




An Ensemble Method for Efficient Classification of Skin Lesion from Dermoscopy Image

B. H. Shekar  and Habtu Hailu 

Mangalore University, Mangalagangothri, Karnataka, India

Abstract. Nowadays, skin cancer is growing-up due to exposure to Ultraviolet (UV) radiation emanating from the sun light. Among several categories of skin lesion, melanoma is the most deadly cancerous kind. Diagnosing skin lesion in its early stage have a great chance to cure the disease. Researchers have proposed several computer-aided diagnosis techniques to detect skin lesions. In this work, we present an ensemble model to classify skin lesion using a pre-trained DenseNet and InceptionV3 algorithms. The fully layered fine-tuned technique is applied to both the algorithms which are previously explored for ImageNet dataset. The fine-tuned algorithms are utilized to train on the HAM10000 dataset. The classification results obtained due to the pre-trained models are concatenated in the average ensemble method. The experimentation on the standard datasets confirm the classification accuracy of 91% and indicates that the proposed approach is a promising as compared to the previously developed approaches.

Keywords: Deep learning · DenseNet · InceptionV3 · Ensemble learning · Skin lesion classification

1 Introduction

Globally in each year, 132,000 new melanoma and 2 to 3 million non-melanoma skin cancer patients appear which shows that the rate of skin cancer incidence is drastically growing-up [1, 2]. The major cause for this is due to ultraviolet (UV) radiation which is the most significant spectrum of sunlight that can destroy the DNA under the skin cell that leads to excess development of skin cells resulting in skin cancer. The main cause of UV to reach our surroundings is the evacuation of the level of the ozone layer [3, 4]. The most usual categories of skin lesions are squamous cell carcinoma, melanoma, basal cell carcinoma, Benign, Actinic keratosis, Melanocytic nevi, Vascular lesions, and Dermatofibroma [5–7]. Melanoma is the most serious cancerous kind of skin lesion, which is the cause for 9000 mortality in 2017 in United States [8] only. If melanoma is diagnosed in its early stage, nearly 95% of the cases have a possibility to cure, especially basal cell and squamous cell carcinomas are highly curable cases [9].

Skin lesion is primarily detected manually by using human naked eyes, which require a magnifying and illuminated skin images. Among several procedural techniques, the most common methods (ABCD) rule, Menzies 7-point checklist and 3-point checklist are used to detect the melanoma in the early stages [10,11]. Reports on the performance of clinical dermatologists on diagnostic accuracy have claimed 80% for a dermatologist who have ten years and more experience, whereas dermatologist who have 3 to 5 years experience were able to reach only 62% [12]. This shows that for detecting skin lesion with a better accuracy, years of experience over difficult situations plays a great role. Applying machine learning techniques on dermoscopic image to classify malignant and benign lesion becomes popular task because of the ability to detect patterns in digital images. Deep learning methods exhibit better performance in detection and classification of various diseases by means of medical image examination [5,13].

Several studies have been prompted to classify skin lesion from dermoscopic images. Barata et al. [14] uses a global and local features for the detection of melanoma in dermoscopy images. They have compared the effect of color and texture features for lesion classification and concluded that a combination of features leads to better performance. In the work of Codella et al. [15] a combination of support vector machine (SVM), sparse coding techniques and deep learning are applied on International Skin Imaging Collaboration (ISIC) dataset to recognize/classify dermoscopy images. A convolutional network with transfer learning is developed by Cicero et al. [16] on a custom dataset of skin image to get better performance in the skin lesion classification task.

In the recent days, we have been witnessing the application of deep learning for many of the medical image analysis problems. Among these, Esteva et al. [17] applied a pretrained CNN technique, GoogleNet and Inception v3 for image classification. In order to tackle the difficulties of classifying skin lesion, Lopez et al. [18] presented a pretrained VGGNet algorithm with the transfer learning method. The ISIC dataset is used for testing the proposed method. In 2017, Krizhevsky et al. [19] applied a deep CNN on a large dataset of ImageNet LSVRC-2010. The number of different classes after classification is 1000. Codella et al. [20] proposed an ensemble of deep residual network and fully CNN in combination with SVM, hand-coded feature extractor and sparse coding method to segment and detect melanoma cases on a dataset of International Symposium on Biomedical Imaging (ISBI). To classify the dermoscopy image dataset of ISBI 2017 into three different classes, Harangi et al. [21] employ an ensemble technique that fuses the classification output of four different deep neural network algorithms. Tan et al. [22] used a feature optimization technique considering Particle Swarm Optimization (PSO) for the purpose of classification of skin lesion into benign and malignant. Dermofit Image Library, PH2, and Dermnet are the datasets used for evaluation.

In addition to these works, most recently Hekler et al. [23] implements a deep learning network for skin lesion classification into malignant melanoma and benign nevus that could help human for the histopathologic melanoma

diagnosis. To enhance the performance of skin lesion classification, a dilated convolution of deep learning technique is applied on four pretrained algorithms (VGG16, VGG19, MobileNet and InceptionV3) by [5]. They have used transfer learning for the extraction of features from the images. Chaturvedi et al. [4] proposed a transfer learning on pretrained MobileNet algorithm and evaluated on HAM10000 dataset to classify into seven different classes. In Pratiwi et al. [13], CNN model is proposed for the detection of skin cancer from HAM10000 dermoscopy image. In the work of Khan et al. [24], an ensemble of pretrained ResNet-50 and ResNet-101 through transfer learning based feature extraction is employed for skin lesion classification. The features extracted are fed to SVM for classification. Even though, several attempts are done for classification of skin lesions, still there is lack of generality in their capability of classification and have not achieved better accuracy because of the complexities in the image itself [4].

