Computers in Biology and Medicine 116 (2020) 103545

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

The impact of patient clinical information on automated skin cancer detection

Andre G.C. Pacheco [a,*], Renato A. Krohling [a,b]

[a] Graduate Program in Computer Science, PPGI, UFES - Federal University of Espírito Santo, Av. Fernando Ferrari 514, Vitória CEP: 29060-270, Brazil
[b] Production Engineering Department, UFES - Federal University of Espírito Santo, Av. Fernando Ferrari 514, Vitória CEP: 29060-270, Brazil

ARTICLE INFO

Keywords:
Skin cancer detection
Deep learning
Data aggregation
Clinical images
Clinical information

ABSTRACT

Skin cancer is one of the most common types of cancer worldwide. Over the past few years, different approaches have been proposed to deal with automated skin cancer detection. Nonetheless, most of them are based only on dermoscopic images and do not take into account the patient clinical information, an important clue towards clinical diagnosis. In this work, we present an approach to fill this gap. First, we introduce a new dataset composed of clinical images, collected using smartphones, and clinical data related to the patient. Next, we propose a straightforward method that includes an aggregation mechanism in well-known deep learning models to combine features from images and clinical data. Last, we carry out experiments to compare the models' performance with and without using this mechanism. The results present an improvement of approximately 7% in balanced accuracy when the aggregation method is applied. Overall, the impact of clinical data on models' performance is significant and shows the importance of including these features on automated skin cancer detection.

1. Introduction

Skin cancer occurs when skin cells are damaged, for example, by overexposure to ultraviolet (UV) radiation from the sun [1]. Unlike other forms of cancer, the reporting of skin cancer diagnoses is not required by most cancer registries worldwide [2]. Regardless, the World Health Organization (WHO) estimates that one in every three instances where a person is diagnosed with cancer is a skin cancer [3]. Over the recent decades, in countries such as Canada, USA, and Australia, the number of people diagnosed with skin cancer has been increasing at a constant rate [1,4,5]. In Brazil, according to the Brazilian Cancer Institute (INCA), skin cancer accounts for 33% of all cancer diagnoses in the country. This is the highest rate among all forms of cancer, and for 2018–2019 it is expected 180 thousand new cases across the nation [6].

There are three main types of skin cancer: basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and melanoma. Melanoma is the rarest type of skin cancer; however, because of the high level of metastasis,[1] it is the most lethal one. On the other hand, BCC and SCC, which are known as keratinocyte cancers, represent the major skin cancer occurrence. As they rarely metastasize, they have low lethality risk [1]. In order to diagnose skin cancer, dermatologists screen the suspicious skin lesion using their experience to identify it. They also take into account clinical information[2] such as the patient's age, where the lesion is located, if the lesion bleeds, among others [7,8]. While these features are essential for dermatologists, distinguishing skin lesions from skin cancer remains challenging. Therefore, to increase the diagnostic reliability, dermatologists use the dermatoscope, a medical instrument that allows the visualization of subsurface structures of the skin revealing lesion details in colors and textures not normally visible to the naked eye [9].

Kittler et al. [10] and Sinz et al. [11] have shown that the use of a dermatoscope can increase diagnostic accuracy. Nonetheless, they also concluded that this improvement significantly depends on the dermatologist's experience, i.e., less experienced examiners do not present improvement using the dermatoscope. This reason is sufficient to justify the need for computer-aided diagnosis (CAD) systems for skin cancer. Additionally, in emerging countries such as Brazil, there is a strong lack of dermatologists and dermatoscopes in most of its countryside cities. Thus, an automated system to assist doctors in skin cancer diagnosis that does not depend on dermoscopic images is very desired. However, designing such a system is a challenging task.

Over the past decades, different computer-aided diagnosis (CAD) systems have been proposed to tackle skin cancer detection. Pioneer

* Corresponding author.
E-mail addresses: agcpacheco@inf.ufes.br (A.G.C. Pacheco), rkrohling@inf.ufes.br (R.A. Krohling).

[1] when damaged cells invade other parts of the body via blood vessels and lymph vessels.

[2] In this paper, the terms clinical information, clinical data, and clinical features are used interchangeably.

https://doi.org/10.1016/j.compbiomed.2019.103545
Received 3 September 2019; Received in revised form 14 November 2019; Accepted 14 November 2019
Available online 18 November 2019

works, such as [12–14], reported the use of low-level handcrafted features to differentiate melanomas and keratinocyte cancers. Later, different computational approaches have been developed based on ABCD(E) rule, pattern analysis, and 7-point checklist, which are common methods used by dermatologists to diagnose skin cancer [15,16]. These approaches mostly use traditional computer vision algorithms to extract various features such as shape, color, and texture [17–21], to feed a classifier, for example, a support vector machine (SVM) [22–24]. Two weaknesses may be found in these approaches. First, the ABCD(E) rule and the 7-point checklist were designed only for pigmented lesions, which means they cannot be used to diagnose BCC or SCC. Second, the handcrafted features extracted by these methods have limited generalization capability [25].

Recently, deep learning models have been achieving remarkable results in different medical image analysis tasks [25,26]. In particular, convolutional neural networks (CNN) have become the standard approach to handle this kind of problem [27,28]. Several deep learning models have been proposed for automated skin cancer detection. Yu et al. [25] presented a very deep CNN and a set of schemes to learn under limited training data. Esteva et al. [29] used a pre-trained GoogleNet CNN architecture [30] to train over 120 thousand images and achieved a dermatologist-level diagnostic. Haenssle et al. [31] and Brinker et al. [32] also used deep learning models to compare their performance to dermatologists. In both works, the models have shown competitive or outperformed the dermatologists. Other efforts using deep learning have been made to detect skin cancer, such as an ensemble of models [33,34], feature aggregation of different models [35], among others [36–40]. Most of these works are based on dermoscopic images, mainly for two reasons: (1) there is an open well-known dataset provided by the International Skin Imaging Collaboration (ISIC) [41,42]; (2) getting a dataset of clinical images of skin cancer is a hard task. Developing CAD systems to work with dermoscopic images is important; however, as mentioned earlier, emerging countries do not have dermatoscopes available in most of their regions, which constraints the use of these systems. In addition, there is a trend in developing CAD systems embedded in smartphones, either for general users or to assist doctors [43,44]. Using smartphones to assist in skin cancer detection is promising; however, this method requires clinical images instead of dermoscopic ones.

Beyond the lack of CAD systems using clinical images, most of the proposed approaches do not consider patient clinical information, which is an important clue towards a more accurate diagnosis [8]. In fact, dermatologists do not trust only on image screening; they also use clinical data to improve their detection accuracy. In this context, Brinker et al. [45] presented a review for deep learning models applied to skin cancer detection and concluded that an improvement in classification quality could be achieved by adding clinical data in the classification process. Based on this idea, Kharazmi et al. [46] proposed a deep learning approach to detect BCC using dermoscopic images and five clinical features. This approach presented a superior performance when these features were considered by the model. However, they did not analyze their effect on model prediction. Also, their method can deal with only one type of skin cancer and does not consider clinical images.

