



A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data

Ginni Arora¹ · Ashwani Kumar Dubey² · Zainul Abdin Jaffery³ · Alvaro Rocha⁴

Received: 19 June 2021 / Accepted: 4 January 2022 / Published online: 25 January 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Among various types of skin diseases, skin cancer is the deadliest form of the disease. This paper classifies seven types of skin diseases: Actinic keratosis and intraepithelial carcinoma, Basal cell carcinoma, Benign keratosis, Dermatofibroma, Melanoma, Melanocytic type, and Vascular lesions. The primary objective of this paper is to evaluate the performance of these deep learning networks on skin lesion images. The lesion classification is implemented through transfer learning on fourteen deep learning networks: AlexNet, GoogleNet, ResNet50, VGG16, VGG19, ResNet101, InceptionV3, InceptionResNetV2, SqueezeNet, DenseNet201, ResNet18, MobileNetV2, ShuffleNet and NasNetMobile. The dataset used for these experiments are from ISIC 2018 of about 10,154 images. The results show that DenseNet201 performs best with 0.825 accuracy and improves skin lesion classification under multiple diseases. The proposed work shows the various parameters, including the accuracy of all fourteen deep learning networks, which helped build an efficient automated classification model for multiple skin lesions.

Keywords Deep learning · DenseNet201 · ISIC · Multi skin lesion classification · Skin cancer

1 Introduction

Human skin is the biggest organ in the body and is loaded with various factors, including sun rays (Ultraviolet), sunburn, lifestyle, smoking, liquor utilization, physical action, viruses, and the workplace. These factors compromise its trustworthiness and have real and dangerous impacts on human health. Diseases that influence the skin directly are the common reasons for every human disease, influencing almost 33% of the total population of around 1.9 billion at once, leading to research in this discipline. Skin diseases added to roughly 1.79% of the global problem of diseases, estimated by balanced life years. In Great Britain, 60% of the population experiences skin diseases for the duration of their lives. Skin issues can be harmful, incendiary, or infectious and influence individuals, all things considered, particularly older people and young children. There are various consequences of skin issues, for example, death (on account of melanoma), loss of connections, effect on day-to-day activity and damage to internal organs. Besides, they likewise represent a danger to humans, affecting them mentally, prompting loneliness, misery, and even rashness. Dermal diseases should be

✉ Ashwani Kumar Dubey
dubeylak@gmail.com

Ginni Arora
garora@amity.edu

Zainul Abdin Jaffery
zjaffery@jmi.ac.in

Alvaro Rocha
amr@iseg.ulisboa.pt

¹ Amity Institute of Information Technology, Amity University Uttar Pradesh, Noida, UP 201313, India

² Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, UP 201313, India

³ Department of Electrical Engineering, Jamia Millia Islamia, New Delhi 110025, India

⁴ ISEG, University of Lisbon, Rua Do Quelhas, N° 6, 1200-781 Lisboa, Portugal

treated at the beginning to decrease related outcomes, costs, mortality, and morbidity rates. As indicated by Dr Macrene Alexiades-Armenakas, cancer and eczema are among the five most basic skin syndromes [1]. In this manner, our fundamental emphasis is on building a system that can automatically diagnose for analyses and grouping these types of diseases.

The visual assessment with the naked eye during a skin disease health check-up makes it hard to contrast skin injuries and typical tissue, producing a misdiagnosis. Indeed, dermoscopy is a mainly solid conceptual method for skin injuries. The input thought of dermoscopy is to obtain an enlarged high-resolution image while removing reflections from the skin's surface. The utilization of dermoscopy imaging tools improves the general perception of skin lesions, sensitivity (correct finding of disease) and specificity (correct finding of doubtful disease) contrasted with visual assessment; further, it needs improvement to analyze skin disease. However, since physical examination from dermoscopy images are frequently time-consuming, error-prone, complex, and personal (i.e., might create different diagnostic outcome). Hence, an automated and consistent computer-aided diagnostic (CAD) framework to recognize skin cancer growth has developed a significant evaluation device that offers dermatologists a subsequent input to help and aid their decisions [2].

There are numerous investigations on diagnosing skin disease utilizing deep learning and machine learning strategies in the literature. In recent times, complex medical problems are handled through deep learning convolutional neural networks (CNN) [3–6], particularly dermoscopic image analysis [7–11] to melanoma recognition. Overall, there are existing and tested convolutional neural network classifiers based on deep learning, for example, AlexNet [12], VGGNet [13], GoogLeNet [14], ResNet [15], and DenseNet [16]. These CNNs go further to expand for more accurately identifying the five challenging tasks [17, 18]. The research goes to a two-course model that utilizes a deep residual network to segment and classify skin lesions [11]. The method of moving a portioned mask of skin lesions, alongside descriptions of clinical rules, such as, shading, texture, and morphological qualities, to the CNN through U-Net [19] has been used. Later, it introduced a two-advance model for the classification of skin lesions [20]. Shading, texture, and shape qualities were removed from images through U-Net and sent to a Support Vector Machine (SVM) classifier to analyze favourable or dangerous dermoscopy pictures. The hybrid system for classifying the skin lesion is introduced, a blend of CNN and SVM with sparse coding(SC) [21]. In recent

years, researchers have also built a deep learning network based on a hybrid approach through encoding local descriptors [11]. The various features, including deep and statistical from ResNet, utilized the SVM classifier and the chi-square kernel to recognize the distinctive skin lesions. The usage of both handcrafted through traditional approach and learned features using deep learning with ResNet-50 for improvement in classification was proposed. It consolidates high-quality conventional capacities through picture handling and ResNet-50 profound, getting the hang of learning capacities [8]. In the latest, the authors detect seven classes of cancer using a support vector machine with an artificial bee colony method. RNA sequencing is also utilized [9].

Apart from the literature, we have classified the seven different skin diseases, including Actinic keratoses and intraepithelial carcinoma, Basal cell carcinoma, Benign keratosis, Dermatofibroma, Melanoma, Melanocytic type, and Vascular lesions, using fourteen types of deep learning networks. The objective of this paper is to evaluate the performance of these networks based on performance measures. These networks are trained, validated, and tested on ISIC 2018 dataset using Matlab.

In the following, we list contributions of this study:

1. We used transfer learning by adding one fully connected layer, softmax layer and cross-entropy layer as pre-trained networks. These three layers in total can classify skin lesions into seven classes.
2. To improve the classification performance, we added pre-processing steps as data normalization, resizing and augmentation.
3. The comparison of fourteen types of deep learning networks is analysed to build an automated classification model.

The remainder of this paper is outlined as follows. Section 2 presents the methodology with description of dataset and detail of fourteen deep learning models. Section 3 shows the experiments and results with implementation. Section 4 deliberates the performance of deep learning networks for classification of multiple skin lesion. Finally, the paper ends with a conclusion in Sect. 5.

2 Methodology

This section briefly explains the dataset description, its pre-processing and different deep convolutional neural networks. Then, all these networks are compared using the same number of epochs, learning rate and batch size.

2.1 Dataset

This study used the ISIC 2018 dataset for dermoscopy images [22, 23]. It contains pigmented lesions in different populations. This data set consists of 10,015 images in total. The image sizes are scaled to $224 \times 224 \times 3$ or $227 \times 227 \times 3$ or as per the network requirement. The diagnostic classes in this dataset are given as follows [22, 23].

- (1) Actinic keratoses and intraepithelial carcinoma abbreviated as 'akiec'.
- (2) Basal cell carcinoma abbreviated as 'bcc'.
- (3) Benign keratosis abbreviated as 'bkl'.
- (4) Dermatofibroma abbreviated as 'df'.
- (5) Melanoma abbreviated as 'mel'.
- (6) Melanocytic type abbreviated as 'nv'.
- (7) Vascular lesions abbreviated as 'vasc'.

Figure 1 illustrate the examples for akiec. Figure 2 shows the examples for bcc. Figure 3 shows the examples for bkl. Figure 4 shows the examples for df. Figure 5 shows the examples for mel. Figure 6 shows the examples for nv. Figure 7 shows the examples for vasc. All examples are taken from [22, 23].

The number of data contained in image set for each class is defined in Table 1.

As shown in Table 1, most of the data in the dataset belong to the sixth grade, and there is no equal distribution between classes. Therefore, the training phase will lead to more learning of the 6th class and the network to learn this information.

2.2 Data pre-processing

In our approach, the pre-processing steps are reduced to guarantee better generalizability when tried on the dermoscopic skin lesion dataset. In this way, just the three standard pre-processing steps are applied, generally utilized for transfer learning. Initially, the images are normalized by taking away the typical RGB values from the dataset. After that, the images are resized utilizing bicubic interpolation for the networks input (227×227 and 224×224). At last, the training set is augmented by arbitrarily flipping the training images along the vertical axis and changing it up to 30 pixels horizontally and

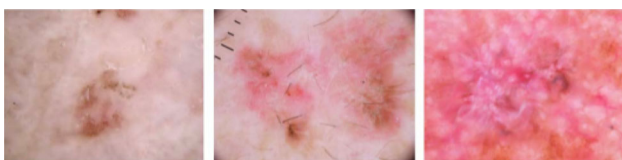


Fig. 1 The dataset for akiec [22, 23]

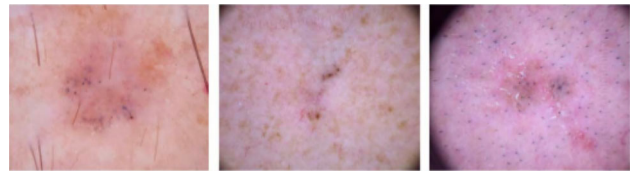


Fig. 2 The dataset for bcc [22, 23]



Fig. 3 The dataset for bkl [22, 23]

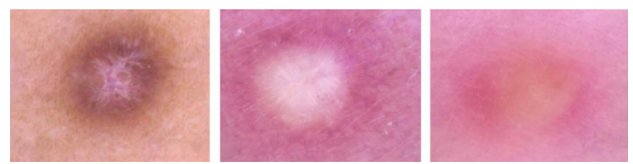


Fig. 4 The dataset for df [22, 23]

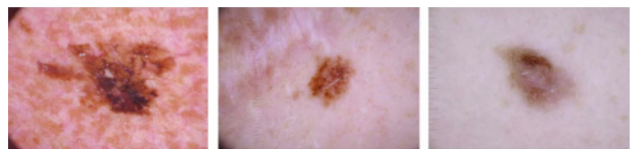


Fig. 5 The dataset for mel [22, 23]

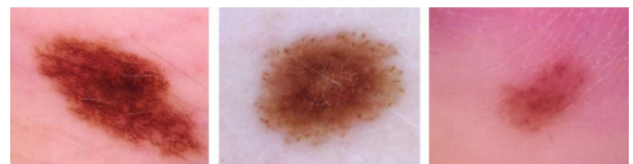


Fig. 6 The dataset for nv [22, 23]



Fig. 7 The dataset for vasc [22, 23]

vertically. Data augmentation keeps the network away from overfitting and memorizing the exact details of the training images. The block diagram of the classification model is shown in Fig. 8.

Table 1 Data indicating how many data each set contains in the dataset

Diagnosis class in the dataset	Total	Train	Validate and test
Akiec	327	229	98
Bcc	514	360	154
Bkl	1099	769	330
Df	115	81	34
Mel	1113	779	334
Nv	6705	4694	2011
Vasc	142	99	43
Total	10,015	7011	3004

2.3 AlexNet

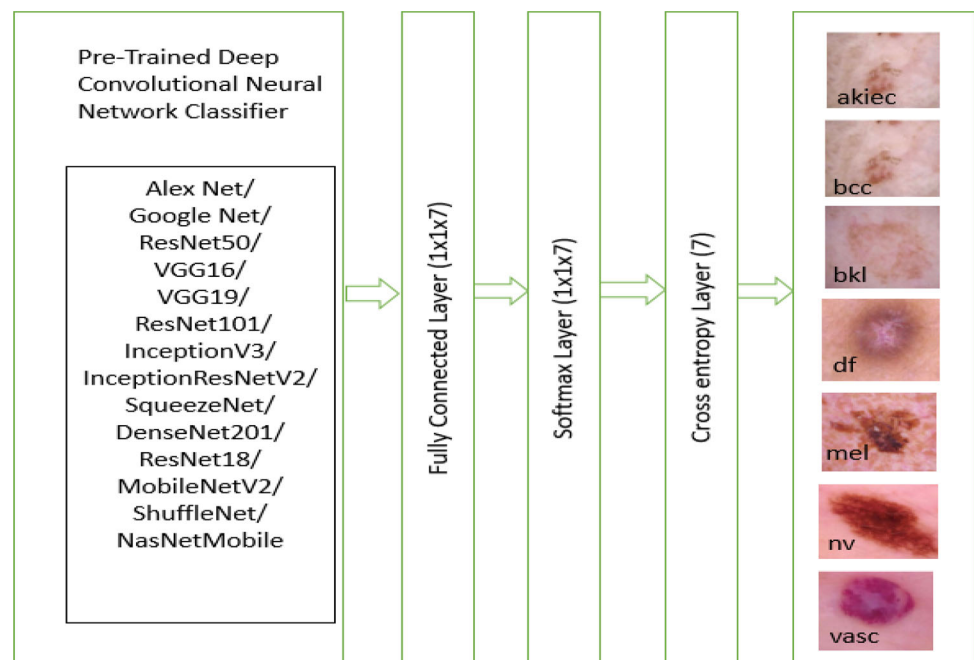
This network contains 8 learned layers (5 convolutional and 3 fully connected layers) [24, 25]. The convolution filters and an activation non-linear function (ReLU) is applied after every layer. Due to fully connected layers, the input size is fixed to RGB images of $224 \times 224 \times 3$, i.e., 150,528 values. In the FCN and cross-entropy layer, there are 4096 neurons; each neuron is a different feature of the image. Dropouts are utilized at regular intervals to avoid overfitting issues.

