Multimedia Tools and Applications (2020) 79:28477–28498
https://doi.org/10.1007/s11042-020-09388-2

A multi-class skin Cancer classification using deep convolutional neural networks

Saket S. Chaturvedi[1] · Jitendra V. Tembhurne[2] · Tausif Diwan[2]

Received: 10 August 2019 / Revised: 22 June 2020 / Accepted: 21 July 2020 /
Published online: 4 August 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
Skin Cancer accounts for one-third of all diagnosed cancers worldwide. The prevalence of skin cancers have been rising over the past decades. In recent years, use of dermoscopy has enhanced the diagnostic capability of skin cancer. The accurate diagnosis of skin cancer is challenging for dermatologists as multiple skin cancer types may appear similar in appearance. The dermatologists have an average accuracy of 62% to 80% in skin cancer diagnosis. The research community has been made significant progress in developing automated tools to assist dermatologists in decision making. In this work, we propose an automated computer-aided diagnosis system for multi-class skin (MCS) cancer classification with an exceptionally high accuracy. The proposed method outperformed both expert dermatologists and contemporary deep learning methods for MCS cancer classification. We performed fine-tuning over seven classes of HAM10000 dataset and conducted a comparative study to analyse the performance of five pre-trained convolutional neural networks (CNNs) and four ensemble models. The maximum accuracy of 93.20% for individual model amongst the set of models whereas maximum accuracy of 92.83% for ensemble model is reported in this paper. We propose use of ResNeXt101 for the MCS cancer classification owing to its optimized architecture and ability to gain higher accuracy.

Keywords Skin Cancer · Dermoscopy · Classification · Deep convolutional neural network

✉ Jitendra V. Tembhurne
jitendra.tembhurne@cse.iiitn.ac.in

Saket S. Chaturvedi
saketschaturvedi@gmail.com

Tausif Diwan
tausif.diwan@cse.iiitn.ac.in

Extended author information available on the last page of the article

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

1 Introduction

The epidermis is the superficial layer of skin mainly consists of three cells: Squamous cells, Basal cells, and Melanocytes, as shown in Fig. 1. The outermost cells are Squamous and lowermost layer cells are Basal cells of the epidermis. Melanocytes protect deeper layers of skin from exposure of sun by producing melanin, a brown pigment substance [10]. When these cells experience excessive ultraviolet light exposure, the DNA mutations induced affects the growth of skin cells and eventually shapes in skin cancer [38, 57]. Squamous Cell Carcinoma, Basal Cell Carcinoma, and Melanoma are the substantial categories of skin cancer usually associated with squamous cells, basal cells, and melanocytes, respectively.

The World Health Organization estimates skin cancer as one-third of all the diagnosed cancers cases globally [76]. Skin Cancer is a global public health issue which causes approximately 5.4 million newly identified skin cancer incidences in the United States each year [63]. However, melanomas are responsible for approximately three-fourth of all skin cancer-related deaths, which count over 10,000 deaths each year in the United States alone. In Europe, over 1,00,000 new diagnosed melanoma cases are reported annually [3]. Australia accounts for nearly 15,229 new cases of melanoma annually [8]. Moreover, past decades had recorded a raise in the incidence rates of skin cancer, which can be observed in the United Kingdom, where raise in melanoma by 119% since the 1990s, or 250% in the United States (starting with 27,600 cases in 1990 to 96,480 in 2019) [67, 68]. This trend is explained not only by the depletion of the ozone layer, but also by the use of solarium and tanning beds [81].

Skin cancer is traditionally diagnosed by examine physical and biopsy. Although, biopsy is one of the simplest methods to diagnose skin cancer, the process is arduous and unreliable. In recent years, the most popular non-invasive instruments that can assist dermatologists in the skin cancer diagnosis are macroscopic and dermoscopic images [25]. Macroscopic images usually have lower quality and resolution problem as they are attained using cameras or mobile phones [55]. Dermoscopy images are high-resolution skin images, derived from the visualization of deeper skin structures to enhance the diagnostic capability of skin cancer [73].

The accurate diagnosis of skin cancer is challenging for dermatologists even with dermoscopy images as multiple skin cancer types may appear similar in initial appearance. Moreover, even the expert dermatologists are limited to their studies and experiences, since

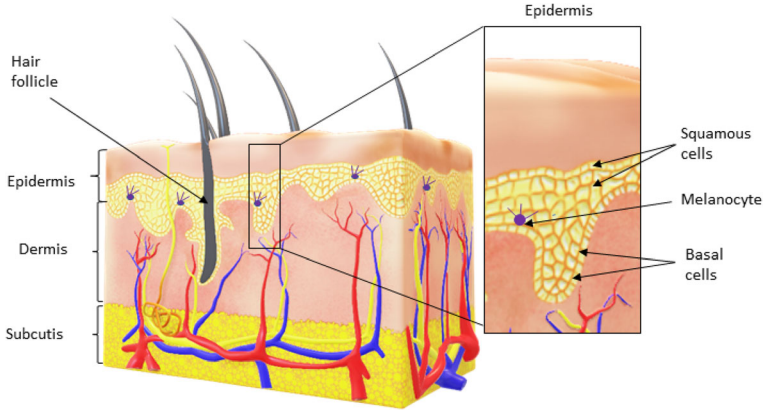

Fig. 1 Skin layers - epidermis, dermis, and subcutis (Epidermis sub-layered: squamous cells, basal cells, and melanocytes)

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

they are only exposed to a subset of all the possible appearances of skin cancer during their lifetime. The dermatologists have an average accuracy of 62% to 80% in skin cancer diagnosis [37, 50]. The reports on the diagnostic accuracy of clinical dermatologist have claimed 62% accuracy for dermatologists with experience of 3 to 5 years. However, dermatologists with experience greater than ten years are able to achieve 80% accuracy. The performance further dropped for less experienced dermatologists [50]. Also, dermoscopy in the hands of inexperienced dermatologists may reduce the accuracy to identify skin cancer [5, 37, 59].

The major drawback of dermoscopy is the requirement of extensive training. The research community has been made significant progress in developing computer-aided diagnosis tools to overcome the issues faced by the dermatologists [39, 55, 58]. The computer-aided diagnosis gets better with more and more data surfaces. Retraining the system with new data is trivial, and the underlying model can also be extended to integrate a plethora of other medical information into its prediction pipeline. The magnificent performance of deep convolutional neural networks (DCNNs) to classify images has made to utilize them for classifying images in medical domain such as skin cancer classification [42].

The researches [27, 31, 48, 51, 80] fail to extend their study for multiple classes in skin cancer classification. Additionally, previous investigations are limited by exploiting limited pre-trained networks [29, 40, 66, 69] or using particular layers of a network for the classification. In this paper, an automated computer-aided diagnostic system for the MCS cancer classification with exceptionally high accuracy is proposed. The proposed method outperformed both expert dermatologists and previously proposed deep learning methods for MCS skin cancer classification. We have conducted a comparative study to analyse the performance of five pre-trained convolutional neural networks and four ensemble models to determine the best method for skin cancer classification on HAM10000 dataset. We performed extensive research in determining the best set-up of hyper-parameters for five models pre-trained on ImageNet [17] namely Xception [15], InceptionV3 [71], InceptionResNetV2 [70], NASNetLarge [83], and ResNetXt101 [77] and their ensembles InceptionV3 + Xception, InceptionResNetV2 + Xception, InceptionResNetV2 + ResNetXt101, and InceptionResNetV2 + ResNetXt101 + Xception. These models are fine-tuned further on HAM10000 dataset [72] using Transfer Learning [56] to learn domain specific features of skin cancers. We preferred not to perform extensive pre-processing in this work. Also, we did not consider hand-crafted feature engineering or lesion segmentation to make the work more generic and reliable.

The paper is structured as follows. The literature review is covers in Section 2. Section 3 discusses the methods used in this research, including dataset, pre-processing, classification models, fine-tuning, feature extraction, and performance matrix. The results and discussions are highlighted in Section 4 and conclusion is presented in Section 5.

2 Related works

In early 1990, the computer-aided diagnosis systems were introduced to overcome the challenges faced by the dermatologists in skin cancer classification [75]. The initial efforts using dermoscopy images were restricted to the classification of benign and melanoma skin cancer lesions [61]. Since then, numerous methods have been published to address this challenging task. Several studies [51], [1, 33, 61] follow the commonly used manual evaluation methods based on the ABCD rules proposed by Nachbar et al. [53]. Moreover, traditional

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

machine learning classifiers such as Super Vector Machines [12], Naive Bayes Classifier [43], K-Nearest Neighbours [4], Logistic Regression [7], Decision Trees [11], and Artificial Neural Networks [30] were also untied for skin cancer classification in a search of more accurate and reliable method. Due to high intra-class and low inter-class variations in melanoma, handcrafted feature based diagnostic performance was found to be unsatisfactory [79].