In this work, we present a study to analyze the impact of patient clinical information on deep learning models applied to skin cancer detection. In order to achieve this goal, we introduce a new skin cancer dataset composed of clinical images and clinical data as well as an approach to aggregate them. In the following, we summarize the main contributions of this work:

- In partnership with the Dermatological Assistance Program (PAD) at the Federal University of Espírito Santo (UFES), which is a nonprofit organization that provides skin lesion treatment for low-income people in Brazil, we developed a smartphone application to collect skin lesion images and clinical data. From this software, we built a new dataset composed of clinical images and clinical data related to the patient. Since collecting this type of data is difficult, we intend to make this dataset available for research purposes. As far as we know, there is no public skin cancer dataset that contains clinical images and patient clinical information available in the literature.
- We use well-known deep learning models to develop an approach to aggregate features from clinical images and clinical data. This approach introduces a straightforward mechanism to control the contribution of each source of data.
- We present a comprehensive study to show the impact of patient clinical information on skin cancer detection. We analyze the models' performance with and without using this information to show how it impacts their predictions.

The rest of this paper is organized as follows: in Section 2, we present the methods and data; in Section 3 is presented experiments and results; in Section 4, we provide a discussion about the results; and in Section 5 we draw our conclusions.

2. Material and methods

In this section, first, we describe the details of the collected dataset. Next, we present a data exploration analysis to understand the patterns of the clinical data. Last, we describe the deep models applied in this work and our method to aggregate features from images and clinical data.

2.1. Dataset

In order to collect clinical images and patient clinical information, we developed a smartphone-based application to the Dermatological Assistant Program (PAD) of the Federal University of Espirito Santo (UFES). Through this application, doctors and medical students attach one or more images of the skin lesion[3] as well as the clinical data related to the patient. Thereby, each sample in this dataset has a clinical diagnosis, an image, and eight clinical features: the patient's age, the part of the body where the lesion is located, if the lesion itches, bleeds or has bled, hurts, has recently increased, has changed its pattern, and if it has an elevation. These features are based on common questions that dermatologists ask patients during an appointment [47,48]. The application also allows tracking of all patient lesions to follow its evolution throughout time. Evolution is an important feature, and it stands for the E in the ABCD(E) rule. Despite its importance, this information will take some years to become available since the lesion may take some time to increase its size and the patient needs to return to the clinic for an assessment. In this context, when dermatologists ask the patient if the lesion has increased and if it has changed its pattern, they are trying to get information about the lesion's evolution. It is important to note that clinicians collect this information by questioning the patient, which may lead to imprecision and uncertainty. Therefore, clinical data must be used to support the diagnosis. The main information is still coming from the image.

Regarding the region of the body where the skin lesion is located, there are over 120 anatomical regions used by the dermatologists. Based on the PAD's dermatologists experience, we grouped all regions in 15 macro-regions that are more frequent and have more potential to raise a skin lesion, they are: face, scalp, nose, lips, ears, neck, chest, abdomen, back, arm, forearm, hand, thigh, shin and foot. As skin lesions have preferences for some regions of the body [7,8], it is an important feature to consider.

[3] we name wounds, moles or spots on the skin as skin lesions. After the diagnosis, the skin cancers will be named as so and the remaining ones will be called by skin diseases.

Table 1
The number of samples for each type of skin lesion present in the PAD dataset.

Clinical diagnosis	N° of images
Actinic Keratosis (ACK)	543
Basal Cell Carcinoma (BCC)	442
Melanoma (MEL)	67
Nevus (NEV)	196
Squamous Cell Carcinoma (SCC)	149
Seborrheic Keratosis (SEK)	215
Total	**1612**

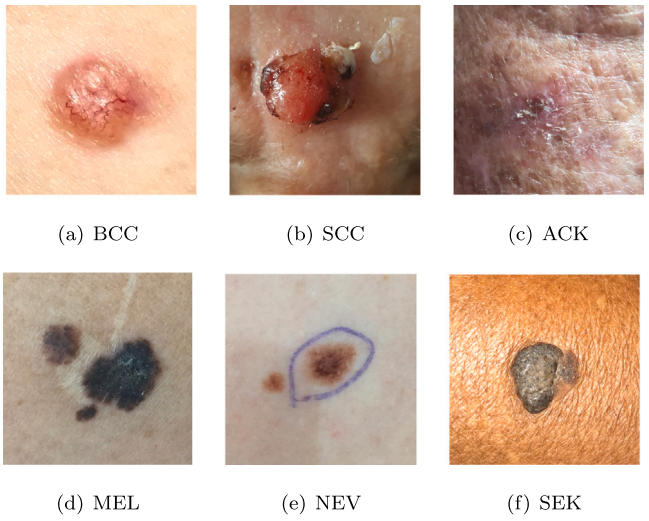


Fig. 1. Samples of each type of skin lesion from the PAD dataset. SCC, BCC, and MEL are skin cancers and NEV, SEK and ACK are skin diseases. In addition, we may note that MEL, NEV, and SEK are pigmented skin lesions and SCC, BCC and ACK are non-pigmented ones.

We have been collecting this dataset for one year and a half. There are over fifty types of skin lesions that were collected by our software. However, most of them are rare and contain only a few samples, which is problematic for deep learning models. For this work, we decided to use the eight most common skin lesions diagnosed at PAD, which are: Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), Actinic Keratosis (ACK), Seborrheic Keratosis (SEK), Bowen's disease (BOD), Lentigo Maligna (LEM), Melanoma (MEL), and Nevus (NEV). As the Bowen's disease and Lentigo Maligna are considered SCC in situ and MEL in situ [7], respectively, we clustered them together, which results in six skin lesions in the dataset. In Table 1 is detailed the number of samples for each type of skin lesion present in the dataset. It is important to note that this dataset has three skin cancers, BCC, SCC and MEL, and three skin diseases, ACK, SEK, and NEV. As stated before, MEL is the most dangerous, but the rarest one, while SCC and BCC are the most common type of skin cancer. The frequency of these samples in the dataset is in agreement with this statement, which makes the PAD dataset imbalanced. In fact, having imbalanced labels is a peculiarity of skin cancer datasets. For instance, the same issue is present in the ISIC dataset [41] and we need to find solutions to deal with this problem.

In Fig. 1 is shown an example for each type of skin lesion present in our dataset. As the images are collected using different smartphone cameras, they come with different sizes and resolutions. As a result, approximately 17% of the collected images were removed from the dataset because of poor resolution, focus, or blur. To the best of our knowledge, this is the first public skin cancer dataset composed of images collected using smartphones and their respective clinical information. This dataset is available upon request.