2.4 GoogleNet

The GoogleNet architecture comprises 22 convolutional layers containing 9 Inception modules [26]. The Inception module has three kernel varieties with various sizes: 5×5 , 3×3 and 1×1 for convolution and 3×3 filters for pooling. These small convolutions help reduce the depth of the feature map, and many inception modules also helps reduce the number of computations. The receptive field size for this network is $224 \times 224 \times 3$ utilizing the RGB colors gap with the predefined parameter. Like different CNNs, this network also learn convolutional filters contributions through stochastic gradient descent (SGD) algorithms throughout the training stage to retrieve compelling features as the image crosses the hierarchical structure of the network.

2.5 ResNet50

This network consists of a stem used for input, followed by four stages and an output layer [27]. It consists of 50 layers and an input image size fixed to 224×224 . The stem is input through a 7×7 convolution by a return of 64 channels besides a pitch of 2, trailed through a 3×3 max pooling layer besides a pitch of 2. The width and height of the input decrease by stem through a factor of 4, and rises a size of channel to 64. The individual stage starts with a downsampling block, further trailed by various residual blocks at the starting of stage2. This block consists of path A and path B. Path A consists of three convolutions with a kernel size of 1×1 , 3×3 , and 1×1 . Path B utilizes a

Fig. 8 Block diagram of Classification model

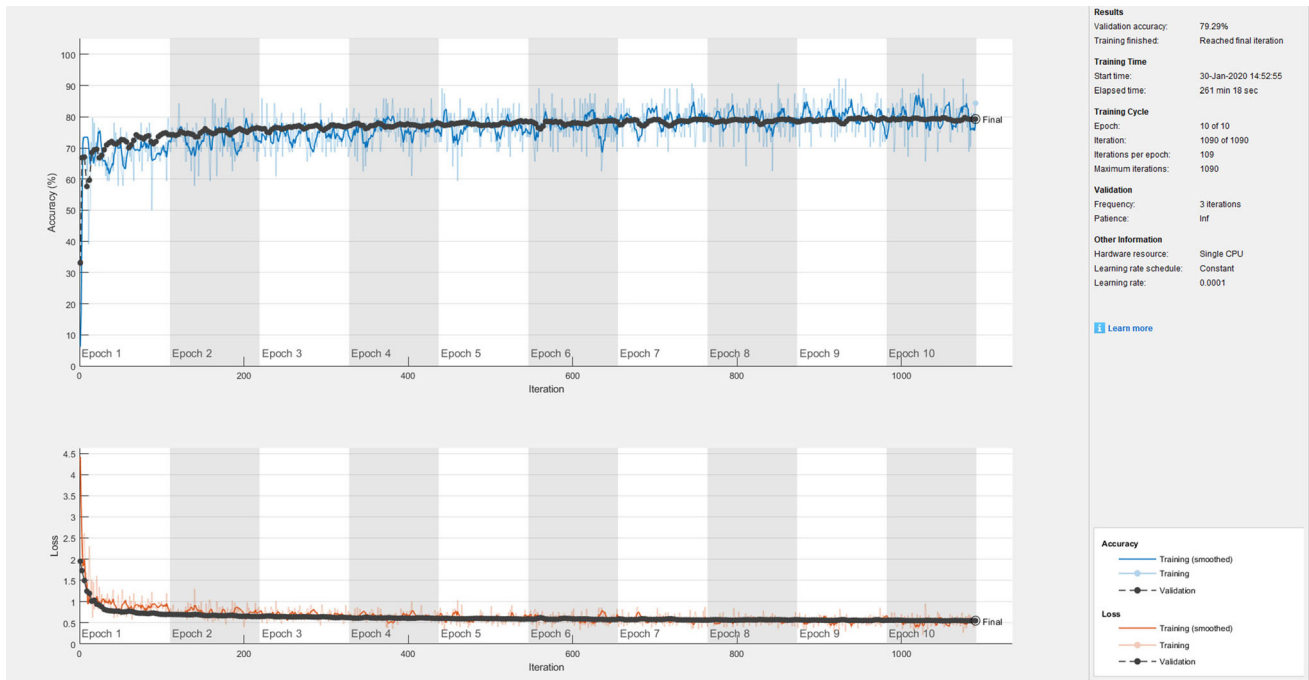


Fig. 9 AlexNet training Progress with accuracy and error loss over 10 Epochs

Fig. 10 Confusion matrix of predictions made by transfer learning on AlexNet

True Class	AKIEC	42	17	18		8	11	2	42.9%	57.1%
	BCC	8	100	6	4	8	26	2	64.9%	35.1%
	BKL	11	9	149		60	99	2	45.2%	54.8%
	DF	2	4	3	3	4	17	1	8.8%	91.2%
	MEL	6	3	25	1	167	131	1	50.0%	50.0%
	NV	1	8	36	1	79	1884	2	93.7%	6.3%
	VASC		2			1	3	37	86.0%	14.0%
			60.0%	69.9%	62.9%	33.3%	51.1%	86.8%	78.7%	
		40.0%	30.1%	37.1%	66.7%	48.9%	13.2%	21.3%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

pitch of 2 with a 1×1 convolution to convert the input into the output of path A. Finally, the sum of these two paths is used to get the output of the downsampling block.

2.6 VGG16/VGG19

The VGG-16 architecture is reasonable for GPUs with local memory [28]. The architecture consists of 16 layers, among which there are 13 convolutional layers, 5 Max Pooling layers and 3 fully connected layers, which sums up to 21 layers but only 16 weight layers. The information

layer is thought to be a $224 \times 224 \times 3$ pixel RGB picture. All convolutional layers are followed by the Relu function with a size of 3×3 . The VGG19 network [29] contains 19 trained layers combining convolutional and FC layers, max-pooling, and dropout layers.

2.7 ResNet101

The ResNet101 network structure consists of residual connections and generally used for classification purposes. The gradients flow throughout the layers directly and

Table 2 Performance measures of AlexNet

Type of disease	Recall	Precision	F1 Score
Akiec	0.438	0.60	0.506
Bcc	0.649	0.699	0.673
Bkl	0.451	0.629	0.525
Df	0.882	0.333	0.483
Mel	0.50	0.511	0.505
Nv	0.937	0.868	0.901
Vasc	0.86	0.787	0.411
Overall accuracy	0.793		

prevent gradient as zero after the chain rule applications [30]. A total of 104 convolutional layers are present in the network. Alongside, there are 33 blocks of layers, and the output of the previous block is used among these 29 blocks directly as residual connections above. To receive the input of the other blocks, these residual connections are being utilized at the termination of every block using the initial value of the sigma operator.

2.8 InceptionV3

This network is based on inception modules and consists of 48 layers deep [31]. These inception blocks contain convolutions corresponding with diverse kernel sizes to abstract features and finally summative the results. Initially, the input image is passed through convolution, batch normalization, Relu and this sequence are followed by pooling and various inception layers for feature extraction. Finally, in the classification part, the dropout layer is

applied to reduce overfitting, and softmax with cross-entropy as output layers.

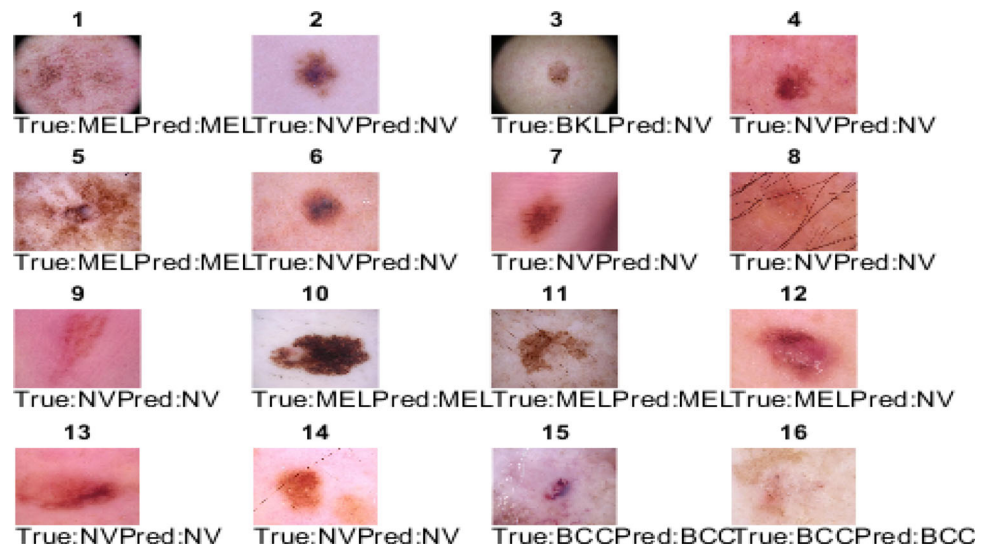
2.9 InceptionResNetV2

This network [32] consists of 164 layers in deep and inception blocks with residual connections. The input size of an image acceptable to the network is of size 299×299 . The network consists of the main module as stem, followed by Inception resnet-A, Reduction-A, Inception resnet-B module, reduction B and inception resnet-C modules. The stem module consists of series of 3×3 convolution, 3×3 maxpool, and filter concat to give input to inception resnet-A block. The Relu activation function is applied in each inception block and the reduction block contains 1×1 , 3×3 convolution with a pooling layer. Inception-ResNet used batch-normalization only on top of the traditional layers but not on top of the summations.

2.9.1 SqueezeNet

This network starts with [33] a separate convolution layer (conv1), trailed by 8 Fire modules (fire2-9), and termination with a final convolution layer (conv10). The filters numbers get increased as per fire module from starting till the end of the network. The fire module is composed with a squeezed convolution layer of 1×1 filters only, serving into an enlarge layer with a combination of 1×1 and 3×3 size convolution filters. The maxpooling operation is applied after convolution1, fire4, fire8, and convolution10 layers.

Fig. 11 Predictions on random samples made by transfer learning on AlexNet



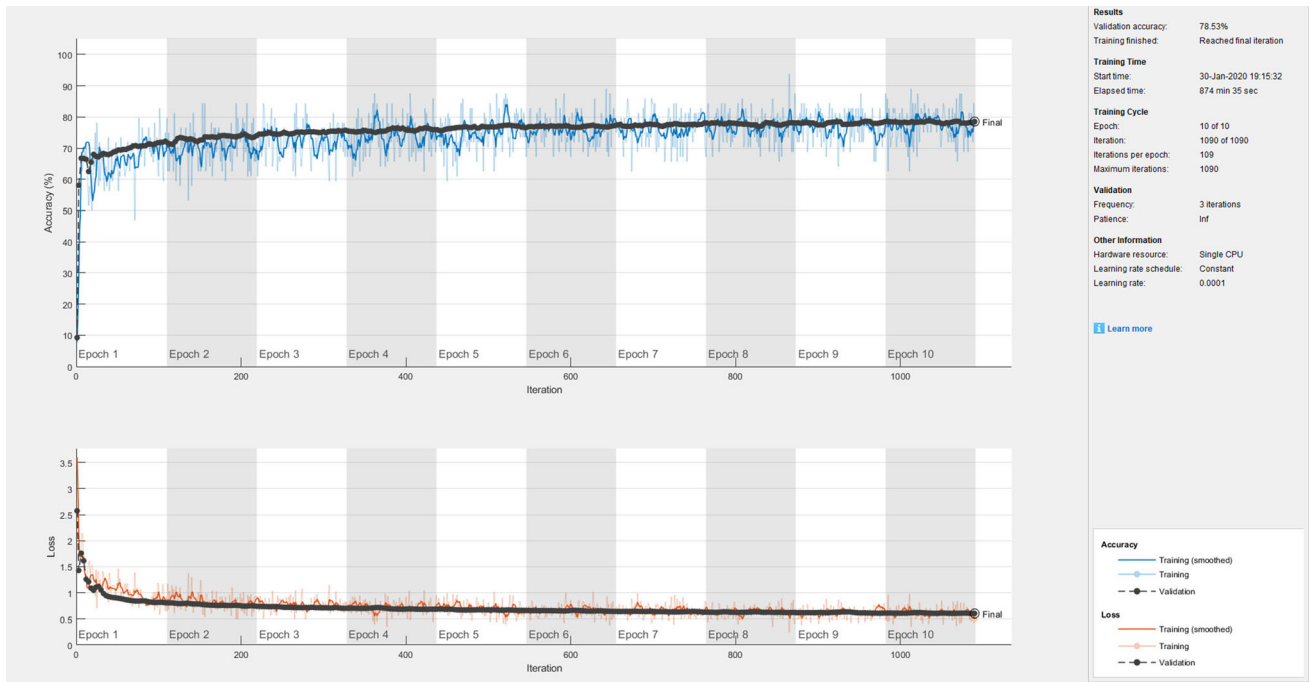


Fig. 12 GoogleNet training progress with accuracy and error loss over 10 Epochs

Fig. 13 Confusion matrix of predictions made by transfer learning on GoogleNet

True Class	AKIEC	BCC	BKL	DF	MEL	NV	VASC		
AKIEC	42	16	20		8	12		42.9%	57.1%
BCC	11	102	14		8	17	2	66.2%	33.8%
BKL	16	17	155		37	103	2	47.0%	53.0%
DF	1	2	9		3	19			100.0%
MEL	12	10	29		119	162	2	35.6%	64.4%
NV	3	17	37		38	1911	5	95.0%	5.0%
VASC		5	1			7	30	69.8%	30.2%
	49.4%	60.4%	58.5%		55.9%	85.7%	73.2%		
	50.6%	39.6%	41.5%		44.1%	14.3%	26.8%		
	AKIEC	BCC	BKL	DF	MEL	NV	VASC		

Table 3 Performance measures of GoogleNet

Type of disease	Recall	Precision	F1 Score
Akiec	0.428	0.494	0.458
Bcc	0.708	0.603	0.651
Bkl	0.469	0.585	0.521
Df	0	0	0
Mel	0.356	0.559	2
Nv	0.950	0.860	0.903
Vasc	0.698	0.732	0.714
Overall accuracy	0.785		

2.9.2 DenseNet201

The connections are directly between the convolutional layers in this network (DenseNet) [34, 35]. In every network layer, the feature map of the continuing layer acts as input, whereas feature maps that are generated are used by way of inputs to the next layer. This network consists of four blocks with layers in the same amount. These blocks have feature maps with various sizes of 56-by-56, 28-by-28, 14-by-14, and 7-by-7. The number of input feature maps are reduced to advance the computational efficiency by using 1 × 1 convolution before each 3 × 3 convolution, known as the bottleneck layer.

