IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Diagnosis of melanoma from dermoscopic images using a deep depthwise separable residual convolutional network

ISSN 1751-9659
Received on 25th December 2018
Revised 11th June 2019
Accepted on 26th June 2019
E-First on 14th August 2019
doi: 10.1049/iet-ipr.2018.6669
www.ietdl.org

Rahul Sarkar[1] ✉, *Chandra Churh Chatterjee*[1], *Animesh Hazra*[1]

[1]*Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India*

✉ *E-mail: rahulsarkar.2798@gmail.com*

Abstract: Melanoma is one of the four major types of skin cancers caused by malignant growth in the melanocyte cells. It is the rarest one, accounting to only 1% of all skin cancer cases. However, it is the deadliest among all the skin cancer types. Owing to its rarity, efficient diagnosis of the disease becomes rather difficult. Here, a deep depthwise separable residual convolutional algorithm is introduced to perform binary melanoma classification on a dermoscopic skin lesion image dataset. Prior to training the model with the dataset noise removal from the images using non-local means filter is performed followed by enhancement using contrast-limited adaptive histogram equilisation over discrete wavelet transform algorithm. Images are fed to the model as multi-channel image matrices with channels chosen across multiple color spaces based on their ability to optimize the performance of the model. Proper lesion detection and classification ability of the model are tested by monitoring the gradient weighted class activation maps and saliency maps, respectively. Dynamic effectiveness of the model is shown through its performance in multiple skin lesion image datasets. The proposed model achieved an ACC of 99.50% on international skin imaging collaboration (ISIC), 96.77% on PH2, 94.44% on DermIS and 95.23% on MED-NODE datasets.

1 Introduction

Melanoma is the most dangerous type of skin cancer usually caused by prolonged exposure to ultraviolet (UV) radiation, which causes mutation in the melanocyte cells [1]. According to the Skin Cancer Foundation of US [2], the statistics reveal that there will be 178,500 new melanoma cases diagnosed in the year 2018. As per Cancer Research UK [3], there were 15,906 additional cases of melanoma skin cancers in 2015 and 2285 deaths from melanoma in 2016. So, early diagnosis of the disease is of supreme priority, because it can be cured by surgery if treated at an early stage, whereas if detected at an advanced stage, treatment merely focuses on the slowing down of its growth. Melanoma is considered to be the most serious variant of skin cancer because it has the potential to spread. Chemotaxis and lymph flow are responsible to spread melanoma metastases, which develop on the lymph nodes. The treatment of melanoma is rather complex as melanocyte cells are resistant to UV light and reactive oxygen species [4].

Dermatologists use the asymmetry, border irregularity, color that is not uniform, diameter greater than 6 mm, evolving size (ABCDE) rule to diagnose melanoma from dermoscopic images containing suspicious moles, where ABCDE stands for asymmetry, border, colour, diameter and evolving. However, it is evident that there is a quite variation in melanoma pictures, which makes it difficult to detect the disease from the images. Also, according to the Melanoma Research Foundation, not all the melanoma images can be diagnosed using the ABCDE guidelines, as shown in Fig. 1. Hence, it is important that we have to differentiate a dysplastic nevus from melanoma moles. We train our model on the international skin imaging collaboration (ISIC) dataset [5], which includes a diverse class of skin lesion images, thereby making the model more accurate and versatile in classifying melanoma. The model's performance is further cross-validated on the PH2 dataset [6], the DermIS dataset [7] and finally the MED-NODE dataset [8], which gives us a scope to verify the dynamic nature of the model.

Fig. 1 *Melanoma images that cannot be diagnosed by ABCDE rule*

Our aim is to enhance the melanoma detection in dermoscopic images at any stage of the disease by introducing a novel and efficient algorithm for building the model in question. The proposed model has been built by incorporating a separable convolution algorithm, which saves the parameter space and boosts up the execution of the network while reducing its time complexity. Minimal error and optimal accuracy (ACC) are achieved by using the residual learning algorithm, which ensure proper transmission of the error throughout the network during the training phase. Additionally, to achieve the best performance results, we merge the red, green and blue (RGB) images with b^* channel from the CIELAB (CIE L* for lightness, a* for green–red component and b* for blue–yellow component of the image) colour space, saturation value from the hue, saturation and value (HSV) image format and an inverted grey-scale format of the image to form a six-channel image matrix. We made sure to preserve the aspect ratio of the core images, in order to conserve the dimension of the lesions, as it is an important factor required for the proper classification of the images. Enhancement and noise removal was performed on the images with optimal parameters to aid the diagnosis. The validity of the proposed method for proper identification and classification of the skin lesion was tested by obtaining gradient weighted class activation maps (GRAD-CAMs) for the convolution layers and saliency maps of the fully connected (FC) layers.

The remaining portion of this paper is arranged as follows: Section 2 provides a detailed overview of the survey of the related works that was performed prior to developing the model. Section 3 explains the methodology proposed, where the preprocessing techniques, the model development and model optimisation phases are discussed in details. Section 4 showcases the results and discussion to support the approach and finally Section 5 concludes the work and discusses its future scopes.

2 Related work

A rigorous study of the existing works toward melanoma diagnosis was performed and the proposed methodology was formulated so

as to validate its novelty and superiority over the existing algorithms. In this section, we summarise all the related works that were studied and present them in a nutshell.

The idea of melanoma detection using a statistical model was started quite a while back. In the year 2006, Tommasi *et al.* [9] proposed a method of melanoma recognition by implementing the two kernel-based classifiers using support vector machine (SVM) and a probabilistic approach using spin glass–Markov random fields. As a discriminative method, they chose SVM and Mercer Kernel so that the scalar products in a linear SVM can be replaced via the kernel function, which improve the recognition rate of the SVM. As the case is our paper, this paper only focuses on melanoma detection alone and its dynamic effectiveness is also not verified. In the same year, Mocellin *et al.* [10] developed an SVM network to predict the status of sentinel nodes in the patients having cutaneous melanoma. The use of SVM was to construct a hyperplane between the classes having optimal separation, namely positive and negative, which establishes the fact that the sentinel node status can be evaluated as a forecasting method to avoid the sentinel node biopsy (SNB). The cost of the proposed system used in health care is the advantage of this paper with the cons that there exists 1% risk of incorrectly submitting patients to observation instead of SNB.

In the year 2012, Shimizu *et al.* [11] proposed a system for automated melanoma screening using the attributes extracted from images, with two classification models for melanoma, where both models use linear classifiers and the second classifier takes the classification output of the first classifier as input. The study explained the concept of a double shot detection model, which splits the classification task into two sub-tasks, hence executing a feature extraction for better performance. There exists a limitation of the number of attributes that can be used for the model. Duarte *et al.* [12] presented a melanoma classification system based on the hidden Markov tree features, enabling the extraction of melanin intensity levels from skin lesions and used Neyman–Pearson SVM to tackle diverse classification for the feature vectors of wavelet transformed images of the pigmented lesions. Like most of the reviewed methods, this one focuses on selective feature extraction; however, in our paper we tackle feature extraction automatically using deep learning. In 2013, Ikuma and Lyatomi [13] explored an innovative screening system for melanoma detection founded on the adaptive fuzzy inference neural network (AFINN), where only 88 fuzzy rules were developed resulting in 88 nodes in the rule layer of the AFINN. The model achieved a sensitivity (SE) of 81.5% and a specificity (SP) of 73.9% but, due to the limited classification ability, the model proves to be inferior as compared with other multi-layer neural networks or non-linear models. This contribution was followed by Cavalcanti *et al.* [14] who developed different systems for melanoma detection and classification based on a two-stage approach using skin lesion detection, which proves to be very robust. Colour and textual descriptors were not used for identification of the dermatological features, which could have increased the robustness of this paper. Mhaske and Phalke [15] used supervised and unsupervised learning which separate the data points into different categories for the SVM to produce high-ACC results. The *K*-means algorithm proved to underperform for this classification task based on the performance scores provided by the author. Ballerini *et al.* [16] proposed a hierarchical *K*-nearest-neighbour (KNN) approach dependent on the texture and colour features obtained from the skin lesion images, which improve the ACC over the flat-KNN approach. The data for the actinic keratosis (AK) and squamous cell carcinoma (SCC) used for the classification were very low, hence resulting in a poor output for the classifications of AK and SCC.

In 2014, Barata *et al.* [17] developed a system for the detection of lesions from dermoscopic images by utilising the colour and texture features and compared the impact of two features in melanoma diagnosis using three classifiers including AdaBoost classifier, SVM and KNN classifiers. The important advantage of this paper is that it focuses on the computation of local features as well as global features for extensive feature extraction with generation of good results for both the computations. In the same year, Abuzaghleh *et al.* [18] proposed a system for advanced detection and prevention of melanoma from the analysis of skin lesions based on the colour and shape geometry feature sets. Classification was done using two-level classifiers with a single classifier in the first and two classifiers in the second level, where SVM was used as a classifier at every level. The paper explained that the two-level classifier outperformed the single-level classifier, thus claiming the robustness of the two-level classifiers over single classifier. A multi-level classification task, as proposed in this paper, always hinders the computation speed of the model, hence, proving to be a major shortcoming.

Valavanis *et al.* [19] in 2015 proposed a robust technique to explore the cutaneous melanoma having diagnostic signatures based on the imaging and genetic data. The important advantage of this paper is the association of the disease descriptors corresponding to imaging features, which were laid to the low-level biological information related to gene expression. Codella *et al.* [20] depicted a system combining SVM, sparse coding and deep learning for melanoma identification in dermoscopic images. The above-mentioned paper was conducted on the ISIC dataset. The beauty of this paper is the use of unsupervised learning within the domain and feature extraction from the natural photographs eliminating the need for annotated data in the classification task. The model proposed in this paper incorporates deep learning, which increases time and space complexity of the model. This specific problem is tackled by our model.

