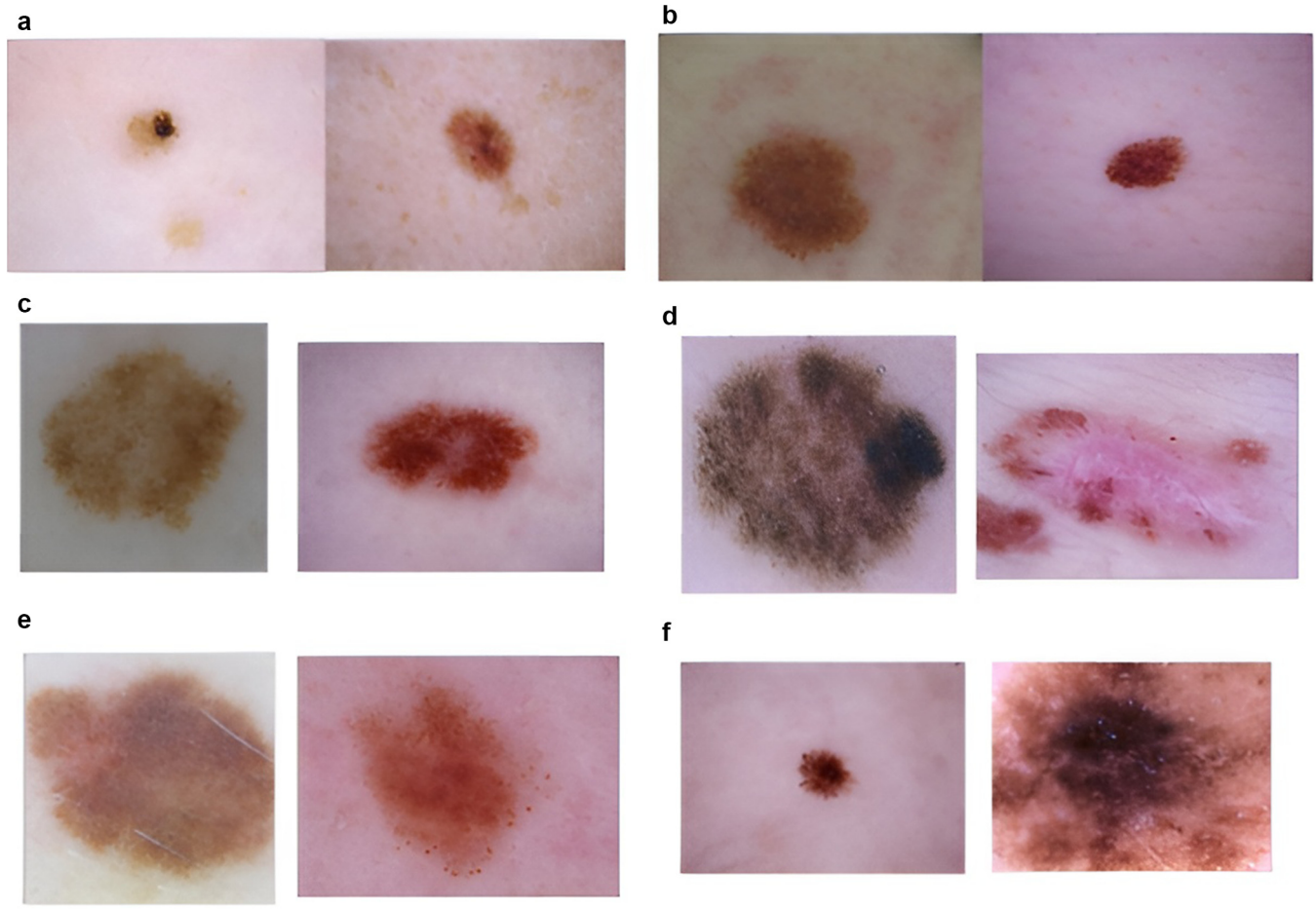


Figure 3. Examples of errors made by one approach but not the other. (**a**) False negatives of the CNN ensemble. (**b**) False positives of the CNN ensemble. (**c**) False negatives of the ANN. (**d**) False positives of the ANN. (**e**) Errors of the CNN ensemble that have been corrected when merging the ANN. Left: a melanoma. Right: a nevus. (**f**) Errors of the ANN corrected by the final fusion. Left: a melanoma. Right: a nevus. Images were sourced from the ISIC 2019 dataset (HAM10000) (Tschandl et al, 2018). ANN, artificial neural network; CNN, convolutional neural network; ISIC, International Skin Imaging Collaboration.

