Medical Image Analysis 75 (2022) 102254

Contents lists available at ScienceDirect

Medical Image Analysis

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

Melanoma classification using light-Fields with morlet scattering transform and CNN: Surface depth as a valuable tool to increase detection rate

Pedro M. M. Pereira[a,b,*], Lucas A. Thomaz[a,c], Luis M. N. Tavora[c], Pedro A. A. Assuncao[a,c], Rui M. Fonseca-Pinto[a,c], Rui Pedro Paiva[b], Sergio M. M. de Faria[a,c]

[a] Instituto de Telecomunicações, Morro do Lena - Alto do Vieiro, Leiria 2411-901, Portugal
[b] University of Coimbra, Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, Pinhal de Marrocos, Coimbra 3030-290, Portugal
[c] ESTG, Polytechnic of Leiria, Morro do Lena - Alto do Vieiro, Leiria 2411-901, Portugal

ARTICLE INFO

Article history:
Received 31 March 2021
Revised 27 July 2021
Accepted 22 September 2021
Available online 7 October 2021

2010 MSC:
92C55
68U10
65T60

Keywords:
Skin lesion
Classification
Light-fields
Wavelet scattering

ABSTRACT

Medical image classification through learning-based approaches has been increasingly used, namely in the discrimination of melanoma. However, for skin lesion classification in general, such methods commonly rely on dermoscopic or other 2D-macro RGB images. This work proposes to exploit beyond conventional 2D image characteristics, by considering a third dimension (depth) that characterises the skin surface rugosity, which can be obtained from light-field images, such as those available in the SKINL2 dataset. To achieve this goal, a processing pipeline was deployed using a morlet scattering transform and a CNN model, allowing to perform a comparison between using 2D information, only 3D information, or both. Results show that discrimination between Melanoma and Nevus reaches an accuracy of 84.00, 74.00 or 94.00% when using only 2D, only 3D, or both, respectively. An increase of 14.29pp in sensitivity and 8.33pp in specificity is achieved when expanding beyond conventional 2D information by also using depth. When discriminating between Melanoma and all other types of lesions (a further imbalanced setting), an increase of 28.57pp in sensitivity and decrease of 1.19pp in specificity is achieved for the same test conditions. Overall the results of this work demonstrate significant improvements over conventional approaches.

1. Introduction

Over the last decades, skin cancer has maintained its position at the top of the most common cancers all over the world (Alliance, 2020). A skin lesion is any kind of skin patch that presents different characteristics when compared to its surrounding area. There are many types of skin lesions, which can be described according to their type, configuration, texture, colour, localisation, and distribution, among other clinical signs. Generally, studies tend to focus on pigmented skin lesions, namely the melanocytic lesions. This type of lesions is primarily denoted as an abnormal proliferation of melanocytes at the basal epidermis or upper dermis layers that may ultimately be classified as benign or malignant (Cichorek et al., 2013). Its classification is typically based on dermatologists visual inspection, with support of dermoscopic imaging and the diagnosis by skin biopsy (Vestergaard et al., 2008).

Around the world, dermatologic work force shortage and the lack of pathology lab facilities set up the main reasons for the lack of access to prompt detection of skin cancer, leading to the increased morbidity and melanoma mortality (Feng et al., 2018). Melanoma diagnosis rates have increased dramatically over the past three decades, outpacing almost all other cancers (Alliance, 2020). As of 2020, in the USA, the risk of developing melanoma was of 1 in 38 (2.6%) for Whites, 1 in 1000 (0.1%) for Blacks, and 1 in 167 (0.6%) for Hispanics (Society, 2020). A classical method to identify melanoma is with parameters known as Asymmetry, Border, Colour, and Diameter – coined the “ABCD” rule (Soyer et al., 2004). This method is based on the principle that melanoma lesions are typically asymmetric, are larger than 6mm in diameter, have irregular borders, and tend to have more than one colour. Additionally, one-third of all melanomas are thought to arise from pre-existing nevus (a similar lesion but of benign ori-

* Corresponding author.
E-mail address: pedrommpereira@co.it.pt (P.M. M. Pereira).

https://doi.org/10.1016/j.media.2021.102254
1361-8415/

gin) – thus detection and removal of such nevus is of utmost importance in the prevention of melanoma (Pampena et al., 2017). The process of lesion identification by specialists is labour intensive, time costly, and error prone, therefore, it could be improved with the use automated methods.

Fortunately, with the advent of Deep Learning (DL), computer-aided diagnosis of cancers seems increasingly possible (Litjens et al., 2017). Indeed, automated DL techniques for skin lesion classification may automate future screening, and enable early detection of skin cancer (Adegun and Viriri, 2020). However, as detailed in Yao et al. (2021), available skin lesion datasets are usually very small in comparison to what is normally used to train DL models. Therefore, many studies prefer to extract hand-crafted features in order to reduce the model learning space and, consequently, its natural capability to overfit (Yang et al., 2018; Satheesha et al., 2017). Datasets used in skin lesion classification use the same type of information as dermatology experts – i.e. dermoscopic imagery (2D/colour). The resulting classification performances are yet to become sufficient to professionally help dermatologists. Despite the limited composition of current datasets, other type of imagery could also be used for this end. This includes other data dimensions, which are fairly unexplored as they are not suited for direct human observation, but can still provide relevant information for computer systems. One of these modalities is 3D imaging (e.g., stereo), which has already proven to enhance skin lesion discrimination performances due to the added depth information (McDonagh et al., 2008; Smith et al., 2011).

In general, image classification requires the use of representations that reduce non-informative intra-class variability, but must preserve discriminative information across classes. In DL, deep neural networks (DNN) build hierarchical invariant representations learned by applying linear and non-linear operators in succession during training. These are learned in a dataset-dependent basis, and most image classification problems have generic learnable representations that are common across fields. When multiple instances of the same element are present in a dataset, translations, rotations, and scaling are common sources of variability for most images. Changes in the object view point and perspective projections of three dimensional surfaces correlate many of the dataset samples. With the use of Wavelet Scattering (WS) (Bruna and Mallat, 2013) it is possible to build neural networks invariant to said translations and rotations (Sifre and Mallat, 2014). These can be implemented as a convolutional neural network (CNN), with successive spatial wavelet convolutions at each layer.

This paper explores the use of depth data from skin lesions combined with colour information by resorting to light-field imagery and semi-automated segmentation masks. Based on a publicly available dataset named SKINL2 (Faria et al., 2019a; 2019b), a DL-based classifier is developed. The DL model relies on Morlet Wavelet-based features that greatly reduce the dimensionality problem, by performing Wavelet Scattering Transforms on the input data (Andén and Mallat, 2014; Bruna and Mallat, 2013; Sifre and Mallat, 2013). These features are used as alternative to a deeper model by providing unique features, invariant to translation, rotation, scale and frequency shifting – a transformation bearing similarities to Gabor filters in initial CNN convolutions (Springenberg et al., 2015; Yosinski et al., 2015). The experimental results show the contribution of these new depth features in comparison to the classification of 2D colour images. Additionally, it is also assessed the extent to which depth information can improve current state-of-the-art (SotA) skin lesion classification systems that only resort to the traditional 2D imagery.

The main contribution of this paper is the exploitation of 3D surface skin data as an alternative data modality for melanoma discrimination. Additionally, the Morlet Wavelet-based features are also introduced for this type of data and compared to the current state-of-the-art results. Because this data is originated from light-field imagery, a comparison to typical colour based classification is possible, as the used dataset provides both colour image and 3D information for every image-pixel data.

The remainder of the paper is organised as follows: Section 2 presents a brief literature review on relevant topics, including previous works that led to the proposed pipeline, as well as details about existing datasets, and the concept of wavelet scattering. Section 3 describes the proposed approach pipeline, including relevant details about the experiment parameters, segmentation, data pre-processing, augmentation, normalisation, model feature extraction, and the DL model. Finally, Section 4 presents and discusses the achieved results while Section 5 highlights the conclusions.

2. Background

Image recognition and classification using Machine Learning (ML) has become a major topic in a wide range of research fields, specially with DL. For instance, in the field of skin lesion classification, CNNs have produced promising results (Gonzalez-Diaz, 2018; Tang et al., 2020). But recently, research based on data-driven models have reported the highest performance measurements ever published across multiple test datasets (Hosny et al., 2019). The use of these pre-trained models is typically accompanied by a Transfer Learning (TL) method (Shin et al., 2016; Barata et al., 2018), which can be additionally aided by manually extracted features (e.g., as in Hagerty et al. (2019)).

In order to properly address various concepts or areas necessary to support this work, the remainder of this section is structured into four subsections, namely: Deep Learning (DL), segmentation, datasets, and Wavelet Scattering (WS).

2.1. Deep learning

Practical implementation of automatic skin cancer classification has significantly improved with deep CNN-based models – DCNN – (Gessert et al., 2019; Xie et al., 2020; Yuan et al., 2017; Esteva et al., 2017; Liu et al., 2020). However, despite the promising research progresses, further improvement in diagnostic accuracy is still hindered by several factors. In particular, skin lesion dataset images usually exhibit low contrast, fuzzy borders and artefacts such as hair, veins, and ruler marks, which hinder lesion classification. As a consequence, large sets of images are necessary for adequately training DCNN models to fit the unknown data features, as exemplified by the use of millions of images in the most used image classification datasets, ImageNet (Deng et al., 2009). Benefiting from an initial training with this large-scale image classification dataset, previous DCNN models have achieved significant classification results, comparable to those of professional dermatologists diagnostics (Esteva et al., 2017; Liu et al., 2020). Additionally, almost all the publicly accessible skin lesion datasets suffer from data imbalance. Samples among different lesion categories have uneven distributions because different types of skin lesions have different occurring rates and image acquisition accessibility. Furthermore, many images have high intra- or low inter-class variations (Yu et al., 2016; Yang et al., 2019). These constrains contribute to an imbalanced dataset and poor metric performance, especially for rare (e.g., melanoma) and similar (e.g., melanoma and nevus) lesion types.

On large-scale image classification tasks, improving the DCNN structure from an initial AlexNet (Krizhevsky et al., 2012) to the recent RegNet (Radosavovic et al., 2020) or increasing the model parameter capacity (Radosavovic et al., 2020; Tan and Le, 2019; He et al., 2016) enables better performances. However, in small-scale

image datasets it is very difficult to increase the performance, as increasing the number of parameters may induce the model to transition from an under-fitting space to space where the over-fitting probability is hi (Belkin et al., 2018). In some works described in the literature, DCNNs are selected without taking into account the new dataset size and the new intra- or inter-class variations – avoiding such issues by using mechanisms that mitigate the over-fitting problem (Esteva et al., 2017; Liu et al., 2020; Yu et al., 2018; Han et al., 2018; Brinker et al., 2019; Hosny et al., 2019). Some of these mechanisms are TL (Hosny et al., 2019), data augmentation (Bisla et al., 2019; Hosny et al., 2019), and multi-target weighted loss functions (Fernando and Tsokos, 2021; Hosny et al., 2019). Alternatively, other data and model adjustments that have been learned from large-scale image classification (such as data normalisation into certain ranges or residual connections between distant layers) can be used (Radosavovic et al., 2020; He et al., 2019; Wu and He, 2018; Ioffe and Szegedy, 2015; Mishkin et al., 2017).