Fig. 14 Predictions on random samples made by transfer learning on GoogleNet

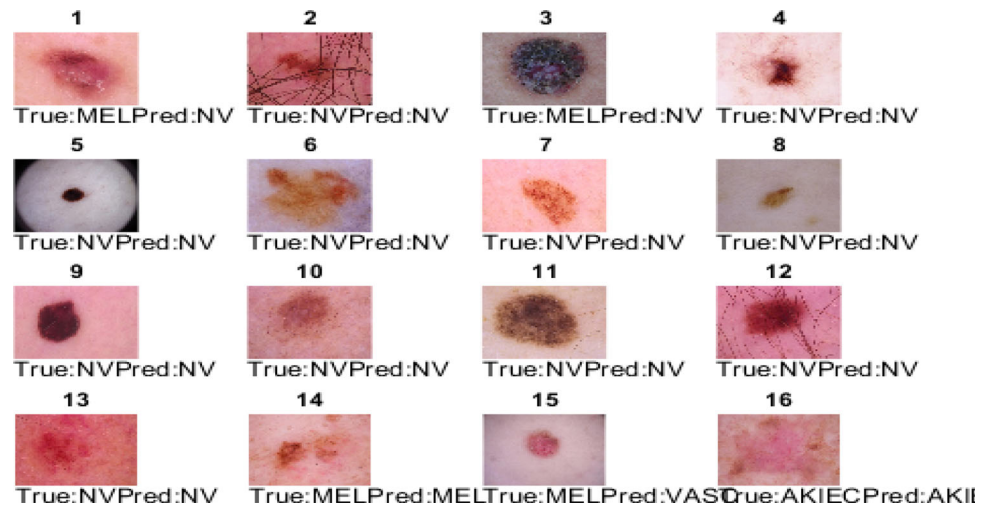


Fig. 15 ResNet50 training with accuracy and error loss over 10 Epochs

2.9.3 ResNet18

This network starts with 7×7 convolution, 3×3 max pooling, followed by 4 convolutional layers in each module [36]. There are total 18 layers which include the first convolution layer and the last FC layer. This network has an image input size of 224-by-224. The initial 2 layers of this network are similar to the GoogLeNet: the 64 output channels with stride 2 based 7×7 convolutional layer trailed by 3×3 max pooling layer, stride 2 and Relu function. The change is the batch wise normalization layer additional next to every convolutional layer in the network.

It consists of residual blocks with 3×3 convolutional layers having a similar amount of output channels.

2.9.4 Mobile NetV2

The MobileNetV2 model [37] is based on the concept of separable convolutions, which can be depth wise (dw) or pointwise (pw) convolution. In the case of the former, a filter is applied to every channel used for input. Then to combine an output of former 1×1 convolution is applied by later convolution. A one-step process is required to get the original block of outputs by filtering and pooling together by convolution. The process is separated in two

Fig. 16 Confusion matrix of predictions made by transfer learning on ResNet50

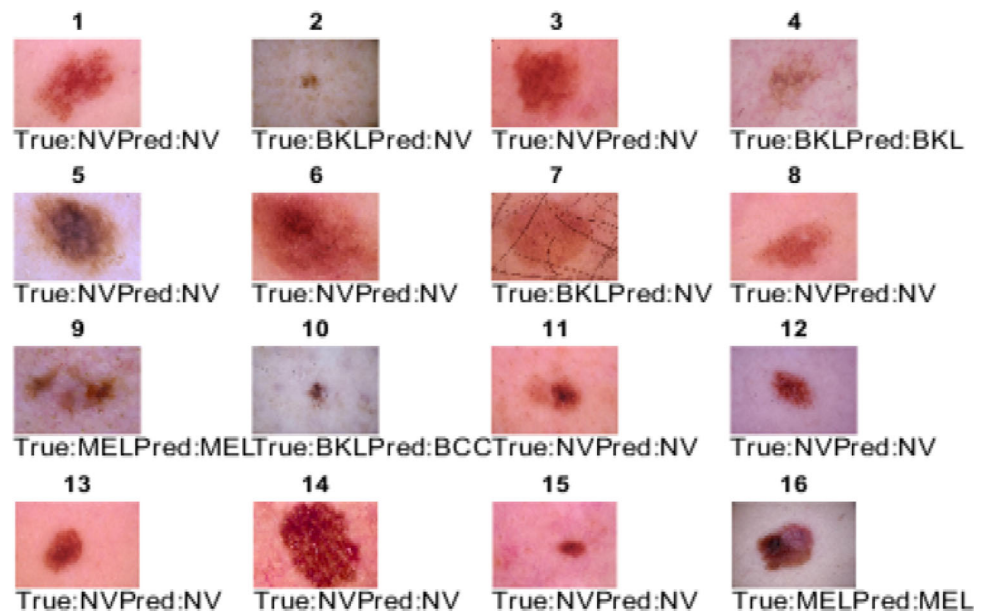
True Class	AKIEC	49	14	20	2	5	8		50.0%	50.0%
	BCC	9	99	17	2	9	17	1	64.3%	35.7%
	BKL	8	8	209	2	30	73		63.3%	36.7%
	DF	2	1	5	3	5	18		8.8%	91.2%
	MEL	4	4	33		149	143	1	44.6%	55.4%
	NV	1	9	48	1	37	1913	2	95.1%	4.9%
	VASC		2	1			10	30	69.8%	30.2%
			67.1%	72.3%	62.8%	30.0%	63.4%	87.7%	88.2%	
		32.9%	27.7%	37.2%	70.0%	36.6%	12.3%	11.8%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC	Predicted Class	

Table 4 Performance measures of ResNet50

Type of disease	Recall	Precision	F1 Score
Akiec	0.50	0.671	0.573
Bcc	0.643	0.723	0.340
Bkl	0.633	0.628	0.630
Df	0.882	0.30	0.448
Mel	0.446	0.634	0.524
Nv	0.951	0.877	0.912
Vasc	0.698	0.882	0.780
Overall accuracy	0.816		

layers, filtering and combining. This factorization results into significantly reducing computation and size.

Fig. 17 Predictions on random samples made by transfer learning on ResNet50



2.9.5 ShuffleNet

The ShuffleNet [38] utilizes pointwise group convolution and channel shuffle to reduce computation cost while maintaining accuracy. By shuffling the channels, ShuffleNet outperforms MobileNetV1. The network contains a heap of network units assembled in three stages. It has two units as bottleneck unit through depthwise convolution and shufflenet unit through pointwise group convolution.

2.9.6 NasNetMobile

Nasnet [39] is an architecture constitute of basic building blocks i.e. cells optimized through reinforcement learning. The network has an image input size of 224-by-224. A cell composed of convolution and pooling operations, depends upon the strength of the network to repeat these operations

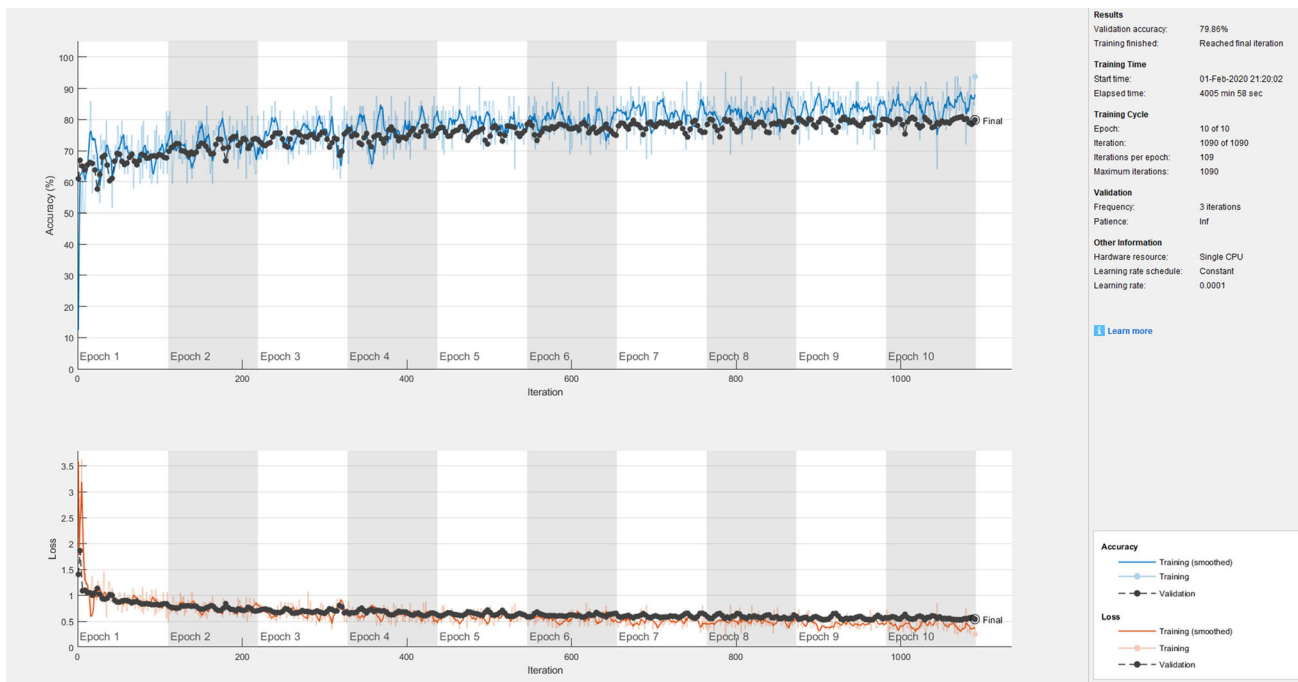


Fig. 18 VGG16 training with accuracy and error loss over 10 epochs

Fig. 19 Confusion matrix of predictions made by transfer learning on VGG16

True Class	AKIEC	45	17	13		18	5		45.9%	54.1%
	BCC	9	115	4		14	8	4	74.7%	25.3%
	BKL	16	9	165	1	80	59		50.0%	50.0%
	DF	3	3	7	3	9	9		8.8%	91.2%
	MEL	5	2	15		219	88	5	65.6%	34.4%
	NV	2	11	37	1	142	1814	4	90.2%	9.8%
	VASC		2			1	2	38	88.4%	11.6%
			56.3%	72.3%	68.5%	60.0%	45.3%	91.4%	74.5%	
		43.8%	27.7%	31.5%	40.0%	54.7%	8.6%	25.5%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

Table 5 Performance measures of VGG16

Type of disease	Recall	Precision	F1 Score
Akiec	0.459	0.562	0.505
Bcc	0.799	0.723	0.759
Bkl	0.50	0.685	0.578
Df	0.882	0.60	0.714
Mel	0.655	0.463	0.542
Nv	0.902	0.914	0.908
Vasc	0.884	0.745	0.808
Overall accuracy	0.799		

at n number of times. This network contains 12 cells and millions of multiply-accumulates (MACs). There are two cells: Normal cell, which returns a feature map of the same dimension and Reduction cell, which returns a feature map where the feature map height and width are reduced by a factor of two. The various combinations of operations or layers applied to the network in the form of identity, convolution (1 × 7, 7 × 1), average pooling (3 × 3), max pooling (5 × 5), depth wise convolution (3 × 3, 7 × 7, 5 × 5) and dilated convolution (3 × 3).

Fig. 20 Predictions on random samples made by transfer learning on VGG16

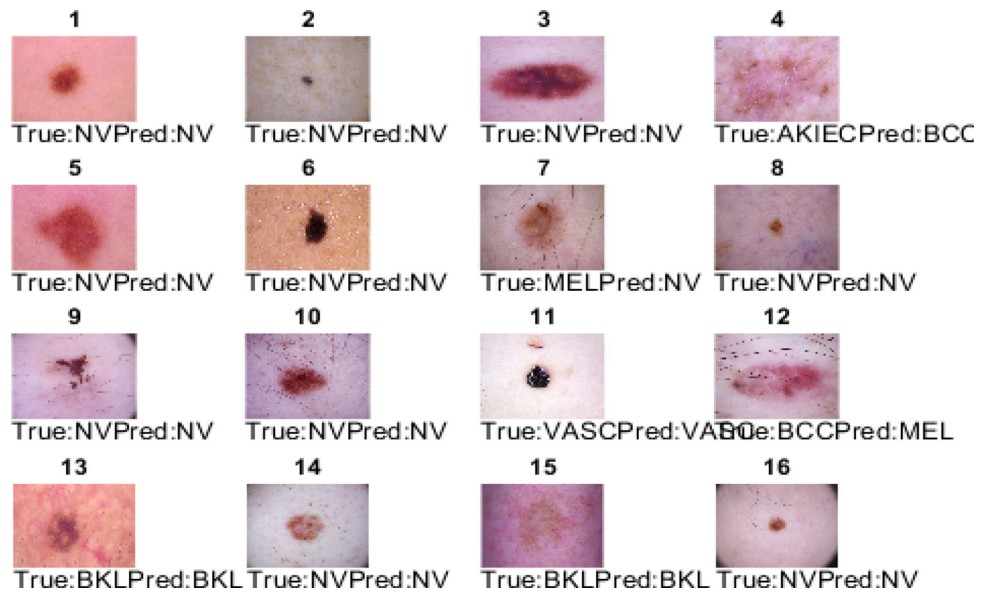


Fig. 21 VGG19 training with accuracy and error loss over 10 epochs

3 Experiments and results

The experiments are performed with Dell compatible computer equipped with a Core i7 processor. The ISIC 2018 dataset defined in RGB space is used. The datasets are divided into seven classes, Actinic keratoses and intraepithelial carcinoma, Basal cell carcinoma, Benign keratosis, Dermatofibroma, Melanoma, Melanocytic type, and Vascular lesions. The fourteen types of convolution deep

learning networks are used; in which the last fully connected layer is replaced with FC7, softmax and cross-entropy for classification. The experiments are performed with all types of networks using 10,015 dermoscopic images.