In this study, we proposed an ensemble method that fuses the two most common pretrained deep convolutional neural networks, namely DenseNet and InceptionV3, which are pretrained on approximately 1.28 million images. In most of the cases these two algorithms outperform in the HAM10000 dataset [25] as we explore from the previous works. We use a fine tuning technique for the feature extraction (discussed in the methodology in detail). The proposed model is trained and tested on HAM10000 dataset [25] that consists of 10015 dermoscopy images. The rest of the paper is outlined as follows. Section 2 introduces the proposed method which details about dataset description, data pre-processing techniques, data augmentation and the proposed architecture. Section 3 present experimental findings. Finally, Discussion and Conclusions are given in Sect. 4.

2 Proposed Methodology

In this section, we present the details of the proposed methodology which include the dataset used for training and evaluation, data pre-processing and augmentation techniques, and the architecture of the proposed method.

2.1 Dataset

To train, validate and test the proposed model, we have used a collection of dermatoscopic images namely Human Against Machine with 10000 training images (HAM10000) dataset which is available publicly on International Skin Imaging Collaboration (ISIC). The dataset accommodates 10015 dermatoscopic images gathered from different populations by using a variety of modalities. The dataset is not equally distributed for each type of lesions, 6705 Melanocytic nevi (nv) images, 1113 Melanoma (mel) images, 1099 Benign keratosis (bkl) images, 514

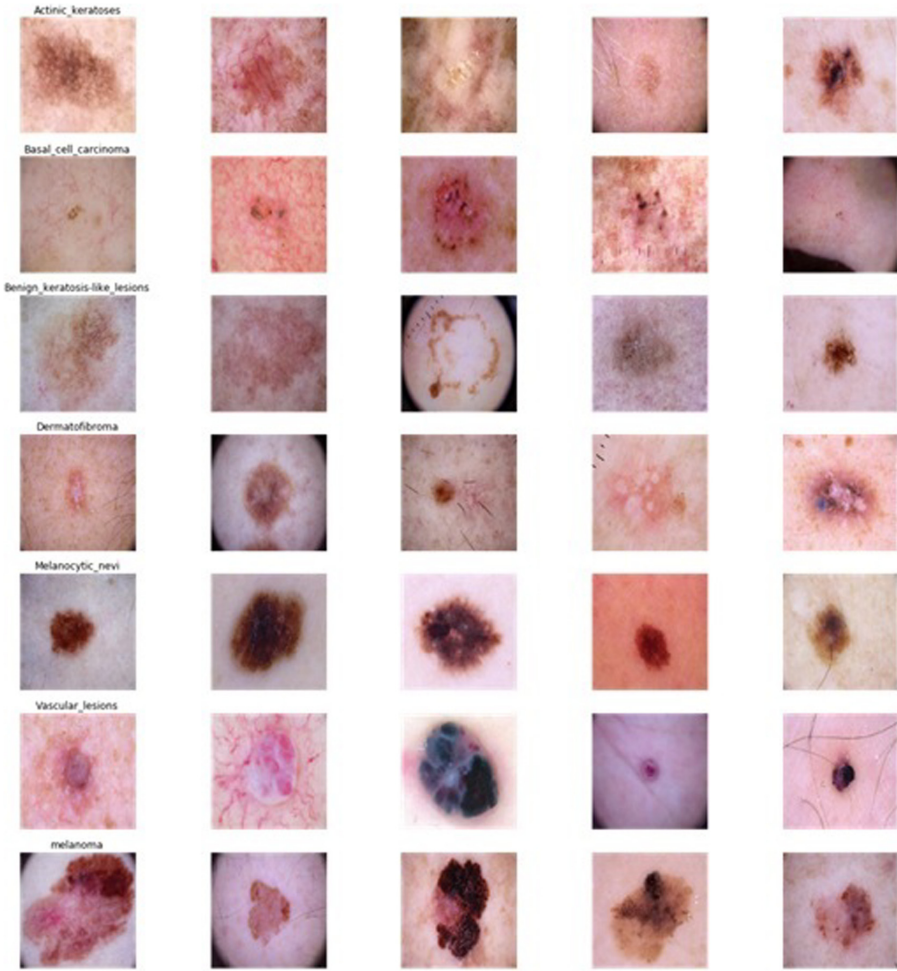


Fig. 1. Randomly selected sample images for each cancer type from the HAM10000 dataset.

Basal cell carcinoma (bcc) images, 327 Actinic keratosis (Akiec) images, 142 Vascular (vasc) images and 115 Dermatofibroma (df) images. All images are stored with 600×450 pixels resolution. This indicates that, more than 50% of the dataset is imbalanced to only one type of lesion namely Melanocytic nevi. Figure 1 shows five sample images from each lesion types.