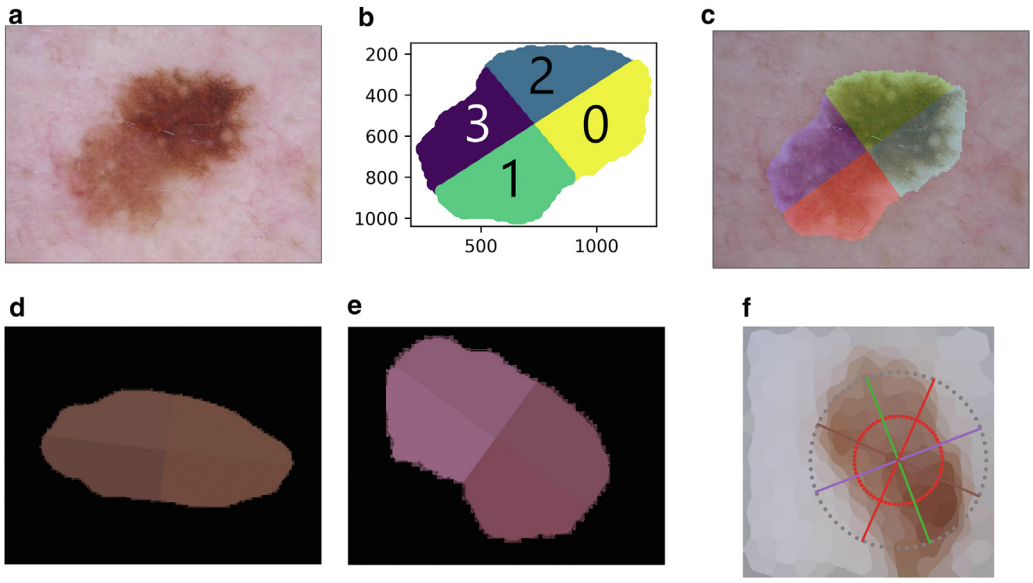


Figure 4. Different steps of the pipeline. (**a**) Original image of a melanoma. **(b)** Computation of the quadrant on the cloud of points. **(c)** Visual representation of the quadrants. (**d**) Mean colors of a nevus. **(e)** Mean colors of a melanoma. **(f)** Visual representation of the pair of points selected to compute asymmetry (same colors are compared). Before comparing the pixels, the image is transformed into a superpixel image. The image has been cropped for visualization purposes. Images were sourced from the ISIC 2019 dataset (HAM10000) (Tschandl et al, 2018). ISIC, International Skin Imaging Collaboration.

also a limitation because some lesions are very difficult to segment when a lot of hair is present in the picture. Algorithms that could remove or deal with these artifacts might make our model more robust. Finally, the computation of all the features takes some time and resources, which could be reduced.

To conclude, developing new ways of diagnosing diseases is an important step toward building a trustful environment for patients and physicians. The ANN based on global asymmetry is inspired by global cognitive approaches from dermatologists and provides insights about how algorithms compute their predictions. We hope that this work will lead to new research projects focusing on other features. Future paths of research could be oriented toward making models focusing on more diverse dermatological knowledge and interfaces that could be used for physicians in real clinical practice.

MATERIALS AND METHODS

In our experiments, all machine learning models were trained on the same dataset. The training was performed on an NVIDIA Tesla K80 using Tensorflow 2.5.0 and Keras 2.5.0 libraries.

Handcrafted features representing melanoma's asymmetry in color

In this section, we will discuss the implementation of our features. The creation of quadrants and the 2 different types of asymmetry (axial and according to the center) are studied. These different calculations are performed only on the lesion, ignoring the skin using a segmentation made with U-Net.

Symmetry axes and quadrants

The 2 principal axes are computed using the principal component analysis on each pixel of the lesion (we ignore skin pixels). Figure 4 shows the different steps of the creation of the quadrants. For all lesions, quadrants are annotated from 0 to 3: quadrant 0 is always symmetrically opposed to quadrant 3 with respect to the center as well as for quadrants 1 and 2.