A combination of such mechanisms is used in Hosny et al. (2019) where the classification of segmented colour skin lesion images of three datasets is performed using TL with a pre-trained AlexNet CNN model. In order to increase the number of dataset samples and lower the model overfit probability, augmentation based on image rotation is performed. Data normalisation is also employed as originally applied for the previously trained ImageNet data (maintaining the same colour feature space). As commonly used in TL methods, the model classification layer is replaced by an appropriate softmax layer for either melanoma and nevus (binary) or melanoma, seborrheic keratosis, and nevus (ternary) discrimination. After fine-tuning the model weights on each dataset, and performing augmentation in both train and test sets, the reported system accuracy performance was measured as 96.86%, 97.70%, and 95.91% for the used MED-NODE, Derm-IS and Derm-Quest, and ISIC datasets, respectively. Without augmentation, the recorded performance was 88.24%, 91.18%, and 87.31% for the same datasets.

Additional information about DNN skin lesion applications using this datasets can be found in Senan and Jadhav (2019).

2.2. Segmentation

In DNN, the majority of works require some form of prior lesion segmentation or location identification (Hosny et al., 2019; Hagerty et al., 2019; Gonzalez-Diaz, 2018; Tang et al., 2020; Li et al., 2018; Ravìet al., 2016; Barata et al., 2018; Navarro et al., 2018; Khan et al., 2019). Surrounding healthy skin information (or image acquisition artefacts) may originate outlier features or expand the dimension of the hyperspace where the parameter search is performed by DL algorithms (e.g., with CNN), urging for a preprocessing step in order to avoiding undesirable outcomes. One example of such method is described in Navarro et al. (2018), where local features guide image segmentation into super-pixels, which are iteratively merged into regions, to form two classes of regions (lesion and non-lesion), while considering a spatial continuity constraint on the super-pixels colour.

2.3. Datasets

To the best of the authors' knowledge, all literature works that experiment on publicly available datasets of skin lesions are restricted to 2D coloured-information, which are either of dermoscopic or macro-photographic images. Significant results have already been achieved using these single modality datasets (Pathan et al., 2018), still, the low granularity of the information might pose limitations to the classification problem, as only planar lesion-information can be retrieved from such data. Other modalities, such as stereoscopic technology (McDonagh et al., 2008; Smith et al., 2011), have already shown alternatives to overcome this limitation by efficiently identifying the type of skin lesion when a third dimension is present. Despite the scarce literature on 3D surface of melanoma or related skin lesions, there are indications that depth provides useful information. For this reason, the study in Satheesha et al. (2017) tries to use artificially generated 3D information to enhance the existing 2D dataset. In order to fill the void of 3D skin lesion data, a recent dataset named Skin Lesion Light-fields (SKINL2) was made public to enable research over skin lesions' 3D surface information (Faria et al., 2019b). Note that this dataset is even smaller than other available 2D datasets like the one used in Yao et al. (2021). At the time of writing, to the authors knowledge, there are no works published by other authors resorting to this recent dataset.

2.4. Wavelet scattering

Also relevant for this work is the concept of Scattering Transform and its early usage in CNN architectures as alternative to initial convolution layers since it provides unique features invariant to translation, rotation, scale, and frequency shifting – allowing the creation of lesser deep models.

In Mallat (2012), the concept of Lipschitz-continuous translation- and rotation-invariant operators for wavelets is presented, where differentiable manifolds are smoothly mapped with invertible functions – diffeomorphism. Lipschitz continuity is the central condition to guarantee the existence and uniqueness of a solution to an optimisation problem. This condition is discarded by CNNs when matching patterns during the training process, allowing similar patterns to exist (even if only initially) and match identical solutions (Bruna and Mallat, 2013). This wavelet-propagating operator is a path-ordered product of nonlinear and not-comparable operators, each one computing the modulus of a wavelet transform. The scattering transform window is generated by a Lipschitz-continuous local integration, which converges to a translation-invariant wavelet scattering transform as the window size increases. The scattering coefficients also provide representations of stationary processes (Mallat, 2012; Waldspurger, 2017).

The Wavelet Scattering (WS) framework is based on this core concept. Which is used as convolution layers for NNs. The convolutions obtained from WS – whose filters are fixed to be wavelet and low-pass averaging filters coupled with modulus non-linearities – compute translation invariant image representations, which are invariant to deformations while preserving high frequency information for classification. While this is true, it is important to note that features acquired with such framework are subject to a level of oscillation, however small. In Bruna and Mallat (2013), the mathematical analysis of wavelet scattering networks explain important properties of DCNN classification, presenting results for handwritten digits and texture discrimination.

Some degree of invariance to translation and diffeomorphism is necessary for many classification or regression tasks. In DL, using CNNs for example, the use of the WS framework can create an initial model which includes one or more layers responsible for transforming the non-linear input into representations invariant to geometric transformations (translations, rotation, scale and frequency shifting), while preserving a high degree of discriminability (Bruna and Mallat, 2013; Waldspurger, 2015; Sifre and Mallat, 2013). These transformations have two main advantages. First, they perform dimensionality reduction to the data, while allowing a structured feature representation to be captured for a given task. Second, the geometric-invariant representation that is mapped into a smaller dimension space allows for simpler model building, especially in

the presence of small training sets (Adel et al., 2017; Bruna and Mallat, 2013; Chudáček et al., 2013).

Across several fields, replacement or augmentation of learnable convolutions is being performed with this WS framework. In short, the scattering transform is defined as a complex-valued CNN whose filters are fixed to be wavelets and the non-linearity is a complex modulus. Because wavelet transform is contractive, as is the complex modulus, so is the whole network, resulting in a reduction of variance and added stability relative to additive noise. Also, since each layer is a wavelet transform that separates the scales of the incoming signal, invariability to deformation of the original signal is also attained. All these aforementioned properties enable the representation of structured signals such as natural images, textures, audio recordings, biomedical signals, and molecular density functions, among others.

3. Proposed approach - Pipeline

As pointed out before, the aim of this work is to improve the accuracy of melanoma discrimination of conventional methods that only use colour (RGB) information, by including an additional dimension (depth) that characterises the skin surface rugosity. To achieve this goal, a pre-processing and classification pipeline is proposed to enable the use of RGB and corresponding depth (Z), which are referred to as *image components* along with a segmentation mask to be computed at the first stage of the pipeline. The influence of depth information in melanoma discrimination is also evaluated when both types of data are simultaneous used (i.e., RGBZ), in comparison with the use of RGB information only. The classification pipeline, in particular, comprises two main stages: a Morlet Scattering Transform, which mimics initial DL convolutions by computing initial features with high discrimination capacity and enabling the use of a shallower model when compared to other DL models like in Hosny et al. (2019) and even Tan and Le (2019); followed by the actual DL model, comprised of learnable convolutions and a softmax output.

Overall, the proposed pipeline has three types of configurations in this study: target classes; target dimensions; and model extensions. The *target classes* configuration, which will be further detailed in Section 4, sets the classification spectrum as either: binary discrimination of melanoma versus nevus samples; or binary discrimination of melanoma versus all other skin lesion types (including nevus). The *target dimensions* configuration sets the data dimensions (e.g. image size after resize) at a given classification study (Section 3.3). Finally, the *model extensions* defines a set of training configurations, which provide extended results to the *target dimensions* and help the interpretation of the model capabilities (Section 3.5 and Section 3.6). Both *target dimensions* and *model extensions* are defined in this section and later exploited in Section 4.2 to define the final configuration of the proposed method for the selected dataset classification targets.

The processing pipeline comprises six stages, as depicted in Fig. 1. Given a RGBZ dataset, where each pixel consists in colour (RGB) and depth (Z) information, a lesion segmentation mask is firstly generated, as described in Section 3.1. After extraction of the lesion segmentation mask, a given dataset sample is comprised of an RGB image, its depth map Z, and the segmentation mask, i.e. a total of five components at the pixel level (RGBZ plus segmentation mask). This dataset undergoes a process of data augmentation by means of random rotations in order to reduce the overfitting probability, as described in Section 3.2. Using the segmentation mask, the minimal lesion-bounding-box is determined and the pixels beyond such box are removed from the data – effectively making the new data a rectangular crop of the segmented lesion area. Concurrently, pixel values belonging to healthy skin in this crop area are set to zero. At this point, as described in Section 3.3, the parameters defined by *target dimensions* configuration define which image components to maintain and to what shape resize the data sample. Then, data is normalised into a defined range (Section 3.4) to feed the Morlet Scattering Transform (Section 3.5), which extracts features to fuel the DL model (Section 3.6). This model increasingly expands the data sample analysis, before feeding the final fully connected layer that provides the softmax discriminative label. Detailed information about each stage is provided in the following six subsections.