2.2. Clinical features analysis

To understand the influence of patient clinical information on skin cancer detection, we performed a data exploration analysis using the eight clinical features described in the previous section. In Fig. 2 is shown the plots of the main findings for this set of data. On the left, we observe two bars plots related to bleeding and pain features. As we can see, they are useful to distinguish the pigmented (NEV, MEL, and SEK) from the non-pigmented lesions (ACK, BCC, and SCC). Further, we observe that the ACK normally does not hurt. In general, only SCC and BCC are painful lesions.

On the right side of Fig. 2, we can see a bar plot for the feature itching and a box plot for the patient's age. For itching, it is possible to observe that pigmented lesions usually itch more than the non-pigmented ones. Regarding the patient's age box plot, we can note that the median age for NEV is lower than MEL and SEK. Thereby, this feature is useful to differentiate these lesions. On the other hand, for ACK, SCC, and BCC, the age presents the same interval of values. Therefore, unlike the pigmented lesions, age does not seem to be a significant predictor for the non-pigmented lesions.

The diagram in the center of the figure is a bar plot that presents the diagnosed skin lesion by body region showing how some lesions may occur more or less frequently on different parts of the body. For instance, ACK appears more in the forearm, NEV in the back and SCC, BCC, MEL, and SEK in the face. We also analyzed the remaining collected features. In summary, we observe that MEL and ACK do not present an elevation on the skin, which distinguishes them from the other lesions;usually, only MEL changes its pattern, which is an important feature to detect this type of cancer; finally, for the feature "has increased recently", it is hard to find a pattern, however, ACK does not present an increase.

We summarize the presented analysis as follows:

- It is expected that these features improve the model performance for pigmented and non-pigmented lesions detection.
- Certain features, such as a change in the lesion pattern and elevation are important for MEL detection.
- In general, SCC and BCC share the same clinical features values. Both bleed, hurt, itch, present elevation, occur in the same age range and have the same preferred region. Thus, it is expected that these features will not be helpful in distinguishing between SCC and BCC.

2.3. Convolutional Neural Networks

The Convolutional Neural Network (CNN) is a special type of Neural Network (NN) that was developed to learn visual features from images. Nowadays, it is the most successful deep learning method to handle the image classification task [49]. The standard CNN is basically composed of three layers: convolutional, pooling, and fully-connected layers. The convolutional layer is the most important and performs most of the computation. It consists of kernels, which are composed of weights that learn visual features from the input images. Each kernel is convolved across the whole image and produces a feature map, which is the output of this layer. The pooling layer is used basically to reduce the feature map size. Consequently, it reduces the number of training parameters for the next layers, helps to control overfit, and, along with a non-linear activation filter, it includes non-linearity to the network. The fully-connected layer is a standard NN that is connected to the last feature map provided by the previous layer. In summary, the composition of convolutional and pooling layers is known as the feature extractor, and the fully-connected layer is the classifier.

Different CNN architectures have been used to deal with skin cancer detection. Successful results have been reported by Steva et al. [29] using GoogleNet [50], Yu et al. [35] with ResNet [51] and Menegola et al. [52] using VGGNet [53]. Nonetheless, for medical tasks,

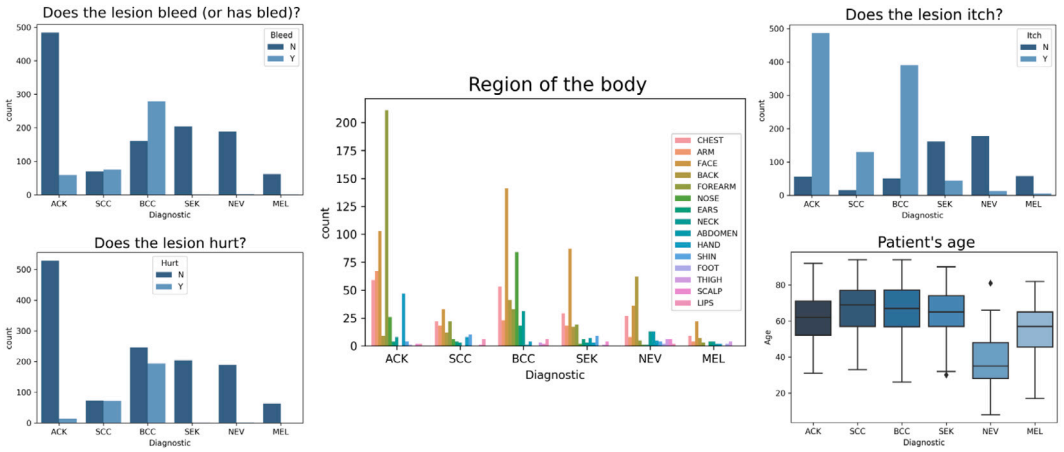


Fig. 2. Exploratory analysis of the patients clinical features.

collecting a large amount of data to train a CNN is challenging. To overcome this issue, all these works used transfer learning, a well-known technique where a model trained for a given source task is partially reused for a new target task [52]. Thereby, the models were initialized using the weights from the ImageNet dataset [54] and then fine-tuned using their own dataset.

In this work, our goal is to investigate the impact of patient clinical information on skin cancer detection. Thus, we decided to use the CNNs previously mentioned, i.e., GoogleNet, ResNet50/101, VGGNet13, and we also included MobileNet [55]. Next, we briefly describe each network:

- **GoogleNet:** this network introduced the inception module, which is an approach based on several tiny convolutions that reduce the number of parameters to be optimized in the training phase [50]. It is composed of 9 stacked inception modules that lead to 22 convolutional layers. Among these layers are applied the pooling operation, batch normalization, and non-linear activations with ReLU. The feature extractor outputs 1024 image features for the classifier.
- **VGGNet-13/19-bn:** this CNN architecture consists of 13/19 layers composed of small convolutional filters [53]. It also includes batch normalization, non-linear activations with ReLU, and pooling layers after two or three convolutions. The feature extractor outputs 25088 image features for the classifier.
- **ResNet-50/101:** this model reformulates the layers as learning residual functions regarding the layer inputs, instead of learning unreferenced functions [51]. It is achieved using an approach that skips connections of some layers and applies batch normalization along with non-linearities (ReLU). In this work, we use two ResNet versions, one containing 49 convolutional layers and another one containing 100. Both feature extractors return 2048 image features for the classifier.
- **MobileNet:** this CNN architecture is based on depth-wise separable convolutions, which is composed of a depth-wise convolution followed by point-wise convolution. This strategy significantly reduces the number of parameters when compared to the networks that use standard convolutions with the same depth [55]. It also introduces two hyper-parameters to control the model's size. In this work, we use the MobileNet with its full size, which leads to 22 convolutional layers. Similar to the others, it also uses ReLU and batch normalization. The feature extractor returns 1024 image features for the classifier.

For all these networks, we kept the feature extractor layers and changed the classifier to include the patient clinical information. This method is described in the next subsection.