Contributions made in 2016 and 2017 include melanoma diagnosis using lesion segmentation by Mishra and Celebi [21], in which lesion segmentation followed by feature segmentation is performed which helps in maximising the classification ACC and is the advantage of this paper. Two-stage segmentation is performed for feature extraction, which is done automatically by our model. A convolutional neural network (CNN) architecture-based model for detection and tracking of the disease was built by Li *et al.* [22], where human interpretable detection has been established. The drawback of this paper is that the network is trained on the synthetic data and tested on real-world data, which decreases the overall validity of this paper. Skin lesion extraction from non-dermoscopic images by applying deep learning was developed by Jafari *et al.* [23], where the images are taken based on the local and global patches, which increased the overall performance of the network. Little importance was given to preprocessing the dataset. Malignant melanoma detection using an artificial neural network (ANN) and adaptive network-based fuzzy inference system, where feature extraction was performed using discrete wavelet transform (DWT) and principal component analysis was developed by Arasi *et al.* [24]. No significant effort was put into making the model time and space efficient, which is of paramount importance when deploying a model for real-time classification. Enhancement of melanoma recognition ACC using Bayesian decision fusion of three standalone parallel SVM classifiers was proposed by Takruri and Abubakar [25]. The use of fusion of multiple classifiers for improving the melanoma detection rate claims the robustness of the methodology. The dataset used for this paper constituted only grey-scale images, which reduce its overall validity as it does not follow the ABCDE rule of lesion classification. Deep learning using ensemble model including deep residual networks, convolutional networks, fully convolutional U-Net architecture as well as sparse coding and hand-coded feature representations was introduced by Codella *et al.* [26]. This paper actually resulted in a state-of-the-art performance for the ensemble model used for the classification task. Melanoma diagnosis using non-linear and linear features from digital images including a blending of Otsu thresholding and *K*-means clustering methods to extract borders of the affected region was explained by Munia *et al.* [27], which used SVM, decision trees, KNN and random forest for classification. Here, the ABCDE rules of the dermoscopic images were not incorporated into the methodology leading to a huge drawback of the system also; the feature extraction performed can further be improved using better techniques.

Improvements in the field of segmentation of melanoma images were brought about by using convolutional and deconvolutional networks which was introduced by Yuan and Lo [28] who depict the efficient use of seven channels of the dermoscopic images

(RGB, HSV and the L channel of the CIELAB colour space). The use of the multiple colour spaces for segmentation of the dermoscopic images inspired the proposed classification model based on the multiple colour models. Skin lesion classification from dermoscopic images using deep learning techniques which involved the VGGNet CNN architecture was introduced by Lopez *et al.* [29], where the volume of the dataset used for training was very less with respect to the classification algorithm that was used (VGGNet convolutional network architecture). Performing transfer learning using a network whose original classification task varies greatly from lesion classification is not a very practical approach. A system of skin lesion detection based on fusing the outputs of the softmax layers of the four different neural network architectures was proposed by Harangi [30], where ensemble learning of different deep CNN (DCNN) architectures were used for the classification task. No extensive feature extraction was performed which is the only drawback of the corresponding paper. In the year 2017, the combination of textual and structural features of dermoscopic images for the melanoma recognition task was explained by Adjed *et al.* [31], where the PH2 dermoscopic database with a total of 200 images was only used for this paper, making it prone to overfitting.

In 2018, Codella *et al.* [32] discussed the several models built for lesion segmentation, feature detection and disease classification in the International Symposium on Biomedical Imaging challenge 2017. Incomplete dermoscopic feature annotations and dataset bias degraded the performance of the system. Melanoma detection using deep learning network, which included two fully convolutional residual networks (FCRN) for disease classification was proposed by Li and Shen [33]. The network proposed in the corresponding paper simultaneously addressed the lesion segmentation and the classification task, where two deep FCRN were used along with two different training sets for this paper. Deep attention network optimised using attention mechanism and fisher criteria is explained by Ma and Yin [34] for melanoma detection. This paper tackles the problem of insufficient training data by using the attention module to learn features of the dermoscopic images. Ensemble modelling using ResNet-50 and Inception V3 discussed in this paper was explained by Shahin *et al.* [35] in the year 2018, for efficient skin lesion classification. No prior segmentation or image preprocessing was performed on the images to optimise the performance of the model. Also, both the networks used to build the ensemble model are used for classification task completely unrelated to lesion classification. Two-step classification of melanocytic and non-melanocytic skin lesions using AlexNet deep neural net architecture is explained in this paper conducted by Kaymak *et al.* [36]. Here, AlexNet was used for both the steps, where in the first step melanocytic and non-melanocytic lesions were distinguished followed by the second level of classification for malignant and benign lesions. The system generated poor results for the classification of melanocytic and non-melanocytic skin lesions as AlexNet was not built to analyse skin images. Yu *et al.* [37] in 2018 explained a combination of deep learning method over local descriptor encoding strategy. The important advantage of this paper is that it can produce many distinctive features and can operate over large variations in the melanoma classes. Multiple CNN models for melanoma classification were tested by Guo *et al.* [38]. No significant preprocessing was performed to aid the detection/classification task. In the year 2018, Jaisakthi *et al.* [39] explained the noise removal algorithms followed by a GrabCut and K-means segmentation techniques for efficient skin lesion segmentation of dermoscopic images. Several algorithms including contrast-limited adaptive histogram equilisation (CLAHE) algorithm for correcting the illumination of the images are described in this paper. A DCNN architecture, which extracts low-dimensional discriminative features for melanoma detection, is described by Sultana *et al.* [40]. Not much attention is provided to the optimisation of the model by tweaking the number of layers or the kernel size in the layers.

A novel regulariser embedded with the CNN for skin lesion classification into benign and malignant categories is described by Albahar [41]. The regulariser that is used is based on the standard deviation of the weight matrix of the classifier using CNN architectures for optimal performance of the network. Fusion between handcrafted and deep learning features for optimal performance on classification task is explained by Hagerty [42]. Other deep classification networks, except the ResNet-50 need to be explored for improving performance of the system. In this paper conducted by Mahbod *et al.* [43], transfer learning on various pre-trained CNN architectures was conducted along with fusion of the deep features for improving the robustness of the system. Transfer learning in this paper was performed using networks which specialise in object classification and is not related to skin lesion analysis. An attention residual learning CNN architecture was described by Zhang *et al.* [44] in the year 2019. Here, feature maps learnt by higher layers have been fed to the lower layers, thus increasing the optimal performance of the model. The attention wrapper also aids the localisation of the affected region while further boosting the performance of the CNN model. The only drawback of the model is its training time and space complexities.

An overview on few of the methodologies mentioned above which were successful in acquiring high-performance scores is summarised in Table 1 as follows.

3 Proposed methodology

Deep learning is generally considered to be a powerful algorithm when it deals with a large dataset of high-resolution images. However, it should also be noted that the deep learning algorithms require considerably high-parameter storage space and computation time.

Table 1 Summary of some of the existing statistical models used to diagnose melanoma

Authors	Year	Employed methodology	ACC, %
Hagerty *et al.* [42]	2019	handcrafted method over deep learning	94.00
Albahar *et al.* [41]	2019	CNN with novel regulariser	97.49
Shahin *et al.* [35]	2018	ensemble learning over dermoscopic images	89.90
Kaymak *et al.* [36]	2018	two-step deep learning classifier	84.00
Yu *et al.* [37]	2018	aggregated deep convolution features	86.81
Guo *et al.* [38]	2018	multiple convolution neural network	85.22
Li and Shen [33]	2018	residual convolution network	85.70
Jaisakthi *et al.* [39]	2018	GrabCut and K-means algorithm	91.93
Arasi *et al.* [24]	2017	network-based fuzzy algorithm and ANN	98.80
Codella *et al.* [26]	2017	ensemble learning	85.50
Yuan and Lo [28]	2017	convolution–deconvolution neural network	93.40
Jafari *et al.* [23]	2017	CNN	98.70
Codella *et al.* [20]	2015	SVM, deep learning and sparse coding	93.10
Abuzaghleh *et al.* [18]	2014	colour and space geometry features of lesions	97.70
Ballerini *et al.* [16]	2013	hierarchical KNN algorithm	93.86
Cavalcanti *et al.* [14]	2013	two-stage melanocytic discrimination	99.34