Convolutional neural networks brought a key breakthrough to existing problems and quickly became the preferred choice for skin cancer classification [16]. The CNNs not only provided high classification accuracy but also alleviate the machine learning expert's burden of "*feature engineering*" by automatically discovering high-level abstractions from the datasets [41]. As CNNs needs a large dataset to get familiar with the problem [47], the current literature [18, 35, 54] mostly employs Transfer Learning to solve large dataset problem, a technique where a model trained for a given source task is partially "*recycled*" for a new target task.

A nearly melanoma detection classifying dermoscopy skin cancer images as malignant or benign was particularly focused in [42]. The proposed solution uses transfer learning along with the VGGNet convolutional neural network and achieved an accuracy of 81.3%, precision of 79.74% and recall as 78.66% evaluated on ISIC archive dataset. However, this method was restricted to binary classification of skin cancer. Harangi et al. [27] analysed, how the ensemble of deep CNNs can be proposed to enhance the accuracy of individual models for the skin cancer classification among three classes. The accuracy of 84.2%, 84.8%, 82.8%, and 81.3% for GoogleNet, AlexNet, ResNet, and VGGNet models are achieved respectively. Further, the best accuracy of 83.8% was achieved with the ensemble of GoogleNet, AlexNet, and VGGNet models. Moreover, recall for their individual models obtained is 59.2%, 51.8%, 52.0%, and 43.4%.

Kawahara et al. [34] demonstrated a linear classifier with a feature extracted from a CNN pre-trained on 1300 natural images dataset, which can distinguish up to ten skin lesions with a higher accuracy. The proposed method neither requires any lesion segmentations nor any complex preprocessing. This approach achieved an accuracy of 85.8% and 81.9% over 5-classes and 10-classes respectively. However, the number of images utilized for the training in the work was insufficient to extract useful features from the dataset. The authors of [35] proposed a novel CNN architecture composed of multiple tracts for skin lesion classification. They converted a CNN, pre-trained on a single resolution, to work for multi-resolution input. Moreover, an entire network was fine-tuned over a public lesion dataset to achieve an accuracy of 79.15% for ten classes.

Esteva et al. [18] utilized InceptionV3 architecture pre-trained on ImageNet [17] for fine-tuning on the dataset of 1,29,450 clinical images, including 3374 dermoscopic images. The authors showed that deep neural network-based method was able to outperform the clinical experts regarding the classification accuracy of the dermoscopy images with a large dataset.

Nyiri and Kiss [54] investigated multiple novel techniques of ensembling deep neural networks with different hyper-parameters and differently pre-processed data for skin lesion classification. The application of ensembling can be surprisingly useful not only for combining different machine learning models but also for combining different hyper-parameter choices for these models. An accuracy of 90.1% is achieved for Xception model evaluated on the seven classes of ISIC2017 and ISIC2018 datasets. Whereas, an accuracy of 80.1% for seven classes using VGG16 model is measured in [46]. A deep neural network-based framework [65] that follows an ensemble approach by combining ResNet-50 and Inception V3 architectures is presented to classify the seven types of skin cancers; an accuracy of 89.9% is reported.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Recently, an efficient seven-way automated MCS cancer classification system [13] is proposed. A pre-trained MobileNet model is utilized to train seven classes of HAM10000 dataset by transfer learning. A categorical accuracy of 83.1% and the precision, recall, and F1-score of 89%, 83%, and 83% is reported respectively. Further, Milton et al. [49] performed extensive study on different deep learning-based methods to skin cancer among seven classes. The experimentation was performed on various neural networks such as PNASNet-5-Large, InceptionResNetV2, SENet154, InceptionV4 on ISIC-2018 challenge dataset. The best accuracy of 76% is achieved for PNASNet-5-Large model. In addition, suggestion for further improvement and optimization of the proposed methods with larger training dataset and carefully selection of hyper-parameters were presented.

In [78], CNN-based features were considered; however, a pre-trained neural network model was trained with only 900 images, which seems insufficient to perform efficient training of a deep learning-based method. They achieved an accuracy of 85.5% with DRN-50 method, 82.6% with VGG-16 method, and 84.7% accuracy with GoogleNet method. Whereas, several studies [28, 54] proposed an ensemble method to achieve higher accuracy for skin cancer classification.

Previous work in dermoscopic computer-aided classification not only lacks in generality capabilities [9, 47, 64], but also failed to achieve higher accuracy for MCS skin cancer classification [18, 28, 35, 46], [6, 44, 45, 52]. Unfortunately, a major part of the earlier studies does not employ large dataset, which is rudimentary for the good performance of deep learning models. In this paper, the proposed method achieves an exceptionally high accuracy for MCS cancer classification using highly accurate and efficient pre-trained models trained on a large HAM10000 datasets over seven classes of skin cancer.

3 Materials and methods

We have proposed the generalized architecture for the multi-class classification of skin cancer, as represented in Fig. 2. Initially, the preprocessing is conducted on the dermoscopic skin cancer images to reconcile an image with the input dimension of the architectures used in this work. The processed images are then fed to the architecture for features extraction and fine-tuning. Finally, the output image is constructed by combining all features and classified among the seven classes of skin cancer, i.e. Melanocytic Nevi, Melanoma, Benign Keratosis, Actinic Keratosis, Vascular Lesions, Dermatofibroma, and Basal Cell Carcinoma. This method is explored for five different architectures such as InceptionV3, ResNeXt101, InceptionResNetV2, Xception, NASNetLarge and their four ensembles i.e. InceptionV3 + Xception, InceptionResNetV2 + Xception, InceptionResNetV2 + ResNetXt101 and InceptionResNetV2 + ResNetXt101 + Xception. The architecture of pre-trained model is shown in Fig. 4.

3.1 Dataset

This research focuses on the dermoscopy images of skin cancer owing to the high impact of dermoscopy over the world [6, 44, 45]. We have utilized HAM10000 dataset [72], a large collection of multi-source dermoscopy images of common pigmented skin lesions. The dataset contains 10,015 dermoscopy images of seven skin cancer types: Melanocytic nevi (6705 images), Melanoma (1113 images), Benign keratosis (1099 images), Basal cell carcinoma (514

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

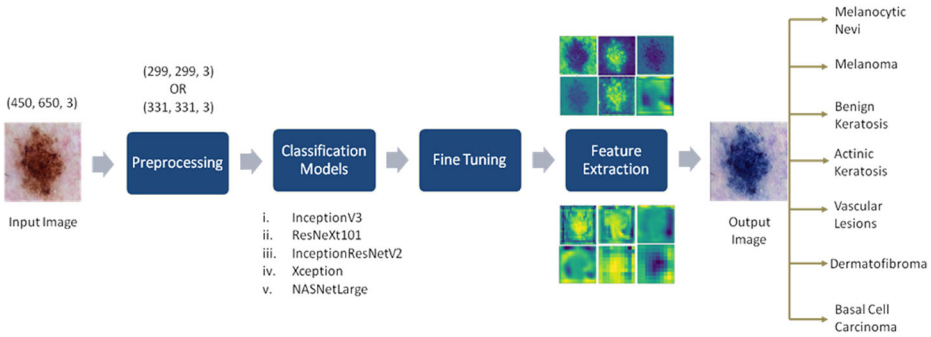

Fig. 2 Proposed Architecture for Multi-class Skin Cancer Classification

images), Actinic keratosis (327 images), Vascular Lesions (142 images), and Dermatofibroma (115 images). Sample images of skin cancer types from HAM10000 are represented in Fig. 3.

The dataset of 10,015 images were split into the training set (8912 images) and validation set (1103 images). The validation data contains unique cases of the dataset (i.e. cases where