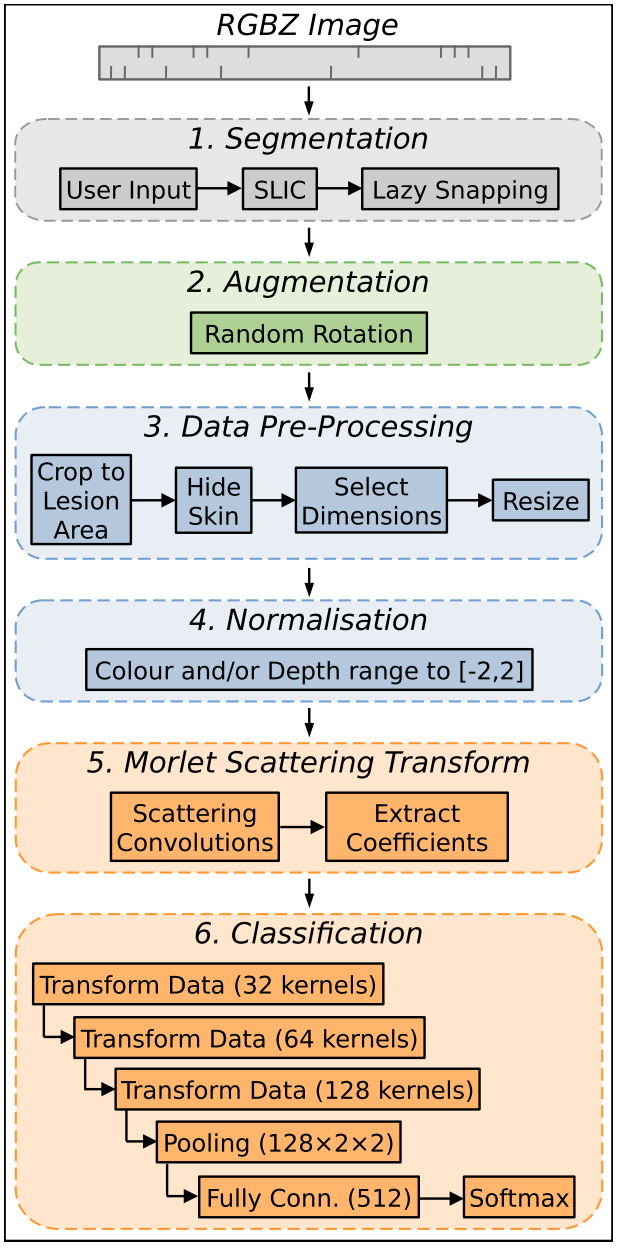

Fig. 1. Proposed pipeline: a given light-field (RGBZ) dataset *1)* ask for user input and perform segmentation using Lazy Snapping empowered by SLIC; *2)* apply augmentation by a random rotation to both RGBZ data and segmentation; *3)* pre-process data by cropping around segmentation lesion area, hide skin information, select which image components to maintain, and resize the cropped image to the target experiment size; *4)* apply normalisation by transforming the cropped image values into a range between $[-2, 2]$; *5)* create the scattering convolutions and extract a set of scattering coefficients; and finally *6)* apply a classification model that sequentially transforms said set into a larger one, which is then reduced through pooling for a final fully connected layer to provide the softmax discrimination label.

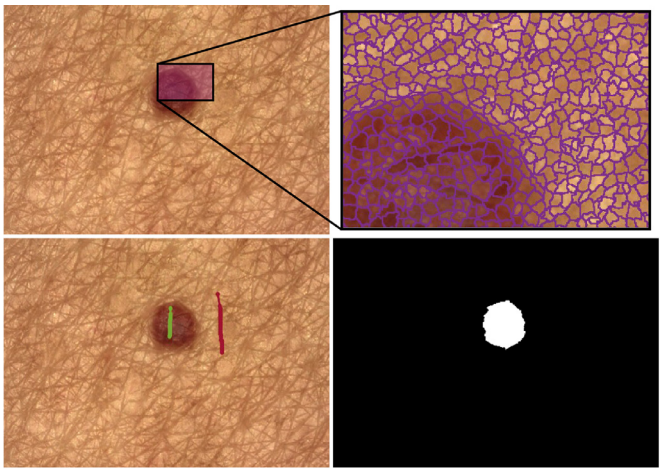

Fig. 2. Lesion segmentation method: given a dataset coloured central-view image (top-left); the image pixels are grouped through super-pixel over-segmentation (top-right); then, visually, some pixels regarding the lesion (in green) and skin region (in red) are marked to help guide the segmentation process (bottom-left); lastly, a skin lesion segmentation mask is generated (bottom-right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1. Segmentation

The segmentation process is based on the algorithm described in Li et al. (2004), dubbed Lazy Snapping, where the method to group similar pixels is substituted by the algorithm proposed by Achanta et al. (2012), which has shown to perform well in coloured skin lesions (Navarro et al., 2018). Given the RGB image data (Fig. 2, top-left), pixels are first grouped into super-pixels (Fig. 2, top-right) using the Simple Linear Iterative Clustering (SLIC) method (Achanta et al., 2012). This pre-processing step reduces the dimension of the problem to fewer image elements (super-pixels) for the subsequent Lazy Snapping algorithm. In this work, the compactness of the SLIC method is set to 10 and its clustering phase is performed for 10 iterations.

The RGB super-pixels are used to construct a graph in the Lazy Snapping algorithm, where each super-pixel is a node that connects to other related super-pixels by weighted edges. The value of the edge weights depend directly on the correlation probability between adjacent nodes. Then, by adaptively cutting edges of smaller weights, the algorithm identifies the object region by maximising the colour similarity within the object. In order to guide the graph-cut algorithm, the user provides information (Fig. 2, bottom-left) about pixels belonging to the lesion (foreground, green points in the figure) and pixels belonging to the non-lesion skin (background, red points in the figure). Given the user input, the separation between foreground object and background elements is generated by the Lazy Snapping algorithm as a segmentation mask (Fig. 2, bottom-right).

3.2. Augmentation

Classification algorithms, as is the case of DCNN, usually require large amounts of data to yield proper performance (class separation) and convergence (feature discovery). The dataset used in this work has a small number of images, therefore it is necessary to expand it by augmenting the existing images. To this aim, all input training image samples are randomly rotated from 0 to 360 degrees prior to be used in the training phase. Additionally, each epoch comprises 72 passes through the training dataset, which allows each image to be analysed at 72 angles before a new epoch begins with another set of 72 random rotations. This is a similar approach to that implemented in Hosny et al. (2019) but in this case the rotation-degree is not restricted and augmentation is not used during the test phase. The selection of this as our only augmentation method was made to fairly compare our results with those obtained with the selected baseline method (detailed in Section 4.4).

At this stage, each input training image (i.e., dataset sample) comprises five components (C): an RGB image, the corresponding depth data, and the lesion segmentation mask. All image channels are geometrically transformed by the same rotation, keeping the information aligned, such that the segmentation mask still provides the correct lesion location in both the RGB and depth information.

3.3. Data pre-Processing

Given an RGBZ dataset sample and its lesion segmentation mask, the pre-processing stage sets the *target dimensions* configuration parameters for the experimental setup. There are two parameterisations: *i)* selection of the image components; and *ii)* model input image size. Besides these options, the image data entering the pre-processing stage is cropped to the bounding limits defined by the lesion segmentation mask. Concurrently, the healthy skin region in this cropped area is removed by setting the corresponding pixel values to 0 (zero). The removal of the surrounding healthy skin region is intended to focus the model on the lesion, not allowing speculations about possible patterns or features of regions outside the lesion area.

In regard to the *image components*, the pipeline can operate in different modes by exploiting either only colour (RGB) data, only depth (Z) data, or both colour and depth (RGBZ) data. Only the selected components are used by the proposed algorithm. The selection of such different operational modes, has obvious impact on the learning process and consequently on the model, allowing to compare the performance between models obtained by learning with different image components.

In regard to the *image size* parameter, three possible resizes are defined: to 32×32, to 64×64, or to 128×128 pixels. This image resize is necessary because the crop of the lesion region generates different area sizes for different images, creating conflicts of input data sizes for the model along the proposed pipeline. Additionally, considering that the original image size may be too large, depending on the number of images available in the dataset, the model resources may be inadequate, for instance, accelerating the model overfit. Therefore, the last step of the pre-processing stage is to resize the existing images to a fixed (smaller) size using bilinear interpolation.

3.4. Normalisation

Given the *image components* entering in this stage, the respective data is normalised to improve the model convergence. This is a usual procedure due to the fact that CNNs, or NN in general, perform better if the input data is constrained to certain ranges.

For the colour components, the normalisation transforms the data to the approximate range $[-2, 2]$ as in other DCNN applications (namely Hosny et al. (2019)). This is, as traditionally applied in ImageNet, normalisation is carried out by subtracting the values of $(0.485, 0.456, 0.406)$ and dividing by the values of $(0.229, 0.224, 0.225)$ for the R, G, and B components, respectively, so that the value range is comprised between $[-2, 2]$. For the depth component, the same operation is performed by subtracting 6.26 and dividing 3.03, in order to constrain it to the range of $[-2, 2]$. This normalisation stage operates on either colour, depth or both components according to the selection made in the previous data pre-processing stage.

3.5. Morlet scattering

At this stage, a dataset sample is represented by either 3, 1, or 4 channels (C) – only RGB, only Z, or RGBZ, respectively. Prior to be processed by the classification model (Section 3.6), unique features invariant to rotation, translation, and scale are extracted using a WS framework with a Morlet wavelet as the mother wavelet (Sifre and Mallat, 2013). In addition to the extraction of unique features, this process also reduces the data volume and, consequently, further prevents model overfitting. This extraction of features can be performed either by calculating only first-order coefficients or by extending to second-order calculations, which are considered as part of the *model extensions* parameters.

The mother wavelet ($\psi(t)$) used in this work is the Morlet wavelet and, to speed up the process, the convolutions are performed in the Fourier domain. The corresponding family of wavelets is generated by dilation and translation from the mother wavelet as in Eq. 1, where a is a scale factor and b is the time index, while the factor $|a|^{1/2}$ is used to ensure energy preservation. In this work, the input data is represented as 2D matrices of $N \times N$ values, where N can only assume the values 32, 64, or 128. Let $x[\mathbf{n}]$ be any signal on this $N \times N$ grid, as $x[n, m]$. The periodic convolution with another signal $y[\mathbf{n}]$ is denoted by $x \circledast y[\mathbf{n}]$. The scattering transform uses a wavelet filter bank for each order greater than zero, that is $\psi^{(1)}_{\lambda_1}[\mathbf{n}]$ for the first-order and $\psi^{(2)}_{\lambda_2}[\mathbf{n}]$ for second-order respectively, where λ_1 and λ_2 are frequency indices in the sets Λ_1 and Λ_2. The low-pass filters are represented by $\phi_J[\mathbf{n}]$, specifying an averaging log-scaling filter of 2^J (which nearly linearises the variations of scattering coefficients), where J is a regulator variable. Input data partitioning is also computed in relation to J as non-overlapping patches of size 2^J, thus producing $N/2^J$ partitions. This logarithmic non-linearity is first applied to invariant scattering coefficients to linearise their power law behaviour across scales. This is similar to the normalisation strategies used with bag of words (Lazebnik et al., 2005) and deep NNs (LeCun et al., 2010). Together with a non-linear function $p(t)$, the filters comprise the scattering transform. The non-linear function employed in this work is the complex modulus $p(t) = |t|$, as in Andén and Mallat (2014); Bruna and Mallat (2013).

$$\psi_{a,b}(t) = |a|^{1/2}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

The zeroth-order scattering coefficient $S_0(x[\mathbf{n}])$ is the local average as given by Eq. 2. The first-order scattering coefficients are obtained from convolution of $x[\mathbf{n}]$ with the first-order wavelet filter bank, as defined in Eq. 3. These are the least computationally expensive coefficients to be used in the classification process. Second-order coefficients are obtained as an extension of the first-order ones, as defined in Eq. 4, where further data structures are captured by decomposing the $p(\cdot)$ results using the second filter bank $\psi^{(2)}_{\lambda_2}$. Note that this is only performed for a subset $\Lambda_{2,*} \subset \Lambda_2$ defined only for the elements of Λ_2 corresponding to elements of Λ_1, since results from the first-order p represent low-frequencies. The Morlet filters are similar to normalised zero-mean Gabor functions and are, therefore, computed as such for simplicity. To reduce computational load, data obtained from $p(t)$ is down-sampled as in Sifre and Mallat (2014).