Features

The descriptors computed from the quadrants are the average color, the SD, the skewness, the kurtosis, and color histograms with 10 bins. Those descriptors are beyond the most commonly used in image analysis (Kavitha and Suruliandi, 2016). Figure 4 shows 2 examples of average color per quadrant for a nevus and a melanoma. The final feature vector is made by concatenating the features from the axial asymmetry and the features from the symmetry according to the center.

Computing color asymmetry using quadrants

For each feature F, the 2 types of asymmetry are defined as follows:

$$A_{axes} = \frac{|F_0 - F_1| + |F_2 - F_3| + |F_0 - F_2| + |F_1 - F_3|}{4}$$

$$A_{center} = \frac{|F_0 - F_3| + |F_1 - F_2|}{2}$$

With F_i being the feature computed on quadrant i of a lesion.

We recall that quadrant 0 is always opposite to quadrant 3 with respect to the center as well as quadrants 1 and 2.

Innovative feature: circles on superpixel images

This feature does not use quadrants and consists of comparing pairs of symmetrical pixels with respect to the center. First, the lesion is transformed into a 200-superpixel image using the SLIC algorithm (Achanta et al, 2012) to regroup similar pixels. Pairs of pixels used are opposite points on circles of different diameters centered on the lesion as shown in Figure 4. The diameter of the circles is defined as follows:

$$D_i = \frac{P_{max}}{(n_{circles}+1)} * i \forall\, i \in [1, n_{circles}]$$

with P_{max} being the furthest distance of a point from the center of the lesion and $n_{circles}$ being the number of circles to use. The furthest distance is not used because it may not contain enough pixels. The comparison is the difference in absolute value between the 2 chosen pixels. Finally, the values are averaged by circle to reduce the final size of the vector without losing much information. The size of the output vector is thus $n_{circles} \times 3$ (3 corresponding to the 3 color channels). For our final vector, we used 3 circles and a total of 6 axes.

Preprocessing of features and color spaces

To achieve better results, we extracted features according to several color spaces such as RGB, HSV, YCbCr, and CieLab. These space colors are widely used in computer vision to highlight different characteristics of the images and potentially increase performance of models (Shaik et al, 2015). HSV describes colors in terms of hue, saturation, and value, aiding the comparison of parts of a lesion with different hue; YCbCr separates luminance from chrominance; and CieLab is designed to mimic human perception, facilitating consistent color comparisons. Several experiments have been conducted, and the use of different color spaces indeed improves the results. Finally, features are normalized using the MinMax algorithm. The values of the final vector are thus between 0 and 1.

Selected model

Our ANN takes inputs of size 714, which corresponds to the concatenation of the features on the different color spaces. Its architecture, presented in Supplementary Figure S1, is composed of a first layer of size 16 with ReLU activation function, then a dropout of 0.2, and a final prediction layer of size 2. The training is performed on 50 epochs, with a batch size of 64 and a weight per class of 4.59 for melanomas and 1 for nevi. The loss function used is the categorical cross-entropy, and the optimizer is Adam, with a learning rate of 0.001. Finally, the model has 11,474 parameters. To obtain our results for each feature in Table 1, we used the same architecture, with the only difference being the size of the input.

Analyzing SHAP values

To visualize SHAP values, we chose a Gradient Boosting model and analyzed the impact of each feature as shown in Figure 3.

CNN approach

For our CNN approach, we chose the EfficientNet family because these models currently perform best in many computer vision problems while having a relatively light architecture in terms of number of parameters.

CNN training

The training is based on 2 parts: all EfficientNets are retrieved with ImageNet weights, and we freeze all layers. The last layer is removed and replaced by a softmax layer for binary classification, which will be the only layer trained in a first training. Once this layer is trained, a second training is performed with the weights unfrozen and a low learning rate to perform the fine tuning. Finally, the weights are saved to reuse these models for the ensemble approach. The first training is performed with 15 epochs, Adam optimizer with a learning rate of 0.001, and the binary cross-entropy loss function. Because of the class imbalance in our dataset, a weight of 4.59 is given to the melanoma class, whereas the nevus class has a weight of 1. In the second training, the learning rate is 10e-4 and lasts 5 epochs. The other parameters are the same as the first training.

Data augmentation

To improve the learning of the model, artificial image augmentation is necessary. It consists of creating new images from the existing ones to improve the generalization of the model. Our method performs random operations on the images before passing them through the model. Selected operations are rotation, translation, mirror image, and contrast change.