2.4. An approach to combine clinical images and patient clinical information

In order to consider the patient clinical information on skin cancer detection, we need to provide a way to combine it with the clinical images. The most common approach to combine metadata with images is through concatenation, i.e., we apply an algorithm to extract the image features and concatenate them along with the metadata [56,57]. Nonetheless, it is problematic to use a similar approach for our task because the number of features that comes from the images is much higher than the clinical ones, which means we need a feature reducer. In this context, we propose an approach that uses an NN to reduce the image features based on a mechanism to control the influence of each set of features.

In Fig. 3 is illustrated the main concept of the proposed aggregation. As we can observe, the CNN feature extractor is kept and we do not change anything in these layers. Next, the last feature map is flattened and outputted as image features. These features are forwarded to the feature reducer block. This block is composed of a standard neural network that is trained to work as a non-linear feature reducer. The architecture of the reducer network varies according to the CNN architecture. Further, the number of features outputted by this block is based on the combination factor (c_f), which is a mechanism to control the amount of nodes/image features that will be used in the next block. Using this mechanism, we control the amount of information that each source contributes to the concatenation and, consequently, to the classifier. Considering N_{img} and N_{cli} the number of features that comes from the images and clinical data, respectively, to compute the total number of features T that will feed the classifier, we need to combine both sets of features according to c_f:

$$T = \lceil c_f N_{img} + (1 - c_f) N_{cli} \rceil \tag{1}$$

where $0 \leq c_f \leq 1$ and $\lceil\rceil$ is the ceil operator. As we can note from Eq. (1), the combination factor c_f controls the amount of feature provided by the image and the clinical data. In this work, we do not intend to reduce the amount of clinical data since it is already small compared to the image features. Thereby, we set $T = \frac{N_{cli}}{(1-c_f)}$, which N_{cli} is the current value of the clinical features, and we compute N_{img} as follows:

$$N_{img} = \left\lceil \frac{N_{cli}}{1 - c_f} - N_{cli} \right\rceil \tag{2}$$

Based on Eq. (2), we are keeping the same value of clinical features and varying the value of image features. In Section 3.3, we present a sensitivity analysis regarding the combination factor.

Since we determined the contribution of each source, both sets of features are concatenated and sent to the classifier, as shown in Fig. 3.

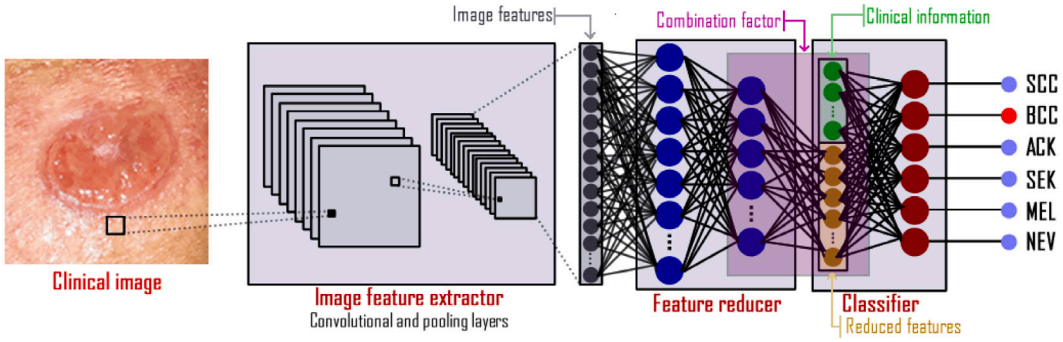


Fig. 3. Illustration of the methodology used to combine the image features with the patient clinical information. First, the image features are extracted by the CNN feature extractor. Next, these features are reduced by the feature reducer block and combined with the clinical data. Finally, the combined features are inputted to the classifier that outputs the final diagnostic.

The classifier is another neural network that assigns the probabilities for each skin lesion. The entire model presented in this section is trained by an end-to-end backpropagation. This is the main reason we employed a neural network as the reducer block, i.e., it can be optimized along with the CNN feature extractor and classifier by including it in the backpropagation [58] computation. We could choose a method such as PCA [59], however, beyond the fact it is linear, the proposed approach is faster and simpler since the backpropagation is already used to train the image feature extractor block. Since the computational cost to train CNNs is high, designing approaches that take advantage of its training process is very desired.

3. Experiments and results

In this section, we present the experiments carried out using the presented dataset and the models described in the previous section. First, we describe how we pre-processed the dataset. Next, we present the common setup for the experiments, a sensitivity analysis regarding the combination factor, and a comparison between the models using only the clinical images and combining them with patient clinical data.

3.1. Data pre-processing

The dataset introduced in Section 2.1 presents similar characteristics as any medical dataset. As we may observe in Table 1, the amount of data is not large and the dataset is imbalanced. In addition, as the dataset is collected using smartphone cameras, it presents fewer details of the skin lesion when compared to dermoscopic images. Also, camera resolution and illumination affect the images' quality. Vasconcelos and Vasconcelos [60] presented a work in which they discuss approaches to handle these types of issues in skin cancer datasets. According to them, we can tackle these problems with transfer learning, up/down-sampling, data augmentation, and using an ensemble of models. Additionally, Barata et al. [20] have shown the benefits of using color constancy algorithms for skin cancer detection. Following their recommendations, we applied the shades of gray method [61] for all images before the training phase. The difference between the clinical images with and without the color constancy pre-processing can be seen in the samples depicted in Fig. 4.

As described in Section 2.3, we use transfer learning for all models. In addition, we applied a strong data augmentation using common image processing operations [62]. We adjust brightness, contrast, saturation, and hue. We also apply horizontal and vertical rotations, translations, re-scale, shear, Gaussian noise, and blur. To tackle the labels' imbalanced issue, we use a weighted loss function based on the label's frequency. The weight for a given label i is computed as follows:

$$w_i = \frac{N}{n_i} \tag{3}$$

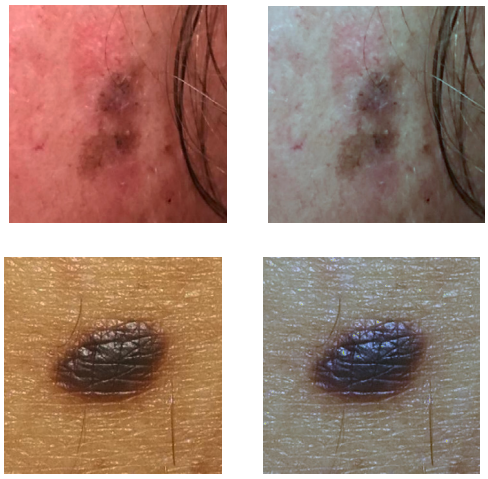


Fig. 4. The difference between the original images (left) and the images after the color constancy pre-processing (right).

where N is the total number of samples in the dataset and n_i is the number of samples of label i. We also tried to use up-sampling to equalize the number of samples for each class. However, it created a high bias for the MEL label, which resulted in a lower performance compared to the weighted loss function approach.