$$S_0(x[\mathbf{n}]) = (x \circledast \phi_J)[\mathbf{n}] \tag{2}$$

$$S_1(x[\mathbf{n}, \lambda_1]) = \left(p\left((x \circledast \psi^1_{\lambda_1})[\mathbf{n}]\right) \circledast \phi_J\right)[\mathbf{n}], \qquad \lambda_1 \epsilon \Lambda_1 \tag{3}$$

$$S_2(x[\mathbf{n}, \lambda_1, \lambda_2]) = \left(p\left(p\left((x \circledast \psi^1_{\lambda_1})[\mathbf{n}]\right) \circledast \psi^2_{\lambda_2}\right)[\mathbf{n}]\right) \circledast \phi_J)[\mathbf{n}], \quad \lambda_1 \epsilon \Lambda_1, \lambda_2 \epsilon \Lambda_2(\lambda_1) \tag{4}$$

In this work, the J regulariser is always set to 2 and a rotation parameter r, which defines how many filter rotations are performed to induce rotation-invariance, is set to 8. Since this rotation parameter is limited in number, some sensitivity to rotation still exists, thus different rotated images produce slightly different features. Assuming the already mentioned $N \times N$ pixel-grid, the Scattering Transform of the WS framework with a scale J and r angles will generate a 3D set of features V_S, as expressed in Eq. 5, for methods configured to use only first-order coefficients, or as expressed in Eq. 6, for methods including second-order coefficients. An input dataset image generates a one-fourth-sized grid $\hat{N}$ of either 8×8, 16×16, or 32×32, with either $K = 17$ or $K = 81$ feature values in each cell, depending if they are configured to use only first-order or both first-order and second-order coefficients. For example, if the experiment is configured to run RGB components (i.e. three pixel-grids, $C = 3$) with first-order coefficients, then three sets are generated, each with $K = 17$ features per cell – a total of three $17 \times \hat{N}$ feature sets per dataset image.

$$V_{S_1K} = 1 + rJ\ ,\ V_{S_1x} = \frac{N}{2^J}\ ,\ V_{S_1y} = \frac{N}{2^J} \tag{5}$$

$$V_{S_2K} = 1 + rJ + \frac{r^2J(J-1)}{2}\ ,\ V_{S_2x} = \frac{N}{2^J}\ ,\ V_{S_2y} = \frac{N}{2^J} \tag{6}$$

Prior to the next stage, feature sets are stacked along the V_{SK} dimension to generate a single feature set $\hat{V}$ of size $KC \times \hat{N}$. This means, for example, if three blocks are produced (as occurs when processing RGB data), then the new set $\hat{V}$ will maintain the second and third dimensions, while the first dimension grows to three times the size – assembling a $\hat{V}$ of $3K \times \hat{N}$ features. Stacking is performed on the first dimension (K), in opposition to other dimensions of size N, so that features regarding the same image location but of different components remain grouped together. That is, when working with the image components, vectors of K features that are extracted from each individual component (in a particular region) are stacked together in order to simplify the visualisation of the feature-information by the subsequent CNN classification model convolutions.

3.6. Classification

As depicted in Fig. 3, given a set of features $\hat{V}$, the classification is performed by a CNN model that, apart from the first convolutions, is a fixed-size network for the whole experiment. The model comprises three main parts: *i)* a fitting part, where input features are convolved with a kernel designed to fit the data to the fixed network dimensions; *ii)* an expanding part, where two repeating blocks process and expand the data; and *iii)* a classification part, where a fully connected layer provides the classification output.

For all convolutions, the value of the stride is 1 pixel, the kernel size is 3×3 unless stated otherwise, and the value of the bias is set to zero. In all batch-normalisation layers (Ioffe and Szegedy, 2015), the running estimates parameter is set to 0.1 and possess learnable affine-transformation parameters, unless stated otherwise. For the remainder of this section, every convolution layer is followed by a batch-normalisation and a Relu activation function, unless stated otherwise.

The fitting part of the network (i) comprises batch-normalisation and a convolution layer of 32 kernels. In this first convolution layer, the feature-set $\hat{V}$, which has an experiment-variable size $KC \times \hat{N}$, is transformed to a fixed size of $32 \times \hat{N}$. The first part of the network has $K \times C \times 288 + 64$ trainable parameters. Additionally, the initial batch-normalisation has no learnable affine-transformation parameters and only exists to further regularise the input data range for the model.

The expanding part of the network (ii) is a structure that repeats twice, each comprising four working blocks (WB) with a

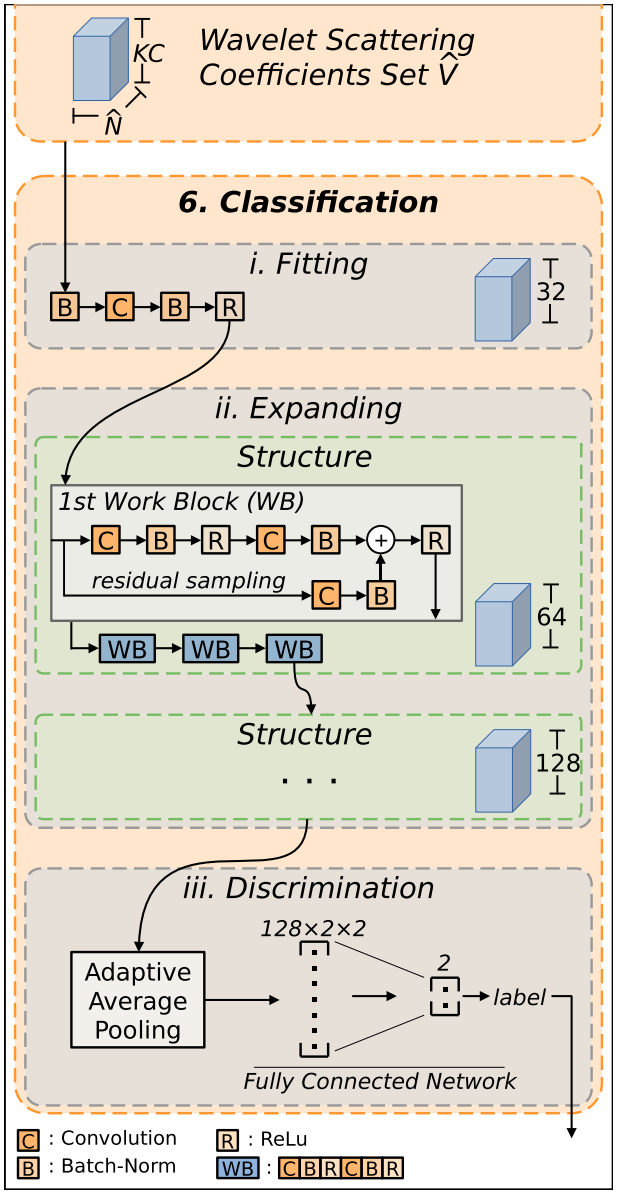

Fig. 3. Model pipeline. Receiving a feature-set $\hat{V}$ of scattering coefficients, train a deep learning model comprised of three main parts: (i) a initial data fitting, (ii) a main processing part with convolutions that expand a given data volume, and finally (iii) a fully connected layer.

residual connection. The only difference from one structure to the next is the target number of kernels in every convolutional layer, which are 64 and 128 for the first and second structure, respectively. Each of the four mentioned working blocks comprises two convolutions. The first working block of each structure has an additional third convolution, which receives the same data as the first convolution (performing the same operations). However, this block's kernels are of size 1×1 and there is no Relu at the end. The output of this third convolution is added to the second convolution batch-normalisation output – before Relu – as residual information. These two-parts of the network structure have 279,680 and 1,116,416 trainable parameters, respectively.

The classification part of the network (*iii*) performs a binary softmax classification with the result of a biased fully connected layer of 512 inputs to two neurons. This layer is adopted, with the traditional sigmoid activations, as it is an universal approximator (Csáji, 2001) for classification problems. Since the set $\hat{V}$ entering the network has size $KC \times \hat{N}$, at this point, after all convolutions, it will have $128 \times \hat{N}$. This means that it will have a variable size in the second and third dimensions – represented by $\hat{N}$. In order to encapsulate this information into a fixed size, so that models compiled for different input sizes remain comparable, an average pooling layer is added before the fully connected layer to adaptably reduce the data volume into a fixed sized $128 \times 2 \times 2$ volume (i.e. the referred 512 input values of the fully connected layer). This last part of the network has 1,026 trainable parameters.

The fully described network is trained using Stochastic Gradient Descent with Nesterov momentum (Sutskever et al., 2013). The learning rate is fixed at 0.001 and the momentum at 0.9. Additionally, weight decay (L2 penalisation) is also performed at 0.0005, in order to exponentially decay weights to zero, limiting the number of free parameters in the model and avoiding rapid over-fitting.

In this work, instead of having the learning rate influencing the new momentum velocity by scaling the gradients, the velocity does not depend on the learning rate. Rather, the learning rate is used when updating the model parameters, scaling the whole velocity equation result (meaning that it also scales the previous momentum-ed velocity). This was performed to smooth the model learning, as to further challenge early overfitting.

Finally, to promote balanced classification-error corrections in the network during training, the model softmax-cross-entropy loss function is weighted (via cost matrix) for a given class, as the number of training samples in the largest class divided by the given class number of training samples. Effectively, this makes one error in the smaller class more significant than one error in the larger class, implicitly balancing the dataset.

4. Experimental assessment

The experimental results presented in this section are expressed in terms of percentage of classification accuracy (*ACC*), sensitivity (*SEN*), and specificity (*SPE*), inline with most of the cited works, where *SEN* represents the successful melanoma identification rate and *SPE* the successful identification of the other class. Since this is an unbalanced problem, the balanced-accuracy (*BAC*) is introduced as defined in Hu et al. (2019), which corresponds to the average value between sensitivity and specificity, as shown in Eq. (7).

$$BAC = \frac{SEN + SPE}{2} \tag{7}$$

These results encompass two main classification experiments (*target classes*), both executed applying 10-fold Cross Validation (CV). The first experiment, named "*MvsN*", refers to melanoma classification against nevus samples while the second experiment, named "*MvsAll*", performs the classification of melanoma versus all other skin lesion types (including nevus). Note that there is no contamination between folds: original images selected for a given training fold are augmented by rotation producing new images used within the same fold; models are reset between CV iterations. No image augmentation is performed in the testing step. Additionally, because CV is used, the previously mentioned metrics can be generated in two ways: at a "*Dataset Level*", by merging all fold-results and calculating the metrics once; or by performing a "*Cross-Validation Average*" of the metric-results attained in each fold. By default, results are presented at a *Dataset Level*, unless stated otherwise. This is due to the fact that, given the size of the dataset, CV folds can have, for example, only one testing image for the melanoma class. As a result, the metric values become either 0% or 100%. Therefore, average values and associated standard deviations are not good performance indicators.