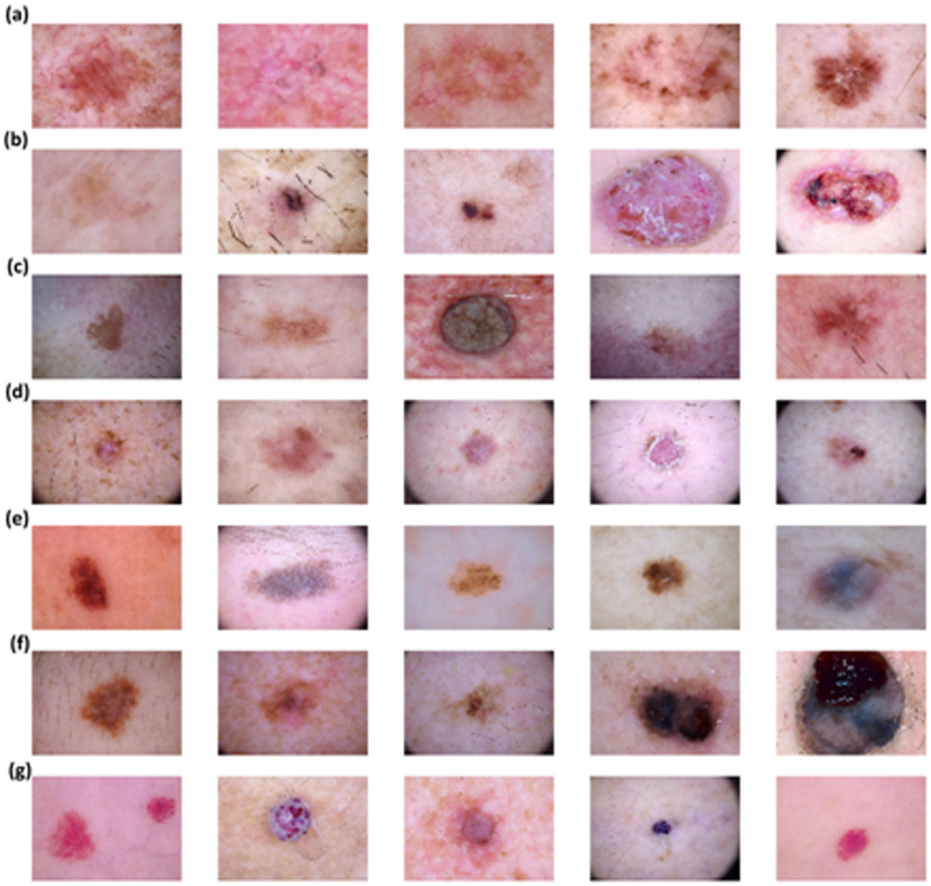

Fig. 3 Sample skin cancer images from HAM10000 dataset (**a**) Actinic keratosis (**b**) Basal cellcarcinoma (**c**) Benign keratosis-like lesions (**d**) dermatofibroma (**e**) Melanocytic nevi (**f**) Melanoma (**g**) Vascular lesions

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

multiple images are associated with the same lesion id were eliminated from the validation set). So that training and validation set must contain a different set of images for the unbiased evaluation of the model's performance.

3.2 Preprocessing

To ensure better generalization, the pre-processing steps were kept minimal for the proposed method. We performed a basic pre-processing step using the built-in pre-processing function of Keras ImageDataGenerator. As the dermoscopy images in the dataset have 450 × 600 pixels resolution, we have downscaled the images to 299 × 299 pixels resolution and 331 × 331 pixels resolution to reconcile with the input image dimension for the models: Xception, InceptionV3, InceptionResNetV2, ResNeXt101, and NASNetLarge.

3.3 Classification Models and Fine Tuning

Recent years has observed the development of more advanced convolutional neural networks to elucidate computer vision problems. Usually, a CNN layer consists of convolutional layer, subsampling layer (*max pooling* or *average pooling*) and optionally fully connected layer. The output at the convolution layer is given by the Eq. 1.

$$A^l{}_j = f\left(\sum_{i=1}^{M^{l-1}} A_i{}^{l-1} * \omega^l{}_{ij} + b^i{}_j\right) \tag{1}$$

Where M^{l-1} is the number of feature maps in the$(l-1)$ layer, $A^i{}_j$ is the activation output at the 1st layer, ω^l_{ij} is the kernel weights from feature map at layer 1 to feature map j at $(l-1)$ layer, and b^i_j is the additional bias parameter.

We employ stochastic gradient descent with momentum (SGDM) [6, 52] and adaptive moment estimation (Adam) [36] optimizers for the loss function to perform the fine-tuning of the models in this work. In each iteration, the SGDM optimizer updates the weights as well as biases of the network to minimize the loss function. The momentum term was utilized to avert the oscillations along the steepest descent path. The SGDM is represented by Eq. 2;

$$\theta_{i+1} = \theta_i - \alpha \nabla E_R(\theta_i) + \gamma(\theta_i - \theta_{i-1}) \tag{2}$$

Where θ is the network's parameter vector, i represents the iteration number, α is the learning rate. In the study, we kept the value of α as 0.0001 and 0.001 for the different networks employed. E_R indicates the loss function, and γ is the momentum term set to 0.9. We utilized categorical cross-entropy loss function while performing the optimisation process using Eq. 3;

$$E_R(\theta) = -\ln\left(\frac{e^{\theta_p}}{\sum_j^C e^{\theta_j}}\right) \tag{3}$$

Where θ_p is the CNN score for the positive class, j represents the iterator number, C is number of classes. The minimization of the loss function using Adam is given by Eq. 4;

$$\theta_{i+1} = \theta_i - \frac{\alpha \nabla E(\theta_i)}{\sqrt{v_i} + \varepsilon} \tag{4}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Where ϑ_i is given by Eq. 5;

$$v_i = \beta_2 \vartheta_{i-1} + (1-\beta_2)[\nabla E(\theta_i)]^2 \quad (5)$$

β_2 is the decay rate which was set to 0.999, ε is very small number which prevents zero in the denominator. The value of ε was set to 0.001.

To enhance the performance of deep learning architectures in the classification of skin cancer, we have performed modifications on the architectures i.e. Xception, InceptionV3, InceptionResNetV2, ResNeXt101, and NASNetLarge. The customizations in the deep learning architecture includes; 1) dense layers with '*relu*' activation, 2) dropout layers and softmax layers at the bottom of the architecture, and 3) improvement in the parameters values. All these customization are applied on the architectures to improve their performance for the skin cancer classification. Further, we have performed fine-tuning of five different CNNs and four ensembles models by adapting SGD and Adam to validate the impact of ensemble methods over seven classes of HAM10000 dataset for MCS cancer classification. The architectures of various models used in this work are represented in Fig. 4.

3.3.1 InceptionV3

InceptionV3 is a well-documented network based on inception modules. Inception modules consist of a series of convolutions in parallel with different kernel sizes to extract features. The InceptionV3 network aims to utilize the added computation as efficiently as possible by suitably factorized convolutions and aggressive regularization. The ability of InceptionV3 to efficiently train over huge dataset makes them popular choice among researchers. Moreover, we have included dense layer with '*relu*' activation, Dropout and Softmax layers with seven outputs at the bottom of the architecture to better fine-tune the model on the dataset. Finally, this architecture is fine-tuned on 8912 sample images for 30 epochs with learning rate as 0.0001 and stochastic gradient descent (SGD) optimizer with momentum as 0.9.

3.3.2 ResNeXt101

ResNet introduced the idea of residual connections as a solution to the problems of accuracy saturation and degradation when increasing the network depth. ResNet comes with different variants such as ResNet-50, ResNet-101, and ResNet-152. The residual learning framework used in the ResNeXt101 architecture eases the training of deeper networks and reformulate the layers to learn residual functions with reference to the layer inputs. This makes ResNeXt101 model easier to converge and can gain accuracy from considerably increased depth. We have included dense layer with '*relu*' activation, dropout and softmax layers with seven outputs as a modification in ResNeXt101 to improve the performance. The modified ResNeXt101 is then fine-tuned on 8912 images (for 30 epochs) with learning rate of 0.0001 and SGD optimizer with momentum as 0.9.

3.3.3 InceptionResNetV2

InceptionResNetV2 has introduced significant simplification to the inception blocks. It is a variation of InceptionV3 model which borrows some ideas from ResNet models. Residual