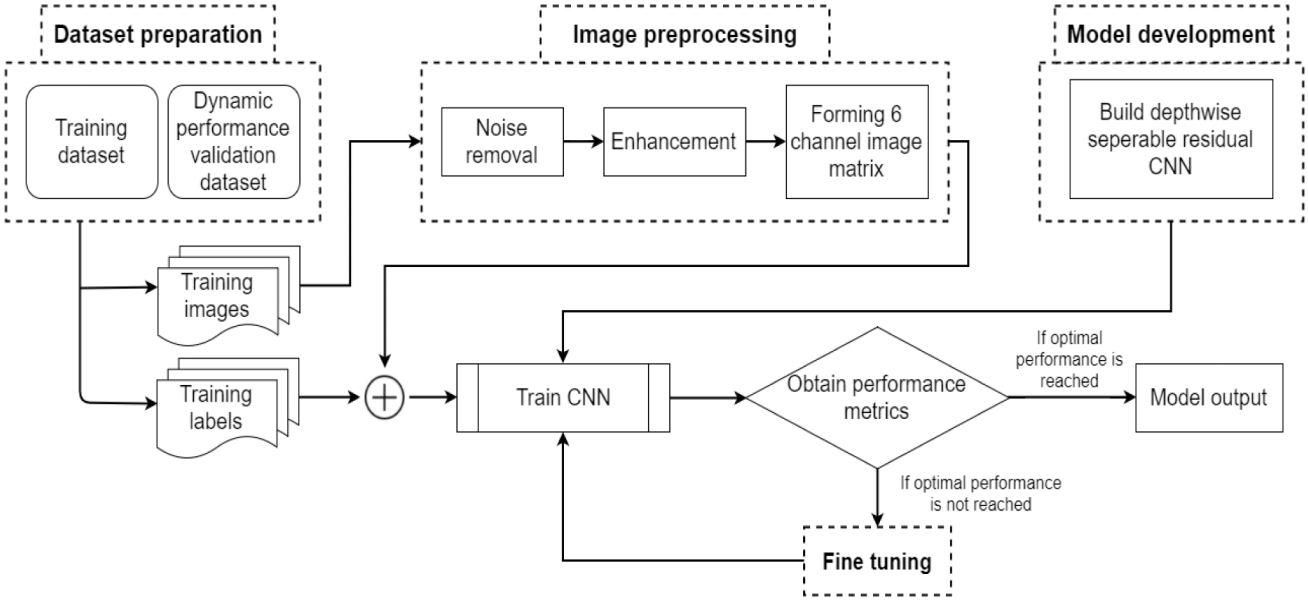


Fig. 2 *Overview of the proposed methodology*

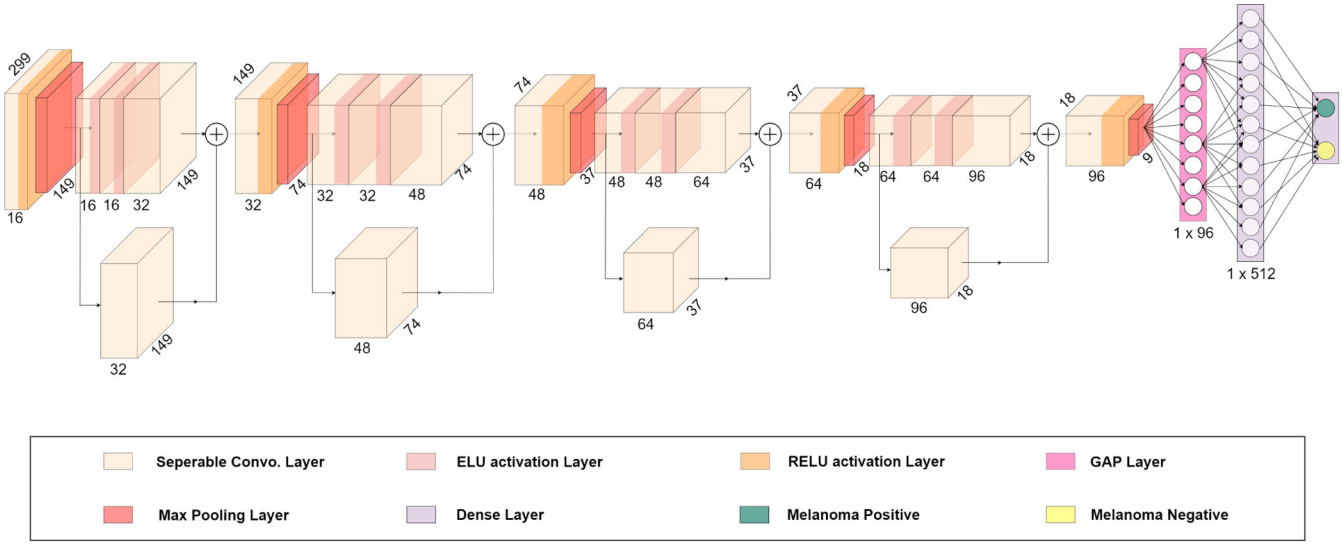


Fig. 3 *Architecture of the proposed model*

In this paper, we use separable convolution algorithm in order to design a model, which has a parameter complexity $O(w \times d)$ as opposed to $O(w^d)$, where the kernel is w elements wide in each dimension and the convolution is of d-dimensions. The reduced parameter storage space allows us to use more channels for each image. The significance of including multiple channels in the image matrix is discussed in Section 3.2.3. Residual learning is also used to boost ACC and reduce error in the classification task. The model thus developed is a highly accurate and versatile in the task of melanoma detection from dermoscopic images.

Fig. 2 shows a brief overview of the proposed methodology and Fig. 3 provides an architectural overview of the network. In the following sections, we discuss in details all the stages involved in developing the classification algorithm.

3.1 Dataset preparation

The model has been trained and validated using the ISIC dataset. It is a publicly available dataset and can be used to build the classification models on melanoma, basal cell carcinoma, SCC and AK. However, to build the model, we select only those images from the dataset, which were classified for melanoma diagnosis purpose only.

An appropriate subset of the ISIC image dataset [5] was selected in order to prevent any memory error. To train and validate the model in question, we use 4000 images (2050 benign images and 1950 malignant images) and splitting the image dataset into a training set of 3400 images and a validation set of 600 images.

To validate the dynamic performance of the model, we obtain the model's score across three different datasets represented as follows:

- PH2 dataset [6] (200 images).
- DermIS dataset [7] (69 images).
- MED-NODE dataset [8] (170 images).

The images in the aforementioned datasets do not undergo same preprocessing steps as performed on the ISIC dataset, due to the variation in hardware involved in developing these images, they differ in brightness and contrast (Fig. 4). Hence, we simply use CLAHE and Gaussian filtering for enhancement and noise removal, respectively. Also, to train the model on the aforementioned datasets, we perform augmentation on the images, up-scaling each dataset to three times its original size.

3.2 Dataset preprocessing

Preprocessing is essential when working with image data. However, while implementing the deep learning algorithms, images are generally lightly preprocessed with little to no effort put

a b c

Fig. 4 *Images from ISIC dataset prior to any processing*
(a)–(c) Images from ISIC dataset prior to any processing

procedure PREPROCESS($D_{set}, filterType, *param$)
  $X_{set} \leftarrow D_{set}.images$
  $y_{set} \leftarrow D_{set}.labels$
  $i \leftarrow 1$
  $E_{set} \leftarrow Null$
  while $i \leq 4000$ **do**
    $E_{set}.append(denoise(X_{set}[i], filterType, *param))$
    $E_{set}[i] \leftarrow dwt\_clahe(E_{set}[i], *param)$
    $i \leftarrow i + 1$
  $Train_{set}, Test_{set} \leftarrow ShuffleSpilt([X_{set}, y_{set}])$
  $FindOptimalModelParam(Train_{set}, Test_{set})$
procedure FINDOPTIMALMODELPARAM($Train_{set}, Test_{set}$)
  $nb\_resBlock \leftarrow 1$
  $M_{Acc} \leftarrow 0$
  $M_{Roc} \leftarrow 0$
  while $nb\_resBlock \leq 4$ **do**
    $k\_size \leftarrow 1$
    while $k\_size \leq 7$ **do**
      $Model \leftarrow BuildModel(nb\_resBlock, k\_size)$
      $Acc_{val}, Roc_{val} \leftarrow TrainModel(Model,$
        $Train_{set}, Test_{set})$
      if $Acc_{val} > M_{Acc}$ && $Roc_{val} > M_{Roc}$ **then**
        $M_{Acc} \leftarrow Acc_{val}$
        $M_{Roc} \leftarrow Roc_{val}$
        $save(Model)$

Fig. 5 *Algorithm 1: Selection of optimal preprocessing technique*

Table 2 Summary of the proposed model's performance after applying various noise removal algorithms on the images

Noise removal algorithm	ACC, %	AUROC score
non-local means denoising	**99.50**	**0.9949**
Gaussian filtering	96.67	0.9655
average filtering	93.33	0.9342
median filtering	95.50	0.9543
bilateral filtering	97.80	0.9731

The bold values define the denoising algorithm chosen by applying algorithm in Fig. 5.

into this phase. The major effort in these cases goes into making the models for extracting the features from the images automatically. In the proposed methodology, we give relatively more effort into preprocessing the raw image data in order to aid feature extraction by the model, the details of which are discussed in the sections below.

3.2.1 Noise removal from images: All the images were primarily subjected to various noise removal algorithms and non-local means denoising filter was ultimately chosen as being most compatible with the model. The model was initially trained on the dataset without applying any preprocessing algorithms on the images. The resulting model achieved an ACC of 94.61% after complete fine-tuning. To improve the performance of the model, we design a six-channel image matrix. A detailed description of the channel selection for the image matrices has been provided in Section 3.2.3. Now, each image matrix represents a denoised and enhanced version of the respective image. In this section, we discuss the selection process, which we follow to obtain the best denoising algorithm.

We consider the five most common and effective denoising techniques to filter the images, namely non-local means denoising, Gaussian filtering, average filtering, median filtering and bilateral filtering. To select the best denoising algorithm for this paper, we follow Algorithm 1 (Fig. 5), where we record the optimal performance of the model [in terms of ACC and area under receiver operating characteristic (AUROC) curve score] against each denoising algorithm. It should also be mentioned that the denoised images are enhanced using CLAHE-DWT(CLAHE over DWT) prior to feeding them to the model. The performance summaries of the various noise removal algorithms have been provided in Table 2.

As shown in Table 2, non-local means denoising algorithm achieves the best result for the noise removal task. The main reason for the application of noise removal algorithms is to reduce the influence of body hair, which is present in a considerably large number of images in the dataset. Also, in certain cases, the lesions are small in size and the textures on the surrounding skin region can cause a misjudgement in the classification task. Hence, blurring is essential in order to reduce the influence of such regions.

The non-local means filter applied for denoising the images can be defined as follows:

$$u(p) = \frac{1}{C(p)} \times \int_{\Omega} v(q) f(p, q) \mathrm{d}q \,. \tag{1}$$

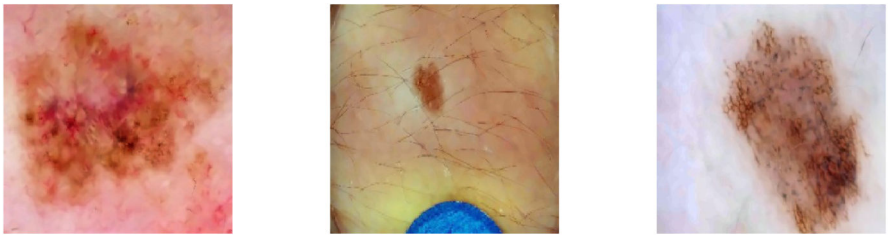
where $u(p)$ is the filtered value of the image at point p; $v(q)$ is the unfiltered value of image at point q; $f(p, q)$ is the weighting function; and the integral is evaluated over $\forall(q) \in \Omega$. $C(p)$ is the normalisation factor given by

$$C(p) = \int_{\Omega} f(p, q) \mathrm{d}q \,. \tag{2}$$

A comparison of the images before and after the noise removal followed by enhancement with CLAHE-DWT using the aforementioned denoising algorithms is shown in Fig. 6.