The learning process was run for 7 epochs in all executions, aiming for approximately 500 dataset passes through the model, as each epoch comprises 72 random rotation of each sample. In

Table 1
10-fold Cross Validation Dataset Distribution.

Dataset	Label	Folds (train | test)		
		1 to 4	5 to 6	7 to 10
MvsN	Melanoma	13 × 72 | 1	13 × 72 | 1	12 × 72 | 2
	Nevus	33 × 72 | 3	32 × 72 | 4	32 × 72 | 4
MvsAll	Melanoma	13 × 72 | 1	13 × 72 | 1	12 × 72 | 2
	All	76 × 72 | 8	76 × 72 | 8	75 × 72 | 9

the *model extensions* configuration, the following seven batch sizes were used for the model: 5, 10, 15, 20, 40, 60 and 80.

The remainder of this section is organized as follows: Section 4.1 describes the dataset and the *target classes* partitioning used in the experimental evaluation. Section 4.2 describes the parametrisation selection of the final model and Section 4.3 shows the effects of excluding image segmentation from the pipeline. Then, results are discussed and compared to current state-of-the-art baseline in Section 4.4 and to other state-of-the-art models in Section 4.5.

4.1. Dataset

The proposed pipeline was evaluated using the publicly available SKINL2 dataset (Faria et al., 2019a). The dataset contains light-field imagery of skin lesions, captured with a Raytrix R42 camera at a hospital facility (Centro Hospitalar de Leiria, Portugal) from patients previously screened by a physician during dermatology clinical appointments. All volunteers received an explanation about the procedure and purpose of the study, and also signed an informed consent form. A health ethics committee evaluated and approved the procedures related to the image acquisition, storage, and publication. Each light-field image comprises 3858 × 2682 pixels per RGBZ component.

In this work, the second (Faria et al., 2019b) and third versions of this dataset were used. Both versions of the dataset can be found in the same online repository[1]. Both versions provide more detailed images, due to their increase in lens magnification of ≈ 30% in comparison to its first version. At the time of publication of this paper, the third version is still under development and the available data was used as an extension of the second version. In total, 98 images were used (70 from the second dataset and 28 from the third). The combined dataset comprises 14 melanomas, 36 nevi, and 48 other lesion types (16 angiomas, 6 basal cell carcinomas, 1 dermatofibroma, 24 seborrheic keratoses, and 1 verruca). Sample RGB and Z image data for different dataset labels (and produced segmentation mask) are presented in Fig. 4.

Therefore, experiment *MvsN* opposes 14 melanoma samples against 36 nevus samples, while experiment *MvsAll* confront 14 melanoma samples against all other 84 non-melanoma samples. For the readers convenience, the number of images used in each fold of the two experiments is depicted in Table 1. Every training set is augmented by ×72, while testing is not.

4.2. Parameter selection

This section discusses the following three parameter configurations: the coefficients order (i.e. either first or second order coefficients), the target size of resized images and the model batch size.

To understand the influence of the batch size on the data components, the results for the more balanced *MvsN* dataset are first analysed. Since these experiments contain different amounts of data samples – 50 for *MvsN* and 98 for *MvsAll* – it is expected

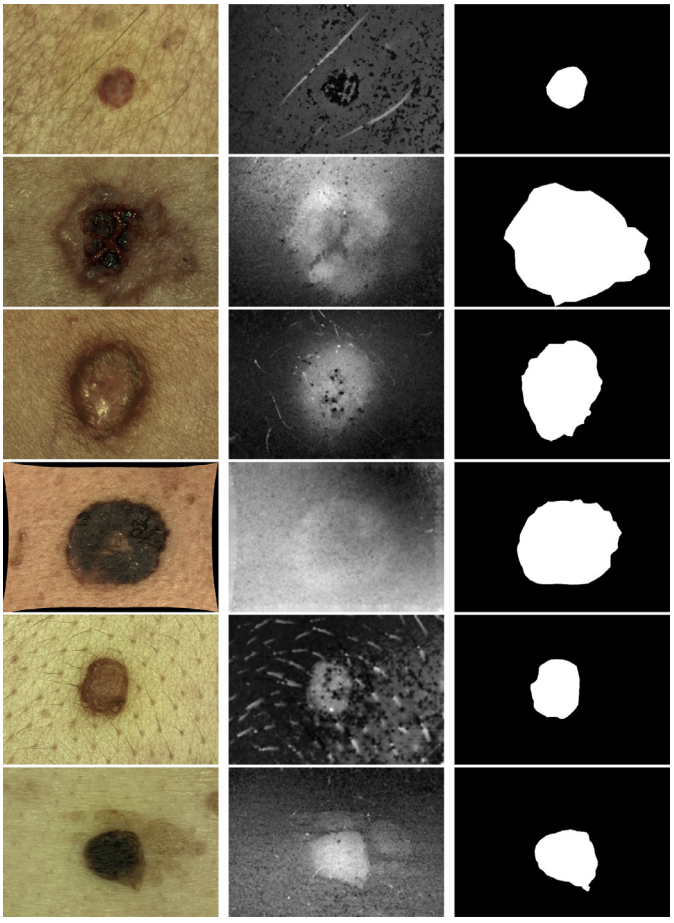

Fig. 4. Sample SKINL2 dataset images. The left column displays RGB images, the middle column shows Z values in grayscale, and the right column contains generated segmentation mask images. From top to bottom, samples show: Angioma, Carcinoma, Dermatofibroma, Melanoma, Nevus, and Seborrheic Keratosis.

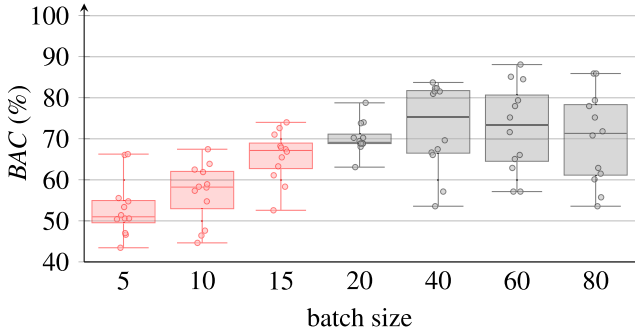

Fig. 5. Box-plot of *BAC* (with its 12 data points) for the different batch sizes in *MvsN* across the remaining parametrisation options.

that the preferred batch size will also differ in a similar ratio. Resorting to a box and whisker plot, Fig. 5 depicts the average *BAC* metric-value for the different batch sizes in *MvsN* independently of the image size, the coefficient order, and the use of either RGB or depth data. This figure also displays the 12 data points generated to build each box plot (a combinatorial execution of two coefficient orders × three images sizes × use of either RGB or depth). An inspection of the results allow us to select the batch size of 40 as the best configuration, due to its average *BAC* performance of 75.30%.

A similar analysis is performed for the *MvsAll* experiment, as depicted in Fig. 6. In this figure, box-plot data-dispersion appears smaller than in Fig. 5, most likely due to the increase in the dataset

[1] Online repository: http://on.ipleiria.pt/plenoisla

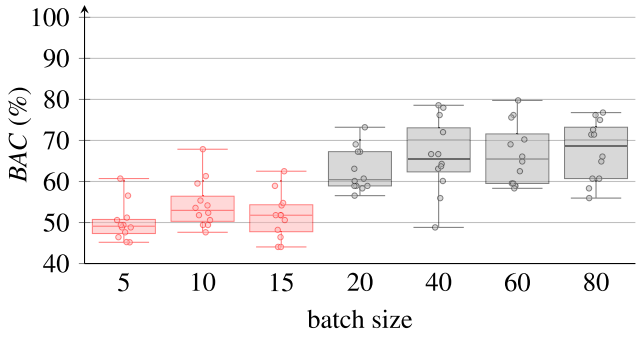

Fig. 6. Box-plot of *BAC* (with its 12 data points) for the different batch sizes in *MvsAll* across the remaining parametrisation options.

Table 2
Average *BAC* for each image resize and for 1st and 2nd order coefficients, over the possible data components - *MvsN* experiment.

Order	Image Size	Data Components	
		RGB	Z
1st	32 × 32	78.08±2.57	64.29±5.76
	64 × 64	**81.25**±6.41	**68.30**±1.76
	128 × 128	79.32±7.03	65.67±1.92
	average	*79.55*	*66.09*
2nd	32 × 32	77.48±3.11	58.28±4.71
	64 × 64	**82.89**±2.70	**67.01**±3.74
	128 × 128	76.36±4.72	59.87±5.56
	average	*78.24*	*61.72*

size. Starting from the left, *MvsAll* results appear initially similar to *MvsN*: a compact spread at batch size 20; an average improvement peaking at 40 with some data points polling down the average performance; then starting to lose performance at batch size 60. In opposition, the average *BAC* performances rise again to a new peak at batch size 80, providing a even better average performance as well as a more compact behaviour than with 40. This is expected since the amount of data samples is almost twice in the *MvsAll* experiment than in the *MvsN*.

Thus, the selected batch size for the *MvsN* and *MvsAll* experiments are 40 and 80, respectively.

The coefficient order and the image size for each experiment can also be determined following the same approach. Table 2 depicts the average *BAC* results for the image size parameter in each data dimension when using either the first- or second-order coefficients. *BAC* values are the averaged results obtained by the different batch sizes. Note, however, that because batch sizes 5, 10, and 15 performed so poorly, they were excluded from the results present in the following tables so to preserve statistical significance. This means each (non-italicized) value is an average of four executions of different batch sizes (including CV).

As can be seen in Table 2, the best average *BAC* performance in each coefficient order (marked in boldface) is achieved by the intermediate image size of 64 × 64, with 81.25% and 68.30% *BAC* performance in the first-order, for RGB and depth respectively, and 82.89% and 67.01% in the second order results. The higher performance in the intermediate image size is expected because using the smaller 32 × 32 image size removes too much information due to the down-sampling. However, using a larger 128 × 128 image size slightly decreases the classification performance as the model quickly overfits on more detailed features provided by the WS framework during the training on this small dataset.

Table 2 also allows one to analyse the average performance, for all images sizes and batch sizes, (marked in italics) for the two different coefficient orders. The best average result is obtained for the first-order coefficients with 79.55% and 66.09%, for RGB and depth respectively, against 78.24% and 61.72% *BAC* when using second-order coefficients.