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

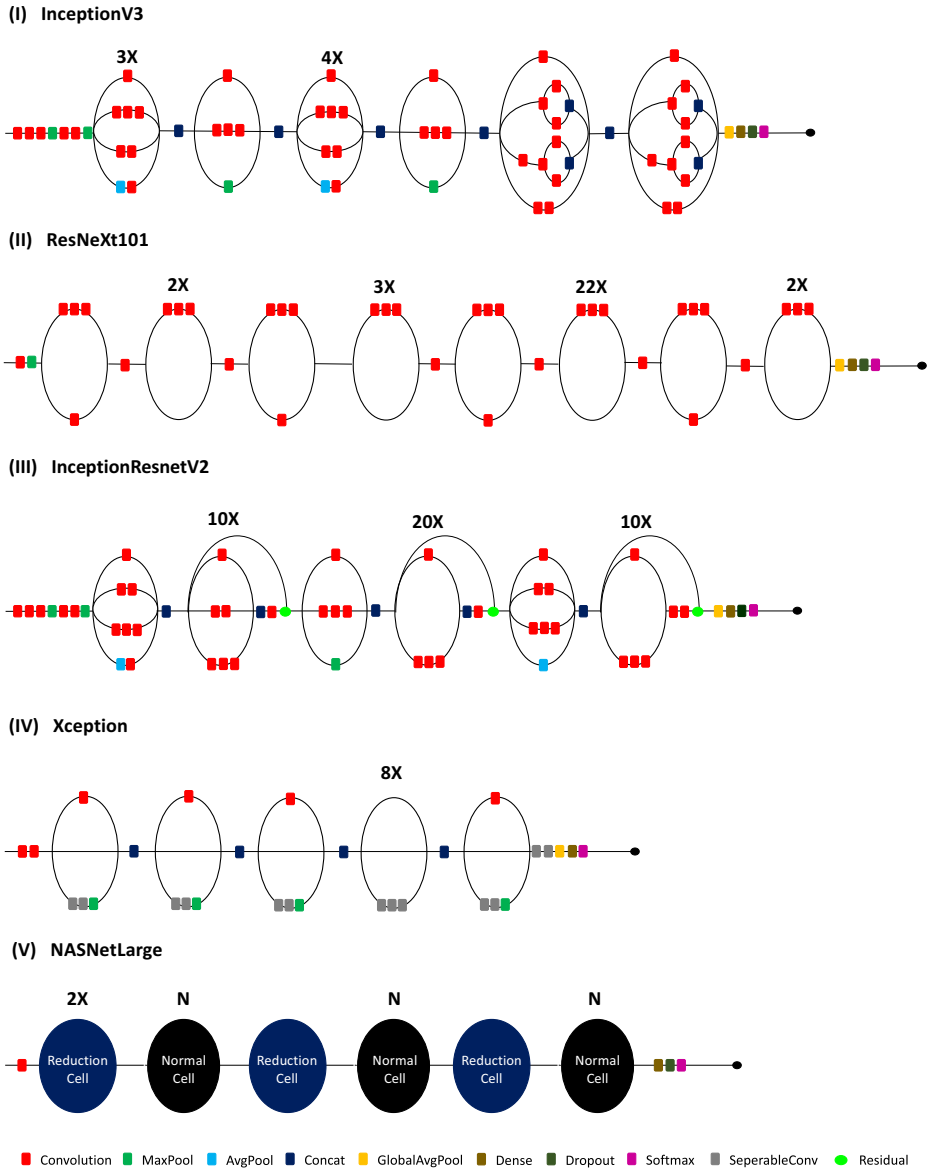

Fig. 4 Architecture of (I) InceptionV3, (II) ResNeXt101, (III) InceptionResnetV2, (IV) Xception, and (V) NASNetLarge

connections allow training much deeper neural networks, which lead to even better performance. The study [70] have shown that InceptionResNetV2 significantly accelerates the training of Inception networks with the help of residual connections. So, we included InceptionResNetV2 as a model after performing certain modifications i.e. added dense layer with '*relu*' activation, dropout and softmax layers with seven outputs. The modified architecture then fine-tuned on 8912 images for 30 epochs wherein the learning rate is 0.0001 and SGD optimizer with momentum is 0.9.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

3.3.4 Xception

Xception architecture is a linear combination of (depth-wise separable convolution) layers with residual connections, which reduces the complexity of architecture. The Xception model focuses more on efficient use of model parameters as compared to other deep convolutional networks and replaces the inception modules with depth-wise separable convolutions.We added dense layer with '*relu*' activation and softmax layer with seven outputs to better fit the architecture of Xception on the dataset. Modified Xception is fine-tuned on 8912 images with learning rate as 0.001 and adaptive moment estimation (Adam) optimizer for the faster optimization of the model. To do the fine-tuning we have used 30 epochs.

3.3.5 NASNetLarge

NASNet architectures introduce a new concept of normal cell and reduction cell, which can be tuned using reinforcement learning search method. NASNetLarge architecture is specifically designed to train over very large datasets. As the training over large dataset is expensive, the search for an architectural building block is conducted on a small dataset and then transfer the block to a larger dataset using NASNet search space. The key aspect of NASNetLarge includes ScheduledDropPath regularization technique which significantly improves generalization in the NASNet models. We have modified the pre-trained architecture of NASNetLarge by adding dense layer with '*relu*' activation, dropout and softmax layers with seven outputs. Finally, this modified architecture is fine-tuned over 8912 images for 25 epochs with learning rate of 0.0001 and SGD optimizer with momentum of 0.9.

3.4 Feature extraction

In this work, we have used Integrated Feature Extractor for the feature extraction of five different pre-trained models. Each approach is effective and save significant time in developing and training a deep convolutional neural network model. The detailed descriptions of five pre-trained models are discussed in section 3.3. Figure 5 highlights the layer-wise processing in Xception model for an input image to produce the output image, for the identification of type of skin cancer. Here, different features are extracted at different layers i.e. layer by layer deeper features are extracted to accumulate more features for accurate prediction. Similarly, other models can also extract the features to identify the type of skin cancer.

The Integrated Feature Extractor uses the concept of transfer learning for the effective feature extraction, where a model trained on a particular problem is used on a different problem after fine-tuning. We found safe to use pre-trained network for this work, as the convolutional layers closer to the input layer learns low-level features such as lines, borders, etc. which can be used for the efficient training for another problem. We decided to integrate the pre-trained model output with few sets of layers at the end. The weights of the pre-trained models were used as the starting point for the training process and fine-tuned to our problem. However, the weights of the pre-trained model were frozen during training so that the pre-trained weights do not modify as the new model is trained.

Each hidden layer of convolutional neural network maps its input data to an internal representation which captures a higher level of abstraction. These learned features evolve

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

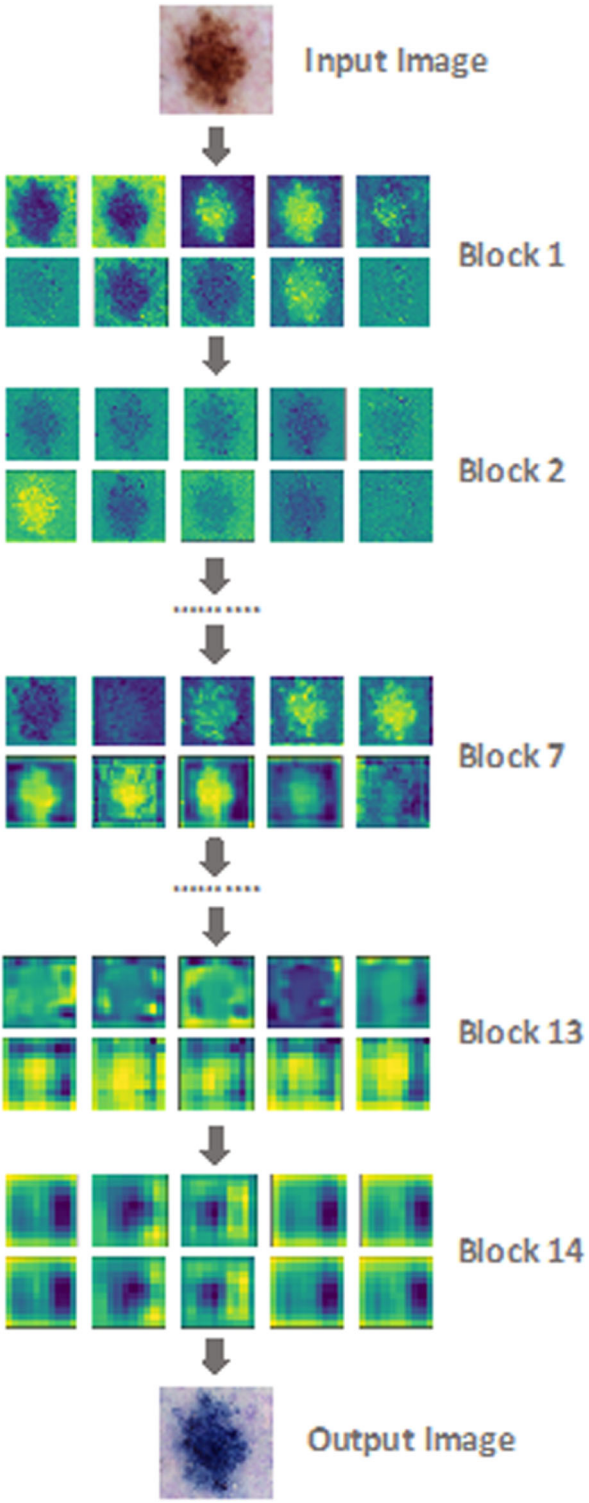

Fig. 5 Feature extraction at different layers of Xception model among five models

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

increasingly more informative as they are passed through the different layers of the network. Ultimately, the individual features of each layer are stored in an image for the classification task. The illustration of the feature extraction process at different layers of the network is shown in Fig. 5, which was performed using the modified Xception model. In [42, 55], feature extraction is done by simply training the images using the pre-trained networks followed by the output of the fully connected (FC) layers. However, we hypothesize that fine-tuning of pre-trained networks on the relevant dataset can contribute to developing higher quality features, which can boost the performance of the pre-trained models.

3.5 Performance matrix

We have validated the performance of models on 1103 images by evaluating recall, precision, accuracy and F1-score. The performance matrix can be evaluated by estimating the predicted image among four subsets: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP represents the number of positive cases classified correctly. TN represents the number of negative cases classified correctly. FP, the number of positive cases classified as inaccurate. FN, the number of negative cases classified incorrectly.