Each experiment consisted of the number of runs where the batch size, epochs and the initial learning rate were 64, 10 and 0.0001, respectively, and these values are fixed for all experiments. The images are randomly divided: 70% of

Fig. 22 Confusion matrix of predictions made by transfer learning on VGG19

True Class	AKIEC	43	10	20		21	4		43.9%	56.1%
	BCC	14	98	10		21	10	1	63.6%	36.4%
	BKL	9	10	195		65	51		59.1%	40.9%
	DF	1	1	5	5	10	12		14.7%	85.3%
	MEL	5		18		221	89	1	66.2%	33.8%
	NV	4	8	43		120	1835	1	91.2%	8.8%
	VASC		2			1	3	37	86.0%	14.0%
			56.6%	76.0%	67.0%	100.0%	48.1%	91.6%	92.5%	
		43.4%	24.0%	33.0%		51.9%	8.4%	7.5%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

Table 6 Performance measures of VGG19

Type of disease	Recall	Precision	F1 Score
Akiec	0.439	0.566	0.494
Bcc	0.636	0.760	0.692
Bkl	0.60	0.670	0.633
Df	0.147	0.100	0.119
Mel	0.662	0.481	0.557
Nv	0.912	0.916	0.914
Vasc	0.860	0.925	0.891
Overall accuracy	0.810		

the images are used for training and 30% are used for validation and testing. All the images are pre-processed using re-sizing, normalization, and augmentation. The networks are evaluated using the following metric described in Eqs. 1, 2, 3 and 4:

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + fn + tn} \tag{1}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{2}$$

$$\text{Precision} = \frac{tp}{tp + fp} \tag{3}$$

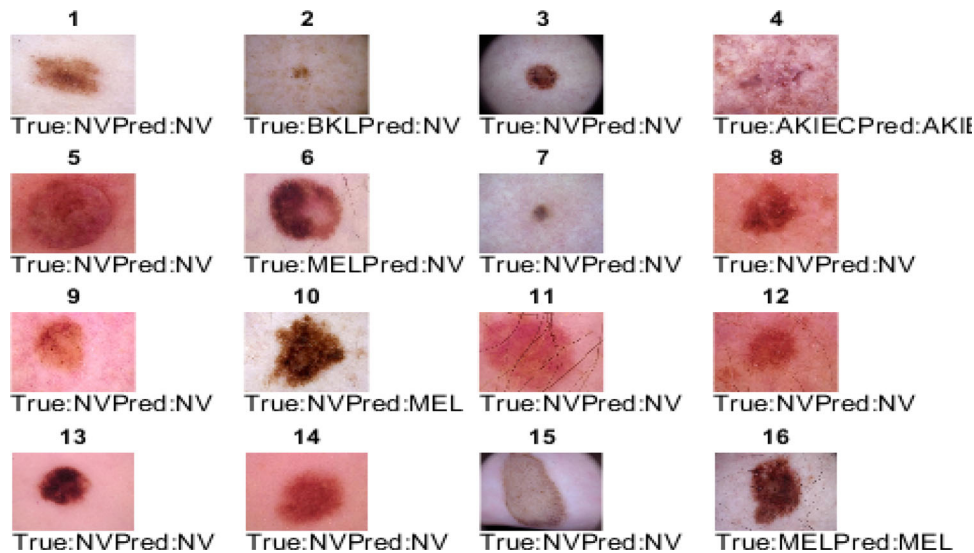
$$F1 \text{ score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \tag{4}$$

where, tp, tn, fp and fn refer to true positive, true negative, false positive and false negative, respectively.

The alexnet training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 9. The performance of alexnet in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 10 and Table 2, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on AlexNet is shown in Fig. 11.

The GoogleNet training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 12. The performance of GoogleNet in which class wise recall,

Fig. 23 Predictions on random samples made by transfer learning on VGG19



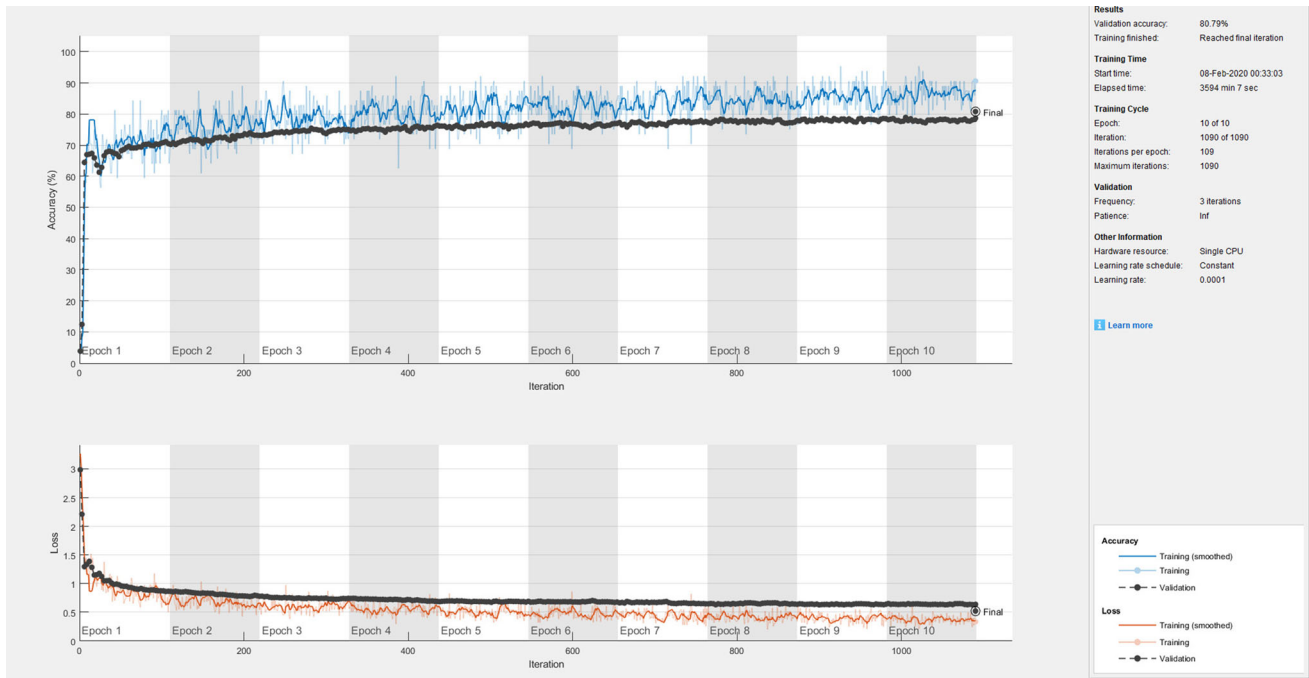


Fig. 24 ResNet101 training with accuracy and error loss over 10 epochs

Fig. 25 Confusion matrix of predictions made by transfer learning on ResNet101

True Class	AKIEC	50	16	19		11	2		51.0%	49.0%	
	BCC	11	113	13	1	7	7	2	73.4%	26.6%	
	BKL	13	16	223		34	44		67.6%	32.4%	
	DF	2	3	5	5	6	13		14.7%	85.3%	
	MEL	11	7	40	1	188	85	2	56.3%	43.7%	
	NV	1	24	70	3	91	1821	1	90.6%	9.4%	
	VASC	1	8				7	27	62.8%	37.2%	
			56.2%	60.4%	60.3%	50.0%	55.8%	92.0%	84.4%		
		43.8%	39.6%	39.7%	50.0%	44.2%	8.0%	15.6%			
		AKIEC	BCC	BKL	DF	MEL	NV	VASC			
		Predicted Class									

Table 7 Performance measures of ResNet101

Type of disease	Recall	Precision	F1 Score
Akiec	0.510	0.562	0.535
Bcc	0.734	0.604	0.663
Bkl	0.676	0.603	0.637
Df	0.147	0.50	0.227
Mel	0.563	0.558	0.560
Nv	0.905	0.920	0.912
Vasc	0.628	0.844	0.720
Overall accuracy	0.808		

precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 13 and Table 3, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on GoogleNet is shown in Fig. 14.

The ResNet50 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 15. The performance of ResNet50 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 16 and Table 4, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on ResNet50 is shown in Fig. 17.

Fig. 26 Predictions on random samples made by transfer learning on ResNet101

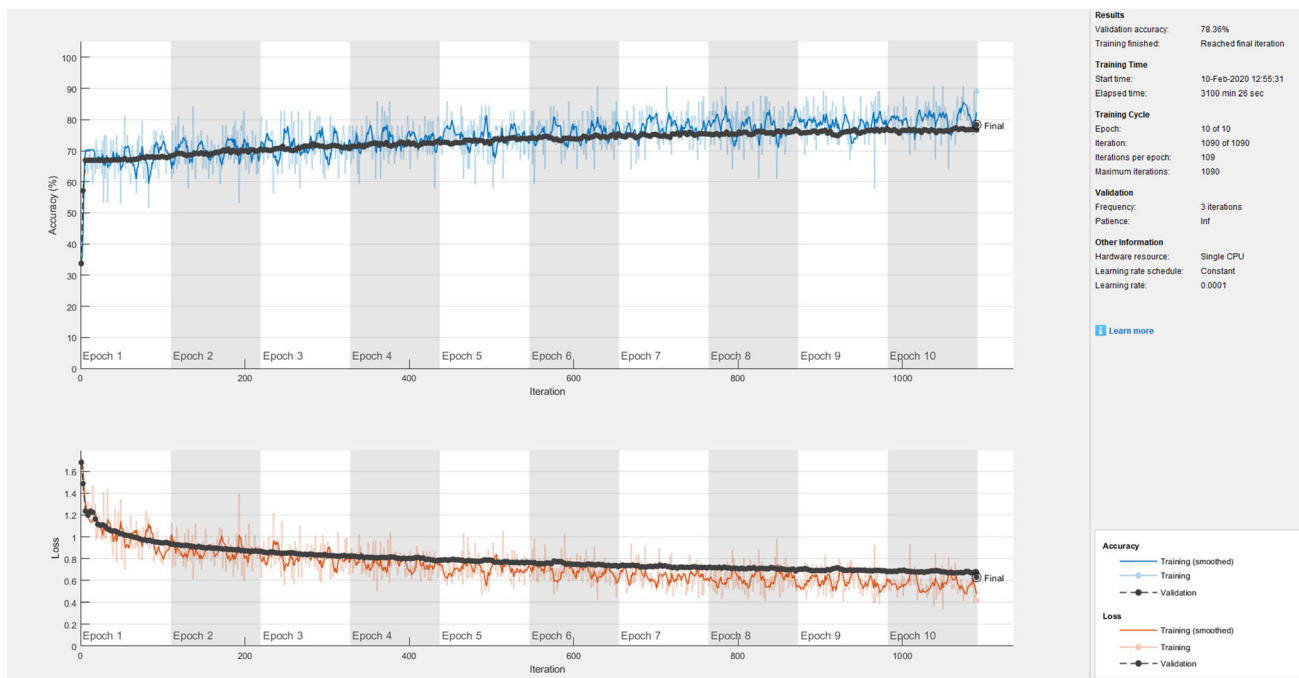
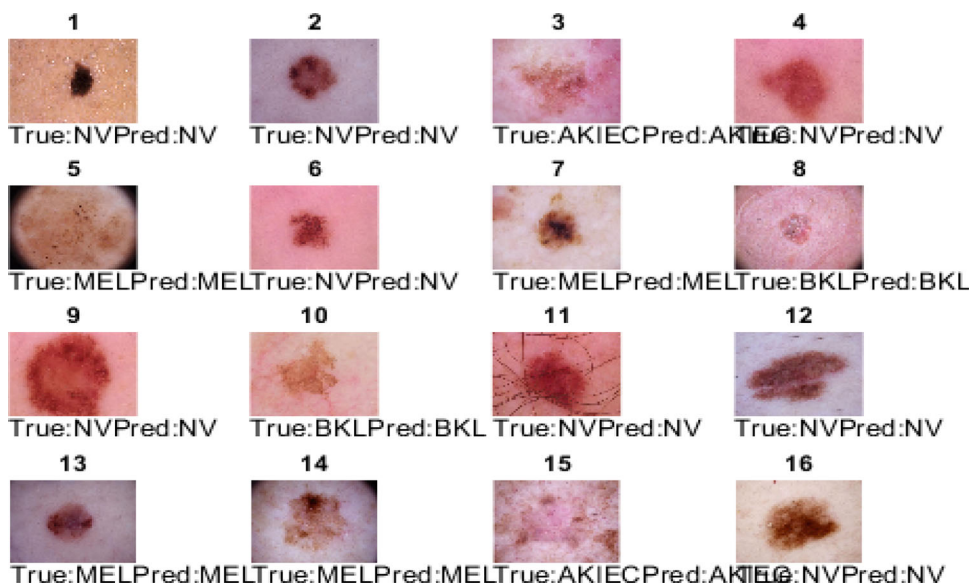


Fig. 27 InceptionV3 training with accuracy and error loss over 10 epochs

The VGG16 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 18. The performance of VGG16 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 19 and Table 5, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on VGG16 is shown in Fig. 20.

The VGG19 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 21. The performance

of VGG19 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 22 and Table 6, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on VGG19 is shown in Fig. 23.