Table 3
Average *BAC* for each image resize and for 1st and 2nd order coefficients over the possible data components - *MvsAll* experiment.

Order	Image Size	Data Components	
		RGB	Z
1st	32 × 32	68.60±1.85	56.99±4.77
	64 × 64	**71.73**±7.73	60.71±4.15
	128 × 128	68.45±5.57	**62.05**±3.45
	average	*69.59*	*59.92*
2nd	32 × 32	66.67±5.37	**62.50**±3.04
	64 × 64	**74.55**±2.47	62.35±3.26
	128 × 128	73.96±7.72	60.71±4.44
	average	*71.73*	*61.86*

The results shown in Table 3 for the *MvsAll* experiment were obtained under the same test conditions. In this case, the best average *BAC* is not achieved for the same image size. Yet, the different results obtained for each image size allows to observe that 64 × 64 offers the best compromise in both coefficient orders. For example, in the first-order coefficient results, selecting 128 × 128 instead of 64 × 64, causes an improvement of 1.3 percentage points (*pp* – unit measure of the arithmetic difference between two percentages) in the Z average *BAC*, while for the RGB the performance drops 3.28*pp*. Therefore, 64 × 64 is preferred, favouring the RGB classification. This analysis also works for the second-order coefficients. If the 32 × 32 image size is selected instead of the 64 × 64, the average *BAC* for Z improved by 0.15*pp*, while for RGB it drops 7.88*pp*. Therefore, 64 × 64 is preferred, also favouring of the RGB classification.

Similarly to the image size, the best coefficient order for the *MvsAll* experiment is not an obvious choice. Resorting to the same rationale as in *MvsN*, in Table 3 the best average *BAC* across image and batch size (marked in italics) is obtained by the second-order coefficients with 71.73% and 61.86%, for RGB and Z respectively, against 69.59% and 59.92% for the first-order coefficients. This can be partially explained due to the added variability in the dataset comprising the *MvsAll* experiment. In this case, there are seven different skin lesion types, instead of only two, creating a broader view of the classification problem and, consequently, requiring more detailed features, as present in second-order coefficients. The difficulty in selecting the best parameters in the case of the *MvsAll* experiment is probably due to the fact that in this experiment classes are even more imbalanced than in *MvsN*.

From these comparisons, it is safe to conclude that a good compromise in terms of the average *BAC* metric performance is achieved when configuring the image size as 64 × 64, using first-order coefficients for the *MvsN* experiment and second-order coefficients for the *MvsAll* experiment.

4.3. Ablation study of the segmentation

In this section an ablation study about the influence of the use of a segmentation method in the pipeline is performed. Having previously determined the best average parameters, this study verifies the impact of not using the lesions segmentation masks in the overall classification results. By not using a segmentation mask to detect the ROI, it becomes impossible to hide skin pixels and resize the ROI information to the target experiment size. Therefore, in this ablation study, the removal of the segmentation information from the pipeline forces the skin pixels to stay in the image, which is later used in the remaining process. This means that when resized to the target image size, the full image is used instead of only

Table 4
BAC results for the ablation study of the use of the segmentation masks paired with the optimal selected parameters.

Dataset	Image Size	Order	Batch Size	Data Components	Segmentation	
					With	Without
MvsN	64 × 64	1	40	RGB	**82.34**	57.54
				Z	**66.67**	60.91
				RGBZ	**93.65**	73.02
MvsAll	64 × 64	2	80	RGB	**71.43**	53.57
				Z	**55.95**	50.00
				RGBZ	**85.12**	61.90

the ROI. Table 4 provides the results for pipeline execution, with and without the segmentation mask. From the attained results, it is clear that using the segmentation mask to clear out healthy skin and make the algorithm focus on the lesion area provides superior results. On average, not using the segmentation mask decreases the *BAC* performance by 17.06*pp* and 15.67*pp* for the *MvsN* and *MvsAll* experiments, respectively.

4.4. Comparison results with baseline method

Using the parameters defined in the previous section, that is: image size of 64 × 64; first-order coefficients and batch size 40 for the *MvsN* experiment; and second-order coefficients and batch size 80 for *MvsAll* – the proposed model achieves the results depicted in Table 5. Results are shown both at a *Dataset Level* and by performing *Cross-Validation Average*. These results were obtained using RGB and Z (depth) components individually – *Proposed (RGB)* and *Proposed (Z)* respectively – and with all components – *Proposed (RGBZ)*. The results are also compared to the state-of-the-art method in Hosny et al. (2019) – named *Baseline (RGB)* – providing classification results for both experiments (*MvsN* and *MvsAll*). This classification method was selected as baseline since it performs comparisons with three well-known 2D datasets and outperforms other 11 state-of-the-art algorithms. Averaging across the three datasets mentioned in its' work, this method reports a 96.8% accuracy performance when using data augmentation and 88.9% when not using it. At the time of writing, to the authors knowledge, there are no other works published by other authors resorting to the SKINL2 dataset, which could be used for comparison.

The *Baseline (RGB)* method was strictly implemented as expressed in Hosny et al. (2019). This means that the pre-trained Alexnet model was used for Transfer Learning after replacing the last three classification layers with new random weights and applying a binary classification softmax layer. Prior to training, all images undergo the lesion segmentation methodology reported by Hosny et al. (2019), which features the manual optimisation of three parameters in each image to find the optimal segmentation mask. The dataset is also augmented by 72 times by performing 72 random rotations (in the range [0,355]) to every image. If an image does not fit the Alexnet model input size, a resizing operation is performed. During the training, back-propagation is used and the Stochastic Gradient Descent algorithm is used to update the weights with a learning rate of 0.001. Additionally, the batch size and number of training epoch are fixed to 10 and 32, respectively. Results are obtained via 10-fold CV.

Using the dataset employed in this work (SKINL2), the baseline method provides a 68.00% and 73.47% accuracy performance with 53.77% and 48.81% *BAC* for the *MvsN* and *MvsAll* experiments, respectively. While the accuracy increases in the *MvsAll* experiment (which has 48 additional samples in comparison with *MvsN*), it is important to point-out that the *SEN* metric decreases by 7.14*pp* even though the number of melanoma samples is the same (14) in both experiments. This decrease represents the misclassification of one additional melanoma, identifying only 2 out 14 in the *MvsN* experiment and 3 in the other. The *SPE* metric is not comparable since the amount of samples differs between experiments. Across the 10-fold CV, the baseline method correctly identifies 31 out of 36 nevus in the first experiment, and 70 out of 84 non-melanoma lesions in the second experiment.

As can be seen in Table 5 for the *MvsN* experiment, the proposed approach (*Proposed (RGBZ)*) achieves 94.00% accuracy and 93.65% *BAC*, an increase of 26.00*pp* and 39.88*pp*, respectively, when compared to the *Baseline (RGB)* method. This improvement comprises the utilisation of both RGB and depth components. If only the RGB data dimension is used, the proposed pipeline achieves only 84.00% accuracy and 82.34% *BAC*, 10.00*pp* and 11.31*pp* lower than the results achieved when using both components, respectively. Also, the use of only the depth component does not perform as well as using RGB component, however its performance is still superior to the baseline method in all metrics except *SPE*.

As expected, the combined use of both RGB and depth components surpasses the individual usage of only one of them, allowing one to infer that the depth component owns discriminative power not present in RGB. For instance, exploring the label predictions performed by the separate RGB and Z models, it is clear that two melanoma samples, which are correctly classified using depth, are not correctly classified when using RGB only. Getting the two components together in the new model (RGBZ) also allows the prediction of the other two melanoma samples, which were wrongly classified using only RGB components, thus supporting the thesis that the skin lesion surface has potential to improve the discrimination between melanoma and nevus.

For the experiment *MvsAll*, the results achieved by the proposed pipeline are also shown in Table 5, where *Proposed (RGBZ)* attains 89.80% accuracy and 85.12% *BAC*, an increase of 16.33*pp* and 36.31*pp* respectively, when compared to the *Baseline (RGB)* method. Like in the *MvsN* experiment, this increase corresponds to the use of both RGB and depth components. When using the RGB component alone, the proposed approach achieves only 86.73% and

Table 5
Proposed Method Results.

Dataset	Method	Dataset Level				Cross-Validation Average							
		ACC	*SEN*	*SPE*	*BAC*	*ACC*		*SEN*		*SPE*		*BAC*	
MvsN	Baseline (RGB)	68.00	21.43	86.11	53.77	65.17	±27.68	15.00	±25.00	83.33	±33.79	49.17	±24.81
	Proposed (RGB)	84.00	78.57	86.11	82.34	84.67	±13.86	85.00	±25.00	85.83	±15.28	85.42	±14.94
	Proposed (Z)	74.00	50.00	83.33	66.67	73.50	±17.72	45.00	±41.67	83.33	±18.99	64.17	±23.21
	Proposed (RGBZ)	**94.00**	**92.86**	**94.44**	**93.65**	**95.00**	±11.79	**95.00**	±16.67	**95.00**	±11.02	**95.00**	±12.67
MvsAll	Baseline (RGB)	73.47	14.29	83.33	48.81	73.33	±9.05	15.00	±35.36	83.06	±13.54	49.03	±16.17
	Proposed (RGB)	86.73	50.00	92.86	71.43	86.97	±6.71	55.00	±43.30	92.78	±8.89	73.89	±19.27
	Proposed (Z)	85.71	14.29	**97.62**	55.95	86.26	±8.56	15.00	±16.67	**97.78**	±7.41	56.39	±9.53
	Proposed (RGBZ)	**89.80**	**78.57**	91.67	**85.12**	**90.10**	±8.97	**75.00**	±44.10	91.94	±10.89	**83.47**	±21.31

Baseline: as in Hosny et al. (2019)

Table 6
EffecientNet Results.

Batch size	EfficientNet version	MvsN				MvsAll			
		ACC	SEN	SPE	BAC	ACC	SEN	SPE	BAC
20	B0	66.00	50.00	72.22	61.11	79.59	42.86	85.71	64.29
	B1	76.00	**78.57**	75.00	76.79	81.63	42.86	88.10	65.48
	B2	78.00	71.43	80.56	75.99	77.55	**57.14**	80.95	69.05
	B3	68.00	42.86	77.78	60.32	77.55	**57.14**	80.95	69.05
	B4	76.00	57.14	83.33	70.24	76.53	42.86	82.14	62.50
	B5	78.00	71.43	80.56	75.99	76.53	**57.14**	79.76	68.45
	B6	70.00	64.29	72.22	68.25	78.57	**57.14**	82.14	69.64
40	B0	78.00	50.00	**88.89**	69.44	82.65	**57.14**	86.90	72.02
	B1	74.00	64.29	77.78	71.03	**84.69**	50.00	**90.48**	70.24
	B2	**80.00**	71.43	83.33	**77.38**	79.59	42.86	85.71	64.29
	B3	74.00	42.86	84.11	64.48	82.65	42.86	89.29	66.07
	B4	78.00	57.14	86.11	71.63	82.65	**57.14**	86.90	72.02
60	B0	72.00	64.29	75.00	69.64	82.65	42.86	89.29	66.07
	B1	76.00	71.43	77.78	74.60	**84.69**	**57.14**	89.29	**73.21**
	B2	78.00	71.43	80.56	75.99	**84.69**	**57.14**	89.29	**73.21**
	B3	74.00	57.14	80.56	68.85	81.63	28.57	**90.48**	59.52
80	B0	68.00	57.14	72.22	64.68	77.55	42.86	83.33	63.10
	B1	72.00	71.43	72.22	71.83	82.65	**57.14**	86.90	72.02
	B2	**80.00**	64.29	86.11	75.25	83.67	**57.14**	88.10	72.62

71.43%, that is 3.07pp and 13.69pp lower than the *Proposed (RGBZ)* results, although still superior to the *Baseline (RGB)* method.