Based on the cardinality of these subsets, the performance matrix can be evaluated [26]. Accuracy is one of the best measures used to interpret the performance of models. The accuracy is expressed by using TP, TN, FP, FN as represented by Eq. 6. The other significant performance metric for multi-class classification are recall, precision and F1-score are expressed using Eqs. 7, 8 and 9 respectively.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{6}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

$$F1{-}score = 2\left(\frac{Precision{*}Recall}{Precision + Recall}\right) \tag{9}$$

4 Results and discussion

The result is derived from the validation data, which consist of 1103 images of seven classes of skin cancer from the HAM10000 dataset. We have used Keras library [14] for implementing

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

the deep models used in this research work. Since, Keras has an ability to run on top of other deep learning libraries such as TensorFlow or Theano. The training of models is done on the Kaggle [32] server with 13GB RAM using Tesla P100-PCIE-16GB and 6 minor GPUs.

We evaluated the performance of five different modelsviz.InceptionV3, ResNetXt101, InceptionResNetV2, Xception, and NASNetLarge for the classification of skin cancer among seven classes: Melanocytic nevi, Melanoma, Benign keratosis, Basal cell carcinoma, Actinic keratosis, Vascular Lesions, and Dermatofibroma. The categorical accuracy for InceptionV3, ResNetXt101, InceptionResNetV2, Xception, and NASNetLarge were found to be 91.56%, 93.20%, 93.20%, 91.47%, and 91.11% respectively. The best accuracy is recorded for ResNetXt101 and InceptionResNetV2.

The training-validation accuracy curves and training-validation loss curves are represented in Fig. 6 for each of the five models. In the initial stage of training for a few epochs, the validation accuracy is higher than training accuracy or validation loss is lower than the training loss; this can be justified in several ways. Firstly, as we have utilized the Dropout layer in the architecture during fine-tuning of the model to make our system less prone to over-fitting, these Dropout layers disable the neurons during training in an attempt to reduce the complexity of the model. In Keras, dropout layers are disabled during testing providing the network full computational power to perform prediction and can lead to better training accuracy for a few epochs while evaluating the model [20]. Secondly, the training loss is the average of the losses over each batch of training data. As the model is evolving with time, the loss over the last batches is generally higher as compared to the starting batches of an epoch. Diversely, the validation loss for a model is computed at the end of an epoch, resulting in a lower loss. This can contribute to lower validation loss as compared to training loss.

The weighted average of recall, precision, and F1-score are also evaluated to check the performance of models with respect to the number of images for each class of validation data. We found that the weighted average of recall, precision, and F1-score for InceptionV3 is 89%, 89%, and 89% respectively. Similarly, the weighted averages of recall, precision, and F1-score for ResNetXt101, InceptionResNetV2, Xception, and NASNetLarge models are also evaluated. The accuracy, weighted average of recall, precision, and F1-score results for the five different models utilized in this paper are summarized in Table 1.

We have also done experimentation for four ensemble models: InceptionV3 + Xception, InceptionResNetV2 + Xception, InceptionResNetV2 + ResNeXt101, and InceptionResNetV2 + ResNeXt101 + Xception. The outputs of the individual models were averaged to develop the required architecture of ensemble models. The accuracy, weighted average of precision, recall, and F1-score results for four ensemble models are shown in Table 2. The training-validation accuracy curves and training-validation loss curves for four ensemble models are represented in Fig. 7.

The categorical accuracy was found to be 91.56% for ensemble model 'InceptionV3 + Xception', 88.66% for 'InceptionResNetV2 + Xception', 92.83% for 'InceptionResNetV2 + ResNeXt101', and 89.66% for 'InceptionResNetV2 + ResNeXt101 + Xception'. We have achieved best results for 'InceptionResNetV2 + ResNeXt101' and 'InceptionV3 + Xception' ensemble methods.

We have observed a trend from the literature that as the number of classification classes increases, the performance of model deteriorates. Since the model needs to perform prediction for multiple classes, there is a greater probability of having an incorrect prediction. Thus,

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

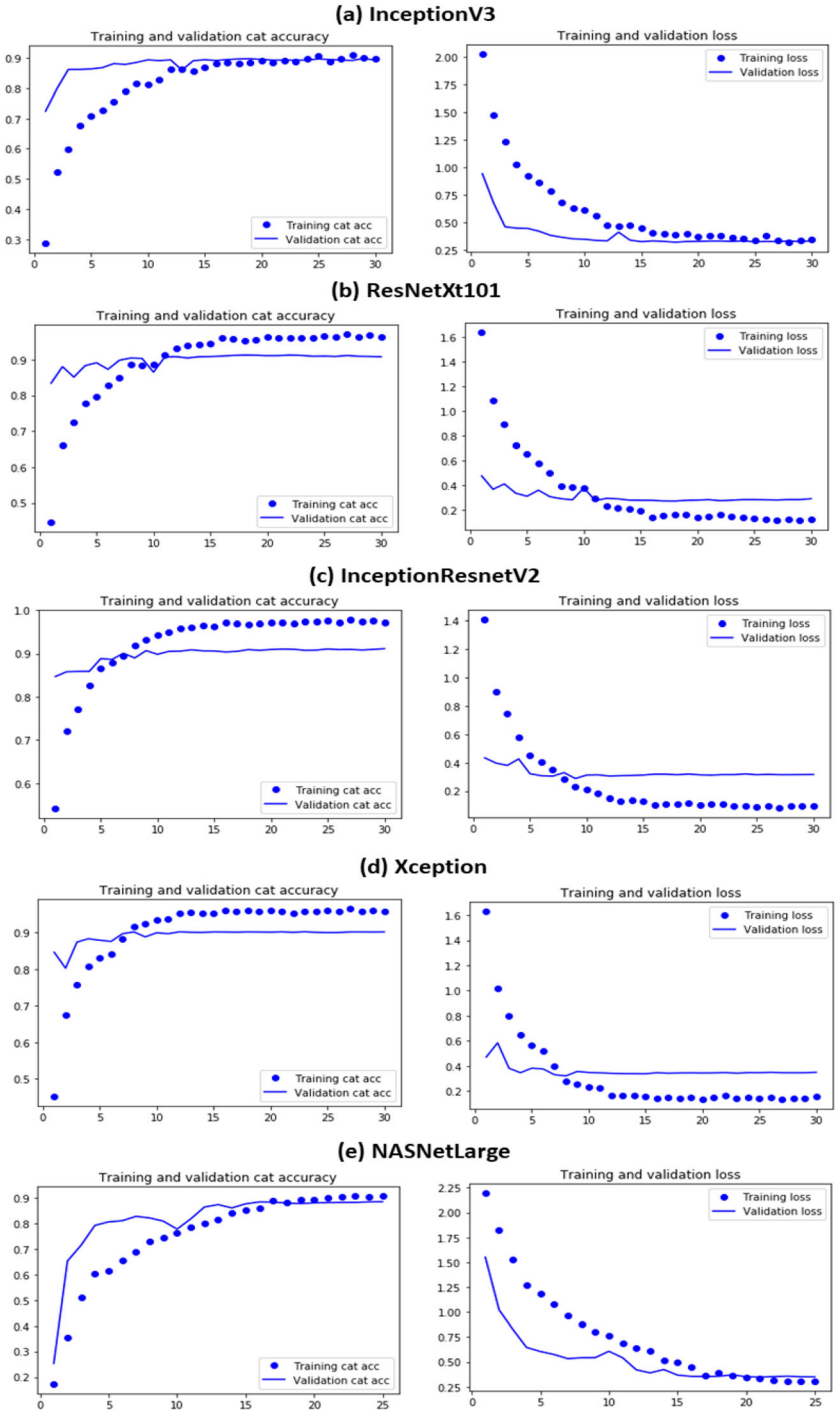

Fig. 6 Training-validation Accuracy and Loss Curves for (**a**) InceptionV3 (**b**) ResNetXt101 (**c**) InceptionResnetV2 (**d**) Xception (**e**) NASNetLarge

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 1 Accuracy, Weighted Average of Precision, Recall and F1-score for HAM10000 dataset of independent model

Method	Accuracy (%)	Weighted Average		
		Precision (%)	Recall (%)	F1-score (%)
InceptionV3	91.56	89	89	89
ResNetXt101	93.20	88	88	88
InceptionResNetV2	93.20	87	88	88
Xception	91.47	89	88	88
NASNetLarge	91.11	86	86	86

model performance tends to decrease with several classification classes. The previous works [27, 42], [18, 35, 46, 54, 65], and [49] lack in performance as compared to the proposed work in this paper.

Although, [18, 27, 42] employ classification on either two or three classes, still their classification accuracy varies from 69.4% to 84.8%. In [18, 70], [46, 54, 65] classification is performed on more than five classes and lacked to achieve higher accuracy, precision and recall. The accuracy, precision, and recall obtained are varying between 48.9% - 90.1%, 78.6% - 84.9%, and 51.8% - 80.0%. Table 3 and Table 4 show the comparison of proposed work with existing models. Moreover, 'N/A' represents the performance matrix was not included in the respective research work.