2.2 Data Pre-processing

The pixel resolutions of all the images used in this study are 600×450 . To make the size of these images compatible with our models (DenseNet and InceptionV3), we downscale the pixel resolution to 256×192 by using Keras ImageDataGenerator. Then normalization of the dataset is performed by dividing the pixel values of the images by 255.0. Finally, we divide the dataset for the training (8111 images), validation (902 images) and testing (1002 images) sets.

2.3 Data Augmentation

For deep learning algorithms, to get a better performance, a large amount of data is required. But still, acquiring an adequate amount of data is the main challenge in the area. One best solution to increase the dataset size is the data augmentation technique as it raises the dataset size without eliminating the structure of the data. In our study, the first data augmentation techniques used is a rotation operation with a range randomly between 0 and 60° . The other concern in image data preparation is, objects of interest in the image may be off-centred by several means. To handle this problem, we apply width shifting and height shifting with a range of 0.2. At last, shear and random zooming operations are applied with a range of 0.2.

2.4 Proposed Architecture

Recently, for image classification, CNN become the state-of-the-art method because it achieved excellent performance on a well-known datasets such as MNIST [26] and ImageNet [27]. There are several varieties of CNN algorithms for image classification task, such as AlexNet [19], GoogLeNet [28], ResNet [29], DenseNet [30], VGGNet [31], InceptionV3 [32] and others. In our work, we perform the classification of skin lesion by using an ensemble method which encompasses well-established deep learning architectures that have shown better accuracy in the previous works, namely DenseNet and InceptionV3. These architectures are available as a pretrained model that were initially trained with ImageNet dataset that contains around 1.28 million natural images with 1000 classes. We use the weights and biases from the pretrained model to initialize learning on our dataset, and then a fine-tuning technique is applied on all the layers in the selected architectures. The details are presented in the following subsections.

The DenseNet Architecture: The DenseNet architecture proposed by Huang et al. [30] contains all the layers which are directly connected to each other to optimize the flow of information between the layers. That means, each layer in the network receive information from all the antecedent layers and feeds its output to all the consequent layers. A concatenation operation is performed in every layer to merge the inputs from the previous layers. Equation 1 presents the

input feature map fetched to the i th layer from all the preceding layers [33]. The connectivity pattern in each layer of a single dense block is illustrated in Fig. 2. DenseNet is one of the best performer method on ImageNet dataset where it performs 0.773 on top-1 and 0.936 on top-5 retrievals.

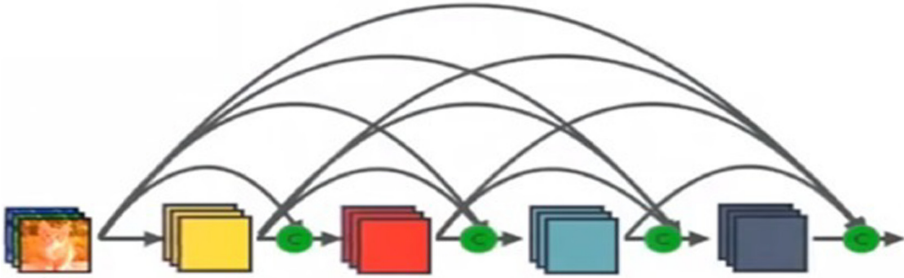


Fig. 2. A layout of single dense block which has 5 layers.

$$x_i = H_i(x_0, x_1, \dots, x_{i-1}) \tag{1}$$

Here, x_i is the output of the i th layer and H_i is the composite function that represent the operations such as rectified linear units (ReLU), Batch Normalization (BN), and Convolution (Conv).

In our work, we have used the variant of DenseNet that is named as DenseNet-201, which has 4 dense blocks and 201 layers. Figure 3 shows a DenseNet architecture with three dense blocks. In each dense block, there is a composition layer which performs sequentially BN and ReLU and then a 3×3 convolution operations. The convolution operation is used to provide the concatenated output feature map, say for example, to transform the input feature maps x_0, x_1, x_2 to output feature map x_3 by using Eq. 1. The Batch Normalization operation is used to normalize the input of each layer [34] in order to decrease the absolute difference between data and make the relative difference higher.

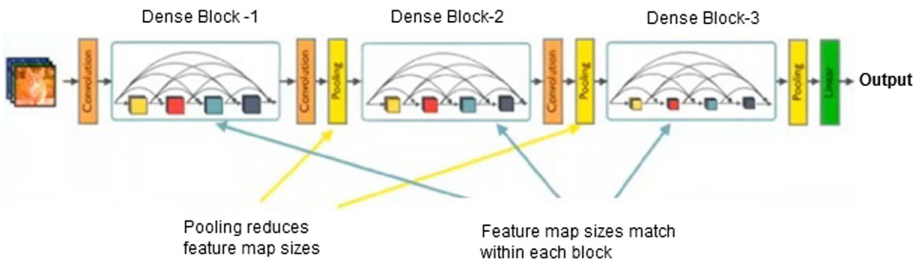


Fig. 3. A deep DenseNet with 3 dense block.