The ResNet101 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 24. The performance of ResNet101 in which class wise recall, precision, f1 score and overall accuracy based on confusion

Fig. 28 Confusion matrix of predictions made by transfer learning on InceptionV3

True Class	AKIEC	26	29	25		12	6		26.5%	73.5%
	BCC	4	103	24		2	21		66.9%	33.1%
	BKL	8	20	195		30	77		59.1%	40.9%
	DF	1	2	6		3	22			100.0%
	MEL	5	12	46		127	144		38.0%	62.0%
	NV		10	52		54	1895		94.2%	5.8%
	VASC		1	3		2	29	8	18.6%	81.4%
			59.1%	58.2%	55.6%		55.2%	86.4%	100.0%	
		40.9%	41.8%	44.4%		44.8%	13.6%			
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

Table 8 Performance measures of inceptionV3

Type of disease	Recall	Precision	F1 Score
Akiec	0.25	0.571	0.348
Bcc	0.669	0.582	0.622
Bkl	0.60	0.555	0.577
Df	0	0	0
Mel	0.380	0.552	0.450
Nv	0.942	0.864	0.901
Vasc	0.186	0.100	0.130
Overall accuracy	0.784		

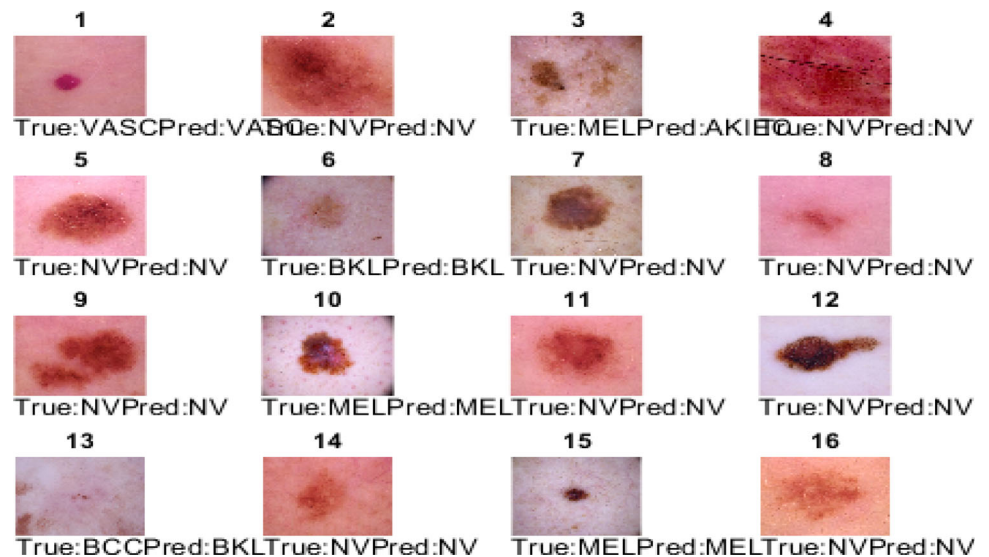
matrix are shown in Fig. 25 and Table 7, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on ResNet101 is shown in Fig. 26.

The InceptionV3 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 27. The performance of InceptionV3 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 28 and Table 8, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on InceptionV3 is shown in Fig. 29.

The InceptionResNetV2 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 30. The performance of InceptionResNetV2 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 31 and Table 9, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on InceptionResNetV2 is shown in Fig. 32.

The SqueezeNet training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 33. The

Fig. 29 Predictions on random samples made by transfer learning on InceptionV3



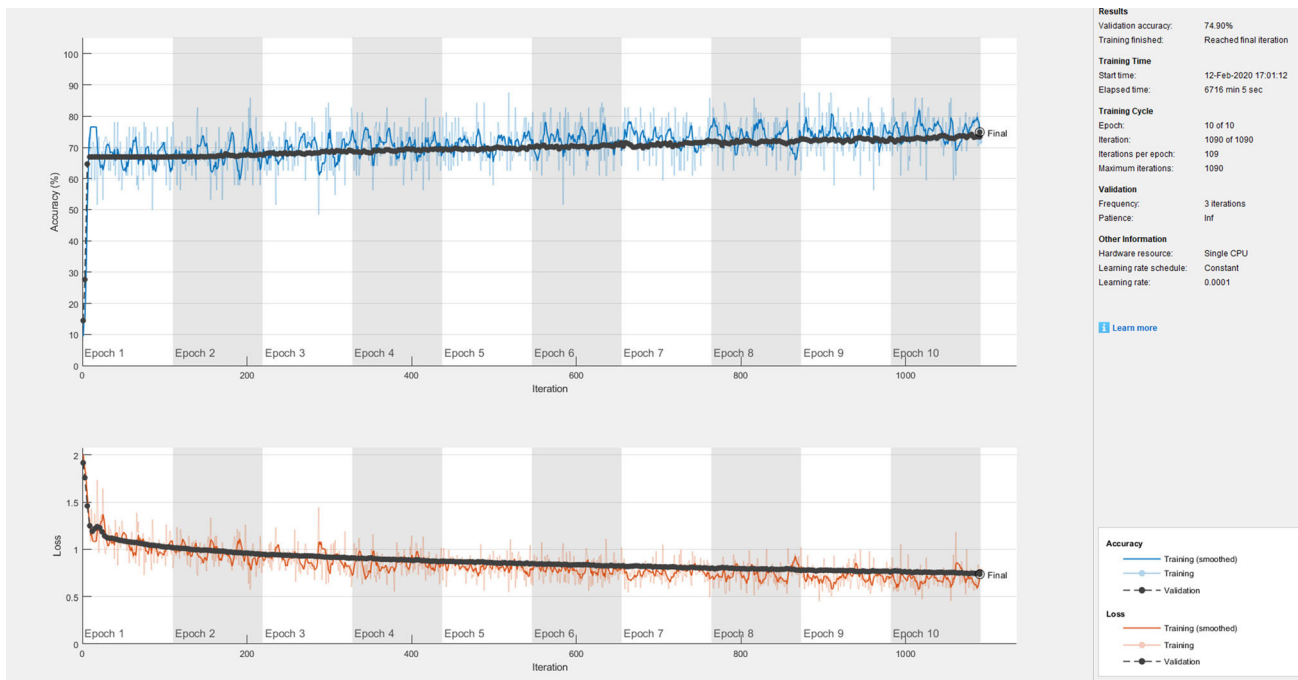


Fig. 30 InceptionResNetV2 training with accuracy and error loss over 10 epochs

Fig. 31 Confusion matrix of predictions made by transfer learning on InceptionResNetV2

True Class	AKIEC	11	5	59		12	11		11.2%	88.8%
	BCC	2	51	59		17	25		33.1%	66.9%
	BKL		4	206		41	79		62.4%	37.6%
	DF			8		5	21			100.0%
	MEL	1	2	71		123	137		36.8%	63.2%
	NV		5	83		64	1859		92.4%	7.6%
	VASC	1		8			34			100.0%
			73.3%	76.1%	41.7%		46.9%	85.8%		
		26.7%	23.9%	58.3%		53.1%	14.2%			
		AKIEC	BCC	BKL	DF	MEL	NV	VASC	Predicted Class	

Table 9 Performance measures of inceptionResNetV2

Type of disease	Recall	Precision	F1 Score
Akiec	0.112	0.733	0.194
Bcc	0.331	0.761	0.461
Bkl	0.624	0.417	0.500
Df	0	0	0
Mel	0.368	0.470	0.413
Nv	0.924	0.858	0.890
Vasc	0	0	0
Overall accuracy	0.750		

performance of SqueezeNet in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 34 and Table 10, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on SqueezeNet is shown in Fig. 35.

The DenseNet201 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 36. The performance of DenseNet201 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 37 and Table 11, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on DenseNet201 is shown in Fig. 38.

Fig. 32 Predictions on random samples made by transfer learning on InceptionResNetV2

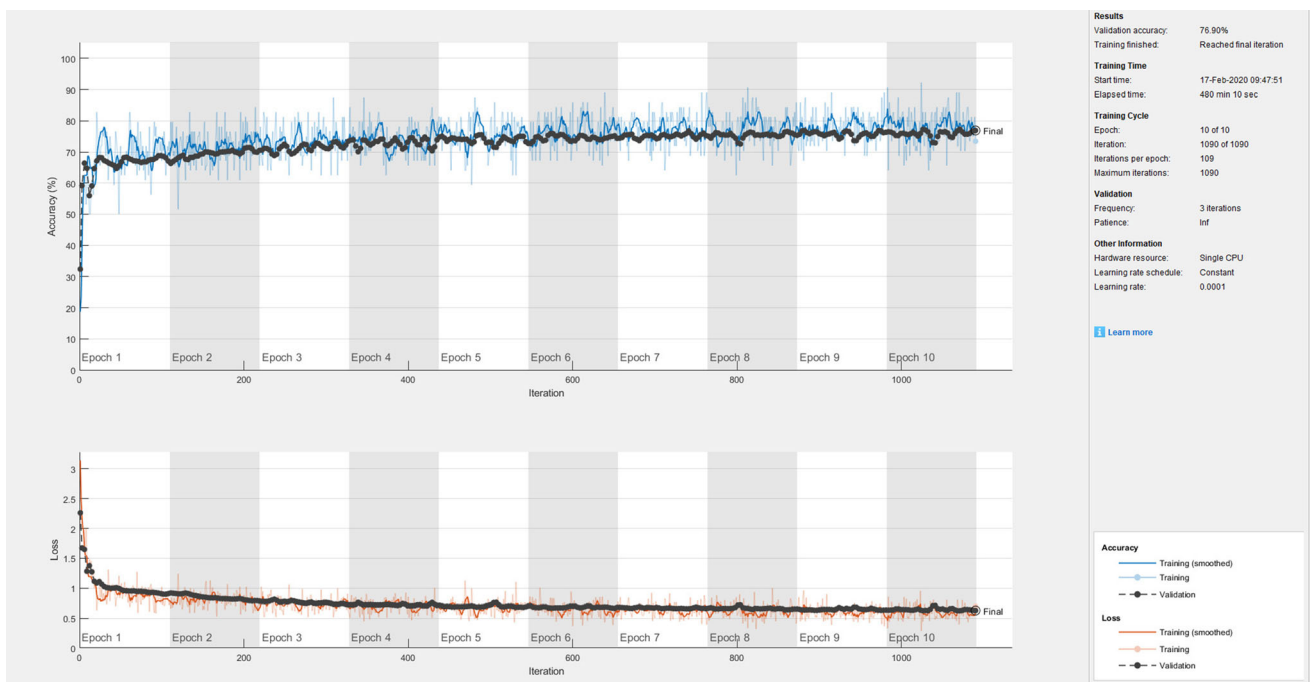
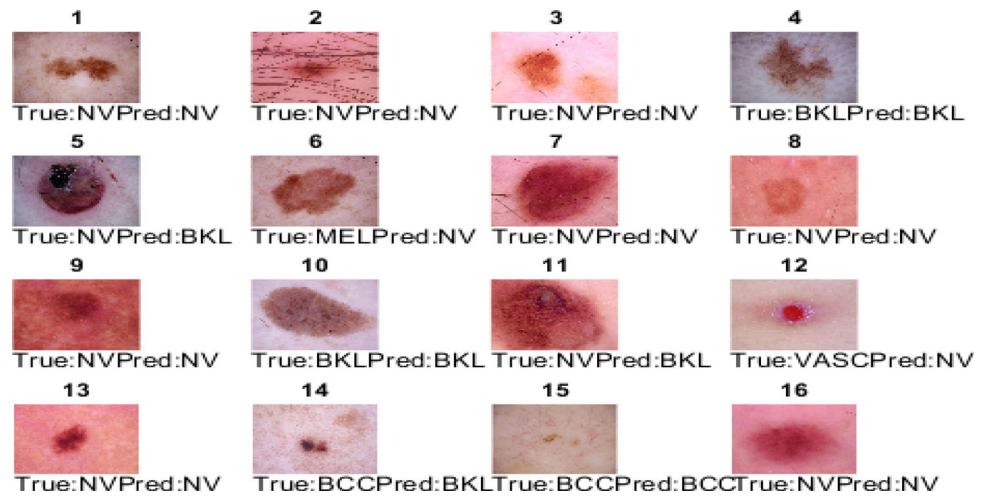


Fig. 33 SqueezeNet training with accuracy and error loss over 10 epochs

Fig. 34 Confusion matrix of predictions made by transfer learning on SqueezeNet

True Class	AKIEC	36	21	22		15	4		36.7%	63.3%
	BCC	12	102	17		6	15	2	66.2%	33.8%
	BKL	17	15	174		34	88	2	52.7%	47.3%
	DF	1	5	7		2	19			100.0%
	MEL	13	3	43		144	130	1	43.1%	56.9%
	NV	7	21	74		80	1826	3	90.8%	9.2%
	VASC		5	1			9	28	65.1%	34.9%
			41.9%	59.3%	51.5%		51.2%	87.3%	77.8%	
		58.1%	40.7%	48.5%		48.8%	12.7%	22.2%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC	Predicted Class	

Table 10 Performance measures of squeezeNet

Type of disease	Recall	Precision	F1 Score
Akiec	0.367	0.419	0.391
Bcc	0.662	0.593	0.626
Bkl	0.527	0.515	0.520
Df	0	0	0
Mel	0.431	0.512	0.468
Nv	0.908	0.873	0.890
Vasc	0.651	0.778	0.709
Overall accuracy	0.769		

The ResNet18 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 39. The performance of ResNet18 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 40 and Table 12, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on ResNet18 is shown in Fig. 41.