If the method uses only the depth component, similarly to the case of *MvsN*, the results are also lower than the *Proposed (RGBZ)* results, yet still superior to the *Baseline (RGB)* results for all metrics. In this *MvsAll* experiment, however, the data imbalance is greater than in *MvsN*. Incorrect melanoma classifications almost go unnoticed by the accuracy metric since, for instance, a classification of all data as non-melanoma image samples immediately achieves 85.71% accuracy. Nevertheless, this would be noticeable because the *BAC* metric would only achieve 50.00%. This means that, although the proposed RGBZ method achieves a similar accuracy performance, the total number of melanoma-misclassification is lower, because the *BAC* performance is 85.12%, accounting for 78.57% *SEN* in this case. This corresponds to the correct classification of 11 out of 14 melanoma samples, nine more than the *Baseline (RGB)*.

In this section, all comparisons with the baseline classification method have shown that the proposed approach provides superior performance results. Accordingly, this can be seen as an indirect benchmark comparison of the proposed method with the works compared in Hosny et al. (2019) and other works that resorted to the same dataset and metrics. In essence, since the baseline method reports results superior to 10 other works, it is expected that the proposed approach could also show results superior to the mentioned works, if they were to be applied to the SKINL2 dataset. This hypothesis may be further extended to other works like Pereira et al. (2020); Tang et al. (2020); Barata et al. (2018); Pathan et al. (2018); Hagerty et al. (2019), that use the same datasets and metrics as the baseline method.

In addition to the discussed results, it is worthwhile to mention some studies that compare the results of computational models with human classification of skin lesions performed by specialists, i.e. dermatologists. This is the case, for instance of Esteva et al. (2017); Marchetti et al. (2018); Haenssle et al. (2018); Brinker et al. (2019), where the *SEN* and *SPE* are evaluated and compared. In Brinker et al. (2019), these comparisons were carried out in regard to the task of performing melanoma versus nevus classification, involving 157 dermatologists that span across 12 German university hospitals. The test dataset used in this experiment comprises 20 melanomas and 80 nevi randomly selected from the ISIC dataset. Indirectly, this enables the comparison of the proposed approach with the results obtained from the 157 dermatologists. A mean of 74.1% for *SEN* and 60% for *SPE* was achieved by the dermatologists with dermoscopic images. This is inferior to the performance reported in Table 5 for the proposed RGBZ approach, which provides an additional 18.76pp in *SEN* and 34.44pp in *SPE*. Furthermore, in Marchetti et al. (2018) and Haenssle et al. (2018), respectively, 8 and 58 dermatologists have also participated in a similar study on another set of 100 images, obtaining 82% and 86.6% for *SEN*, and 59% and 71.3% for *SPE*. Again, on average, the proposed approach outperforms these classification results obtained by specialists.

Although the results obtained in Table 5 cannot be directly compared with the studies cited above, they establish a valuable reference for the expected classification performance made by specialists in the same *MvsN* dataset. Therefore, it is possible to infer that, on average, the proposed Morlet Scattering approach would outperform the human-based classification.

4.5. Comparison results with other sota methods

In addition to comparisons made in the previous section, it is also relevant to compare the proposed RGB classification pipeline with other state-of-the-art methods. As such, the EfficientNets family was selected (Tan and Le, 2019).

Following the same constraints used to attain the previous models, EfficientNet-B0 to -B7 were trained (using Transfer Learning) under the exact same conditions as performed for the other experiments. This includes the 72 random rotations for the data augmentation, the image resize, the use of back-propagation and Gradient Descent algorithms, the 0.001 learning rate, and the same amount of training epochs. All of which were repeated on a 10-fold CV execution scheme. However, given the growth rate of the amount of trainable parameters from one version of the network to the next many combinations of batch size and EfficientNet versions were not feasible, due to hardware limitations, namely GPU memory space (12GB).

Results depicting the EfficientNets family executions are presented in Table 6. EfficientNets have more trainable parameters than our proposed model, which causes them to quickly overfit in the small training dataset. Most EfficientNet executions achieved +95% training accuracy by the third epoch. In terms of test accuracy, EfficientNets topped at 80.00% and 84.69% for the *MvsN* and *MvsAll* experiments, respectively.

From the table, the best accuracy EffcientNet results were achieved with batch sizes of 40 and 80 with EfficientNet-B2 in the *MvsN* experiment. In the *MvsAll* experiment, the best accuracy results correspond to the configurations using batch sizes 40 and 60 with EfficientNet-B1 and batch size 60 with EfficientNet-B2. In comparison, the results obtained with the *Proposed (RGB)* method, in Table 5, achieve 84.00% and 86.73%, respectively for each experiment.

Using the *BAC* metric, which takes into account both the sensitivity and specificity, in particular to the *MvsN* experiment, none of the EfficientNet executions outperformed the *Proposed (RGB)*, must less the *Proposed (RGBZ)* – also in Table 5. Using this metric, the best EfficientNet execution (batch size 40 and EfficientNet-B2) is $4.96pp$ lower than the *Proposed (RGB)*, and $16.27pp$ lower than the *Proposed (RGBZ)*.

In particular to the *MvsAll* experiment, it is possible to see that six of the executions outperform the *Proposed (RGB)* by correctly classifying one extra melanoma to the detriment of some non-melanoma lesion images. In any of these cases, the *BAC* results increase was limited to a maximum of $1.78pp$, when checked against to the *Proposed (RGB)* – 71.43%. In comparison, the use of depth information (RGBZ) adds $11.91pp$ on top of that. Meaning that the proposed pipeline always outperforms the EfficientNet models if supplied with depth information.

5. Conclusions and future work

Automated melanoma discrimination is crucial to aid dermatologists improving their diagnostic accuracy. The pursuit for a solution to automatically identify melanoma has been under study for decades. Still, discrimination of melanoma, even with Deep Learning methods, remains a challenging problem and current systems are yet to achieve satisfactory sensitivity performances. Rather than continuously attempting to improve algorithms by using the same data as commonly used by dermatology experts, other dimensions and modalities, as the skin lesion 3D surface, should be explored. In order to go beyond current state-of-the-art results, more reliable solutions might include merging 2D data together with other dimensional aspects, such as surface, which has potential to provide extended melanoma discrimination capabilities.

Taking advantage of the recently introduced technology of the light-field cameras, apart from the proposed pipeline, the main contribution of this paper is the evaluation of the skins' 3D surface data as an alternative data modality when performing melanoma classification, as well as its comparison to current state-of-the-art results. This is done resorting to a recent dataset of multi-dimensional imaging, which was specifically acquired for this goal. Because the data originates from light-field imagery, every image-pixel data comprises both dimensions, enabling the creation of a proposed pipeline which operates in a same comparable setting.

Despite the large class imbalance (often present in medical image datasets) and limited data samples, the attained classification results appear to surpass the sensitivity and specificity to discriminate melanomas from nevi, not only of the state-of-the-art algorithms, but also of human specialists. In the proposed approach pipeline (*RGBZ*), the melanoma discrimination against nevus was achieved with 94.00% accuracy (comprising 92.86% sensitivity and 94.44% specificity) when combining 2D data with depth, a $26.00pp$ accuracy increase in relation to the state-of-the-art baseline method.

In a similar setting, for the discrimination of melanomas against all other available skin lesions, the proposed approach achieved 89.80% accuracy (comprising 78.57% sensitivity and 91.67% specificity), an increase of $16.33pp$ in relation to the state-of-the-art baseline method.

The experimental assessment allows to conclude that image classification problems, including melanoma skin lesion classification, can be further improved by including 3D information, such as surface depth data.

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.

CRediT authorship contribution statement

Pedro M.M. Pereira: Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Lucas A. Thomaz:** Methodology, Formal analysis, Writing – review & editing, Supervision. **Luis M.N. Tavora:** Validation, Formal analysis, Writing – review & editing. **Pedro A.A. Assuncao:** Validation, Writing – review & editing. **Rui M. Fonseca-Pinto:** Supervision, Funding acquisition. **Rui Pedro Paiva:** Writing – review & editing, Supervision, Funding acquisition. **Sergio M. M. de Faria:** Conceptualization, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT), Portugal, under PhD Grant SFRH/BD/128669/2017, Programa Operacional Regional do Centro, project PlenoISLA POCI-01-0145-FEDER-028325 and by FCT/MCTES through national funds and when applicable co-funded by EU funds under the project UIDB/EEA/50008/2020.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC Superpixels compared to state-of-the-art superpixel methods. Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.

Adegun, A., Viriri, S., 2020. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. Artif. Intell. Review 1–31.

Adel, T., Cohen, T., Caan, M., Welling, M., group, A.s., Initiative, A.D.N., et al., 2017. 3D Scattering transforms for disease classification in neuroimaging. NeuroImage: Clin. 14, 506–517.

Alliance, M. R., 2020. Melanoma statistics. https://www.curemelanoma.org/about-melanoma/melanoma-statistics-2/ Accessed: 2020-02-10.

Andén, J., Mallat, S., 2014. Deep scattering spectrum. Trans. on Signal Process. 62 (16), 4114–4128.

Barata, C., Celebi, M.E., Marques, J.S., 2018. A survey of feature extraction in dermoscopy image analysis of skin cancer. J. Biomed. Health Inform. 23 (3), 1096–1109.

Belkin, M., Hsu, D., Ma, S., Mandal, S., 2018. Reconciling modern machine learning practice and the bias-variance trade-off 1–23 arXiv preprint arXiv:1403.1687.

Bisla, D., Choromanska, A., Berman, R.S., Stein, J.A., Polsky, D., 2019. Towards automated melanoma detection with deep learning: Data purification and augmentation. In: IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit. Workshops, pp. 2720–2728. Long Beach, CA, USA

Brinker, T.J., Hekler, A., Enk, A.H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., Holland-Letz, T., et al., 2019. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur. J. Cancer 113, 47–54.