The other operation which is part of composition layer is the ReLU, which is applied in DenseNet architecture. Equation 2 describes how the ReLU operation works.

$$ReLU(x) = \begin{cases} x, & \text{if } x > 0. \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The changing of feature map size due to the down-sampling layers in convolutional networks makes difficult to perform the concatenation operation. To facilitate down-sampling, the Densenet architecture separates the whole network into various dense blocks that are connected densely. As shown in Fig. 3, there are transition layers between these densely connected dense blocks used to perform convolution and pooling operation. In this work, these transition layers include three different operations namely batch normalization, 1×1 convolution and a 2×2 average pooling operation [30]. Beside, a bottleneck layer is incorporated within the dense blocks before a 3×3 convolution layer. It consists of BN, ReLU and a 1×1 convolution layer. The 1×1 convolution operation in this layer makes the network computationally efficient by reducing the number of input feature maps to the 3×3 convolution operation in the dense block. This layer make DenseNet method effective by reducing the complexity and size of the model. The main benefits of DenseNet, when compared to other methods, are presented below.

- *Only a few parameters:* Since the feature maps from the preceding layers acts like an input for the current layer, many feature maps can be reused to learn by some convolution kernels.
- *Capability to reduce over fitting:* The dense connection in the DenseNet network built short paths from the beginning to the end layers. Due to this, the loss function provides additional guidance for each layer. Consequently, the dense connection protects the over-fitting problem in a better way, particularly it is a good choice for learning from the small size of data.
- *Layers are deeper:* Because all layers are linked directly to each other, the network has highly deep architecture.

The InceptionV3 Architecture: By enhancing the GoogleNet [28] network, Szegedy et al. [32] proposed an algorithm called InceptionV3. The major enhancement is reducing the size of the parameters by concatenating the convolutional filters which have different sizes into a new filter. Consequently the computational complexity of the model is decreased. Figure 4 illustrates the architecture of InceptionV3. This model scores an error rate of 3.5% on top-5 and 17.3% on top-1 of ImageNet dataset.

In this network, the number of parameters is reduced by replacing the convolution filters of size greater than 3×3 (e.g. 5×5 or 7×7) by a sequence of 3×3 convolution layers. The computational cost of a large spatial filter convolution is expensive [32]. In addition to this, spatial factorization into asymmetric convolutions is applied. This means, replacing an $n \times n$ filters by two layer asymmetric filters of $n \times 1$ followed by $1 \times n$. The InceptionV3 network has 42 layers and the detail of the network that is shown in Fig. 4 is presented in Table 1.

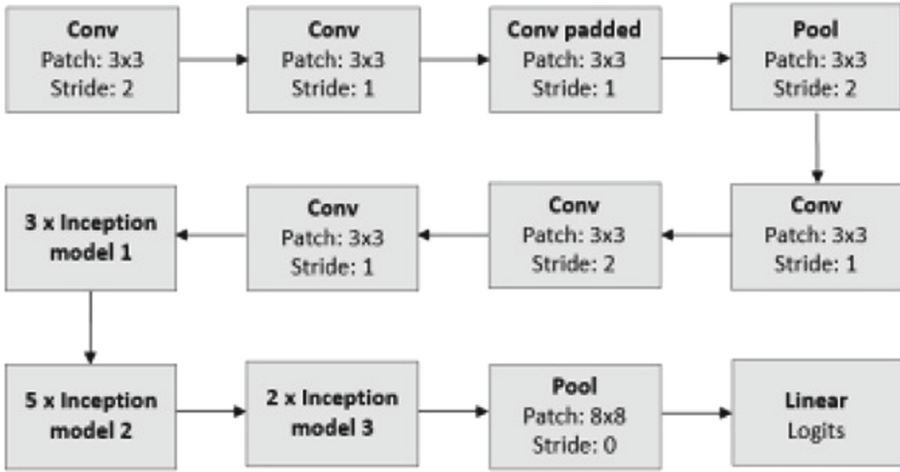


Fig. 4. The architecture of InceptionV3.

Table 1. The detailed outline of InceptionV3 architecture. The input size column also represents the output size of the previous layer.

Type of layer	Patch size/stride	Input size
Conv	$3 \times 3/2$	$299 \times 299 \times 3$
Conv	$3 \times 3/1$	$149 \times 149 \times 32$
conv padded	$3 \times 3/1$	$147 \times 147 \times 32$
Pool	$3 \times 3/2$	$147 \times 147 \times 64$
Conv	$3 \times 3/1$	$73 \times 73 \times 64$
Conv	$3 \times 3/2$	$71 \times 71 \times 80$
Conv	$3 \times 3/1$	$35 \times 35 \times 192$
Pool	8×8	$8 \times 8 \times 2048$
Linear	Logits	$1 \times 1 \times 2048$
Softmax	Classifier	$1 \times 1 \times 1000$

Fine Tuned Ensemble: To increase the image classification accuracy on a dataset which do not have sufficient amount of annotated images, an ensemble of DNN is a powerful technique, which makes a decision by combining the prediction results from multiple models [21]. In our work, we have explored an ensemble of two well-known pretrained CNN algorithms, DenseNet and InceptionV3. Firstly, the two methods (DenseNet and InceptionV3) are fine-tuned and trained on our dataset individually, and then the best performed model is saved. The fine tuning technique is applied by freezing all the layers of the

networks prior to the final fully connected layer. A fully connected layer of the pretrained networks is removed and replaced by a new fully connected layer that have seven neurons which is equivalent to the number of classes in the prediction task. Finally, classification is performed by fusing saved models by averaging technique. An averaging of a models' prediction is an ensemble learning technique that predicts based on the predictions obtained by each model. It considers each model equally for average calculation and used to bring down the variance in the final neural network model [21]. Figure 5 presents the model architecture of the current work.