3.2.2 Enhancement of images: To enhance the edges of the skin lesion and the contrast of the lesion region, image enhancement technique is used which aids in feature extraction by the network. This is essential in case of moles, which cannot be identified as malignant by the ABCDE rule and for dysplastic nevi, which can wrongly be diagnosed as malignant moles but are actually benign. It should also be noted that the enhancement algorithms can cause the benign lesions to fall under malignant categories, thus contrast enhancement must be limited extensively.

Enhancement was performed after removing the noise from the images. If enhancement algorithms are applied to the images prior to noise removal, then the noise in the images will be enhanced and noise removal will then result in over-blurring of the images causing a loss in the information. For these reasons, we implement CLAHE-DWT, which enhances the low-frequency features obtained on applying discrete Haar wavelet transform on the images using CLAHE [45]. This method serves as an improvement to the traditional CLAHE algorithm, and is more suited for the



a *b* *c*

Fig. 6 *Images from Fig. 4 after undergoing noise removal and successive enhancement*
(a)–(c) Images in Fig. 5 after noise removal and successive enhancement operation

enhancement task at hand. Here, the enhancement technique involves decomposing each channel of the RGB images into low-frequency and high-frequency coefficients using DWT performed using the Haar wavelet. The lower-frequency components of each channel are then enhanced using CLAHE, whereas the high-frequency components are kept unchanged. Inverse DWT is then performed to obtain the enhanced version of each channel, and the channels are then merged to obtain the transformed image. A weighted average of the transformed image and the original image is then performed to give the final enhanced image. The equation for the weighted averaging is defined in [45] as follows:

$$I_{\mathrm{E}} = I_{\mathrm{O}} * H + \beta * I_{\mathrm{R}} * (M_1 - H)\,. \tag{3}$$

where * denotes a point-to-point multiplication operation. I_{O}, I_{R} and I_{E} are the original, reconstructed and final enhanced images, respectively; β is the brightness compensation factor, which is used to compensate the decreased luminance which occurs on performing the weighted average. In our method, the value of β is chosen to be 1.5. $\boldsymbol{M}_1$ is the matrix of all ones. H and $\boldsymbol{M}_1 - H$ are the weighting coefficients of I_{O} and I_{R}, respectively. H is defined as follows:

$$H = [f(I_{\mathrm{O}}(i,\ j))^{\alpha}]_{m\times n}, \quad 1 \le i \le m, \quad 1 \le j \le n\,. \tag{4}$$

Considering the image to be of dimensions $m \times n$, $I_{\mathrm{O}}(i,\ j)$ represents the intensity value at pixel coordinate $(i,\ j)$. We consider α to be a regulatory exponent, whose optimal value gives rise to optimal enhancement as described in [45]. The function f is defined as follows:

$$f(I_{\mathrm{O}}(i,\ j)) = \frac{I_{\mathrm{O}}(i,\ j) - I_{\mathrm{Omin}}}{I_{\mathrm{Omax}} - I_{\mathrm{Omin}}}\,. \tag{5}$$

where I_{Omax} and I_{Omin} denote the maximum and minimum intensities values of original image I_{O}, respectively.

CLAHE, as compared with other enhancement algorithms, produce less noise as it divides the image into smaller regions or tiles and perform contrast-limited enhancement on these regions. Bilinear interpolation is implemented by CLAHE to eliminate the region boundaries, which make the regions look smoother. Hence, to enhance the image, we apply CLAHE on only the low-frequency components of the image as explained above using CLAHE-DWT technique. We show the algorithm followed in order to obtain the best possible denoising algorithm in Algorithm 1 (Fig. 5).

In Algorithm 1 (Fig. 5), we use two major functions to find the model parameters, which provide the best performance, in terms of AUROC score and ACC. In the function Preprocess(), we aim to find the best noise removal filter, we do this by executing the Preprocess(D_{set}, filterType, *param) for all the five filter types under consideration (see Table 2 for the list) on the dataset D_{set} with their respective parameters represented by (*param) in the function call. Now, the images in the dataset D_{set}. images are first extracted and then denoised using the function call denoise(< image > , < filterType > , ∗ param), successive enhancement is then performed using CLAHE-DWT, which is represented by the function call dwt_clahe(< image > , ∗ param). Finally, the optimal parameters for the model are found and saved using the procedure FindOptimalModelParam($\mathrm{Train}_{\mathrm{set}}$, $\mathrm{Test}_{\mathrm{set}}$), where the parameters represent the training and testing datasets, respectively. The local variables in the function FindOptimalModelParam(), namely nb_resBlock and k_{size} represent the number of residual blocks and the kernel size used in the model, respectively. Again, M_{Acc} and M_{Roc} represent the maximum ACC and AUROC scores, respectively, which are the decision parameters (initially set to 0) used in the model.

For a particular denoising algorithm, we preprocess the images and then try to find the best model parameters for it. As seen in the algorithm described in Algorithm 1 (Fig. 5), we iterate through each combination of residual block and kernel size and only store the maximum in each case. In our paper, we found the optimal scores to be obtained when the model has four residual blocks and has a convolution kernel of size (4×4). Also, non-local means denoising proved to be the best while following this algorithm. For the proposed model, we observe a sharp fall in performance for kernel sizes going beyond (7×7); hence, we stop increasing the kernel size beyond this point. Similar problems emerge when increasing the number of residual blocks beyond four, so, we consider four to be the maximum possible number of residual block to be present in the model in our paper. The performances thus obtained against the various denoising algorithms have been shown in Table 2.

3.2.3 Selecting multiple channels: For the proposed model, we merge certain channels with the image matrix, which is in RGB colour space. Primarily, a ten-channel image matrix of mixed colour space was fed to the model, namely channels of CIELAB colour space, channels of HSV colour space, channels of the RGB colour space and finally, the grey-scale format of the image. However, this resulted in a considerable reduction in the size of training subset. Thus, only selected channels, which produced the best performance were included. An optimal image matrix with dimensions $(299 \times 299 \times 6)$ was obtained for which training and validation sets of reasonable sizes were selected. The six channels in question are the channels of RGB colour space, saturation channel of the HSV colour space, b* channel of the CIELAB colour space and finally, the single channel from the inverted grey-scale colour space. This matrix is considered to be optimal, because it achieved the maximum increase in ACC from 94.61% (where the image matrix contained channels from the RGB colour space only) to 99.50%.

To obtain the optimal image matrix, we initially considered a matrix consisting of the RGB channels and an extra channel from the aforementioned colour spaces. The channels having the best performance were then chosen to be merged with the RGB image matrix. The summary of the model's performance for the different channels is shown in Table 3.

One of the basic grey-level transformation, the inverted grey-scale format of an image was obtained by using the equation defined as follows:

$$s = (L - 1) - r\,. \tag{6}$$

where s is the intensity value of each pixel in the transformed image; r is the pixel intensity value in the source image; and L is the number of intensity levels. The transformed image in this case showed better contrast between the lesion and its surroundings.

Table 3 Summary of model's performance on merging individual channels to the RGB image matrix

Colour channel	Colour space	ACC, %
single channel	inverted grey scale	96.96
single channel	grey scale	95.60
L* channel	CIELAB	96.42
a* channel	CIELAB	95.76
b* channel	CIELAB	97.15
H channel	HSV	95.36
S channel	HSV	97.52
V channel	HSV	96.27

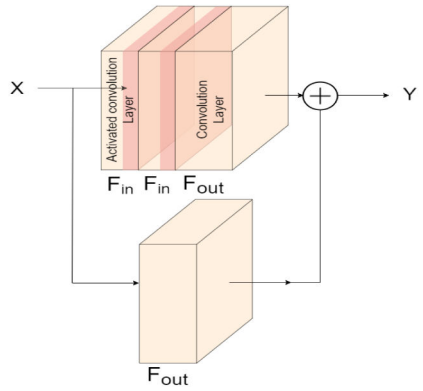


Fig. 7 *Architecture of the basic residual module used in our model*

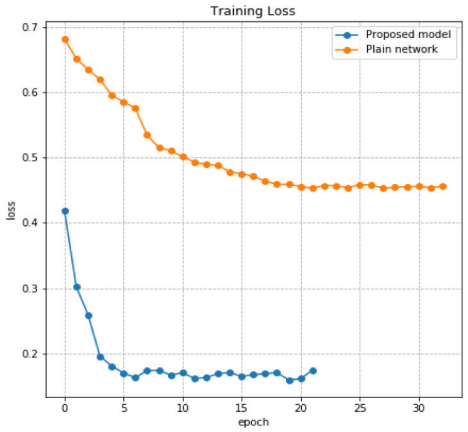


Fig. 8 *Loss curve comparison between our proposed model and a plain network*

However, it should be mentioned that this image was sufficiently blurred but not enhanced.

3.2.4 Normalisation and resizing of images: The images are normalised following the enhancement operation on the images. Normalisation of the pixel intensity values causes contrast stretching, which increases the image visibility. After applying the contrast stretching, the transformed image can be expressed as follows:

$$I_{\mathrm{N}} = (I - \mathrm{Min})\frac{\mathrm{newMax} - \mathrm{newMin}}{\mathrm{Max} - \mathrm{Min}} + \mathrm{newMin}\,. \tag{7}$$

where (Min, Max) are the existing limits for intensity range of I and (newMin, newMax) are the new limits for the new intensity range of I_N. In this paper, the images are normalised to have intensity values within the range of 0–1.

Finally, the images are resized without stretching the core images to the dimensions (299, 299). The aspect ratio of the images are maintained by resizing them to a size lower than the size mentioned and the remaining region is padded with zeroes. It was done in order to uphold the importance of the ABCDE rule.

3.3 Model development

The architecture for the proposed model is shown in Fig. 3. The model is designed based on the two principal approaches, namely residual learning and separable convolution. Initially, 16 feature maps are extracted using separable convolution algorithm. We follow the aforementioned feature maps with four residual modules in the model. The functioning of these modules and their significance in the classification task is discussed in the next section.