The MobileNetV2 training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 42. The performance of MobileNetV2 in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 43 and Table 13, the best result is highlighted in bold. The Predictions on Random samples

Fig. 35 Predictions on random samples made by transfer learning on SqueezeNet

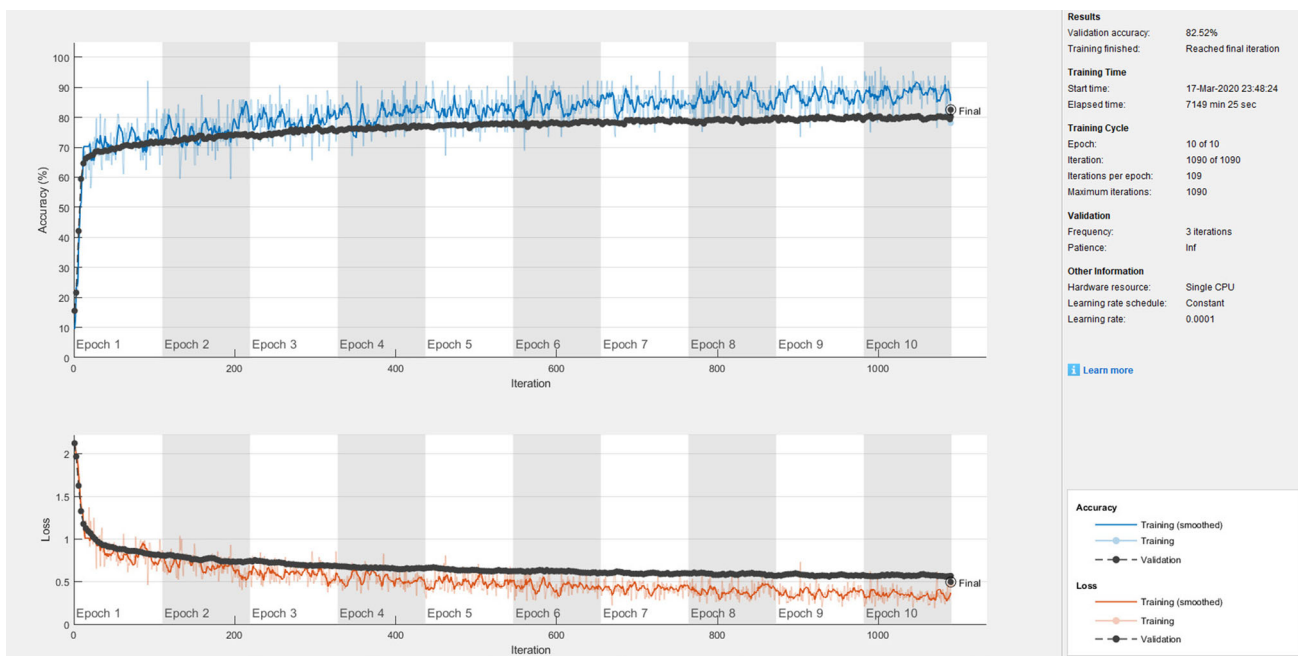
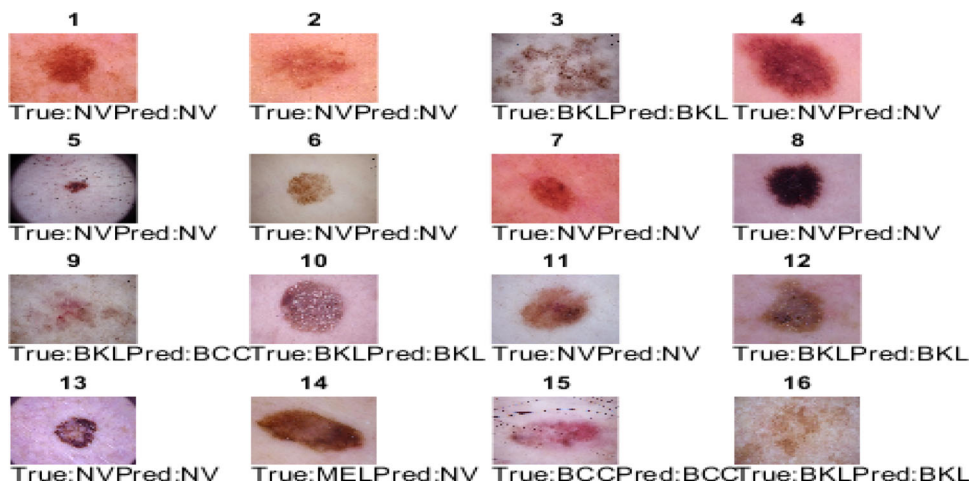


Fig. 36 DenseNet201 training with accuracy and error loss over 10 epochs

Fig. 37 Confusion matrix of predictions made by transfer learning on DenseNet201

True Class	AKIEC	50	19	18		7	4		51.0%	49.0%
	BCC	6	119	12	3	7	6	1	77.3%	22.7%
	BKL	8	12	231		35	44		70.0%	30.0%
	DF	2	2	5	7	7	11		20.6%	79.4%
	MEL	10	11	29		177	107		53.0%	47.0%
	NV	4	19	61	1	65	1860	1	92.5%	7.5%
	VASC		2	1			5	35	81.4%	18.6%
		62.5%	64.7%	64.7%	63.6%	59.4%	91.3%	94.6%		
		37.5%	35.3%	35.3%	36.4%	40.6%	8.7%	5.4%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

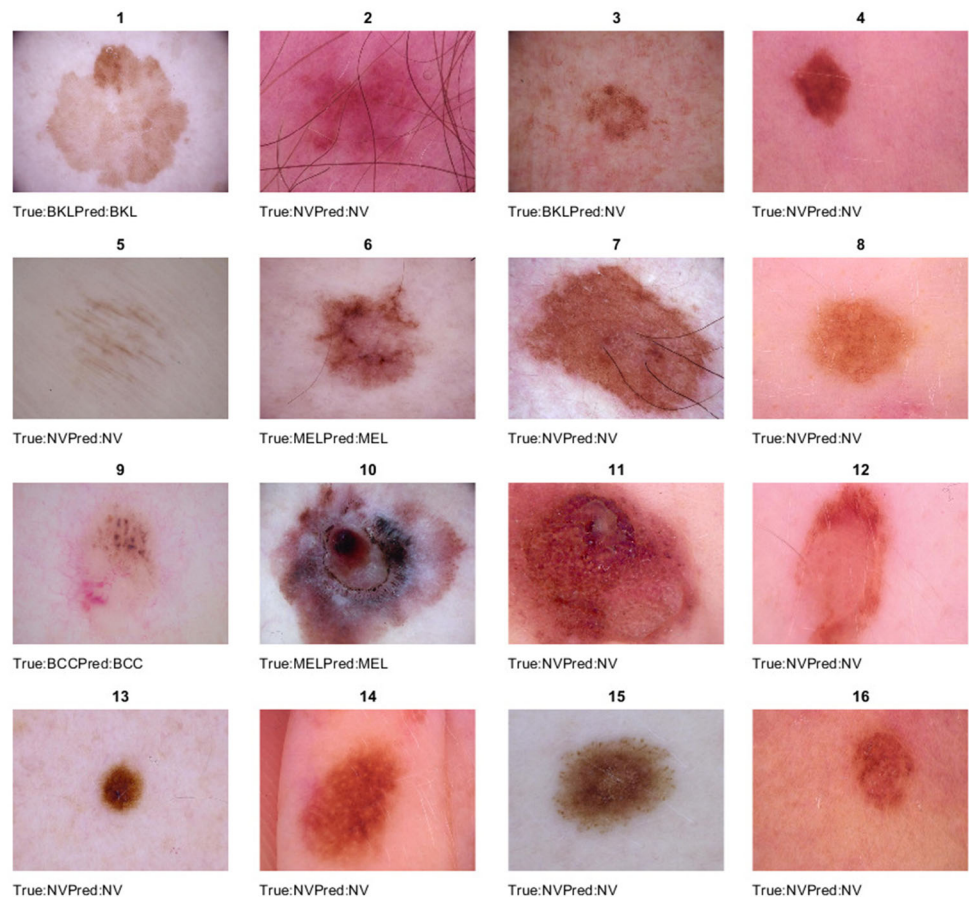
Table 11 Performance measures of denseNet201

Type of disease	Recall	Precision	F1 Score
Akiec	0.51	0.625	0.562
Bcc	0.773	0.647	0.352
Bkl	0.70	0.647	0.672
Df	0.206	0.636	0.311
Mel	0.530	0.594	0.560
Nv	0.925	0.913	0.460
Vasc	0.814	0.945	0.875
Overall accuracy	0.825		

made by Transfer Learning on MobileNetV2 is shown in Fig. 44.

The ShuffleNet training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 45. The performance of ShuffleNet in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 46 and Table 14, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on ShuffleNet is shown in Fig. 47.

Fig. 38 Predictions on random samples made by transfer learning on DenseNet201



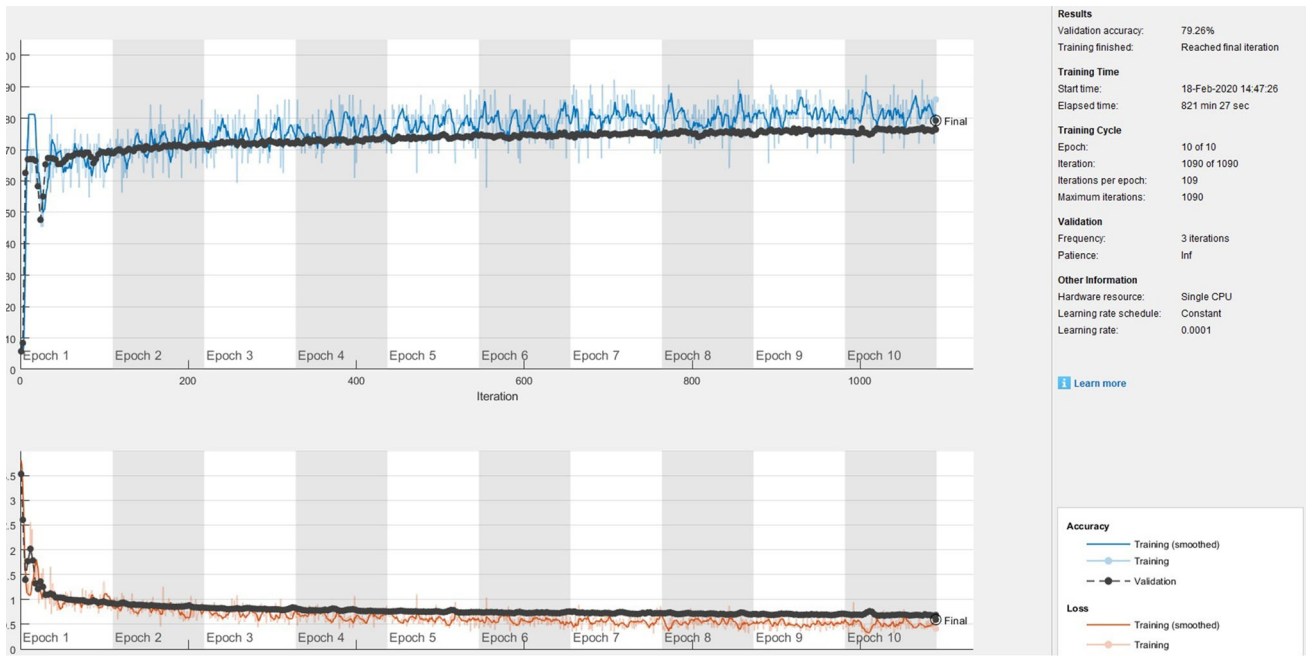


Fig. 39 ResNet18 training with accuracy and error loss over 10 epochs

Fig. 40 Confusion matrix of predictions made by transfer learning on ResNet18

True Class	AKIEC	36	24	26		10	2		36.7%	63.3%
	BCC	7	130	7	1	3	3	3	84.4%	15.6%
	BKL	7	20	221		35	47		67.0%	33.0%
	DF	1	6	8	3	5	11		8.8%	91.2%
	MEL	7	10	48		155	111	3	46.4%	53.6%
	NV	6	31	78		83	1804	9	89.7%	10.3%
	VASC		3				8	32	74.4%	25.6%
			56.3%	58.0%	57.0%	75.0%	53.3%	90.8%	68.1%	
		43.8%	42.0%	43.0%	25.0%	46.7%	9.2%	31.9%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

Table 12 Performance measures of resNet18

Type of disease	Recall	Precision	F1 Score
Akiec	0.367	0.562	0.444
Bcc	0.844	0.580	0.687
Bkl	0.670	0.570	0.616
Df	0.882	0.75	0.811
Mel	0.464	0.533	0.496
Nv	0.897	0.908	0.902
Vasc	0.744	0.681	0.711
Overall accuracy	0.793		