Bruna, J., Mallat, S., 2013. Invariant scattering convolution networks. Trans. Pattern Anal. Mach. Intell. 35 (8), 1872–1886.

Chudáček, V., Andén, J., Mallat, S., Abry, P., Doret, M., 2013. Scattering transform for intrapartum fetal heart rate variability fractal analysis: a case-control study. Trans. Biomed. Eng. 61 (4), 1100–1108.

Cichorek, M., Wachulska, M., Stasiewicz, A., Tymińska, A., 2013. Skin melanocytes: biology and development. Adv. in Dermatol. and Allergol. 30 (1), 30–41.

Csáji, B.C., 2001. Approximation with artificial neural networks. Faculty of Sciences, Etvs Lornd University, Hungary Ph.D. thesis.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: IEEE Conf. on Comput. Vis. and Pattern Recognit., pp. 248–255. Miami, FL, USA

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nat. 542 (7639), 115–118.

Faria, S., Filipe, J., Pereira, P., Tavora, L., Assuncao, P., Santos, M., Fonseca-Pinto, R., Santiago, F., Dominguez, V., Henrique, M., 2019. Light field image dataset of skin lesions. In: Int. Conf. of the IEEE Eng. in Med. and Biol. Soc., pp. 3905–3908. Berlin, Germany

Faria, S., Santos, M., Assuncao, P., Tavora, L., Thomaz, L., Pereira, P., Fonseca-Pinto, R., Santiago, F., Dominguez, V., Henrique, M., 2019. Dermatological imaging using a focused plenoptic camera: the SKINL2 light field dataset. In: Conf. on Telecommun., pp. 1–4. Lisbon, Portugal

Feng, H., Berk-Krauss, J., Feng, P.W., Stein, J.A., 2018. Comparison of dermatologist density between urban and rural counties in the united states. JAMA Dermatol. 154 (11), 1265–1271.

Fernando, K.R.M., Tsokos, C.P., 2021. Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks. IEEE Trans. Neural Netw. Learn. Syst. Early Access, 1–12.

Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., Werner, R., Schlaefer, A., 2019. Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting. Trans. Biomed. Eng. 67 (2), 495–503.

Gonzalez-Diaz, I., 2018. DermaKNet: incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. J. Biomed. Health Inform. 23 (2), 547–559.

Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A.B.H., Thomas, L., Enk, A., et al., 2018. 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), 1836–1842.

Hagerty, J.R., Stanley, R.J., Almubarak, H.A., Lama, N., Kasmi, R., Guo, P., Drugge, R.J., Rabinovitz, H.S., Oliviero, M., Stoecker, W.V., 2019. Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images. J. Biomed. Health Inform. 23 (4), 1385–1391.

Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I., Chang, S.E., 2018. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. J. Investig. Dermatol. 138 (7), 1529–1538.

He, K., Girshick, R., Dollár, P., 2019. Rethinking ImageNet pre-training. In: IEEE/CVF Int. Conf. on Comput. Vis., pp. 4918–4927. Seoul, South Korea

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conf. on Comput. Vis. and Pattern Recognit., pp. 770–778. Las Vegas, NV, USA

Hosny, Kassem, K.M., Foaud, M.A., Mohamed, M., 2019. Classification of skin lesions using transfer learning and augmentation with alex-net. PLoS ONE 14 (5), 1–17.

Hu, K., Niu, X., Liu, S., Zhang, Y., Cao, C., Xiao, F., Yang, W., Gao, X., 2019. Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model. Biomed. Signal Process. Control 51, 200–209.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: PMLR Int. Conf. on Mach. Learn., pp. 448–456. Lille, France

Khan, M.Q., Hussain, A., Rehman, S.U., Khan, U., Maqsood, M., Mehmood, K., Khan, M.A., 2019. Classification of melanoma and nevus in digital images for diagnosis of skin cancer. IEEE Access 7, 90132–90144.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. in Neural Inf. Process. Syst. 25, 1097–1105.

Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine regions. Trans. Pattern Anal. and Mach. Intell. 27 (8), 1265–1278.

LeCun, Y., Kavukcuoglu, K., Farabet, C., 2010. Convolutional networks and applications in vision. In: IEEE Int. Symp. on Circuits and Syst., pp. 253–256. Paris, France

Li, Y., Sun, J., Tang, C.-K., Shum, H.-Y., 2004. Lazy snapping. Trans. Graph. 23 (3), 303–308.

Li, Z., Zhang, X., Müller, H., Zhang, S., 2018. Large-scale retrieval for medical image analytics: a comprehensive review. Med. Image Anal. 43, 66–84.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Liu, Y., Jain, A., Eng, C., Way, D.H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S., et al., 2020. A deep learning system for differential diagnosis of skin diseases. Nat. Med. 26 (6), 900–908.

Mallat, S., 2012. Group invariant scattering. Commun. on Pure and Appl. Math. 65 (10), 1331–1398.

Marchetti, M.A., Codella, N.C.F., Dusza, S.W., Gutman, D.A., Helba, B., Kalloo, A., Mishra, N., Carrera, C., Celebi, M.E., DeFazio, J.L., et al., 2018. Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J. the American Acad. of Dermatol. 78 (2), 270–277.

McDonagh, S., Fisher, R., Rees, J., 2008. Using 3D information for classification of non-melanoma skin lesions. In: Med. Image Underst. and Anal., pp. 164–168. Dundee, United Kingdom

Mishkin, D., Sergievskiy, N., Matas, J., 2017. Systematic evaluation of convolution neural network advances on the imagenet. Comput. Vis. and Image Underst. 161, 11–19.

Navarro, F., Escudero-Viñolo, M., Bescós, J., 2018. Accurate segmentation and registration of skin lesion images to evaluate lesion change. J. Biomed. Health Inform. 23 (2), 501–508.

Pampena, R., Kyrgidis, A., Lallas, A., Moscarella, E., Argenziano, G., Longo, C., 2017. A meta-analysis of nevus-associated melanoma: prevalence and practical implications. J. Am. Acad. Dermatol. 77 (5), 938–945.

Pathan, S., Prabhu, K., Siddalingaswamy, P., 2018. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions a review. Biomed. Signal Process. Control 39, 237–262.

Pereira, P., Fonseca-Pinto, R., Paiva, R., Assuncao, P., Tavora, L., Thomaz, L., Faria, S., 2020. Skin lesion classification enhancement using border-line features - the melanoma vs nevus problem. Biomed. Signal Process. Control 57, 101765.

Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P., 2020. Designing network design spaces. In: IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit., pp. 10428–10436. Virtual

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G.-Z., 2016. Deep learning for health informatics. J. Biomed. Health Inform. 21 (1), 4–21.

Satheesha, T., Satyanarayana, D., Prasad, M., Dhruve, K., 2017. Melanoma is skin deep: a 3D reconstruction technique for computerized dermoscopic skin lesion classification. J. Transl. Eng. Health Med. 5, 1–17.

Senan, E.M., Jadhav, M.E., 2019. Classification of dermoscopy images for early detection of skin cancer – a review. Int. J. of Comput. Appl. 975, 8887.

Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. Trans. Med. Imag. 35 (5), 1285–1298.

Sifre, L., Mallat, S., 2013. Rotation, scaling and deformation invariant scattering for texture discrimination. In: IEEE Conf. on Comput. Vis. and Pattern Recognit., pp. 1233–1240. Portland, OR, USA

Sifre, L., Mallat, S., 2014. Rigid-motion scattering for texture classification 1–9 arXiv preprint arXiv:1403.1687.

Smith, L., Smith, M., Farooq, A., Sun, J., Ding, Y., Warr, R., 2011. Machine vision 3D skin texture analysis for detection of melanoma. Sens. Rev. 31 (2), 111–119.

Society, A. C., 2020. Melanoma skin cancer statistics. https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html Accessed: 2020-02-10.

Soyer, H.P., Argenziano, G., Zalaudek, I., Corona, R., Sera, F., Talamini, R., Barbato, F., Baroni, A., Cicale, L., Di Stefani, A., et al., 2004. Three-point checklist of dermoscopy. Dermatol. 208 (1), 27–31.

Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2015. Striving for simplicity: the all convolutional net 1–14 arXiv preprint arXiv:1412.6806.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning. In: PMLR Int. Conf. on Mach. Learn., pp. 1139–1147. Atlanta, GA, USA

Tan, M., Le, Q., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. In: PMLR Int. Conf. on Mach. Learn., pp. 6105–6114. Long Beach, CA, USA

Tang, P., Liang, Q., Yan, X., Xiang, S., Zhang, D., 2020. GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification. J. Biomed. Health Inform. 24 (10), 2870–2882.

Vestergaard, M.E., Macaskill, P., Holt, P.E., Menzies, S.W., 2008. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. Br. J. of Dermatol. 159 (3), 669–676.

Waldspurger, I., 2015. Wavelet transform modulus: phase retrieval and scattering. Ecole doctorale 386: Sciences Mathématiques de Paris Centre Ph.D. thesis.

Waldspurger, I., 2017. Exponential decay of scattering coefficients. In: IEEE Int. Conf. on Sampl. Theory and Appl., pp. 143–146. Tallin, Estonia

Wu, Y., He, K., 2018. Group normalization. In: Eur. Conf. on Comput. Vis., pp. 3–19. Munich, Germany

Xie, Y., Zhang, J., Xia, Y., Shen, C., 2020. A mutual bootstrapping model for automated skin lesion segmentation and classification. Trans. Med. Imaging 39 (7), 2482–2493.

Yang, J., Sun, X., Liang, J., Rosin, P.L., 2018. Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. In: IEEE Conf. on Comput. Vis. and Pattern Recognit., pp. 1258–1266. Salt Lake City, UT, USA

Yang, J., Wu, X., Liang, J., Sun, X., Cheng, M.-M., Rosin, P.L., Wang, L., 2019. Self-paced balance learning for clinical skin disease recognition. Trans. Neural Netw. Learn. Syst. 31 (8), 2832–2846.

Yao, P., Shen, S., Xu, M., Liu, P., Zhang, F., Xing, J., Shao, P., Kaffenberger, B., Xu, R.X., 2021. Single model deep learning on imbalanced small datasets for skin lesion classification 1–10 arXiv preprint arXiv:1403.1687.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding neural networks through deep visualization 1–12 arXiv preprint arXiv:1506.06579.

Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A., 2016. Automated melanoma recognition in dermoscopy images via very deep residual networks. Trans. Med. Imaging 36 (4), 994–1004.

Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T., 2018. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. Trans. Biomed. Eng. 66 (4), 1006–1016.

Yuan, Y., Chao, M., Lo, Y.-C., 2017. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. Trans. Med. Imaging 36 (9), 1876–1886.