We outperformed both dermatologists and the current deep learning methods in multiclass skin cancer classification with seven architectures used in this work; InceptionV3, ResNeXt101, InceptionResNetV2, Xception, NASNetLarge, and ensemble models 'InceptionV3 + Xception', 'InceptionResNetV2 + ResNet101' by avoiding the use of extensive pre-processing, and data augmentation methods.

We have observed that ResNeXt101 model emerge as an optimized architecture which makes training easier and can gain higher accuracy for skin cancer classification. ResNeXt101 achieves the best results hence; we propose the use of ResNeXt101 for the MCS cancer

Table 2 Accuracy, Weighted Average of Precision, Recall and F1-score for HAM10000 dataset of ensemble models

Ensemble	Accuracy (%)	Weighted Average		
		Precision (%)	Recall (%)	F1-score (%)
InceptionV3 + Xception	91.56	82	84	83
InceptionResNetV2 + Xception	88.66	80	82	81
InceptionResNetV2 + ResNeXt101	92.83	83	84	84
InceptionResNetV2 + ResNeXt101 + Xception	89.66	83	85	84

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

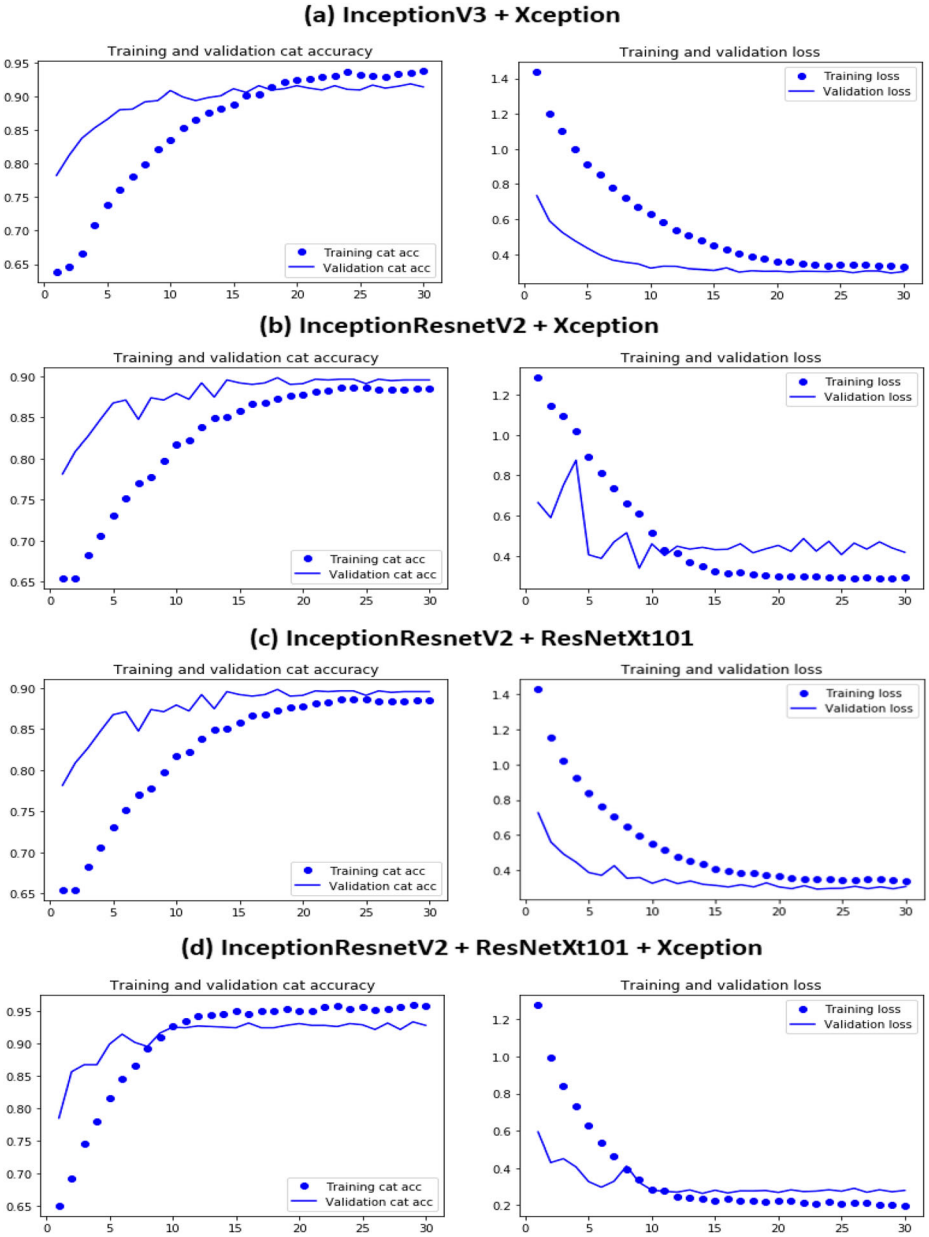

Fig. 7 Training-validation Accuracy and Loss Curves for (**a**) InceptionV3 + Xception (**b**) InceptionResnetV2 + Xception (**c**) InceptionResnetV2 + ResNeXt101 (**d**) InceptionResnetV2 + ResNeXt101 + Xception

classification. Additionally, we have noted better results without using ensemble methods for skin cancer classification on HAM1000 dataset which is listed in Table 1. Although, ensemble method is used globally to increase the accuracy in the classification task, but it drastically increases the architectural complexity leading to much longer training time for the model.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 3 Comparison with Other Deep Learning Models

Ref.	Method	Number of Classes	Accuracy (%)	Precision (%)	Recall (%)
[42]	VGGNet	Two	81.3	79.74	78.66
[27]	GoogleNet	Three	84.2	N/A	59.2
	AlexNet		84.8		51.8
	ResNet		82.8		52.0
	VGGNet		81.3		43.4
[34]	Fully Convolutional Network	Five	85.8	N/A	N/A
		Ten	81.8		
[35]	Multi-tract CNN	Ten	79.15	N/A	N/A
[18]	CNN	Three	69.4	N/A	N/A
	CNN-PA	Nine	72.1		
	CNN		48.9		
	CNN-PA		55.4		
[54]	VGG16	Seven	75.6	N/A	N/A
	ResNet50		86.6		
	DenseNet121		89.2		
	Xception		90.1		
	InceptionV3		74.3		
	DenseNet161		88.7		
	InceptionResNetV2		86.1		
[46]	VGG16	Seven	80.1	N/A	N/A
	GoogleNet		79.7		
[65]	ResNet50	Seven	87.1	78.6	77.0
	InceptionV3		89.7	84.9	80.0
[13]	MobileNet	Seven	83.1	89.0	83.0
[49]	InceptionResnetV2	Seven	70.0	N/A	N/A
	PNASNet-5-Large		76.0		
	SENet154		74.0		
	InceptionV4		67.0		
[23]	Triple-Net + CAM-BP	Two	82.0	N/A	N/A
[62]	Dilated VGG16	Seven	87.42	87.0	87.0
	Dilated VGG19		85.02	85.0	85.0
	Dilated MobileNet		88.22	89.0	88.0
	Dilated InceptionV3		89.81	89.0	89.0
[2]	IRRCNN	Seven	87.0	N/A	N/A
[60]	CNN	Seven	77.0	N/A	N/A
	CNN (One vs All)		92.90		
Our	**InceptionV3**	**Seven**	**91.56**	**89.0**	**89.0**
	ResNeXt101		**93.20**	**88.0**	**88.0**
	InceptionResnetV2		**93.20**	**87.0**	**88.0**
	Xception		**91.47**	**89.0**	**88.0**
	NASNetLarge		**91.11**	**86.0**	**86.0**

5 Conclusions

As the incidence rates of skin cancer have been raising over the past decades, there is an urgent need to address this global public health issue. The magnificent performance of deep CNNs for medical image classification has made to utilize them for the skin cancer classification. Although, various researches have been done before for the classification of skin cancer, they failed to extend their study for multiple classes of skin cancer with high performance. In this paper, we outperformed both dermatologists and current deep learning methods for MCS cancer classification. The performance is analyzed for five

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 4 Comparison with Other Ensemble Deep Learning Models

Ref.	Ensemble Method	Number of Classes	Accuracy (%)	Precision (%)	Recall (%)
[27]	AlexNet + VGGNet	Three	79.9	N/A	N/A
	GoogleNet + AlexNet		80.7		
	GoogleNet + VGGNet		81.2		
	GoogleNet + AlexNet+ VGGNet		83.8		
[46]	VGG16 + GoogleNet	Seven	81.5	N/A	N/A
[65]	ResNet50 + InceptionV3	Seven	89.9	86.2	79.6
Our	**InceptionV3 + Xception**	**Seven**	**91.56**	**82.0**	**84.0**
	InceptionResnetV2 + Xception		**88.66**	**80.0**	**82.0**
	InceptionResnetV2 + ResNeXt101		**92.83**	**83.0**	**84.0**
	InceptionResnetV2 + ResNeXt101 + Xception		**89.66**	**83.0**	**85.0**

pre-trained CNNs and four ensemble models to determine the best method for skin cancer classification on HAM10000 dataset. We performed extensive research to determine the best set-up of hyper-parameters for five models pre-trained on ImageNet namely Xception, InceptionV3, InceptionResNetV2, NASNetLarge, ResNetXt101 and their ensembles InceptionV3 + Xception, InceptionResNetV2 + Xception, InceptionResNetV2 + ResNetXt101, InceptionResNetV2 + ResNetXt101 + Xception. The ResNetXt101 shows significant improvement in the performance as compared to previously proposed deep learning models. Hence, we propose the use of ResNetXt101 for MCS cancer classification. Moreover, we conclude that the training of deep learning models with best-setup of hyper-parameter can perform better than even ensemble models. Although, ensemble methods are used globally to increase the accuracy in the classification task, they not only drastically increase the architectural complexity of the model but may not have a significant role in improving the performance of deep learning models tuned with the best hyper-parameters.