Fig. 41 Predictions on random samples made by transfer learning on ResNet18

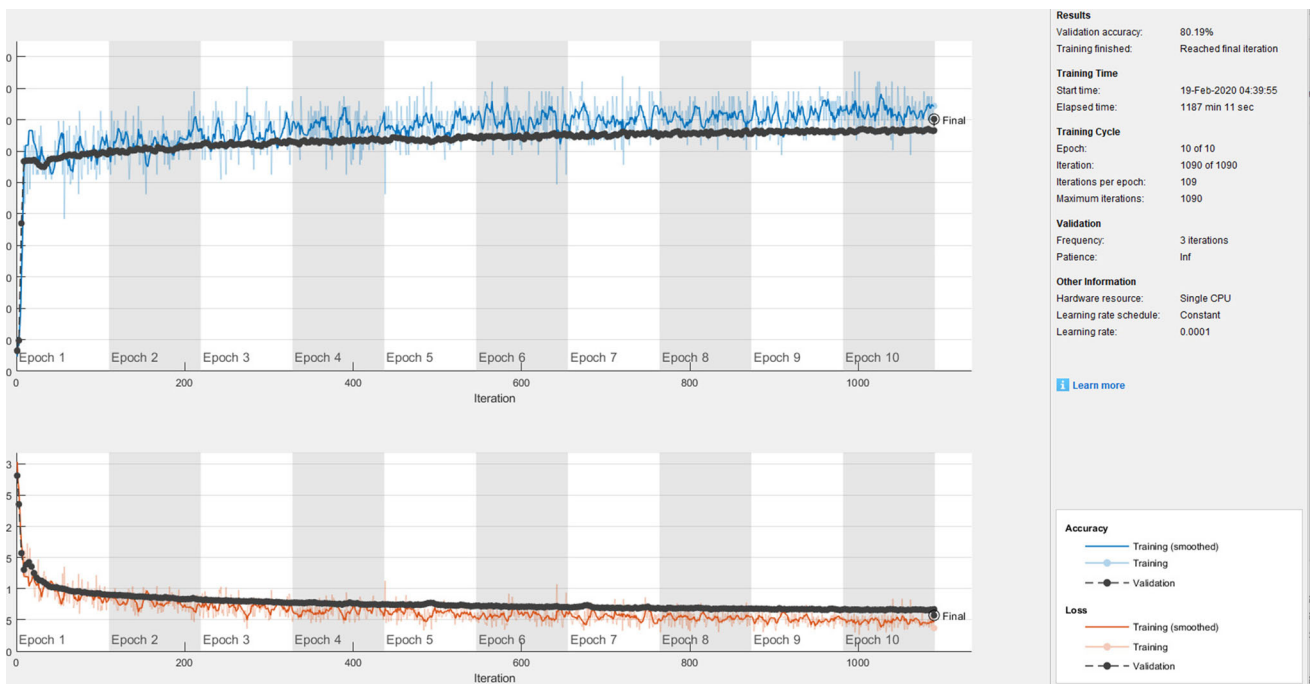
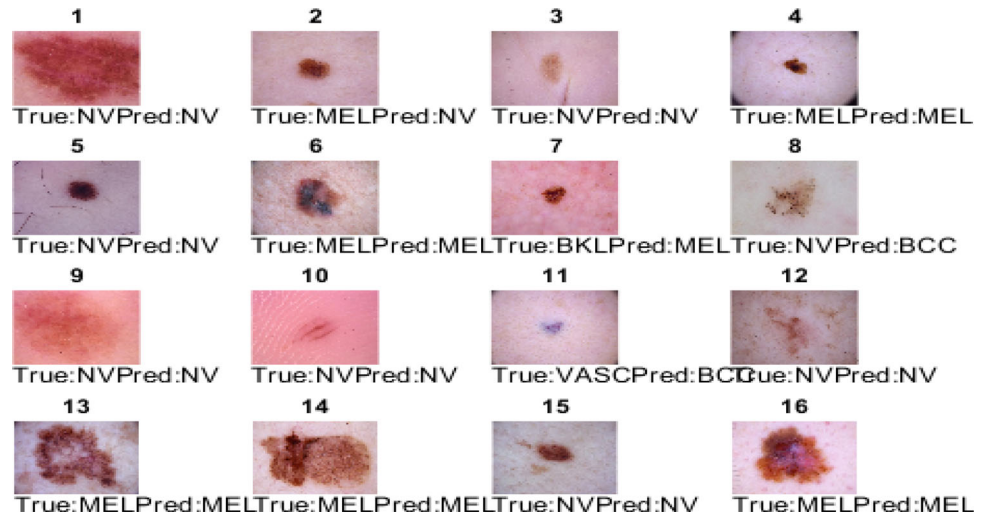


Fig. 42 MobileNetV2 training with accuracy and error loss over 10 epochs

Fig. 43 Confusion matrix of predictions made by transfer learning on MobileNetV2

True Class	AKIEC	40	17	19		16	6		40.8%	59.2%
	BCC	11	106	15	1	7	11	3	68.8%	31.2%
	BKL	8	7	204	2	44	64	1	61.8%	38.2%
	DF	1	2	5	2	2	22		5.9%	94.1%
	MEL	1	6	46		168	113		50.3%	49.7%
	NV	1	13	62	1	72	1862		92.6%	7.4%
	VASC	1	2	3		1	9	27	62.8%	37.2%
			63.5%	69.3%	57.6%	33.3%	54.2%	89.2%	87.1%	
		36.5%	30.7%	42.4%	66.7%	45.8%	10.8%	12.9%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

Table 13 Performance measures of mobileNetV2

Type of disease	Recall	Precision	F1 Score
Akiec	0.408	0.635	0.497
Bcc	0.688	0.693	0.690
Bkl	0.618	0.576	0.596
Df	0.588	0.333	0.425
Mel	0.503	0.542	0.522
Nv	0.926	0.896	0.911
Vasc	0.794	0.880	0.835
Overall accuracy	0.802		

Fig. 44 Predictions on random samples made by transfer learning on MobileNetV2

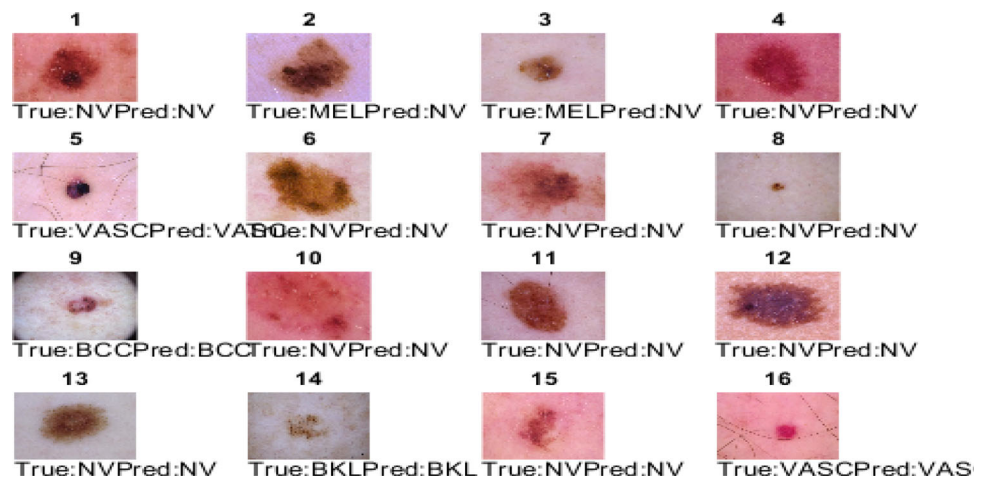


Fig. 45 ShuffleNet training with accuracy and error loss over 10 epochs

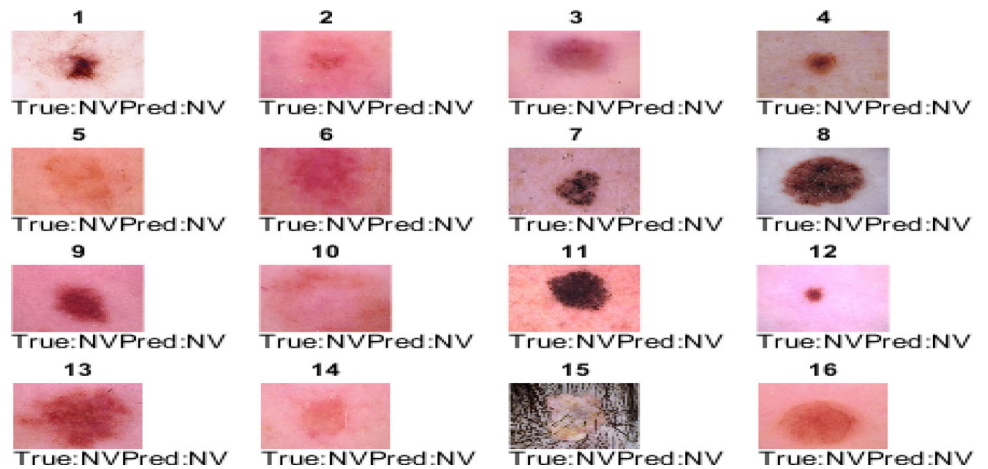
Fig. 46 Confusion matrix of predictions made by transfer learning on ShuffleNet

True Class	AKIEC	37	12	23		20	6		37.8%	62.2%
	BCC	9	98	11	2	15	16	3	63.6%	36.4%
	BKL	9	7	192	2	39	79	2	58.2%	41.8%
	DF	2	3	4	5	5	15		14.7%	85.3%
	MEL	6	4	30	2	164	127	1	49.1%	50.9%
	NV	1	9	41	2	85	1871	2	93.0%	7.0%
	VASC		2	1		1	9	30	69.8%	30.2%
			57.8%	72.6%	63.6%	38.5%	49.8%	88.1%	78.9%	
		42.2%	27.4%	36.4%	61.5%	50.2%	11.9%	21.1%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC	Predicted Class	

Table 14 Performance measures of shuffleNet

Type of disease	Recall	Precision	F1 Score
Akiec	0.377	0.578	0.456
Bcc	0.636	0.726	0.678
Bkl	0.582	0.636	0.608
Df	0.147	0.385	0.213
Mel	0.491	0.498	0.494
Nv	0.930	0.881	0.905
vasc	0.698	0.789	0.741
Overall accuracy	0.798		

Fig. 47 Predictions on random samples made by transfer learning on ShuffleNet



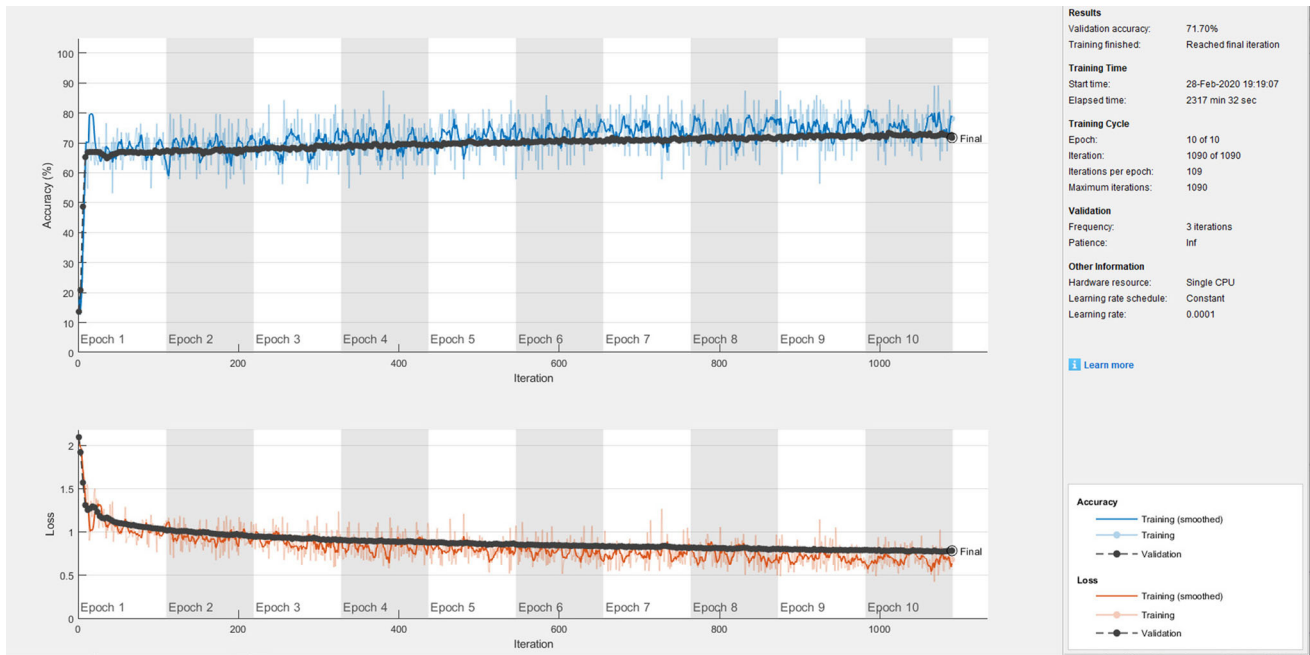


Fig. 48 NasNetMobile training with accuracy and error loss over 10 epochs

Fig. 49 Confusion matrix of predictions made by transfer learning on NasNetMobile

		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
True Class	AKIEC	9	17	44		21	7		9.2%	90.8%
	BCC	3	59	36		32	24		38.3%	61.7%
	BKL	3	14	165		69	78	1	50.0%	50.0%
	DF		1	8		2	23			100.0%
	MEL	3	10	47		139	135		41.6%	58.4%
	NV	1	23	69		142	1776		88.3%	11.7%
	VASC		1	2		3	31	6	14.0%	86.0%
		47.4%	47.2%	44.5%		34.1%	85.6%	85.7%		
		52.6%	52.8%	55.5%		65.9%	14.4%	14.3%		
		AKIEC	BCC	BKL	DF	MEL	NV	VASC		
		Predicted Class								

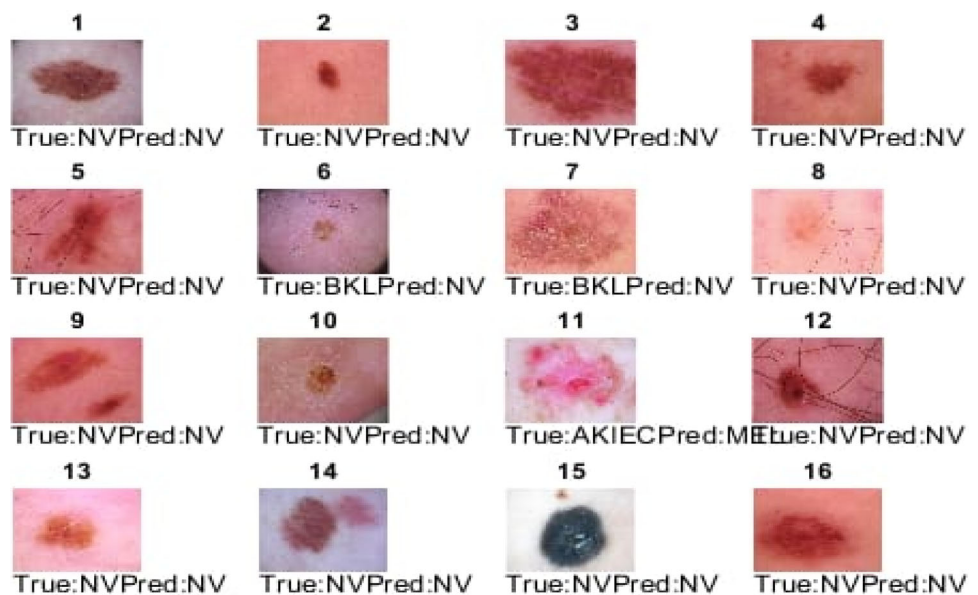
Table 15 Performance measures of nasNetMobile

