

IMPLEMENTATION OF AUTOMATIC SKIN LESIONS DIAGNOSIS—A DEEP LEARNING ENSEMBLING APPROACH

Tajjour SALWAN¹, Patel SHOBHIT K.^{2,3,*}, Parmar CHANDRASINH^{1,2},
Vigneswaran DHASARATHAN⁴

¹Information and Communication Technology Department, Marwadi University,
Rajkot-360003, India

²Electronics and Communication Department, Marwadi University,
Rajkot-360003, India

³Department of Computer Engineering (AI and Big Data), Marwadi University,
Rajkot-360003, India

⁴Department of Physics, University College of Engineering, Ramanathapuram campus,
Tamilnadu, India

*Corresponding Author: Patel SHOBHIT K.

(Email: shobhitkumar.patel@marwadieducation.edu.in)

(Received: 15-Oct-2020; accepted: 19-Dec-2020; published: 31-Dec-2020)

DOI: <http://dx.doi.org/10.25073/jaec.202044.306>

Abstract. *Skin is the largest organ of the human body, it protects them from the external environment. However, the number of new skin cancer cases has been going up in the last few years. Fortunately diagnosing skin cancer in earlier stages could increase the probability of treating cancer. Due to machine learning (ML) and deep learning (DL) has achieved great success in different fields and has shown an outstanding performance in computer vision applications, lots of ML Applications is concerned in this domain. The main goal of this study is to find a computer-aided diagnosis system that can be powerful in early detection, which saves time and money. A simple pipeline is used to train different types of convolutional neural networks (CNNs), choosing the best four networks and ensemble them together without any prior segmentation or repainting. By using the HAM10000 dataset for training and testing, a high-accurate model is achieved which can classify seven kinds of skin lesions very precisely.*

Keywords

Skin lesion, machine learning, deep learning, medical images, ensemble modeling.

1. Introduction

Skin cancer became more common nowadays because of the Ozone hole, Global warming and the high-level of pollution, especially in developing countries. According to American Cancer Society statistics in 2019, the number of patients is increasing, whereas 96,480 new melanomas (the deadliest skin cancer) will be diagnosed and 7,230 people are expected to die because of melanoma [1]. Dermatologists prefer to start with non-invasive methods to detect the lesions (such as ABCDE, 7-points checklist and more methods) to avoid excisional biopsies which are invasive, painful, costly and time-consuming. Dermatoscopy is a non-invasive examination technique that is used widely by der-

matologists to improve diagnosis better than examination with the naked eyes [2]. Figure 1 displays some dermoscopic images for kinds of lesions.

Due to the imbalanced numbers of dermatologists and patients, and the huge amounts of money expensing annually at treatments, computer-aided diagnosis systems (CAD) become much involved in this domain. In the last few years, Deep Learning takes a significant role in healthcare applications and especially in Skin Lesion applications, due to the increasing amount of available datasets, which are classified as a medium-size dataset. Indeed, the state-of-the-art image-based CAD achieved significant results that can be compared to dermatologist experts who have more than 10 years of experience. Moreover, researchers have compared between machine-learning algorithms predictions and dermatologists diagnoses on the HAM10000 dataset and they found that the best machine-learning algorithms achieved a mean of (6.65 and 7.94) more correct diagnoses than experts and average human readers respectively [2].

Regarding the previous work in this domain and because of the complex surface of the skin, the process of diagnosis can be split up into many steps, starting with image acquisition and ending with classification. Image process techniques are used to remove noise from images or to take the best features from the image, such as DullRazor or other techniques to remove hair and rulers [3]-[5], resizing images to fit networks, remove black boundaries [6] and colour space transformation [7], etc. In addition to that, image segmentation is a very important step that removes the normal skin area, which is not required. There are three approaches to do segmentation, one by using image-processing techniques such as Otsu thresholding [8], active contours, GrabCut [9] and relative fuzzy membership degree (RFMD) [4, 10]. The second approach is to use Deep Learning to train networks that can segment the lesion effectively, such as U-net, ResUNet [11, 12], etc. The third approach is to use a hybrid technique such as to use YOLO with GrabCut [9]. Attribute detection is also used to detect attributes of the lesion such as globules and streaks, Deep CNNs are used for this task such as Multi-Task U-Net

[13]. Finally classification task, one approach is to extract some features manually form the images, then using these features to feed a neural network [5], or other classification methods such as KNN, SVM, decision tree [14, 15], etc. Another approach is to use CNNs which performed well in other fields and fine-tuning them to fit the skin cancer datasets [16, 17] or using those pre-trained networks to extract features from the dataset and classify them using classical classification methods such as SVM [15, 18]. Furthermore, ensembling techniques performed better than others [15]-[17], [19].