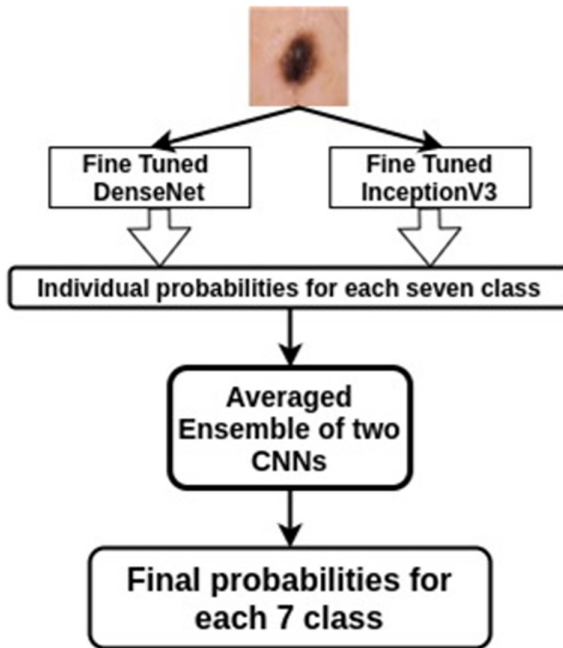


Fig. 5. The proposed ensemble architecture.

3 Experimental Details

In this section, we present the details of the experimental setup, the evaluation metrics and results with a detailed discussion on the experimentation.

3.1 Experimental Setup

In the current study, we perform a classification task using two fine-tuned DNN methods and then we have employed an ensemble technique. The proposed models are trained for 40 epochs with a batch size of 32 on our dataset. The evaluation of the model is done using 902 sample images from the validation set and 1002 sample images from the test set. We have considered Adam optimizer to optimize learning with a learning rate of 0.0001, the minimum learning rate is set to 0.1×10^6 . We have used the Top 1 accuracy, which is the standard performance measure in CNN studies [19]. Firstly, we have compared the classification accuracy obtained on the HAM10000 dataset considering

- Fine-tuned denseNet algorithm
- Fine-tuned InceptionV3 algorithm and
- An ensemble method of DenseNet and InceptionV3.

In addition, the best accuracy result from our evaluation is compared with other previous studies which have better performance and done on HAM10000 dataset. To program the model for our computer aided diagnosis (CAD) system, we use python programming language. We built it on top of Keras deep learning framework for neural networks [35] with the tensorflow [36] back-end. The training is performed on Google Collaborator, which come up with a single 12 GB NVIDIA Tesla K80 GPU and 12 GB RAM.

3.2 Evaluation Metrics

The overall testing of the proposed model is performed using 1002 unseen test dataset. To assess the achievement of the proposed model, we use several evaluation metrics, namely precision, recall, accuracy, and F1-Score. For each of the 7 groups' precision, Recall, and F1-score are determined. Also, the weighted average which is a good measure for unbalanced dataset is also calculated for recall, precision and f1-score.

3.3 Experimental Results

This part narrate the assessment outcomes of the proposed method empirically and graphically on the HAM10000 dataset. As shown in Tables 2, 3 and 4, the experimental result for DenseNet, InceptionV3, and the Ensemble of the two models respectively is presented with precision, recall, and f1score on the HAM10000 dataset for seven classes. Accordingly, for individual models, namely DenseNet and InceptionV3, the Melanocytic nevi class which has the maximum number of the test sample (678 out of 1002) scores the highest precision, recall, and f1-score. For Melanocytic nevi class, DenseNet scores 95%, 96% and 95%

Table 2. Precision, recall, and F1-score for each class due to DenseNet model.

Class name	Precision	Recall	F1-score
akiec	0.94	0.57	0.71
bcc	0.92	0.79	0.85
bkl	0.73	0.83	0.78
df	0.92	0.73	0.81
nv	0.95	0.96	0.95
mel	0.93	0.93	0.93
vasc	0.71	0.71	0.71

Table 3. Precision, recall, and F1-score for each class due to Inceptionv3 model.

Class name	Precision	Recall	F1-score
akiec	0.70	0.57	0.63
bcc	0.93	0.74	0.82
bkl	0.66	0.78	0.71
df	0.83	0.67	0.74
nv	0.84	0.95	0.95
mel	0.81	0.93	0.87
vasc	0.74	0.69	0.71