The future work may deal with the development of more robust deep learning computer-aided systems for skin cancer diagnosis by including clinical images as additional inputs along with the dermoscopy images to deep learning models by extending the concept of saliency objects or features detection [19, 21–24, 74, 82] which have been effectively utilized in the past for skin cancer diagnosis. The combination of these two clinical and dermoscopy modalities can provide complementary visual features that can develop highly accurate and efficient computer-aided systems for skin cancer classification.

References

1. Abbas Q, Emre Celebi M, Garcia IF, Ahmad W (2013) Melanoma recognition framework based on expert definition of ABCD for dermoscopic images. Skin Research And Technology 19(1):e93–e102. https://doi.org/10.1111/j.1600-0846.2012.00614.x
2. Alom, MZ, Aspiras, T, Taha, TM, & Asari, VK (2020). Skin cancer segmentation and classification with improved deep convolutional neural network. In: Medical Imaging 2020: Imaging informatics for healthcare, research, and applications, vol. 11318, pp. 1131814. International Society for Optics and Photonics. doi: https://doi.org/10.1117/12.2550146.
3. Australian Government (2018). Melanoma of the skin statistics. https://melanoma.canceraustralia.gov.au/statistics. Accessed 19 June 2019.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

4. Ballerini L, Fisher RB, Aldridge B, Rees J (2013) A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In: Color medical image analysis. Springer, Dordrecht, pp 63–86
5. Binder M, Schwarz M, Winkler A, Steiner A, Kaider A, Wolff K, Pehamberger H (1995) Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Arch Dermatol 131(3):286–291
6. Bishop CM (2006) Pattern recognition and machine learning. Springer
7. Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C (2004) Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. Br J Dermatol 151(5):1029–1038. https://doi.org/10.1111/j.1365-2133.2004.06210.x
8. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424. https://doi.org/10.3322/caac.21492
9. Burroni M, Corona R, Dell'Eva G, Sera F, Bono R, Puddu P, Rubegni P (2004) Melanoma computer-aided diagnosis: reliability and feasibility study. Clin Cancer Res 10(6):1881–1886. https://doi.org/10.1158/1078-0432.CCR-03-0039
10. Cancer Facts and Figures 2016 - American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html. Accessed 31March 2019.
11. Celebi ME, Iyatomi H, Stoecker WV, Moss RH, Rabinovitz HS, Argenziano G, Soyer HP (2008) Automatic detection of blue-white veil and related structures in dermoscopy images. Comput Med Imaging Graph 32(8):670–677. https://doi.org/10.1016/j.compmedimag.2008.08.003
12. Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH (2007) A methodological approach to the classification of dermoscopy images. Comput Med Imaging Graph 31(6): 362–373. https://doi.org/10.1016/j.compmedimag.2007.01.003
13. Chaturvedi, SS, Gupta, K, Prasad, P (2019). Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. arXiv preprint arXiv:1907.03220.
14. Chollet, F. (2015). GitHub - keras-team/keras: Deep Learning for humans. https://github.com/keras-team/keras. Accessed 24 June 2019.
15. Chollet, F (2017). Xception: deep learning with depthwise separable convolutions. In: IEEE conference on computer vision and pattern recognition, pp. 1251–1258.
16. Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR (2015, October) Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: International workshop on machine learning in medical imaging. Springer, Cham, pp 118–126
17. Deng, J, Dong, W, Socher, R, Li, LJ, Li, K, Fei-Fei, L (2009). Imagenet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, pp. 248–255. doi: https://doi.org/10.1109/CVPRW.2009.5206848.
18. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118. https://doi.org/10.1038/nature21056
19. Fan, DP, Cheng, MM, Liu, JJ, Gao, SH, Hou, Q, Borji, A (2018). Salient objects in clutter: bringing salient object detection to the foreground. In: proceedings of the European conference on computer vision (ECCV), pp. 186-202. Springer, Cham. Doi: https://doi.org/10.1007/978-3-030-01267-0_12.
20. FAQ - Keras Documentation (2019). https://keras.io/getting-started/faq/#why-is-the-training-loss-much-higher-than-the-testing-loss. Accessed 29 June 2019.
21. Fu, K, Fan, DP, Ji, GP, Zhao, Q (2020). JL-DCF: joint learning and densely-cooperative fusion framework for RGD-D salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3052–3062.
22. Fu K, Zhao Q, Gu IYH, Yang J (2019) Deepside: a general deep framework for salient object detection. Neurocomputing 356:69–82
23. Ge Z, Demyanov S, Chakravorty R, Bowling A, Garnavi R (2017) Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 250–258. https://doi.org/10.1007/978-3-319-66179-7
24. Gong, C, Tao, D, Liu, W, Maybank, SJ, Fang, M, Fu, K, Yang, J (2015). Saliency propagation from simple to difficult. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2531–2539.
25. Goodson AG, Grossman D (2009) Strategies for early melanoma detection: approaches to the patient with nevi. J Am Acad Dermatol 60(5):719–735. https://doi.org/10.1016/j.jaad.2008.10.065

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

26. Google Developers (2019). Machine Learning Glossary. https://developers.google.com/machine-learning/glossary. Accessed 24 June 2019.
27. Harangi B (2018) Skin lesion classification with ensembles of deep convolutional neural networks. J Biomed Inform 86:25–32. https://doi.org/10.1016/j.jbi.2018.08.006
28. Harangi, B, Baran, A, Hajdu, A (2018). Classification of skin lesions using an ensemble of deep neural networks. In: IEEE 40th annual international conference of the IEEE engineering in medicine and biology society - EMBC'2018, pp. 2575–2578. doi: https://doi.org/10.1109/EMBC.2018.8512800.
29. He, K, Zhang, X, Ren, S, Sun, J (2016). Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition, pp. 770–778.
30. Iyatomi H, Oka H, Saito M, Miyake A, Kimoto M, Yamagami J, Argenziano G (2006) Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. Melanoma Res 16(2):183–190. https://doi.org/10.1097/01.cmr.0000215041.76553.58
31. Jana, E, Subban, R, Saraswathi, S (2017). Research on skin Cancer cell detection using image processing. In: IEEE international conference on computational intelligence and computing research - ICCIC'2017, pp. 1–8. doi: https://doi.org/10.1109/ICCIC.2017.8524554.
32. Kaggle: Your Home for Data Science (2019). https://www.kaggle.com/. Accessed 31 March 2019.
33. Kasmi R, Mokrani K (2016) Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Process 10(6):448–455. https://doi.org/10.1049/iet-ipr.2015.0385
34. Kawahara, J, BenTaieb, A, Hamarneh, G (2016). Deep features to classify skin lesions. In: IEEE 13th international symposium on biomedical imaging - ISBI'2016, pp 1397-1400). doi: https://doi.org/10.1109/ISBI.2016.7493528.
35. Kawahara J, Hamarneh G (2016) Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In: International workshop on machine learning in medical imaging. Springer, Cham, pp 164–171. https://doi.org/10.1007/978-3-319-47157-0_20
36. Kingma, DP, Ba, J (2014). Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.
37. Kittler H, Pehamberger H, Wolff K, Binder MJTIO (2002) Diagnostic accuracy of dermoscopy. The lancet oncology 3(3):159–165. https://doi.org/10.1016/S1470-2045(02)00679-4
38. Koh HK, Geller AC, Miller DR, Grossbart TA, Lew RA (1996) Prevention and early detection strategies for melanoma and skin cancer: current status. Arch Dermatol 132(4):436–443
39. Korotkov K, Garcia R (2012) Computerized analysis of pigmented skin lesions: a review. Artif Intell Med 56(2):69–90. https://doi.org/10.1016/j.artmed.2012.08.002
40. Krizhevsky, A, Sutskever, I, Hinton, GE (2012). Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105.
41. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
42. Lopez, AR, Giro-i-Nieto, X, Burdick, J, Marques, O (2017). Skin lesion classification from dermoscopic images using deep learning techniques. In: IEEE 13th IASTED international conference on biomedical engineering – BioMed'2017, pp 49-54. doi: https://doi.org/10.2316/P.2017.852-053.
43. Maglogiannis I, Doukas CN (2009) Overview of advanced computer vision systems for skin lesions characterization. IEEE Trans Inf Technol Biomed 13(5):721–733. https://doi.org/10.1109/TITB.2009.2017529
44. Mahbod A, Schaefer G, Ellinger I, Ecker R, Pitiot A, Wang C (2019) Fusing fine-tuned deep features for skin lesion classification. Comput Med Imaging Graph 71:19–29. https://doi.org/10.1016/J.COMPMEDIMAG.2018.10.007
45. Mahbod, A, Schaefer, G, Wang, C, Ecker, R, Ellinge, I (2019). Skin lesion classification using hybrid deep neural networks. In: IEEE international conference on acoustics, speech and signal processing - ICASSP'2019, pp. 1229–1233.
46. Majtner, T, Bajić, B, Yildirim, S, Hardeberg, JY, Lindblad, J, Sladoje, N (2018). Ensemble of convolutional neural networks for dermoscopic images classification. arXiv preprint arXiv:1808.05071.
47. Masood A, Ali Al-Jumaily A (2013) Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. International journal of biomedical imaging 2013:323268–323222. https://doi.org/10.1155/2013/323268
48. Mhaske, HR, & Phalke, DA (2013). Melanoma skin cancer detection and classification based on supervised and unsupervised learning. In: IEEE international conference on circuits, controls and communications - CCUBE'2013, pp 1-5. doi: https://doi.org/10.1109/CCUBE.2013.6718539.
49. Milton, MAA (2019). Automated skin lesion classification using ensemble of deep neural networks in ISIC 2018: skin lesion analysis towards melanoma detection challenge. arXiv preprint arXiv:1901.10802.
50. Morton CA, Mackie RM (1998) Clinical accuracy of the diagnosis of cutaneous malignant melanoma. Br J Dermatol 138(2):283–287