Ensemble method

The state of the art consists of training several classifiers independently and then grouping them together to make predictions. To achieve this, 2 late fusion methods have been performed: average predictions and voting. The average predictions consist of computing the average of the predictions of all models, whereas the voting method predicts the most predicted class. The models used for this method are EfficientNets from B0 to B7. The training was the same for all of them except for the input size of the images. The instances are resized for each model to their corresponding size. As for the number of parameters, they vary between 5 million for B0 and 66 million for B7, unlike the previous ANN with just over 11,000 parameters.

DATA AVAILABILITY STATEMENT

The original dataset of International Skin Imaging Collaboration 2019 is available on the official website: https://challenge.isic-archive.com/landing/2019/. We created a subset of the International Skin Imaging Collaboration 2019 dataset by excluding images that contained artifacts. This subset is accessible through the following link: https://drive.google.com/file/d/1SiTLTzxhI60fj1fFxaBri1SiM-VgrMZL/view?usp=drive_link

ORCIDs

Jules Collenne: http://orcid.org/0000-0002-7540-0610
Jilliana Monnier: http://orcid.org/0000-0002-8747-5011
Rabah Iguernaissi: http://orcid.org/0000-0002-1728-3532
Motasem Nawaf: http://orcid.org/0000-0003-2156-8589
Marie-Aleth Richard: http://orcid.org/0000-0002-0870-9132
Jean-Jacques Grob: http://orcid.org/0000-0002-0667-153X
Caroline Gaudy-Marqueste: http://orcid.org/0000-0001-6955-5117
Séverine Dubuisson: http://orcid.org/0000-0001-7306-4134
Djamal Merad: http://orcid.org/0000-0001-9233-0020

CONFLICT OF INTEREST

The authors state no conflict of interest

ACKNOWLEDGMENTS

The project leading to this publication has received funding from the Excellence Initiative of Aix-Marseille University - A*MIDEX, a French "Investissements d'Avenir" programme (AMX-19-IET-002). This research was funded by the French National Research Agency through the DIAMELEX project (https://anr.fr/Projet-ANR-20-CE45-0026, accessed on April 11, 2023). The work was done in Marseille, France.

AUTHOR CONTRIBUTIONS

Conceptualization: JC, JM, RI, DM; Methodology: JC, RI, DM; Investigation: JC, JM, RI; Visualization: JC, JM, RI, DM; Funding acquisition: JM, DM, RI; Project administration: RI, DM; Supervision: JM, RI; Writing — Original Draft Preparation: JC, JM; Writing — Review and Editing: RI, DM, JM, SD

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2023.09.289.

REFERENCES

Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 2012;34:2274–82.

Chandra TG, Nasution AMT, Setiadi IC. Melanoma and nevus classification based on asymmetry, border, color, and GLCM texture parameters using deep learning algorithm. AIP Conf Proc 2019;2193:050016.

Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: an overview. Int J Cancer 2021;149:778–89.

Forsea AM. Melanoma epidemiology and early detection in Europe: diversity and disparities. Dermatol Pract Concept 2020;10:e2020033.

Hagerty JR, Stanley RJ, Almubarak HA, Lama N, Kasmi R, Guo P, et al. Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images. IEEE J Biomed Health Inform 2019;23:1385–91.

Han SS, Park I, Eun Chang SE, Lim W, Kim MS, Park GH, et al. Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. J Invest Dermatol 2020;140:1753–61.

Kavitha JC, Suruliandi A. Texture and color feature extraction for classification of melanoma using SVM. A paper presented at: 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16). January 2016; Kovilpatti, India.

Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Paper presented at: 31st Conference on Neural Information Processing System (NIPS 2017). January 2017; Long Beach, CA.

Milczarski P, Stawska Z. Classification of skin lesions shape asymmetry using machine learning methods. In: Barolli L, Amato F, Moscato F, Enokido T, Takizawa M, editors. Artificial intelligence and network applications. Cham, Switzerland: Springer; 2020. p. 1274–86.

Nancy Jane Y, Charanya SK, Amsaprabhaa M, Jayashanker P, Nehemiah HK. 2-HDCNN: a two-tier hybrid dual convolution neural network feature fusion approach for diagnosing malignant melanoma. Comput Biol Med 2022;152:106333.

Nezhadian FK, Rashidi S. Melanoma skin cancer detection using color and new texture features. Paper presented at: 2017 Artificial Intelligence and Signal Processing Conference (AISP). October 2017; Shiraz, Iran.