Developing an automatic system that could address the techniques, technologies, and terminology used in skin imaging to help dermatologists to diagnose seven kinds of skin lesions. However, due to the complexity of the dataset, the existed models still lack the accuracy, although many efforts were applied to find the best solution to this problem. Therefore, the main objective of this study is to find a highly accurate model that could classify these seven skin lesions very precisely without applying any kind of segmentation or denoising to the images.

The proposed idea is to find multiple networks that perform well on the task of skin lesion classification, then fuse them together to get better results. The first step is implementing some of the convolutional neural networks that performed well on different tasks of classification, then retrain them with the HAM10000 data, by adding class weights to deal with the imbalanced data and observing their behaviour on this data, after training it is clear that each network is sensitive for some classes more than the others. The networks are chosen due to their sensitivity to all classes. Then the networks are ensembled together by averaging them. The final model is tested by using 2003 images from the holdout test set (20% of the dataset) and the accuracy is found to be increased by 3% to 4% in comparison to the accuracy of each network individually.

2. HAM10000 dataset

Unlike in similar studies, which treated this problem as a binary classification of malign-

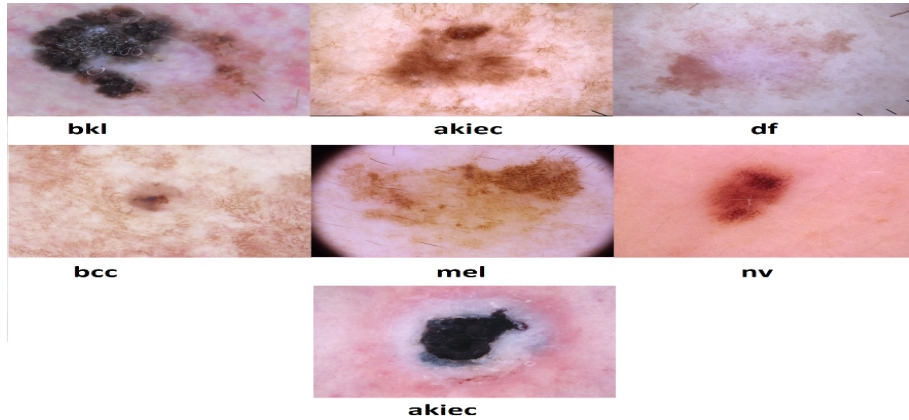


Fig. 1: Seven kinds of lesions images taken by Dermatoscopy.

nant or benign, this study will take seven kinds of skin lesions to classify, this will make the task more complex. The data set used in this study is HAM10000, it contains seven classes (actinic keratosis / Bowen's disease (intraepithelial carcinoma) (AKIEC), Melanoma (MEL), melanocytic nevus (NV), actinic keratosis / Bowen's disease (intraepithelial carcinoma) (AKIEC), benign keratosis (solar lentigo / seborrheic keratosis/lichen planus-like keratosis) (BKL), dermatofibroma (DF), basal cell carcinoma (BCC), and vascular lesion (VASC) with more than 10000 dermoscopic images. More than 50% of lesions are confirmed through histopathology, the ground truth for the rest of the cases is either follow-up examination, expert consensus, or confirmation by in-vivo confocal microscopy [20]. As shown in Fig. 2 some classes have less than 200 samples (such as DF and VASC) where the NV class has more than 6000 samples which make the training process difficult because the predictions of these networks are more likely to be NV.

This data suffers from variance, noise, and imbalance. Moreover, adding segmentation steps in front of classification would give undesirable results if the segmentation is not as perfect as it should be, this will add more noise to the data. In addition to that, removing the hair from images using the DullRazor technique will affect the texture of the images, due to the median filter and repainting the hair's pixels in the original images with neighbour pixels. This technique also relies on the black thin hair only. So that

this study will avoid any kinds of de-noising or segmentation steps.

3. CNN architecture

A. Convolutional layer

It is the fundamental layer in CNN and it is a mathematical computation to combine two groups of data. Convolution filter on the input will give the feature map, by concatenating these feature maps we get the convolution layer. The filters weights have random values at the beginning of training during training the optimizer change these weights to improve the results of the model in the backpropagation process.