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

51. Moura N, Veras R, Aires K, Machado V, Silva R, Araújo F, Claro M (2019) ABCD rule and pre-trained CNNs for melanoma diagnosis. Multimed Tools Appl 78(6):6869–6888. https://doi.org/10.1007/s11042-018-6404-8
52. Murphy KP (2012) Machine learning: a probabilistic perspective. MIT press
53. Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, Plewig G (1994) The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. J Am Acad Dermatol 30(4):551–559
54. Nyíri T, Kiss A (2018) Novel Ensembling methods for dermatological image classification. In: International conference on theory and practice of natural computing. Springer, Cham, pp 438–448
55. Oliveira RB, Papa JP, Pereira AS, Tavares JMR (2018) Computational methods for pigmented skin lesion classification in images: review and future trends. Neural Comput & Applic 29(3):613–636. https://doi.org/10.1007/s00521-016-2482-6
56. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359. https://doi.org/10.1109/TKDE.2009.191
57. Parkin DM, Mesher D, Sasieni P (2011) 13. Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010. Br J Cancer 105(2):S66–S69. https://doi.org/10.1038/bjc.2011.486
58. Pathan S, Prabhu KG, Siddalingaswamy PC (2018) Techniques and algorithms for computer aided diagnosis of pigmented skin lesions-a review. Biomedical Signal Processing and Control 39:237–262. https://doi.org/10.1016/j.bspc.2017.07.010
59. Piccolo D, Ferrari A, Peris KETTY, Daidone R, Ruggeri B, Chimenti S (2002) Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: a comparative study. Br J Dermatol 147(3):481–486. https://doi.org/10.1046/j.1365-2133.2002.04978.x
60. Polat K, Koc KO (2020) Detection of skin diseases from Dermoscopy image using the combination of convolutional neural network and one-versus-all. Journal of Artificial Intelligence And Systems 2(1):80–97. https://doi.org/10.33969/ais.2020.21006.
61. Ramteke NS, Jain SV (2013) ABCD rule based automatic computer-aided skin cancer detection using MATLAB. International Journal of Computer Technology and Applications 4(4):691
62. Ratul AR, Mozaffari MH, Lee WS, Parimbelli E (2019) Skin Lesions Classification Using Deep Learning Based on Dilated Convolution bioRxiv:860700. https://doi.org/10.1101/860700
63. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM (2015) Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. JAMA dermatology 151(10):1081–1086. https://doi.org/10.1001/jamadermatol.2015.1187
64. Rosado B, Menzies S, Harbauer A, Pehamberger H, Wolff K, Binder M, Kittler H (2003) Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. Arch Dermatol 139(3):361–367
65. Shahin, AH, Kamal, A, Elattar, MA (2018). Deep ensemble learning for skin lesion classification from dermoscopic images. In: IEEE 9th Cairo international biomedical engineering conference - CIBEC'2018, pp 150-153. doi: https://doi.org/10.1109/CIBEC.2018.8641815.
66. Sharif Razavian, A, Azizpour, H, Sullivan, J, & Carlsson, S (2014). CNN features off-the-shelf: an astounding baseline for recognition. In: IEEE conference on computer vision and pattern recognition workshops, pp. 806–813.
67. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. CA Cancer J Clin 69(1):7–34. https://doi.org/10.3322/caac.21551
68. Silverberg E, Boring CC, Squires TS (1990) Cancer statistics, 1990. CA Cancer J Clin 40(1):9–26
69. Simonyan, K, Zisserman, A (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
70. Szegedy, C, Ioffe, S, Vanhoucke, V, & Alemi, AA (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence, pp. 4278–4284.
71. Szegedy, C, Vanhoucke, V, Ioffe, S, Shlens, J, Wojna, Z (2016). Rethinking the inception architecture for computer vision. In: IEEE conference on computer vision and pattern recognition, pp. 2818–2826.
72. Tschandl P, Rosendahl C, Kittler H (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data 5:180161. https://doi.org/10.1038/sdata.2018.161
73. Vestergaard ME, Macaskill PHPM, Holt PE, Menzies SW (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 Dermatol 159(3):669–676. https://doi.org/10.1111/j.1365-2133.2008.08713.x
74. Wei, J, Wang, S, & Huang, Q (2019). F3Net: fusion, feedback and focus for salient object detection. arXiv preprint arXiv:1911.11445.
75. White R, Rigel DS, Friedman RJ (1991) Computer applications in the diagnosis and prognosis of malignant melanoma. Dermatol Clin 9(4):695–702

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

76. WHO (2017). Skin cancers. https://www.who.int/uv/faq/skincancer/en/index1.html. Accessed 19 June 2019.
77. Xie, S, Girshick, R, Dollár, P, Tu, Z, & He, K (2017). Aggregated residual transformations for deep neural networks. In: IEEE conference on computer vision and pattern recognition, pp. 1492–1500.
78. Yu L, Chen H, Dou Q, Qin J, Heng PA (2016) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging 36(4):994–1004. https://doi.org/10.1109/TMI.2016.2642839
79. Yu, Z, Ni, D, Chen, S, Qin, J, Li, S, Wang, T, Lei, B (2017). Hybrid dermoscopy image classification framework based on deep convolutional neural network and Fisher vector. In: IEEE 14th international symposium on biomedical imaging - ISBI'2017, pp 301-304. doi: https://doi.org/10.1109/ISBI.2017.7950524.
80. Zaqout I (2016) Diagnosis of skin lesions based on dermoscopic images using image processing techniques. International Journal Of Signal Processing, Image Processing And Pattern Recognition 9(9):189–204. https://doi.org/10.14257/ijsip.2016.9.9.18.
81. Zhang M, Qureshi AA, Geller AC, Frazier L, Hunter DJ, Han J (2012) Use of tanning beds and incidence of skin cancer. J Clin Oncol 30(14):1588–1593. https://doi.org/10.1200/JCO.2011.39.3652
82. Zhao, JX, Liu, JJ, Fan, DP, Cao, Y, Yang, J, Cheng, MM (2019). EGNet: edge guidance network for salient object detection. In: proceedings of the IEEE international conference on computer vision, pp. 8779–8788.
83. Zoph, B, Vasudevan, V, Shlens, J, Le, QV (2018). Learning transferable architectures for scalable image recognition. In: IEEE conference on computer vision and pattern recognition, pp. 8697–8710.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Saket S. Chaturvedi[1] · **Jitendra V. Tembhurne**[2] · **Tausif Diwan**[2]

[1] Department of Computer Science & Engineering, PIET, Nagpur, India

[2] Department of Computer Science & Engineering, Indian Institute of Information Technology, Nagpur, India

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com