3.3.1 Residual module: Residual modules in the model are used to incorporate the idea of deep residual learning. The model in this paper uses gradient-based learning technique, so it is prone to the problem of gradient degradation. Gradient degradation or vanishing gradient problem arises during backpropagation. The basic residual module implemented in our network is shown in Fig. 7.

During the training phase, the network has the weights of its units modified by propagating the gradient of the error or loss function throughout the network using backpropagation. However, during this phase, the error thus propagated becomes vanishingly small, and hence the weights are hardly modified. The error saturates far too soon with an extraordinarily high value, thus giving rise to an inaccurate classification result. In addition to this, an increase in the depth of the model gives rise to an issue of degradation, in which case the ACC is found to be decreased on stacking additional layers to an already deep model. To overcome these problems, He *et al.* [46] proposed the deep residual learning approach, where identity mappings or *shortcut connections* are used in order to propagate the error throughout the network. These mappings allow the error gradient to propagate without considerable degradation.

Unlike in case of plain network, where

$$y = f(x)\,. \tag{8}$$

residual networks use the identity mappings, in which case y can be redefined as follows:

$$y = f(x) + x\,. \tag{9}$$

The residual module used in the proposed model as shown in Fig. 7 consists of three convolution blocks and the shortcut connection consists of a single convolution block. We include the convolution block within the identity mapping in order to match the shape of feature maps and the number of kernels. This in fact degrades the gradient of the loss function. However, it minimises the loss far more as compared with a plain network with no identity mappings. The comparison of the loss curves of a plain network and the proposed model is shown in Fig. 8.

Forward propagation through the merge layer following the residual modules via the exponential linear unit (ELU) activation layers is defined as follows:

$$a^{\mathrm{l}} := \alpha(W^{\mathrm{l}-1,\,1}\,.\,a^{\mathrm{l}-1} + b^{\mathrm{l}} + W^{\mathrm{l}-3,\,1}\,.\,a^{\mathrm{l}-3})\,. \tag{10}$$

Considering the normalised convolution layer and the activation layer as a single layer, say l, we can assume a^{l} as the activation applied to the units in the lth layer. Here, α is the activation function applied to layer l which is described in (18), and finally, $\boldsymbol{W}^{\mathrm{l}-k,\,1}$ is the weight matrix between layers l - k and l. Now, during backpropagation, the change in the weight matrix between layers l - k and l can be defined as follows:

$$\Delta w^{\mathrm{l}-k,\,1} := -\eta\frac{\partial E^{\mathrm{l}}}{\partial w^{\mathrm{l}-k,\,1}}\,. \tag{11}$$

Here, $\Delta w^{\mathrm{l}-k,\,1}$ is the change caused in the weight matrix of layer l during backpropagation; η is the learning rate, and finally, E is the loss function.

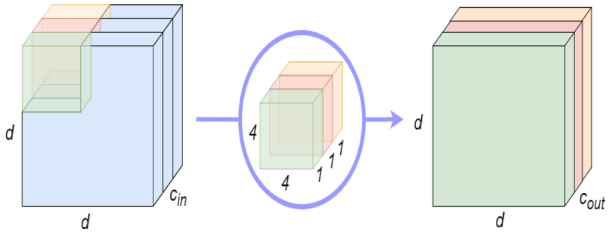


Fig. 9 *Working mechanism of a depthwise separable convolution kernel*

Table 4 Performance of the proposed model against various activation functions within each residual branch

Activation function	ACC, %	AUROC score	F_1 score
ELU	**99.50**	**0.9949**	**0.9948**
leaky ReLU	98.49	0.9839	0.9821
parametric ReLU	98.75	0.9865	0.9877

The bold values signify the activation function chosen in the residual blocks used to build the model.

It is clear that the change caused to the weight matrix of the merge layers in our architecture (see Fig. 3) is influenced more by the convolution layer in the identity branch rather than the last convolution layer on the residual branch. The loss gradient gets degraded three times as compared with the single degradation through the identity block. Thus, the identity mapping plays a significant role in the minimisation of error throughout the network.

3.3.2 Convolution layer: Instead of a traditional three-dimensional (3D) convolution kernel, a depthwise separable 3D convolution [47] is used in the proposed model. The convolution operation being separable in nature is faster and uses less parameter space. To justify this, we consider the traditional convolution algorithm, where the parameter cost is defined as

$$\rho = f \times f \times c_{\text{in}} \times c_{\text{out}}. \tag{12}$$

Moreover, the computation cost is given by

$$\kappa = f \times f \times c_{\text{in}} \times d_{\text{in}} \times d_{\text{in}} \times c_{\text{out}}.3.3.2 \tag{13}$$

In (12) and (13), it is considered that the input layer has *half* or *same* zero padding. The input layer and the kernel have square spatial dimensions and f represents the dimension of the kernel; c_{in} represents the number of input channels; c_{out} represents the number of output channels; and finally, d_{in} is the spatial dimension of the input layer. Considering the first convolution layer of the model, i.e. $f = 4$, $c_{\text{in}} = 6$, $d_{\text{in}} = 299$ and $c_{\text{out}} = 16$, we obtain $\rho = 1536$ and $\kappa \simeq 1.37 \times 10^8$. It should be noted that the filters in a traditional convolution operation are 3D in nature, and there is one such filter for each feature map; hence, the equation for ρ is expressed in the form of (12). Computation cost is calculated by considering the entire input (in other words $\rho \times d_{\text{in}} \times d_{\text{in}}$). However, if we consider depthwise separable convolution algorithm, then the parameter cost will be as follows:

$$\rho_{\text{s}} = f \times f \times c_{\text{in}} + c_{\text{in}} \times c_{\text{out}}. \tag{14}$$

Moreover, the computation cost can be expressed as

$$\kappa_{\text{s}} = f \times f \times c_{\text{in}} \times d_{\text{in}} \times d_{\text{in}} + d_{\text{in}} \times d_{\text{in}} \times c_{\text{in}} \times c_{\text{out}}. \tag{15}$$

From (14), we can calculate the parameter cost in case of depthwise convolution operation to be $\rho_{\text{s}} = 192$ and the computation cost to be $\kappa_{\text{s}} \simeq 1.2 \times 10^7$. Therefore, depthwise separable convolution is faster and more efficient than its traditional counterpart. This makes our network more suitable to be implemented in practise.

Fig. 9 provides a brief insight into the working of a depthwise separable convolution kernel. The convolution operation in case of a normal convolution kernel can be expressed as follows:

$$\begin{aligned} y[m, n] &= h[m, n] \times x[m, n]. \\ &= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} h[i, j]. x[m-i, n-j]. \end{aligned} \tag{16}$$

where $x[m, n]$ is the input signal and $h[m, n]$ is the impulse signal. However, in case of depthwise separable convolution, we first have a spatial convolution operation across each channel of the input signal followed by a pointwise convolution operation. Hence, (16) is converted to (17) which is defined as follows:

$$y[m, n] = \sum_{j=-\infty}^{\infty} h_2[j] \left(\sum_{i=-\infty}^{\infty} h_1[i]. x[m-i, n-j] \right). \tag{17}$$

It should be mentioned that all the convolution layers in the proposed model have the value of their units to be normalised. Also, the convolution layers have *half* or *same* zero padding, hence the input and output feature maps have same spatial dimensions.

Spatially separable convolution is avoided as not all kernels can be factored into smaller kernels. Hence, depthwise separable convolution is performed which works with such 'inseparable' kernels. Unlike spatially separable convolution operation, depthwise separable convolution separates the convolution operation itself into a spatial convolution operation followed by a pointwise or depthwise convolution operation instead of factoring a larger kernel.

3.3.3 Activation functions: Throughout the model, two types of activation functions are used, namely the rectified linear unit (ReLU) and the ELU. The convolution layers on the residual branches use ELU activation exclusively and the merge layers use the ReLU activation. The purpose of using ELU activation in the residual branches is to simply hasten the learning process. ELU activation tends to bring the mean of activation closer to zero and consequently fixes the variance of the activation, which in turn speeds up the learning phase of the model. This is essential in case of residual branches as they tend to maximise the loss of the network. This objective can also be fulfilled by using leaky ReLU or parametric ReLU activation. However, ELU obtained the best possible result, as shown in Table 4; therefore, it was chosen as the activation function for the convolution layers in the residual branches.

It should also be mentioned that ELU activated layers are self-normalised, unlike ReLU activated layers. Considering x as the input signal, the ELU activation function can be defined as follows:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(\mathrm{e}^x - 1), & \text{otherwise}(\alpha\text{being constant}) \end{cases} \tag{18}$$

In case of the merge layers, which are ReLU activated, error minimisation and learning rate are maintained by the identity mappings and hence ELU activation is not exclusively required. The ReLU activation is defined as follows:

$$f(x) = x^+ = \max(0, x). \tag{19}$$

3.3.4 Pooling layers: Throughout the network max pooling algorithm is used to downsize the weight matrices obtained by using the depthwise separable convolution algorithm. The reason for using this algorithm is to retain the most prominent features. Similar to regular convolution networks, we place this layer after the detector stage or the activation stage in our network. The pooling algorithm makes the network invariant to small translations of the input. We are interested in the features of the lesion and not on their specific locations, hence max pooling has been incorporated in our proposed model.

Table 5 Comparison of the performances of the model against various kernel sizes and number of residual blocks

Number of residual blocks	Kernel size	Validation ACC	Validation loss	AUROC score	F_1 score
three residual blocks	1×1	0.9633	0.1298	0.9635	0.9623
	2×2	0.9783	0.0876	0.9783	0.9776
	3×3	0.9800	0.0776	0.9800	0.9793
	4×4	0.9783	0.0708	0.9783	0.9776
	5×5	0.9800	0.0797	0.9800	0.9793
	6×6	0.9800	0.0858	0.9799	0.9792
	7×7	0.9566	0.1316	0.9564	0.9549
four residual blocks	1×1	0.9817	0.0856	0.9819	0.9812
	2×2	0.9816	0.0733	0.9817	0.9810
	3×3	0.9867	0.0686	0.9866	0.9862
	4×4	**0.9950**	**0.0704**	**0.9949**	**0.9948**
	5×5	0.9799	0.0756	0.9800	0.9793
	6×6	0.9917	0.0589	0.9915	0.9913
	7×7	0.9533	0.1083	0.9537	0.9523

The bold values signify the architecture of the model with the best performance.