Regarding the clinical features, we applied the one-hot strategy to encode the features represented by strings. Thus, instead of 8 values, after the one-hot encoding, the total number of clinical features (N_{cli}) is an array of 28 values. As described in Eq. (2), the size of this array will be used to compute N_{img}.

3.2. Experiments setup

To verify the impact of the patient clinical information on the deep learning models, we carried out the experiments considering two scenarios:

- Scenario 1: the models consider only the clinical images
- Scenario 2: the models combine the clinical images and patient clinical information

As noted earlier, for each model, we use its original image feature extractor block. Thus, this block is the same for both scenarios. For scenario 2, we need to apply the reducer block described in Section 2.4. For all models, except VGGnet, the reducer block is composed of one layer, where the number of neurons is defined by the combination factor. As the number of image features outputted by VGGnet is over 25 thousand, the only difference for this model is that we add an

Table 2
The number of features from both sources varying c_f.

c_f	N_{img}	N_{cli}	**Total**
0.5	28	28	56
0.6	42	28	70
0.7	66	28	94
0.8	112	28	140
0.9	252	28	280

intermediate layer containing 1024 neurons. For scenario 1, we tested all models with and without the feature reducer block. Both approaches presented the same results. Thus, for simplicity, we keep the reducer block for this scenario. Also, the reducer block uses the ReLU activation function and a dropout rate equal to 0.5.

For both scenarios, the classifier's input neurons are defined by the reducer block and the output for the number of labels, which is 6. All models were trained using two phases. First, we freeze the convolutional layers and train only the fully connected one. Next, we reduce the training rate and fine-tune the entire model. For all models, we use Adam optimizer with a learning rate equal to 0.0001 for the first phase and 0.00001 for fine-tuning. As a loss function, we use the weighted cross-entropy computing the weights as described in Eq. (3). The first and second phases are trained for 50 and 100 epochs, respectively. We reduce the learning rate by a rate of 0.1 if the model does not improve for 10 consecutive epochs. We also use early stopping if the model does not improve for 15 consecutive epochs. For a fair comparison, both experiments for scenarios 1 and 2 were carried out in the same way and using the same hyper-parameters. All procedures were implemented using PyTorch and performed on Nvidia Titan X and RTX 2080 ti. The code is available upon request.

For the following experiments, we use 5-fold cross-validation and present the average and standard deviation for the following metrics: accuracy (ACC), balanced accuracy (BACC), weighted precision (P), weighted recall (R), weighted F1 score (F1) and area under the curve (AUC). To compare the experiments results, we perform the non-parametric Friedman test following by the Wilcoxon test (if applicable), using $p_{value} = 0.05$ and $p_{value} = 0.01$, respectively [63].

3.3. Experiment 1: sensitivity analysis of the combination factor

The goal of this section is to evaluate the contribution of each source of features to skin cancer detection. As described in the previous section, the number of clinical features (N_{cli}) is 28. To vary only the number of image features (N_{img}), we keep N_{cli} and vary c_f from 0.5 to 0.9. In Table 2 is described the amount of features for each source according to c_f. For instance, considering $c_f = 0.7$, it means the reducer block will output 66 features from the image and these features will be concatenated with the 28 clinical features. Thus, 94 features are used as inputs to the classifier, in which 70% come from the images and 30% from clinical information. This is how c_f controls the contribution of each source.

In order to test each value of c_f presented in Table 2, we applied the ResNet-50 for each folder of our dataset. The result for each metric is detailed in Table 3. The Friedman test returned $p_{value} < 0.05$, which means we need to apply a pair-wise comparison. We applied the Wilcoxon test, which pointed out many differences among the pairs. Based on the test assessment and on the results presented in Table 3, we conclude that the best value for c_f is either 0.8 or 0.7. There is no statistical difference between this pair. On the other hand, forc_f equal to 0.5 and 0.9, the results are slightly worse, which is also identified by the test.

According to the presented analysis, we decided to use $c_f = 0.8$ for the next experiment. As discussed before, the clinical features should be used as a support source of information. Clinical images are still the main evidence, which is in accordance with the results presented in this section.

Table 3
The ResNet-50 performance for each value of c_f.

c_f	ACC	BACC	P	R	F1	AUC
0.5	0.759 ± 0.025	0.718 ± 0.032	0.776 ± 0.021	0.758 ± 0.028	0.764 ± 0.024	0.948 ± 0.010
0.6	0.773 ± 0.026	0.739 ± 0.015	0.790 ± 0.022	0.774 ± 0.023	0.780 ± 0.023	0.948 ± 0.005
0.7	0.763 ± 0.032	0.733 ± 0.022	0.788 ± 0.016	0.762 ± 0.029	0.766 ± 0.032	0.955 ± 0.004
0.8	0.788 ± 0.025	0.750 ± 0.033	0.800 ± 0.028	0.788 ± 0.025	0.790 ± 0.027	0.958 ± 0.007
0.9	0.736 ± 0.036	0.710 ± 0.035	0.740 ± 0.046	0.726 ± 0.031	0.734 ± 0.021	0.949 ± 0.013

3.4. Experiment 2: the impact of the patient clinical information

In this section, our main goal is to compare the performance of the deep learning models for both scenarios described in Section 3.2. According to our previous analysis, we set $c_f = 0.8$ and apply all models described in Section 2.3 to the presented dataset. In Tables 4 and 5 are presented the results, in terms of mean and standard deviation, considering the 5-folders for scenarios 1 and 2, respectively. As we can note from both tables, there is a notable improvement for all metrics from scenario 1 to scenario 2. In general, the clinical features impacted positively in all models. In terms of balanced accuracy, the average model was improved by approximately 7%. The overall improvement considering all metrics is also around 7%. Although the improvement is notable, we also applied the statistical test for this experiment. As expected, the test shows that all models from scenario 2 are statistically different from those of scenario 1.

Observing only the results for scenario 2 in Table 5, in terms of balanced accuracy, we note that all models present almost the same performance, except ResNet-50, which is around 4% above compared to the rest of the models. Thereby, in Fig. 5 is depicted the confusion matrix and ROC curve for this model considering both scenarios.[4] In general, it is possible to note an improvement for all labels.

To conclude, we analyze the effect of the clinical features in the probability distribution outputted by the ResNet-50 model. In Fig. 6 is depicted the distributions obtained by the same model for each lesion in both scenarios. As we may see, the distributions for ACK, MEL, NEV, and SEK are improved from scenario 1 to scenario 2. The model increases the probability of the correct label while decreasing the incorrect ones. On the other hand, the distributions for SCC and BCC are almost the same, which is in accordance with the previous results.