Oukil S, Kasmi R, Mokrani K. Colors skin lesions detection for melanoma discrimination. Paper presented at: The Electrical Engineering International Conference EEIC;'19. December 2019; Bejaia, Algeria.

Pham TC, Luong CM, Hoang VD, Doucet A. AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function. Sci Rep 2021;11:17485.

Rastgoo M, Garcia R, Morel O, Marzani F. Automatic differentiation of melanoma from dysplastic nevi. Comput Med Imaging Graph 2015;43:44–52.

Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Med Image Comput Comput Assist Interv MICCAI Navab N, Hornegger J, Wells WM, Frangi AF, editors, Medical image computing and computer-assisted intervention – MICCAI 2015 lecture notes in computer science. Cham, Switzerland: Springer; 2015.

Shaik KB, Ganesan P, Kalist V, Sathish BS, Jenitha JMM. Comparative study of skin color detection and segmentation in HSV and YCbCr color space. Procedia Comput Sci 2015;57:41–8.

Shekar BH, Hailu H. Fusion of features extracted from transfer learning and handcrafted methods to enhance skin cancer classification performance. In: Tistarelli M, Dubey SR, Singh SK, Jiang X, editors. Computer vision and machine intelligence. Lecture Notes in Networks and Systems. Singapore, Singapore: Springer; 2023. p. 243–57.

Sies K, Winkler JK, Fink C, Bardehle F, Toberer F, Buhl T, et al. Past and present of computer-assisted dermoscopic diagnosis: performance of a conventional image analyser versus a convolutional neural network in a prospective data set of 1,981 skin lesions. Eur J Cancer 2020;135:39–46.

Srivastava R, Ong EP, Lee BH, Tan LS, Tey HL. Quantitative comparison of color asymmetry features for automatic melanoma detection. Annu Int Conf IEEE Eng Med Biol Soc 2021;2021:3753–6.

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.

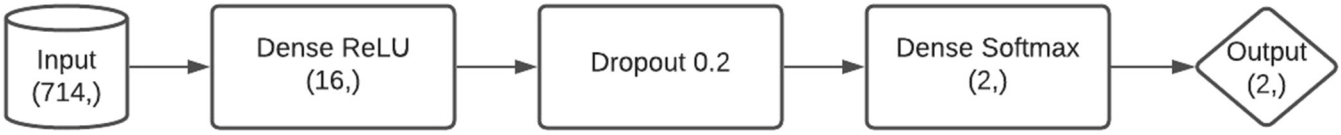
Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. Nat Med 2020;26:1229–34.

Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018;5:180161.

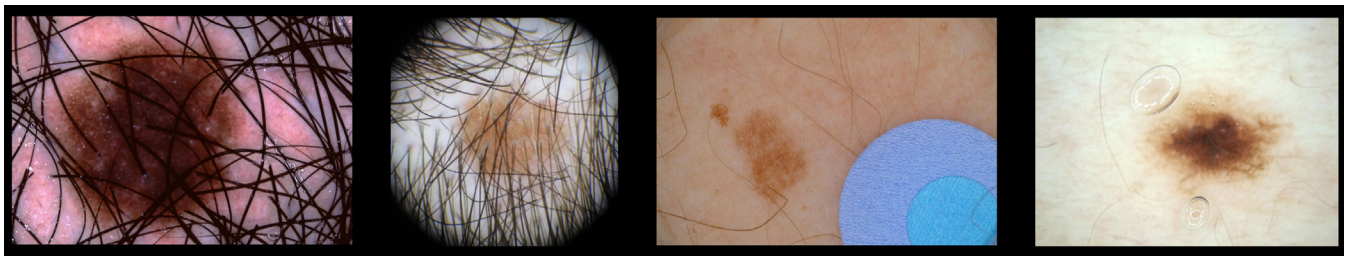
Zghal NS, Kallel IK. An effective approach for the diagnosis of melanoma using the sparse auto-encoder for features detection and the SVM for classification. Paper presented at: 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). September 2020; Sfax, Tunisia.

SUPPLEMENTARY REFERENCE

Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018;5:180161.



Supplementary Figure S1. Architecture of the selected ANN for handcrafted features. ANN, artificial neural network.



Supplementary Figure S2. Example of removed images containing artifacts (presence of hair, dark halo, colored dot, and bubbles). Images were sourced from the ISIC 2019 dataset (HAM10000) (Tschandl et al, 2018). ISIC, International Skin Imaging Collaboration.