Finally, to obtain a 1D feature vector from the 3D tensors, we apply global average pooling (GAP) [48]. GAP layer can result in a huge loss of features. However, we downsized the feature maps throughout the network to obtain a max pooling layer with feature maps f_k having dimensions (9×9), where $k \in \{1, 2, 3, \ldots, 96\}$. These feature maps get resized into a vector of dimension $(1 \times 1 \times 96)$. GAP layer makes the network invariant of minute changes in the input. Also, the GAP-CNN models are known to perform object localisation and this is verified by computing the class activation maps of the last convolution layer.

3.3.5 FC layers: The proposed model contains a feedforward network to approximate the feature maps obtained from the convolution layers. This feedforward network can be expressed in the form of $Y = f(X)$, where $f(X)$ can be expressed as follows:

$$f(X) = f_{\mathrm{SM}}(f_{\mathrm{D}}(f_{\mathrm{R}}(X))). \tag{20}$$

where the function f_{SM} is the output layer that uses softmax activation on the feature vector obtained from the previous layer, which, in this case is the ReLU activation layer f_{R} that activates the units of the preceding multi-layer perceptron (MLP) layer expressed as X. The softmax activation can be defined as

$$\mathrm{softmax}(\boldsymbol{Z})_i = \frac{\exp(\boldsymbol{Z}_i)}{\Sigma_j \exp(\boldsymbol{Z}_j)}. \tag{21}$$

where $\boldsymbol{Z}$ is a k-dimensional feature vector obtained from the previous MLP layer. Thus, considering the previous layer, which is an FC feedforward network, now we can express Z as follows:

$$Z = X \,.\, W + B. \tag{22}$$

where $\boldsymbol{X}$ represents the input matrix; the variable $\boldsymbol{W}$ holds the weight matrix or the layer's kernel; and the variable B is the bias vector of the layer. However, for our model, we activate the aforementioned function Z using ReLU activation before feeding it into the softmax layer. We also allow random dropping of 50% of the units obtained from the previous feedforward layer in order to regularise the model. Thus, we can reformulate (22) as follows:

$$Y = Z^{+} = \max(0,\, X \,.\, W + B). \tag{23}$$

For our model X represents an input vector from the activation layer f_{R} of dimension (1×512) (after the dropout using f_{D}) is obtained from the GAP layer of dimension (1×96). The difference $Y - Y^{*}$ is then optimised to obtain optimum performance from the model, where Y^{*} is the actual output and Y is the predicted output.

3.4 Fine-tuning

We tested the model for various kernel sizes and its summary is provided in Table 5. Initially, a base model was defined, which contained a single residual module, that is, shown in Fig. 7. The optimal number of kernels and the optimal kernel size were then obtained for this module. The parameters of this module were optimised by monitoring the standard performance measures along with GRAD-CAM for each layer which visualises the major regions under attention. Additionally, saliency maps were used for visualisation of the prediction layer that verifies the correctness of the feature identification ability of the model. Finally, the classification ACC of the model was found and the model was considered to be completely tuned if the ACC obtained was optimal; otherwise, the model was retrained using the next higher kernel size. If the model showed stagnation (which is observed by monitoring the loss curve), we retrain the model after adding a new residual module to it. The optimisation phase was then performed again. This process was repeated until we obtained the maximum possible score. Fig. 10 shows a summary of the training performances of the various kernel sizes for the model with three and four residual blocks.

The above model was optimised using the image matrix from the RGB colour space. This optimised model was then retrained on the modified version of the image matrix, which uses six image channels. An optimal kernel dimension of (4×4) was obtained for every layer and four residual modules were found to be best suited for the task.

Additionally, ELU activation was found to be better than parametric ReLU and leaky ReLU for the activated layers within the residual module. A summary of the model's performance using the various activation functions in the residual module is shown in Table 4.

4 Results and discussion

In this section, we discuss the performance of the model across various parameters and also provide a comparative study, which showcases the superiority of the proposed methodology over the existing melanoma diagnosis algorithms.

4.1 Experimental setup

We trained the model on a system having 2.8 GHz processor with 16 GB of memory and a GPU support of 4 GB. The model was built using a Keras framework on a Python 3 kernel and trained using RMSprop optimiser. We use *early stopping* technique to halt the continued execution as well as to save the model parameters for which it has the best performance, i.e. highest ACC and lowest cross-entropy loss. To avoid the further stagnation in learning phase, we reduce the learning rate of the model whenever it fails to achieve an ACC higher than those obtained in the previous two

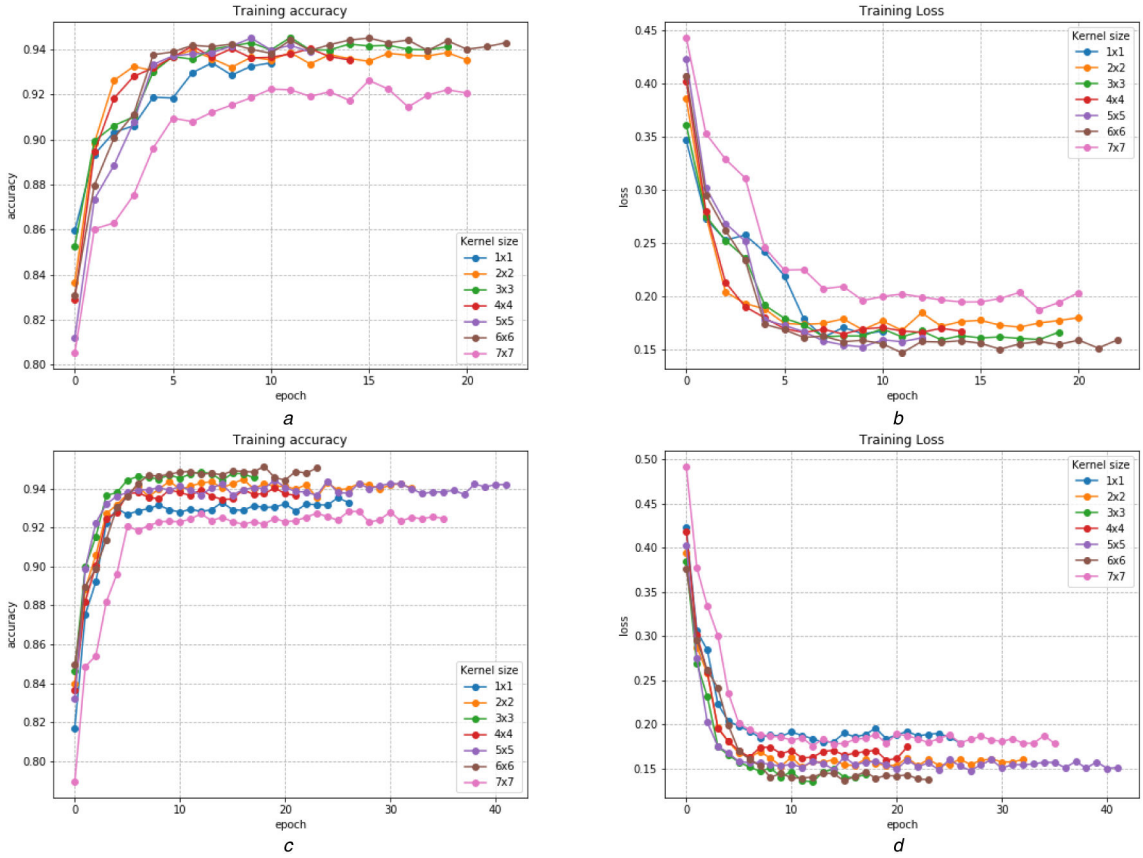


Fig. 10 *Summary of the training performances of the various kernel sizes for the model with three and four residual blocks*
(a) Training ACC curve for various kernel sizes in the model with three residual blocks,
(b) Training loss curve for various kernel sizes in the model with three residual blocks,
(c) Training ACC curve for various kernel sizes in the model with four residual blocks,
(d) Training loss curve for various kernel sizes in the model with four residual blocks

epochs. A total of 4000 images were selected, out of which 600 images were used for validation and 3400 images were used for training the model.

4.2 Results

The performance of the model is illustrated in Fig. 11, which shows the plots for training ACC, training loss, ROC curve and the precision (P)–recall curve.

The confusion matrix of the model is also shown in Fig. 12, from which we can derive the various performance metrics of the model.

Table 6 summarises the performance of the proposed model using best parameters. It becomes very clear from the above table that the model is a state-of-the-art model as per its performance. However, in this paper, we validate the ability of the model to localise the lesions using GRAD-CAM visualisation along with saliency maps to visualise the features, which the model consider as essential.

As already mentioned, if the model fails to properly identify the lesions, it needs to be fine-tuned because in that case it most probably is overfitting the dataset. Fig. 13 shows the processed images and their respective GRAD-CAM and saliency map visualizations(as obtained by the proposed model) have been shown in Fig. 14 and Fig. 15.

We also show the performance of the model on various datasets. It should be mentioned that the images in these datasets differ from the ISIC dataset in terms of contrast, brightness and other features, which are probably a result of the difference in the hardware used to develop those images. The proposed model, however, showed impressive results on these datasets, which are tabulated in Table 7.

4.3 Comparative analysis

It is mentioned that the model was tested against various square kernel sizes and the summary of the performances for the following kernel sizes are cited in Table 5. Kernel sizes in the range [1, 7] were tested for the model containing [1, 4] residual blocks. In this table, the results against all kernel sizes with the aforementioned range with three and four residual blocks are shown. Results for the model containing a single or two residual blocks are not shown as they were not even close to the optimal score. Additionally, increasing the depth of the model resulted in a decrease in the ACC and AUROC score.