and InceptionV3 scores 94%, 95% and 95% precision, recall, and f1-score respectively. The ensemble method scores the highest precision of 100% for Melanoma and Dermatofibroma, the highest recall of 96% for Melanocytic nevi, and 96% f1-score for melanoma classes. Quantitatively, Table 5 demonstrates the accuracy, weighted precision, weighted recall and weighted f1-score of our ensemble network and the two fine-tuned network for test dataset. A weighted average of precision, recall, and f1-score for DenseNet 90%, 89% and 89%; for InceptionV3 88%, 88% and 88%; and for Ensemble model 91%, 91% and 91% in the given order is recorded. We have also computed the training-validation accuracy curves for the proposed method. The training-validation accuracy curve for DenseNet and InceptionV3 models are demonstrated in Fig. 6 and Fig. 7 respectively. The graphs exhibit that there is a high increasing rate of accuracy until 25th epoch and there-after 25th epoch, the graph becomes converge in both models. Another evaluation metric that is applied in this study is to visualize the classification performance using confusion matrix which is described in terms of correctly classified and wrongly classified test samples. The proposed method performance is presented in Figs. 8, 9 and 10 for DenseNet, InceptionV3, and Ensemble models respectively on HAM10000 dataset that contain seven classes. The diagonals of a confusion matrix from top-left to bottom-right are correctly classified samples, and all other cells out of this diagonal represent wrongly classified samples.

Table 4. Precision, recall, and F1-score for each class due to Ensemble approach.

Class name	Precision	Recall	F1-score
akiec	0.87	0.71	0.78
bcc	0.98	0.86	0.92
bkl	0.74	0.86	0.80
df	1.00	0.73	0.85
nv	0.95	0.96	0.95
mel	1.00	0.93	0.96
vasc	0.77	0.75	0.76

Table 5. The evaluation metrics(%) of the proposed methods.

ModelName	WeightedPrecision	WeightedRecall	WeightedF1Score	Accuracy	Loss
DenseNet	90	89	89	89.42	53.36
InceptionV3	88	88	88	87.82	55.55
Ensemble	91	91	91	90.91	38.33

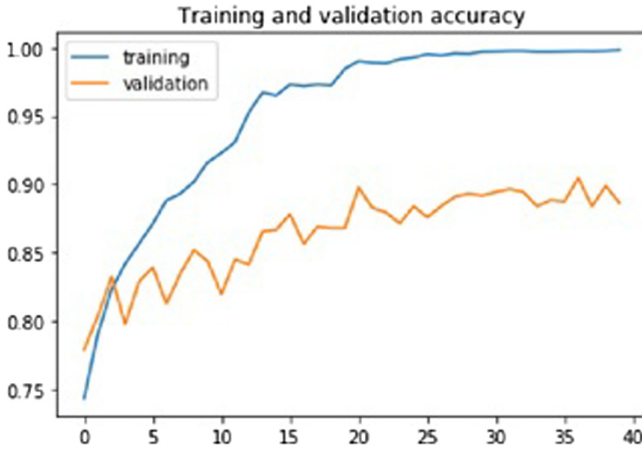


Fig. 6. The classification accuracy curve due to DenseNet model.

Finally, we have made a comparative study with the existing state-of-the-art methods that are validated on the HAM10000 dataset as shown in Table 6. The highest outcome of proposed architecture is indicated by making bold. The comparison in the table indicates that the proposed method achieves better when compared to existing algorithms.

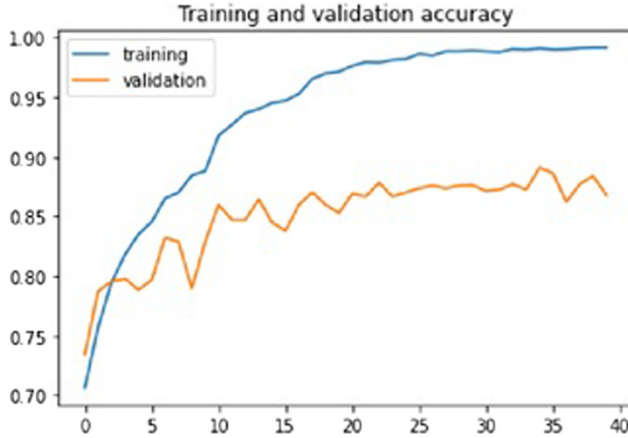


Fig. 7. The classification accuracy curve due to Inceptionv3 model.

Table 6. The Comparative analysis of the proposed method with the existing methods.

Method	Model used	Precision	Recall	F1 score	Accuracy
Khan et al. [24]	Resnet-50 and Resnet-101	90.14	89.71	-	89.9
SS Chaturvedi et al. [4]	MobileNet	89	83	83	83.15
Shahin et al. [37]	Ensemble (ResNet and InceptionV3)	86.2	79.6	-	89.9
MAR Ratul et al. [5]	InceptionV3	89	89	89	89.81
RA Pratiwi et al. [13]	VGG19	78.21	96.40	86.36	87.64
Our method (ensemble)	DenseNet and InceptionV3	91	91	91	90.91

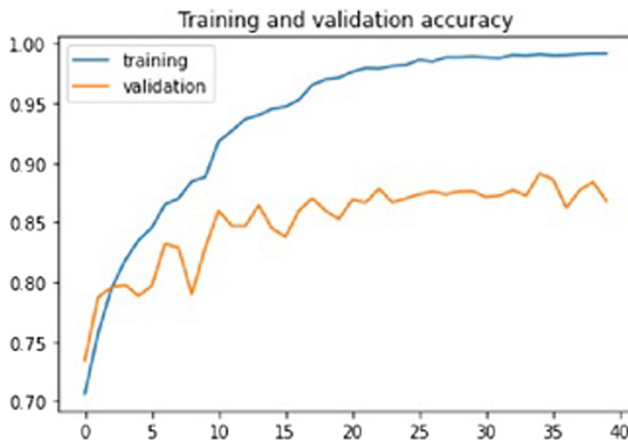


Fig. 8. The confusion matrix due to DenseNet model.