4. Discussion

The results presented in the previous section confirm the hypothesis raised by Brinker et al. [45], i.e., patient clinical information is important to improve deep learning performance for skin cancer detection. In general, it is possible to note a performance improvement for all skin lesions. However, the model is still confusing SCC and BCC fairly often. This result is in accordance with the analysis provided in Section 2.2, in which we show that both lesions share almost the same values of clinical features. In fact, even dermatologists get confusing regarding these two lesions. As shown in Figs. 1(a) and 1(b), both lesions are very similar and distinguishing them is a challenging task even using a dermatoscope. Nonetheless, confusing SCC and BCC is not quite a problem, since both are skin cancer and need to be biopsied. The real problem is confusing them with ACK, which is just a minor skin disease that is treated without a surgical process.

The results achieved in this work are comparable to other state-of-the-art deep learning approaches reported in the literature. Han et al. [36] applied a ResNet-152, which is similar to ResNet-50, to classify 12 skin lesions considering two different datasets. They reported an average AUC of 0.91 and 0.89 for each dataset, respectively. Steva et al. [29] validated a GoogleNet model for benign/malignant

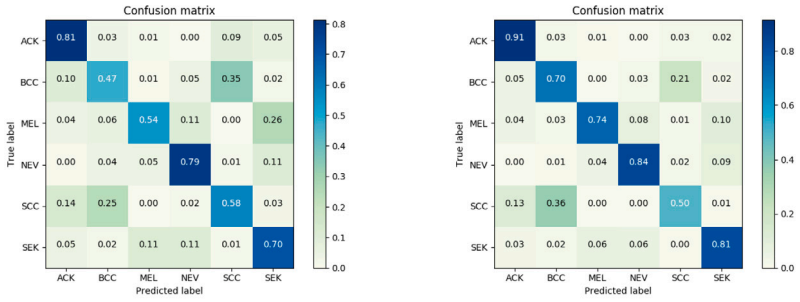
[4] As we would have 12 confusion matrices and 12 ROC curves, we decided to present a thorough analysis only for the best model. However, the remaining results are similar.

Table 4
The results for all models in scenario 1, i.e, considering only clinical images.

Model	ACC	BACC	P	R	F1	AUC
ResNet-50	0.671 ± 0.041	0.649 ± 0.047	0.720 ± 0.041	0.670 ± 0.041	0.678 ± 0.037	0.927 ± 0.017
ResNet101	0.691 ± 0.039	0.651 ± 0.035	0.736 ± 0.028	0.692 ± 0.039	0.700 ± 0.042	0.938 ± 0.008
GoogleNet	0.704 ± 0.024	0.652 ± 0.019	0.714 ± 0.024	0.702 ± 0.025	0.706 ± 0.023	0.927 ± 0.011
MobileNet	0.691 ± 0.024	0.663 ± 0.027	0.720 ± 0.014	0.690 ± 0.025	0.698 ± 0.018	0.932 ± 0.008
VGGNet-13	0.707 ± 0.028	0.658 ± 0.045	0.734 ± 0.029	0.708 ± 0.028	0.710 ± 0.029	0.932 ± 0.010
VGGNet-19	0.679 ± 0.020	0.628 ± 0.012	0.696 ± 0.020	0.678 ± 0.019	0.680 ± 0.021	0.919 ± 0.009
AVG	0.690 ± 0.029	0.650 ± 0.031	0.720 ± 0.026	0.690 ± 0.030	0.695 ± 0.028	0.929 ± 0.011

Table 5
The results for all models in scenario 2, i.e, considering both clinical features and images.

Model	ACC	BACC	P	R	F1	AUC
ResNet-50	0.788 ± 0.025	0.750 ± 0.033	0.800 ± 0.028	0.788 ± 0.025	0.790 ± 0.027	0.958 ± 0.007
ResNet101	0.757 ± 0.021	0.711 ± 0.019	0.784 ± 0.021	0.756 ± 0.021	0.766 ± 0.022	0.953 ± 0.003
GoogleNet	0.779 ± 0.011	0.714 ± 0.028	0.780 ± 0.011	0.780 ± 0.009	0.778 ± 0.007	0.948 ± 0.008
MobileNet	0.762 ± 0.040	0.717 ± 0.020	0.774 ± 0.031	0.762 ± 0.042	0.762 ± 0.037	0.948 ± 0.013
VGGNet-13	0.746 ± 0.027	0.704 ± 0.007	0.758 ± 0.013	0.748 ± 0.029	0.744 ± 0.023	0.937 ± 0.011
VGGNet-19	0.750 ± 0.013	0.709 ± 0.023	0.776 ± 0.005	0.748 ± 0.013	0.756 ± 0.010	0.946 ± 0.004
AVG	0.764 ± 0.023	0.718 ± 0.022	0.779 ± 0.018	0.764 ± 0.023	0.766 ± 0.021	0.948 ± 0.008



(a) Scenario 1 (b) Scenario 2

(c) Scenario 1 (d) Scenario 2

Fig. 5. The confusion matrices (above) and ROC curves (below) for ResNet-50 for both scenarios.

classification. They achieved an AUC equal to 0.96, which is on par with the dermatologist level. These two works also use clinical images; however, they do not consider the patient clinical information. Kharazmi et al. [46] took into account the clinical information, but for dermoscopic images. The authors proposed an approach based on sparse autoencoder (SAE) [64] to combine clinical features and dermoscopic images. They achieved an improvement in AUC from 0.847 to 0.911, when the clinical data were included. Even though none of these works consider clinical images and patient clinical information, the results achieved in this work are in accordance with them.

Despite the positive results achieved by the proposed model, there is still room for improvement. Our approach considers only one image per lesion; however, it is possible to have more than one from different angles. Thus, it is desirable to extend it to take into account more images per lesion. In addition, the model cannot handle missing data in the clinical features. For example, if one of the eight clinical features is missing, the model cannot deal with it. Finally, it is also valid to include a mechanism to detect out-of-distribution samples to identify lesions that are not present in a given dataset.

To conclude, it is important to note that the PAD dataset is composed of clinical diagnosis samples. All diagnoses are provided by a board of three senior dermatologists (over 20 years of experience) who work for the public health department of the Espírito Santo state in Brazil. It means that the samples are not proved by biopsy. It is a recurrent issue in skin cancer datasets. For example, consensus diagnosis is also applied to the dataset used by Steva et al. [29], Han et al. [36], and in ISIC archive [42]. In order to reduce the number of unnecessary biopsies and consequently,the costs, most of the skin