B.Pooling layer

Pooling is used to reduce the dimensions, just height and width and keep the depth as it. This will help in reducing the number of parameters so that training will be easier, faster and combat overfitting.

C. Fully-connected layer

It is a normal dense layer where each node is connected to all nodes in the previous layer. The convolutional layers output will be the features of the images so that FC layers are added to classify these features into a number of classes. Before connecting the features with the FC layers, it is better to flatten these 3D features to make

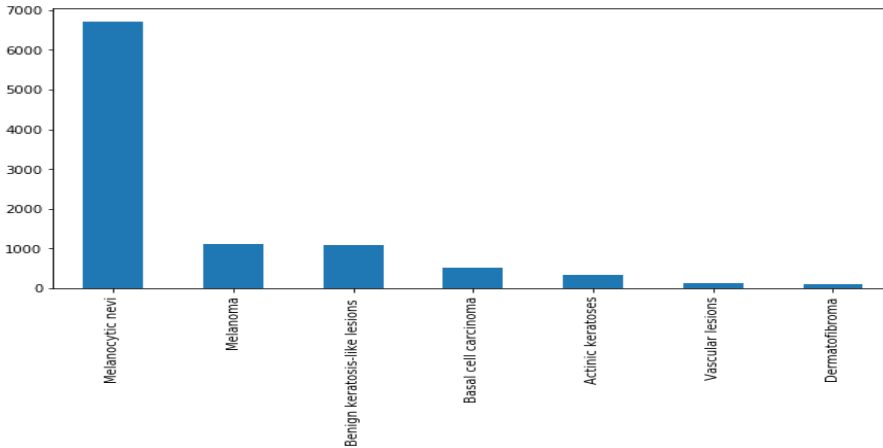


Fig. 2: The distribution of the dataset where the x-axis displays the classes and the y-axis displays the number of samples in each class.

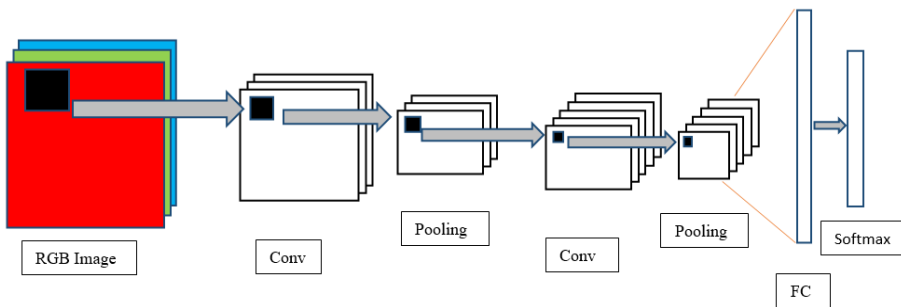


Fig. 3: CNN architecture.

them single-dimensional vector without losing any information.

D. Softmax layer

It is a dense layer with a number of nodes denotes to the classes, and Softmax activation function is applied to it. This function computes the probability of each class to adapt the results into values between 0 and 1 using this equation:

$$P(y = j|\theta^i) = \frac{e^{\theta^{(i)}}}{\sum_{j=0}^k e^{\theta^{(i)}}}, \tag{1}$$

where $\theta = \sum_{i=0}^k w_i x_i = W^T * X$

4. Proposed Methodology

The pipeline shown in Fig. 4 explains the followed procedure for building the classification model. In the beginning, the HAM10000 dataset is divided into five folds, as shown in Tab. 1, in order to apply cross-validation. Indeed to ensure maintaining enough number of samples in each class for testing, only five folds were chosen because some classes have less than 150 samples so that at least 19 samples will be held for testing in these classes. Whereas at every single check, only one fold will be used for testing and the others for training. This helps in finding the actual accuracy of each network and prevents testing using biased data, which could be a powerful technique for monitoring the behaviour of the model and for generalization.