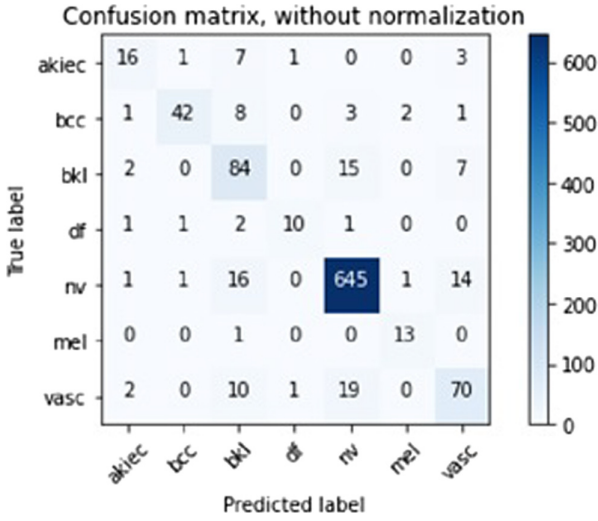


Fig. 9. The confusion matrix due to Inceptionv3 model.

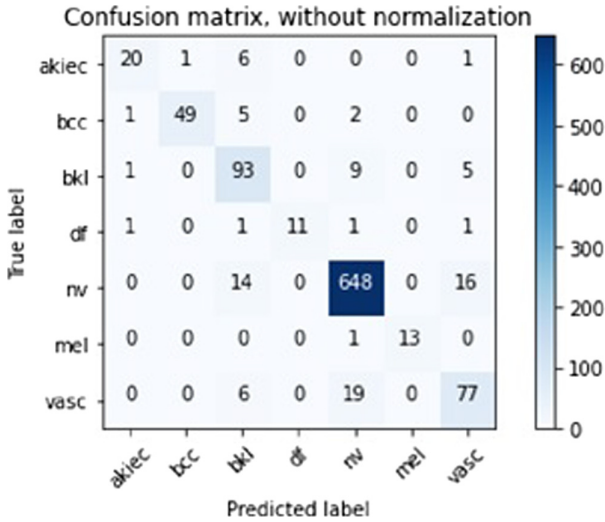


Fig. 10. The confusion matrix due to Ensemble model.

4 Discussion and Conclusions

The diagnosis and detection of skin cancer is the complex task due imbalanced number of training samples. In the previous section, we have presented the experimental results with well-known metrics numerically and graphically. It is evident from the experimental results (see Tables 2 and 3) that the DenseNet and InceptionV3 scores the highest precision, recall and f1-score with respect to Melanocytic nevi class which has large number of samples. This class has 678 samples out of which 67% is considered for the test dataset. In the Ensemble model, there is no single class that possess a good score than the other classes with respect to all the evaluation metrics as indicated in the Table 4. But, almost in all evaluation metrics for all individual classes, the ensemble model achieves a better performance than DenseNet and InceptionV3 models. By observing the classification rate of every individual class obtained from the experimental analysis, it is clear that almost all classes that have large categories have better classification rates whereas those images from small classes are highly misclassified.

In addition, as the multi-class classification report shows in Table 5, the ensemble model records a better classification accuracy when compared with the individual models in terms of accuracy, loss, and weighted average of precision, recall, and f1-score. It achieves an accuracy of 91%, loss of 38.33% and an equal score for weighted precision, recall, and f1-score of 91% for the unseen test datasets. We have observed that concatenating two or more models together by using various ensembling techniques can improve the prediction capability and generalization ability of a classification model.

The confusion matrix also gives a clear illustration by comparing the True label and Predicted label for each sample in the test set. Even if most of the images are classified correctly, due to the presence of high similarity in the inter-class and the variability due to intra-class between images in some classes in the training data makes it impossible to reach high classification capability for each class. The comparative process indicates that the proposed method achieves better performance in terms of precision, f1-score, and accuracy. In terms of recall, our model is in the second rank next to Pratiwi et al. '[13]'. But our model is much better than this model in view of precision and hence better f1-score is registered by the proposed study.

In summary, in the current work, we have employed a new ensemble method for skin lesion classification by using deep learning. In our proposed method, the fine-tuned technique is applied on DenseNet and InceptioV3 networks. These algorithms were pretrained on ImageNet dataset which is a large image dataset with 1000 different classes. For our task, we remove the last fully-connected layer of the algorithms and replace it with a new fully-connected layer that is appropriate for our classification task on the HAM10000 dataset which has 7 classes. After we train the algorithms on our dataset separately, we concatenate the results by the average ensemble technique. Experimentation is performed on the test dataset and achieved an accuracy of 89.42%, 87.82% and 91% for DenseNet, InceptionV3 and ensemble models respectively. From the experimental result,

we observe that through the fusion of two methods, the model scores better performance than individual architectures. Moreover, the comparative study shows that the result of the proposed method achieves better performance in most of the parameters, when compared to the existing state-of-the-art methods. Although the proposed method has improved accuracy, it still needs improvement to tackle the overfitting problem and to increase the accuracy by using different regularization techniques.