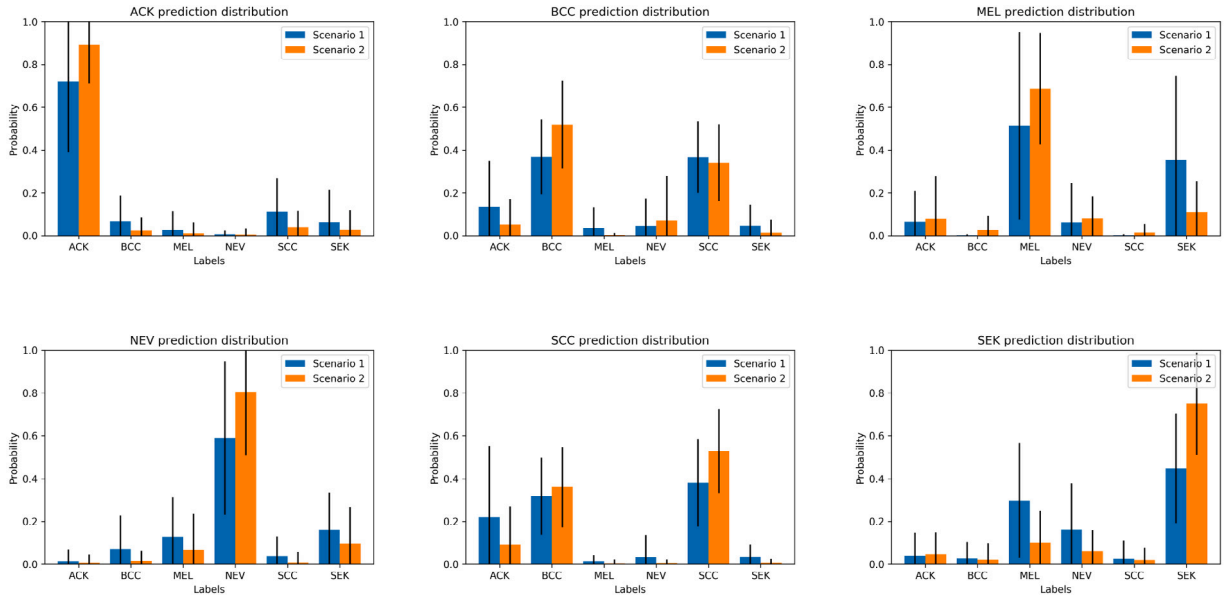


Fig. 6. The probability distribution generated by the ResNet-50 for each lesion in both scenarios.

lesions are not referred to histopathology. Thus, it is complex to build a dataset in which all images are confirmed by biopsy. In any case, we are already working on it. Soon, we will include biopsy proved images and compare the clinical diagnosis provided by the board of dermatologists and the models. Meanwhile, the achieved results represent an advance towards automated skin cancer detection. However, it is still necessary to increase the dataset and perform exhaustive tests before deploying such a system.

5. Conclusion

In this paper, we presented a study to analyze the impact of patient clinical information on skin cancer detection using deep learning models. First, we introduced a new dataset composed of clinical images, collected from smartphones, and patient clinical information. Next, we presented a straightforward approach to combine the image and clinical features using convolutional neural networks. We implemented this approach for different CNN models and applied them for the presented dataset. The results indicated that clinical features provided a substantial performance improvement for all investigated models. Nonetheless, the clinical features used are not helpful for all kinds of lesions. As we showed, they could not improve the classification of SCC/BCC lesions since their clinical features share similar values. In general, this work demonstrated the importance of clinical features on skin cancer detection and confirms the hypothesis that patient clinical information is important for this task. We also presented a discussion about limitations regarding our approach and the presented dataset. We are already working to improve the discussed weaknesses, and we intend to present these improvements in the near future.

Declaration of consent

The dataset was collected along with the Dermatological Assistance Program (PAD) of the Federal University of Espírito Santo. The program is managed by the Department of Specialized Medicine and was approved by the university ethics committee. In addition, all data is collected under patient consent and the patient's privacy is preserved.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; the Conselho Nacional de Desenvolvimento Científico e Tecnólogico (CNPq), Brazil – grant n.309729/2018-1 – and the Fundação de Amparo a Pesquisa e Inovação do Espírito Santo (FAPES), Brazil — grant n. 575/2018. We also thank all the members of the Dermatological Assistance Program (PAD-UFES) and the support of NVIDIA Corporation, USA with the donation of a Titan X GPU used for this research.

References

[1] CCA, UnderstandIng Skin Cancer - A Guide for People with Cancer, their Families and Friends, Cancer Council Australia, 2018, https://www.cancer.org.au/about-cancer/types-of-cancer/skin-cancer.html. (Accessed 15 May 2019).

[2] R.L. Siegel, K.D. Miller, A. Jemal, Cancer. statistics, Cancer statistics 2019, CA Cancer J. Clin. 69 (1) (2019) 7–34.

[3] WHO, World Health Organization (WHO), 2019, https://www.who.int/uv/faq/skincancer/en/index1.html. (Accessed 15 May 2019).

[4] CCSsACoC, Canadian cancer statistics 2014 - special topic: Skin cancers, Canadian Cancer Society's Advisory Committee on Cancer Statistics.

[5] ACS, Cancer Facts & Figures 2019, American Cancer Society Atlanta, 2019.

[6] INCA, The cancer incidence in Brazil, National Institute of Cancer José Alencar Gomes (INCA), 2018, available on http://www1inca.gov.br/estimativa/2018/estimativa-2018pdf (8).

[7] K. Wolff, R.A. Johnson, A.P. Saavedra, E.K. Roh, Fitzpatrick'S Color Atlas and Synopsis of Clinical Dermatology, eighth ed., McGraw-Hill Education, New York, USA, 2017.

[8] R.D. Azulay, Dermatologia, seventh ed., Guanabara Koogan, Rio de Janeiro, Brazil, 2017.

[9] G. Argenziano, H.P. Soyer, Dermoscopy of pigmented skin lesions – a valuable tool for early, Lancet Oncol. 2 (7) (2001) 443–449.

[10] H. Kittler, H. Pehamberger, K. Wolff, M. Binder, Diagnostic accuracy of dermoscopy, Lancet Oncol. 3 (3) (2002) 159–165.

[11] C. Sinz, P. Tschandl, C. Rosendahl, B.N. Akay, G. Argenziano, A. Blum, R.P. Braun, H. Cabo, J.-Y. Gourhant, J. Kreusch, et al., Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin, J. Am. Acad. Dermatol. 77 (6) (2017) 1100–1109.

[12] S.E. Umbaugh, R.H. Moss, W.V. Stoecker, G.A. Hance, Automatic color segmentation algorithms with application to skin tumor feature identification, IEEE Eng. Med. Biol. Mag. 12 (3) (1993) 75–82.

[13] F. Ercal, A. Chawla, W.V. Stoecker, H.-C. Lee, R.H. Moss, Neural network diagnosis of malignant melanoma from color images, IEEE Trans. Biomed. Eng. 41 (9) (1994) 837–845.