As mentioned in Section 3.3.1, an increase in the depth of the model makes it prone to the problem of vanishing gradient. Moreover, surely, from the plot of the loss curves of our model with its plain counterpart, we can clearly see that the proposed model attains a saturation at a much lower value of loss as compared with the plain network.

The identity mapping in our model makes the error propagation more efficient, and hence we obtain a lower loss and higher ACC by preventing early saturation in training of the model. The plain network contains all the layers as in the proposed model only excluding the merge layers. The proposed model attains saturation at a loss of 0.0584 with an ACC of 99.50% while its plain counterpart saturates at a loss of 0.4324 with an ACC of 82.13%. It is clear that due to the inefficient loss propagation, the model gets stagnated during the training phase. The difference in the loss curves of the proposed model and the plain model is shown in Fig. 8.

We can see from Table 5 that the model performs best with four residual modules and a kernel size of dimensions (4×4). To achieve this score, we use the architecture, which is described in Fig. 3. However, the model was tested with activation functions other than ELU to activate the weight layers within the residual branch. The performance of the model for these activation

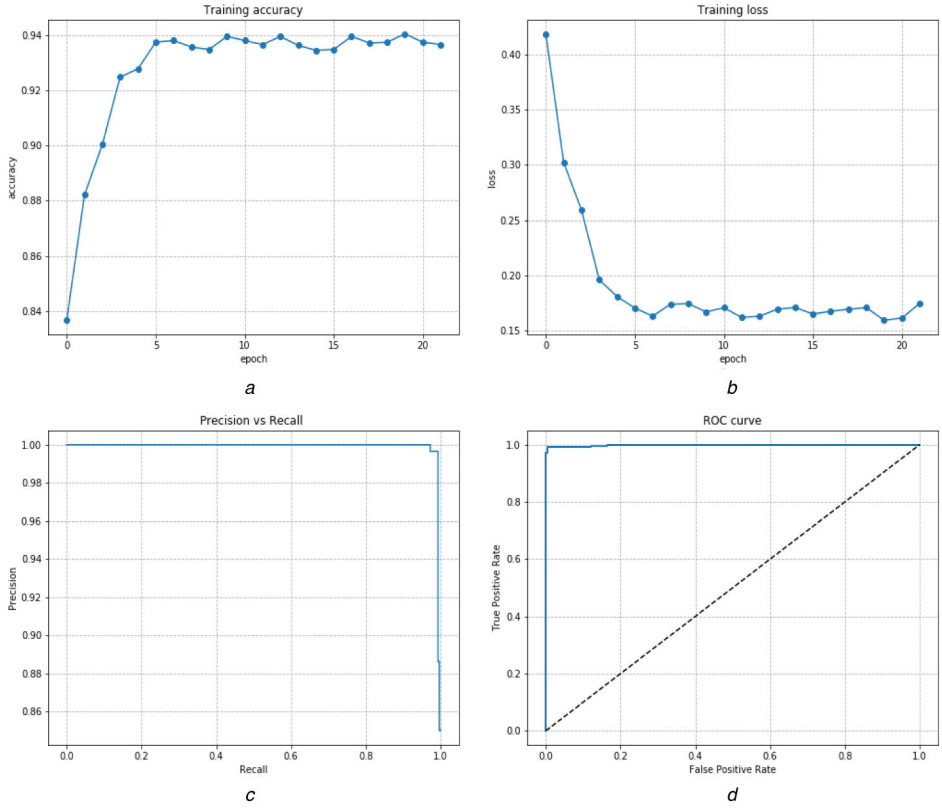


Fig. 11 *Plots for training ACC, training loss, ROC curve and the precision (P)–recall curve*
(a) Training ACC curve of the proposed model,
(b) Training loss curve of the proposed model,
(c) P–recall curve of the proposed model,
(d) ROC curve of the proposed model

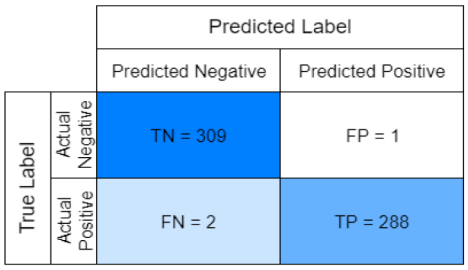


Fig. 12 *Confusion matrix of the proposed model*

Table 6 Various performance metrics and their corresponding values for the proposed model

Performance metrics	Expression of the metrics	Value
ACC	$\frac{TP+TN}{TP+TN+FP+FN}$	0.9950
P	$\frac{TP}{TP+FP}$	0.9965
recall (SE)	$\frac{TP}{TP+FN}$	0.9931
SP	$\frac{TN}{TN+FP}$	1.0000
F_1 score	$\frac{2\times TP}{2\times TP+FP+FN}$	0.9948
AUROC score	—	0.9949

functions is shown in Table 4. In Fig. 10, we see the comparison of the learning curves of the various kernel sizes for the proposed model containing three and four residual blocks.

We also cited a comparison of performances for our model against the existing algorithms in Table 8. It is evident from this table that the proposed model is superior to the existing algorithms in terms of both ACC of classification and the area under curve or AUROC score. The proposed model clearly has a SP and SE trade-off better than the existing algorithms. Hence, it can be considered as close to an ideal model in this context.

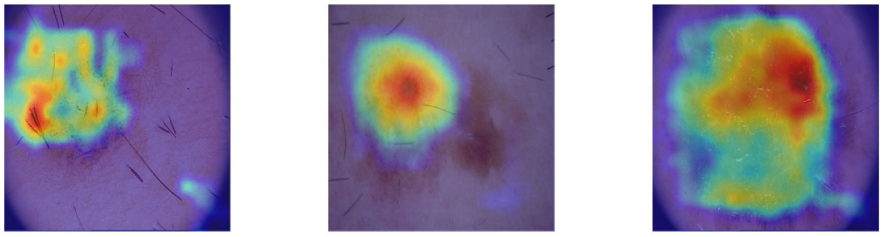
5 Conclusion and future scope

In this paper, we develop a parameter cost-effective and time-efficient state-of-the-art model for melanoma diagnosis from dermoscopic images using the concepts of residual learning and separable convolution network. The proposed model is a deep depthwise separable residual convolution model which is significantly more accurate for the purpose of diagnosing the disease as compared with contemporary methods. The proposed methodology introduces a concept to increase the performance of the model by building a multiple channel image matrix with six channels selected across different colour spaces, which enhances the visibility of the lesion for the network. Optimal blurring and enhancement of the images were performed to reduce the influence of unrelated regions from the images and to enhance the region enclosing the lesion. The model presented here is dynamic in nature and achieves high-performance scores when validated against several benchmark datasets. We provide the comparative analysis of the performance of our model with contemporary methods in Table 8.

Melanoma diagnosis performed by this model is limited to only dermoscopic images. However, it is not always a feasible option for individuals to clinically obtain the dermoscopic images of the suspicious moles. Hence, we plan to further explore the study on diagnosing melanoma from non-dermoscopic images. In near future, we intend to implement this model for diagnosing the remaining three major classes of skin cancers, i.e. basal cell carcinoma, SCC and actinic keratoses, thereby making this a role model for diagnosis of skin cancer. Also, the preprocessing applied to the model makes it demand more processing power from the CPU. Hence, we must consider improvement of the network itself, so that it can perform the classification without the help of any preprocessing techniques that have been applied to the images.

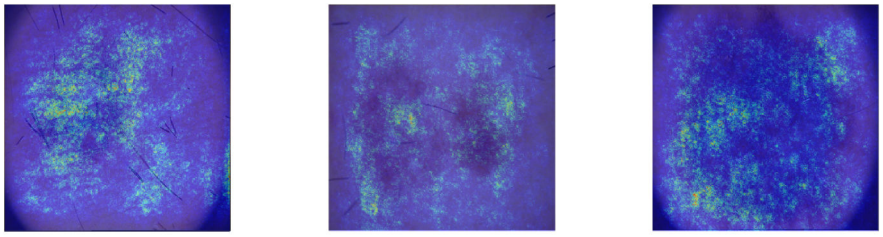
a *b* *c*

Fig. 13 *Processed images*
(a)–(c) Images from ISIC dataset after preprocessing phase



a *b* *c*

Fig. 14 *GRAD-CAM of the last convolution layer*
(a)–(c) GRAD-CAMs of the last convolution layer for images in Fig. 13



a *b* *c*

Fig. 15 *Saliency map visualisation of the penultimate MLP or FC layer of the model*
(a)–(c) Saliency maps of the penultimate FC layer for images in Fig. 13

Table 7 Performance of the proposed model on some well known datasets

Dataset	ACC, %	P	Recall	AUROC score
ISIC	99.50	0.9965	0.9931	0.9949
PH2	96.77	0.9090	1.0000	0.9762
MED-NODE	95.23	1.0000	0.9233	0.9444
DermIS	94.44	0.9166	1.0000	0.9286

Table 8 Performance comparison of the proposed methodology with the existing methodologies

Authors	ACC, %	AUROC score	SP, %	SE, %
Hagerty *et al.* [42]	94.00	0.9000	—	—
Albahar *et al.* [41]	97.49	0.9800	93.60	94.30
Shahin *et al.* [35]	89.90	—	—	79.60
Kaymak *et al.* [36]	84.00	—	63.50	83.90
Yu *et al.* [37]	86.81	0.8520	—	—
Guo *et al.* [38]	85.22	0.8100	—	—
Li and Shen [33]	91.20	0.9120	96.10	49.00
Jaisakthi *et al.* [39]	91.93	—	98.48	80.48
Arasi *et al.* [24]	86.10	—	92.40	57.30
Arasi *et al.* [24]	98.80	—	100.0	97.80
Codella *et al.* [26]	85.50	0.8040	93.10	54.70
Yuan and Lo [28]	93.40	—	97.50	82.50
Jafari *et al.* [23]	98.70	—	99.00	95.20
Codella *et al.* [20]	93.10	—	94.90	92.80
Cavalcanti *et al.* [14]	99.34	—	97.78	100.0
proposed methodology	**99.50**	**0.9949**	**100.0**	**99.31**

The bold values signify the performance of the proposed model.