Type of disease	Recall	Precision	F1 Score
Akiec	0.918	0.474	0.625
Bcc	0.383	0.472	0.423
Bkl	0.50	0.445	0.471
Df	0	0	0
Mel	0.416	0.341	0.375
Nv	0.883	0.856	0.869
Vasc	0.139	0.857	0.239
Overall accuracy	0.717		

Table 16 Comparison of fourteen deep convolutional networks

Network	Depth	Size	Parameters (Millions)	Image input size	Time ($\times 10^5$ s)	Accuracy	Recall	Precision	F1 Score
AlexNet	8	227 MB	61.0	227-by-227	0.1575	0.793	0.673	0.632	4.004
Google Net	22	27 MB	7.0	224-by-224	0.5262	0.785	0.515	0.547	0.749
ResNet 50	50	96 MB	25.6	224-by-224	1.2764	0.816	0.679	0.673	0.601
VGG16	16	515 MB	138	224-by-224	2.4092	0.798	0.726	0.670	0.688
VGG19	19	535 MB	144	224-by-224	2.8894	0.810	0.608	0.631	0.614
ResNet 101	101	167 MB	44.6	224-by-224	2.1734	0.808	0.595	0.656	0.608
Inception V3	48	89 MB	23.9	299-by-299	1.8746	0.784	0.432	0.460	0.432
Inception ResnetV2	164	209 MB	55.9	299-by-299	4.0600	0.749	0.337	0.463	0.351
Squeeze Net	18	4.6 MB	1.24	227-by-227	0.2892	0.769	0.506	0.527	0.515
Dense Net201	201	77 MB	20.0	224-by-224	4.2896	0.825	0.637	0.715	0.542
ResNet 18	18	44 MB	11.7	224-by-224	0.4973	0.793	0.695	0.655	0.667
Mobile NetV2	53	13 MB	3.5	224-by-224	0.7178	0.802	0.646	0.651	0.639
Shuffle Net	50	6.3 MB	1.4	224-by-224	0.3882	0.798	0.551	0.642	0.585
NasNet Mobile	Do not consist of linear modules	20 MB	5.3	224-by-224	1.3905	0.717	0.462	0.492	0.429

Fig. 50 Predictions on random samples made by transfer learning on NasNetMobile



The NasNetMobile training Progress with accuracy and error loss over 10 Epochs is shown in Fig. 48. The performance of NasNetMobile in which class wise recall, precision, f1 score and overall accuracy based on confusion matrix are shown in Fig. 49 and Table 15, the best result is highlighted in bold. The Predictions on Random samples made by Transfer Learning on NasNetMobile is shown in Fig. 50.

The fourteen deep convolutional networks with number of layers, size, parameters, input size, time, accuracy, recall, precision and f1 score, the average out for a single

matrix, are presented in Table 16. DenseNet201 performs better in performance measures and highlighted bold in the Table 16.

4 Discussion

It is noted that the results obtained from the fourteen deep convolutional networks are comparable in terms of accuracy, recall, precision and f1 score performance measures. By comparing the performance measures, DenseNet201

network can perform best with 0.825 accuracy. Furthermore, this network gives good efficiency by eliminating the vanishing gradient problem, supporting feature propagation, encouraging reuse of features, and reducing the number of parameters. Besides improved efficiency, another best point is that the information flow and gradients movement is smooth in this network, which helps them in training and reduces the overfitting problem.

The performance of the networks is evaluated on the ISIC 2018 dataset consists of 10,015 dermoscopic images. The training progress, confusion matrix for each type of class and testing images results show the best performance of DenseNet201. The ResNet50 model takes the second position, which is taking less time and good accuracy. Similarly, the other model's performance is measurable with time parameters complexity.

5 Conclusion

This study compared the performance of fourteen deep convolutional neural networks for classifying seven types of diseases. This model's performance is measured using ISIC 2018 dataset. The pre-trained networks are modified by replacing their last fully connected layer with our fully connected-7, softmax and cross-entropy layer for classification into seven classes. The results are summarized in Table 16. The results show the best performance of DenseNet201 for classification. In the upcoming years, computer-aided diagnostic systems will be more for classifying a skin lesion by looking at the inspiring results of our study and the former studies using deep learning techniques. The models designed with this baseline are expected to give high accuracy with minimum computation time. In future, our goal is to achieve more effective results by increasing the number of epochs and skin lesions categories.

Author contributions Ginni Arora: Conceptualization, Methodology, Software, Data curation, Validation, Writing- Original draft preparation. Ashwani Kumar Dubey: Conceptualization, Methodology, Supervision, Reviewing and Editing. Zainul Abidin Jaffery: Conceptualization, Supervision, Reviewing and Editing. Alvaro Rocha: Reviewing and Editing.

Funding None.

Data Availability ISIC database.

Declarations

Conflict of interest There is no conflict of interest.

Consent to participate I consent to participate.

Consent for publication I consent for publication.

Code availability Custom code.

References

- Hameed N, Shabut AM, Ghosh MK, Hossain MA (2020) Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Syst with Appl* 141:112961. <https://doi.org/10.1016/j.eswa.2019.112961>
- Al-Masni MA, Kim DH, Kim TS (2020) Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Programs Biomed* 190:105351. <https://doi.org/10.1016/j.cmpb.2020.105351>
- Nasiri S, Helsper J, Jung M, Fathi M (2020) DePicT Melanoma Deep-CLASS: a deep convolutional neural networks approach to classify skin lesion images. *BMC bioinform* 21:1–3. <https://doi.org/10.1186/s12859-020-3351-y>
- Mukherjee S, Adhikari A, Roy M (2019) Malignant melanoma classification using cross-platform dataset with deep learning CNN architecture. In: Bhattacharyya S, Pal SK, Pan I, Das A (eds) *Recent trends in signal and image processing 2019*. Springer, Singapore, pp 31–41
- Seeja RD, Suresh A (2019) Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). *Asian Pac J of Cancer Prev APJCP* 20(5):1555
- Chung YM, Hu CS, Lawson A, Smyth C (2019) Toporesnet: A hybrid deep learning architecture and its application to skin lesion classification. *arXiv preprint, arXiv:1905.08607*
- Al-Antari MA, Al-Masni MA, Kim TS (2020) Deep learning computer-aided diagnosis for breast lesion in digital mammogram. In: Lee G, Fujita H (eds) *Deep Learning in Medical Image Analysis*. Springer, Cham, pp 59–72
- Ali R, Hardie RC, De Silva MS, Kebede TM (2019) Skin lesion segmentation and classification for ISIC 2018 by combining deep CNN and handcrafted features. *arXiv preprint, arXiv:1908.05730*
- Al-Obeidat F, Rocha Á, Akram M et al (2021) (CDRGI)-Cancer detection through relevant genes identification. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-021-05739-8>
- Yilmaz E, Trocan M (2020) Benign and malignant skin lesion classification comparison for three deep-learning architectures. In: Thanh Nguyen N, Jearanaitanakij K, Selamat A, Trawiński B, Chittayasothorn S (eds) *Asian Conf on Intell Inf and Database Syst*. Springer, Cham, pp 514–524
- Saba T, Khan MA, Rehman A, Marie-Sainte SL (2019) Region extraction and classification of skin cancer: a heterogeneous framework of deep CNN features fusion and reduction. *J Med Syst* 43:289. <https://doi.org/10.1007/s10916-019-1413-3>
- Amin J, Sharif A, Gul N, Anjum MA, Nisar MW, Azam F, Bukhari SA (2020) Integrated design of deep features fusion for localization and classification of skin cancer. *Pattern Recognit Lett* 131:63–70. <https://doi.org/10.1016/j.patrec.2019.11.042>
- Delibasis K, Georgakopoulos SV, Tasoulis SK, Maglogiannis I, Plagianakos VP (2020) On image prefiltering for skin lesion characterization utilizing deep transfer learning. In: Iliadis L, Parvanov Angelov P, Jayne C, Pimenidis E (eds) *International conf on engineering appl of neural networks*. Springer, Cham, pp 377–388
- Ballester P, Araujo R (2016) On the performance of GoogLeNet and AlexNet applied to sketches. In: *Proceedings of the AAAI Conference on Artificial Intelligence*: 30(1)
- Chen M, Chen W, Chen W, Cai L, Chai G (2020) Skin cancer classification with deep convolutional neural networks. *J of Med*

- Imaging and Health Info 10(7):1707–1713. <https://doi.org/10.1166/jmhi.2020.3078>
16. Gessert N, Nielsen M, Shaikh M, Werner R, Schlaefler A (2020) Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX* 7:100864. <https://doi.org/10.1016/j.mex.2020.100864>
 17. Carcagni P, Leo M, Cuna A, Mazzeo PL, Spagnolo P, Celeste G, Distante C (2019) Classification of skin lesions by combining multilevel learnings in a DenseNet architecture. In: Ricci E, Bulò SR, Snoek C, Lanz O, Messelodi S, Sebe N (eds) *International Conf on Image Analysis and Processing*. Springer, Cham, pp 335–344. https://doi.org/10.1007/978-3-030-30642-7_30
 18. Chen EZ, Dong X, Li X, Jiang H, Rong R, Wu J (2019) Lesion attributes segmentation for melanoma detection with multi-task u-net. In: *Proceedings of the 2019 IEEE 16th international symp on biomedical imaging (ISBI 2019)*, pp 485–488, IEEE. <https://doi.org/10.1109/ISBI.2019.8759483>
 19. Yang J, Sun X, Liang J, Rosin PL (2018) Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. In: *Proceedings of the IEEE conf on computer vision and pattern recognition*, pp 1258–1266, IEEE.
 20. Nazi ZA, Abir TA (2020) Automatic skin lesion segmentation and melanoma detection: Transfer learning approach with u-net and dcnn-svm. In: Uddin MS, Bansal JC (eds) *Proc of international joint conf on computational intelligence*. Springer, Singapore, pp 371–381. https://doi.org/10.1007/978-981-13-7564-4_32
 21. Rastgoo M, Lemaître G, Morel O, Massich J, Garcia R, Meriaudeau F, Marzani F, Sidibé D (2016) Classification of melanoma lesions using sparse coded features and random forests. *Med Imaging 2016: Comput-Aided Diagn* 9785:97850
 22. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, Helba B, Kalloo A, Liopyris K, Marchetti M, Kittler H (2019) Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*
 23. Tschandl P, Rosendahl C, Kittler H (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 5(1):1–9. <https://doi.org/10.1038/sdata.2018.161>
 24. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
 25. Negrete PDM, Iano Y, Monteiro ACB, França RP, Gomes G, de Oliveira D, Pajuelo (2021) Classification of dermoscopy skin images with the application of deep learning techniques. In: Iano Y, Arthur R, Saotome O, Kemper G, Monteiro ACB (eds) *Proc of the 5th Brazilian technology symp*. Springer, Cham, pp 73–81. https://doi.org/10.1007/978-3-030-57566-3_7
 26. Balazs H (2018) Skin lesion classification with ensembles of deep convolutional neural networks. *J of Biomed Inf* 86:25–32. <https://doi.org/10.1016/j.jbi.2018.08.006>
 27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conf on computer vision and pattern recognition*, pp 1–9
 28. Quang NH (2017) Automatic skin lesion analysis towards melanoma detection. In: *proceedings of the 2017 21st Asia Pacific symposium on intelligent and evolutionary systems (IES)*, pp 106–111. <https://doi.org/10.1109/iesys.2017.8233570>
 29. Kwasigroch A, Mikołajczyk A, Grochowski M (2017) Deep neural networks approach to skin lesions classification—A comparative analysis. In: *proceedings of the 2017 22nd int conf on methods and models in automation and robotics (MMAR)*, pp 1069–1074. <https://doi.org/10.1109/mmar.2017.8046978>
 30. Demir A, Yilmaz F, Kose O (2019) Early detection of skin cancer using deep learning architectures: Resnet-101 and inception-v3. In *2019 Medical Technologies Congress (TIPTEKNO)*, pp 1–4, IEEE. <https://doi.org/10.1109/tiptekno47231.2019.8972045>
 31. Shahin AH, Kamal A, Elattar MA (2018) Deep ensemble learning for skin lesion classification from dermoscopic images. In: *proceedings of the 2018 9th cairo international biomedical engineering conference (CIBEC)*, pp. 150–153. <https://doi.org/10.1109/cibec.2018.8641815>
 32. Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the AAAI Conf on Artificial Intelligence*:31(1)
 33. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*
 34. Al-Masni MA, Kim DH, Kim TS (2020) Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Programs Biomed* 190:105351. <https://doi.org/10.1016/j.cmpb.2020.105351>
 35. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017). Densely connected convolutional networks. In: *Proceedings of the IEEE conf on computer vision and pattern recognition*, pp 4700–4708
 36. Czum JM (2020) Dive into deep learning. *J of the American College of Radiology* 17(5):637–638. <https://doi.org/10.1016/j.jacr.2020.02.005>
 37. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*
 38. Zhang X, Zhou X, Lin M, Sun J (2018) Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE conf on computer vision and pattern recognition*, pp 6848–6856
 39. Zoph B, Vasudevan V, Shlens J, Le QV (2018) Learning transferable architectures for scalable image recognition. In: *Proceedings of the IEEE conf 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.