[14] A. Green, N. Martin, J. Pfitzner, M. O'Rourke, N. Knight, Computer image analysis in the diagnosis of melanoma, J. Am. Acad. Dermatol. 31 (6) (1994) 958–964.
[15] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, M. Delfino, Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis, Arch. Dermatol. 134 (12) (1998) 1563–1570.
[16] A. Masood, A. Ali Al-Jumaily, Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms, Int. J. Biomed. Imaging 2013 (2013) 1–22, (ID 323268).
[17] M.E. Celebi, H.A. Kingravi, B. Uddin, H. Iyatomi, Y.A. Aslandogan, W.V. Stoecker, R.H. Moss, A methodological approach to the classification of dermoscopy images, Comput. Med. Imaging Graph. 31 (6) (2007) 362–373.
[18] P. Wighton, T.K. Lee, H. Lui, D.I. McLean, M.S. Atkins, Generalizing common tasks in automated skin lesion diagnosis, IEEE Trans. Inf. Technol. Biomed. 15 (4) (2011) 622–629.
[19] I. Maglogiannis, K.K. Delibasis, Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy, Comput. Methods Programs Biomed. 118 (2) (2015) 124–133.
[20] C. Barata, M.E. Celebi, J.S. Marques, Improving dermoscopy image classification using color constancy, IEEE J. Biomed. Health Inf. 19 (3) (2014) 1146–1152.
[21] R.B. Oliveira, J.P. Papa, A.S. Pereira, J.M.R. Tavares, Computational methods for pigmented skin lesion classification in images: Review and future trends, Neural Comput. Appl. 29 (3) (2018) 613–636.
[22] J. Scharcanski, M.E. Celebi, Computer Vision Techniques for the Diagnosis of Skin Cancer, Springer Science & Business Media, 2013.
[23] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, J.R. Smith, Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2015, pp. 118–126.
[24] P. Spyridonos, G. Gaitanis, A. Likas, I.D. Bassukas, Automatic discrimination of actinic keratoses from clinical photographs, Comput. Biol. Med. 88 (2017) 50–59.
[25] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (4) (2017) 994–1004.
[26] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[27] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
[28] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312.
[29] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115.
[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[31] H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A.B.H. Hassen, L. Thomas, A. Enk, 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 (8) (2018) 1836–1842.
[32] T.J. Brinker, A. Hekler, A.H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, T. Holland-Letz, et al., Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task, Eur. J. Cancer 113 (2019) 47–54.
[33] N.C. Codella, Q.-B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, J.R. Smith, Deep learning ensembles for melanoma recognition in dermoscopy images, IBM J. Res. Dev. 61 (4/5) (2017).
[34] B. Harangi, Skin lesion classification with ensembles of deep convolutional neural networks, J. Biomed. Inform. 86 (2018) 25–32.
[35] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, T. Wang, Melanoma recognition in dermoscopy images via aggregated deep convolutional features, IEEE Trans. Biomed. Eng. 66 (4) (2019) 1006–1016.
[36] S.S. Han, M.S. Kim, W. Lim, G.H. Park, I. Park, S.E. Chang, Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm, J. Invest. Dermatol. 138 (7) (2018) 1529–1538.
[37] M. Attia, M. Hossny, S. Nahavandi, A. Yazdabadi, Skin melanoma segmentation using recurrent and convolutional neural networks, IEEE 14th International Symposium on Biomedical Imaging, 2017, pp. 292–296.
[38] N. Nida, A. Irtaza, A. Javed, M.H. Yousaf, M.T. Mahmood, Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy c-means clustering, Int. J. Med. Inform. 124 (2019) 37–48.
[39] G.G. de Angelo, A.G.C. Pacheco, R.A. and Krohling, Skin lesion segmentation using deep learning for images acquired from smartphones, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8.
[40] S. Serte, H. Demirel, Gabor wavelet-based deep learning for skin lesion classification, Comput. Biol. Med. (2019) 103423.
[41] P. Tschandl, N. Codella, B.N. Akay, G. Argenziano, R.P. Braun, H. Cabo, D. Gutman, A. Halpern, B. Helba, R. Hofmann-Wellenhof, et al., Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: An open, web-based, international, diagnostic study, Lancet Oncol. 20 (7) (2019) 938–947.
[42] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Nat. Sci. Data 5 (2018) 180161.
[43] E. Chao, C.K. Meenan, L.K. Ferris, Smartphone-based applications for skin monitoring and melanoma detection, Dermatol. Clin. 35 (4) (2017) 551–557.
[44] A. Ngoo, A. Finnane, E. McMeniman, H.P. Soyer, M. Janda, Fighting melanoma with smartphones: A snapshot of where we are a decade after app stores opened their doors, Int. J. Med. Inform. 118 (2018) 99–112.
[45] T.J. Brinker, A. Hekler, J.S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A.H. Enk, C. von Kalle, Skin cancer classification using convolutional neural networks: Systematic review, J. Med. Internet Res. 20 (10) (2018) e11936.
[46] P. Kharazmi, S. Kalia, H. Lui, Z. Wang, T. Lee, A feature fusion system for basal cell carcinoma detection through data-driven feature learning and patient profile, Skin Res. Technol. 24 (2) (2018) 256–264.
[47] A. Marzuka, S. Book, Basal cell carcinoma: Pathogenesis, epidemiology, clinical features, diagnosis, histopathology, and management, Yale J. Biol. Med. 88 (2) (2015) 167–179.
[48] N.A. Kasparian, R. Bränström, Y.-m. Chang, P. Affleck, L.G. Aspinwall, A. Tibben, E. Azizi, O. Baron-Epel, L. Battistuzzi, W. Bruno, et al., Skin examination behavior: The role of melanoma history, skin type, psychosocial factors, and region of residence in determining clinical and self-conducted skin examination, Arch. Dermatol. 148 (10) (2012) 1142–1151.
[49] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, Neural Comput. 29 (9) (2017) 2352–2449.
[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[52] A. Menegola, M. Fornaciali, R. Pires, F.V. Bittencourt, S. Avila, E. and Valle, Knowledge transfer for melanoma screening with deep learning, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017, pp. 297–300.
[53] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.
[54] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[55] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861.
[56] G. Ma, X. Yang, B. Zhang, Z. Shi, Multi-feature fusion deep networks, Neurocomputing 218 (2016) 164–171.
[57] S. Sun, Y. Liu, L. Mao, Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features, Inf. Fusion 50 (2019) 43–53.
[58] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Cogn. Model. 5 (3) (1988) 1.
[59] H. Abdi, L.J. Williams, Principal component analysis, Comput. Stat. 2 (4) (2010) 433–459.
[60] C.N. Vasconcelos, B.N. Vasconcelos, Experiments using deep learning for dermoscopy image analysis, Pattern Recognit. Lett. (2017) in press.
[61] G.D. Finlayson, E. Trezzi, Shades of gray and colour constancy, Color Imaging Conf. 2004 (1) (2004) 37–41.
[62] F. Perez, C. Vasconcelos, S. Avila, E. Valle, Data augmentation for skin lesion analysis, in: OR 20 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Springer, 2018, pp. 303–311.
[63] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18.
[64] A. Ng, et al., Sparse autoencoder, CS294A Lecture Notes 72 (2011) 1–19.