6 References

[1] Aim At Melanoma Foundation: 'Melanoma stats, facts, and figures'. Available at https://www.aimatmelanoma.org/about-melanoma/melanoma-stats-facts-and-figures, accessed 2018

[2] Skin Cancer Foundation: 'Melanoma'. Available at https://www.skincancer.org/skin-cancer-information/melanoma, accessed 2018

[3] Cancer Research UK: 'Melanoma skin cancer statistics. Available at https://www.cancerresearchuk.org/healthprofessional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer, accessed 2018

[4] Mishra, R., Patel, H., Yuan, L., *et al.*: 'Role of reactive oxygen species and targeted therapy in metastatic melanoma', *Cancer Res. Front.*, 2018, **4**, pp. 101–130

[5] ISIC Archive: 'Melanoma project'. Available at https://www.isic-archive.com, accessed 2018

[6] ADDI Project: 'PH2 database'. Available at http://www.fc.up.pt/addi/, accessed 2018

[7] DermIS: 'DermIS dataset'. Available at https://www.dermis.net, accessed 2018

[8] MED-NODE: 'Med-node dataset'. Available at http://www.cs.rug.nl/~imaging/databases/melanoma_naevi/, accessed 2018

[9] Tommasi, T., La.Torre, E., Caputo, B.: 'Melanoma recognition using representative and discriminative kernel classifiers', *Comput. Vis. Approaches Med. Image Anal.*, 2006, **4241**, pp. 1–12

[10] Mocellin, S., Ambrosi, A., Montesco, M.C., *et al.*: 'Support vector machine learning model for the prediction of sentinel node status in patients with cutaneous melanoma', *Ann. Surg. Oncol.*, 2006, **13**, (8), pp. 1113–1122

[11] Shimizu, K., Iyatomi, H., Norton, K., *et al.*: 'Extension of automated melanoma screening for non-melanocytic skin lesions'. 2012 19th Int. Conf. Mechatronics and Machine Vision in Practice (M2VIP), Auckland, New Zealand, 2012, pp. 16–19

[12] Duarte, M.F., Matthews, T.E., Warren, W.S., *et al.*: 'Melanoma classification from hidden markov tree features'. 2012 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012, pp. 685–688

[13] Ikuma, Y., Lyatomi, H.: 'Production of the grounds for melanoma classification using adaptive fuzzy inference neural network'. 2013 IEEE Int. Conf. Systems, Man, and Cybernetics, Manchester, UK, 2013, pp. 2570–2575

[14] Cavalcanti, P.G., Scharcanski, J., Baranoski, G.V.G.: 'A two-stage approach for discriminating melanocytic skin lesions using standard cameras', *Expert Syst. Appl.*, 2013, **40**, (10), pp. 4054–4064

[15] Mhaske, H.R., Phalke, D.A.: 'Melanoma skin cancer detection and classification based on supervised and unsupervised learning'. 2013 Int. Conf. Circuits, Controls and Communications (CCUBE), Bengaluru, India, 2013, pp. 1–5

[16] Ballerini, L., Fisher, R.B., Aldridge, B., *et al.*: 'A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions', *Color Med. Image Anal.*, 2013, **6**, pp. 63–86

[17] Barata, C., Ruela, M., Francisco, M., *et al.*: 'Two systems for the detection of melanomas in dermoscopy images using texture and color features', *IEEE Syst. J.*, 2014, **8**, (3), pp. 965–979

[18] Abuzaghleh, O., Barkana, B.D., Faezipour, M.: 'Automated skin lesion analysis based on color and shape geometry feature set for melanoma early detection and prevention'. IEEE Long Island Systems, Applications and Technology (LISAT) Conf. 2014, Farmingdale, NY, USA, 2014, pp. 1–6

[19] Valavanis, I., Maglogiannis, I., Chatziioannou, A.A.: 'Exploring robust diagnostic signatures for cutaneous melanoma utilizing genetic and imaging data', *IEEE. J. Biomed. Health Inf.*, 2015, **19**, (1), pp. 190–198

[20] Codella, N., Cai, J., Abedini, M., *et al.*: 'Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images', *Mach. Learn. Med. Imaging*, 2015, **9352**, pp. 118–126

[21] Mishra, N.K., Celebi, M.E.: 'An overview of melanoma detection in dermoscopy images using image processing and machine learning', ArXiv e-Print, 2016, pp. 1–15, arXiv:1601.07843

[22] Li, Y., Esteva, A., Kuprel, B., *et al.*: 'Skin cancer detection and tracking using data synthesis and deep learning', ArXiv e-Print, 2016, pp. 1–4, arXiv:1612.01074

[23] Jafari, M.H., Nasr Esfahani, E., Karimi, N., *et al.*: 'Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma', *Int. J. Comput. Assist. Radiol. Surg.*, 2017, **12**, (6), pp. 1021–1030

[24] Arasi, M.A., El Horbaty, E.M., Salem, A.M., *et al.*: 'Computational intelligence approaches for malignant melanoma detection and diagnosis'. 2017 Eighth Int. Conf. Information Technology (ICIT), Amman, Jordan, 2017, pp. 55–61

[25] Takruri, M., Abubakar, A.: 'Bayesian decision fusion for enhancing melanoma recognition accuracy'. 2017 Int. Conf. Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, United Arab Emirates, 2017, pp. 1–4

[26] Codella, N.C.F., Nguyen, Q.B., Pankanti, S., *et al.*: 'Deep learning ensembles for melanoma recognition in dermoscopy images', *IBM J. Res. Dev.*, 2017, **61**, (4/5), pp. 5:1–5:15

[27] Munia, T.T.K., Alam, M.N., Neubert, J., *et al.*: 'Automatic diagnosis of melanoma using linear and nonlinear features from digital image'. 2017 39th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, South Korea, 2017, pp. 4281–4284

[28] Yuan, Y., Lo, Y.: 'Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks', *IEEE. J. Biomed. Health Inf.*, 2018, **23**, (2), pp. 519–526

[29] Lopez, A.R., i-Nieto, X.G., Burdick, J., *et al.*: 'Skin lesion classification from dermoscopic images using deep learning techniques'. 2017 13th IASTED Int. Conf. Biomedical Engineering (BioMed), 2017, pp. 49–54

[30] Harangi, B.: 'Skin lesion detection based on an ensemble of deep convolutional neural networks', *Comput. Vis. Pattern Recognit.*, 2017, pp. 1–4, arXiv:1705.03360

[31] Adjed, F., Safdar Gardezi, S.J., Ababsa, F., *et al.*: 'Fusion of structural and textural features for melanoma recognition', *IET Comput. Vis.*, 2018, **12**, (2), pp. 185–195

[32] Codella, N.C.F., Gutman, D., Celebi, M.E., *et al.*: 'Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)'. Int. Symp. Biomedical Imaging, Washington, D.C., USA, 2018, pp. 168–172

[33] Li, Y., Shen, L.: 'Skin lesion analysis towards melanoma detection using deep learning network', *Sensors (Basel)*, 2018, **18**, pp. 1–16

[34] Ma, Z., Yin, S.: 'Deep attention network for melanoma detection improved by color constancy'. 2018 Ninth Int. Conf. Information Technology in Medicine and Education (ITME), Hangzhou, China, 2018, pp. 123–127

[35] Shahin, A.H., Kamal, A., Elattar, M.A.: 'Deep ensemble learning for skin lesion classification from dermoscopic images'. 2018 Ninth Cairo Int. Biomedical Engineering Conf. (CIBEC), Cairo, Egypt, 2018, pp. 150–153

[36] Kaymak, S., Esmaili, P., Serener, A.: 'Deep learning for two-step classification of malignant pigmented skin lesions', 2018 14th Symp. Neural Networks and Applications (NEUREL), Belgrade, Serbia, 2018, pp. 1–6

[37] Yu, Z., Jiang, X., Zhou, F., *et al.*: 'Melanoma recognition in dermoscopy images via aggregated deep convolutional features', *IEEE Trans. Biomed. Eng.*, 2019, **66**, (4), pp. 1006–1016

[38] Guo, Y., Ashour, A.S., Si, L., *et al.*: 'Multiple convolutional neural network for skin dermoscopic image classification'. 2018 IEEE Int. Symp. Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, USA, 2018, pp. 365–369

[39] Jaisakthi, S.M., Mirunalini, P., Aravindan, C.: 'Automated skin lesion segmentation of dermoscopic images using GrabCut and *k*-means algorithms', *IET Comput. Vis.*, 2018, **12**, (8), pp. 1088–1095

[40] Sultana, N.N., Mandal, B., Puhan, N.B.: 'Deep residual network with regularised fisher framework for detection of melanoma', *IET Comput. Vis.*, 2018, **12**, (8), pp. 1096–1104

[41] Albahar, M.A.: 'Skin lesion classification using convolutional neural network with novel regularizer', *IEEE Access*, 2019, **7**, pp. 38306–38313

[42] Hagerty, J., Stanley, J., Almubarak, H., *et al.*: 'Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images', *IEEE. J. Biomed. Health Inf.*, 2019, **23**, (4), pp. 1385–1391

[43] Mahbod, A., Schaefer, G., Wang, C., *et al.*: 'Skin lesion classification using hybrid deep neural networks'. ICASSP 2019 – 2019 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 1229–1233

[44] Zhang, J., Xie, Y., Xia, Y., *et al.*: 'Attention residual learning for skin lesion classification', *IEEE Trans. Med. Imaging*, 2019, DOI: 10.1109/TMI.2019.2893944

[45] Lidong, H., Wei, Z., Jun, W., *et al.*: 'Combination of contrast limited adaptive histogram equilisation and discrete wavelet transform for image enhancement', *IET Image Process.*, 2015, **9**, (10), pp. 908–915

[46] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778

[47] Chollet, F.: 'Xception: deep learning with depthwise separable convolutions'. 2017 IEEE Conf. Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 1800–1807

[48] Lin, M., Chen, Q., Yan, S.: 'Network in network', *CoRR*, 2013, doi: arXiv:1312.4400