References

1. WHO: Skin cancers WHO (2017)
2. Narayanan, D.L., Saladi, R.N., Fox, J.L.: Ultraviolet radiation and skin cancer. *Int. J. Dermatol.* **49**(9), 978–986 (2010)
3. Parkin, D., Mesher, D., Sasieni, P.: 13 cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010. *Br. J. Cancer* **105**(2), S66–S69 (2011)
4. Chaturvedi, S.S., Gupta, K., Prasad, P., et al.: Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. arXiv preprint [arXiv:1907.03220](https://arxiv.org/abs/1907.03220) (2019)
5. Ratul, A.R., Mozaffari, M.H., Lee, W.S., Parimbelli, E.: Skin lesions classification using deep learning based on dilated convolution, bioRxiv, p. 860700 (2019)
6. Gandhi, S.A., Kampp, J.: Skin cancer epidemiology, detection, and management, medical. *Clinics* **99**(6), 1323–1335 (2015)
7. Damsky, W., Bosenberg, M.: Melanocytic nevi and melanoma: unraveling a complex relationship. *Oncogene* **36**(42), 5771–5792 (2017)
8. Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. *JAMA Dermatol.* **151**(10), 1081–1086 (2015)
9. Thörn, M., Ponté, F., Bergström, R., Sparén, P., Adami, H.O.: Clinical and histopathologic predictors of survival in patients with malignant melanoma: a population-based study in Sweden. *JNCI: J. Natl. Cancer Inst.* **86**(10), 761–769 (1994)
10. Binder, M., Schwarz, M., Winkler, A., Steiner, A., Kaider, A., Wolff, K., et al.: Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch. Dermatol.* **131**(3), 286–291 (1995)
11. Gachon, J., Beaulieu, P., Sei, J.F., Gouvernet, J., Claudel, J.P., Lemaitre, M., et al.: First prospective study of the recognition process of melanoma in dermatological practice. *Arch. Dermatol.* **141**(4), 434–438 (2005)
12. Morton, C., Mackie, R.: Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *Br. J. Dermatol.* **138**(2), 283–287 (1998)
13. Pratiwi, R.A., Nurmaini, S., Rini, D.P.: Skin lesion classification based on convolutional neural networks. *Comput. Eng. Appl. J.* **8**(3), 203–216 (2019)
14. Barata, C., Ruela, M., Francisco, M., Mendonça, T., Marques, J.S.: Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Syst. J.* **8**(3), 965–979 (2013)
15. Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds.) *MLMI 2015. LNCS*, vol. 9352, pp. 118–126. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24888-2_15

16. Cicero, F.M., Oliveira, A.H.M., Botelho, G.M., da Computação, C.D.C.: Deep learning and convolutional neural networks in the aid of the classification of melanoma. In: Conference on Graphics, Patterns and Images, SIBGRAPI (2016)
17. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115–118 (2017)
18. Lopez, A.R., Giro-i Nieto, X., Burdick, J., Marques, O.: Skin lesion classification from dermoscopic images using deep learning techniques. In: 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), pp. 49–54. IEEE (2017)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012)
20. Codella, N.C., Nguyen, Q.B., Pankanti, S., Gutman, D.A., Helba, B., Halpern, A.C., et al.: Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J. Res. Dev.* **61**(4/5), 1–5 (2017)
21. Harangi, B.: Skin lesion classification with ensembles of deep convolutional neural networks. *J. Biomed. Inform.* **86**, 25–32 (2018)
22. Tan, T.Y., Zhang, L., Neoh, S.C., Lim, C.P.: Intelligent skin cancer detection using enhanced particle swarm optimization. *Knowl.-Based Syst.* **158**, 118–135 (2018)
23. Hekler, A., Utikal, J.S., Enk, A.H., Berking, C., Klode, J., Schadendorf, D., et al.: Pathologist level classification of histopathological melanoma images with deep neural networks. *Eur. J. Cancer* **115**, 79–83 (2019)
24. Khan, M.A., Javed, M.Y., Sharif, M., Saba, T., Rehman, A.: Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. In: 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1–7. IEEE (2019)
25. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multisource dermoscopic images of common pigmented skin lesions. *Sci. Data* **5**, 161–180 (2018)
26. LeCun, Y.: The MNIST database of handwritten digits (1998). <http://yann.lecun.com/exdb/mnist/>
27. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)
28. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
29. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
30. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
31. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:14091556](https://arxiv.org/abs/1409.1556) (2014)
32. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
33. Zhang, J., Lu, C., Li, X., Kim, H.J., Wang, J.: A full convolutional network based on DenseNet for remote sensing scene classification. *Math. Biosci. Eng.* **16**(5), 3345–3367 (2019)

34. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint [arXiv:150203167](https://arxiv.org/abs/1502.03167) (2015)
35. Chollet, F., et al.: Keras (2015)
36. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al.: Tensorflow: large scale machine learning on heterogeneous distributed systems. arXiv preprint [arXiv:160304467](https://arxiv.org/abs/1603.04467) (2016)
37. Shahin, A.H., Kamal, A., Elattar, M.A.: Deep ensemble learning for skin lesion classification from dermoscopic images. In: 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), pp. 150–153. IEEE (2018)