A. Pre-Processing

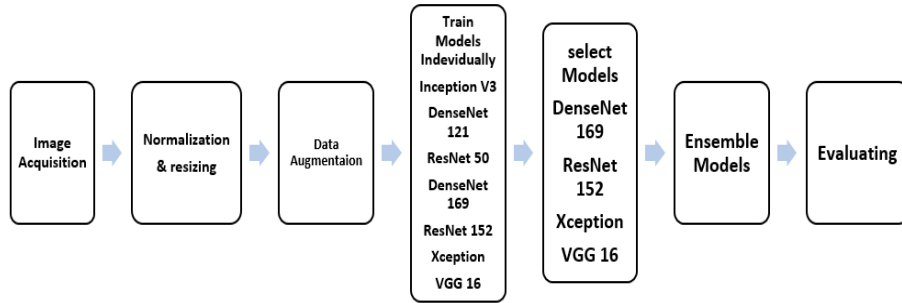


Fig. 4: The proposed pipeline for lesion classification.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
AKIEC	69	64	62	71	61
BCC	93	118	107	89	107
BKL	228	230	230	187	224
DF	28	26	20	22	19
MEL	226	228	219	229	211
NV	1338	1314	1332	1366	1355
VASC	21	23	33	39	26

Tab. 1: The distribution of the folds.

The original size of the images in HAM10000 dataset is 600x450. Therefore, two preprocessing steps are applied: first, resizing all images to 256x192 using bicubic interpolation, this step is important to reduce the computational cost and to decrease the probability of overfitting, besides, 256x192 was chosen to maintain the difference between width and height in the origin images. The second step is to normalize the images to range between zero and one to feed into the Networks by dividing each pixel by 255, which is the maximum value that pixels can have (pixels original range is between 0 and 255). Normalization has a good influence to increase converging time since all data dimensions will have the same scale.

B. Data Augmentation

Increasing the number of images in the dataset has a good influence on the training process by reducing the overfitting and helping the model to generalize well. By taking into consideration that the augmentation methods will not damage the useful information in the images. These are the selected methods: horizontal flipping, vertical flipping, randomly rotate images in a range between +10 and -10 degrees, randomly

zooming image by zooming factor 0.1, and randomly shifting images horizontally and vertically by shifting factor 0.1.

C. Evaluating metrics

Many kinds of metrics are used to evaluate the proposed models and to compare them with the previous works, and the used evaluating metrics are shown in Eqs. 2-7:

Precision:

$$Pre = \frac{TP}{TP + FP} \tag{2}$$

Recall:

$$Rec = \frac{TP}{TP + FN} \tag{3}$$

Accuracy:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

F1-Score:

$$f1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{5}$$

AUC: Area Under Receiver Operating characteristic (ROC) Curve: it is generated by plotting the TPR (y-axis) against FPR (x-axis).

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

Where TP: True Positives; TN: True Negatives; FP: False Positives; FN: False Negatives; TPR: True Positive Rate; FPR: False Positive Rate.

D. Deep Learning Models

This paper investigates the performance of some CNNs that is achieved a good performance on ImageNet. The predictions of models for the holdout test set are reported in Tab. 2.

All of the chosen networks are trained on more than 1.5 million images in the ImageNet dataset to classify 1000 different classes. Therefore, using transfer learning can help to train the networks with a small dataset and a small number of epochs.

All networks were prepared and trained as following: First, taking the pre-trained networks with their original weights and replace the original fully connected (FC) layers by new FC layers, and then initialized their weights using Glorot_Uniform for the kernel weights and zeros for biases weights. Global Average Pooling technique is used to connect between the last convolutional layer and the new FC layer with 512 neurons, because it is very effective in flattening the convolutional layers and reducing the dimension of the output by taking kernel with the same dimensions of the feature map, then replacing each channel with the average value. This has many benefits such as reducing the number of parameters, resisting overfitting and making the model less sensitive to the location of the lesion in the source images in addition to the down-sampling in each layer in the original architecture. A dropout layer is added after the FC layers with a 0.5 dropout factor to resist overfitting by ignoring some unites from the FC layers which become sparse. Finally, Softmax layers are added on the top of each network to obtain the classification results. Adam optimizer is used to train models with an initial learning rate 0.001, then reducing its value by 0.5 factor after 3 epochs if the accuracy did not get better. Before training the entire network, it is necessary to train just the new layers for 3 epochs while the other layers are frozen. This step has a good influence on training because the loss will be very

big at the beginning of the training because of the new layers, whereas pushing the weights of the pre-trained layer as big as the weights of new layers will defect the trained layers. Weighted categorical cross-entropy represented in formula (8) is used to deal with data imbalance.

$$WCE(p, q) = -wi \sum_x p(x) \log(q(x)) \quad (8)$$

Where wi is the weight of class i , $p(x)$ are the ground truth label and $q(x)$ is the predicted value from the Softmax layer. The actual weights of classes are $wi = [5.0, 4.0, 2.0, 14.0, 2.0, 1.0, 10.0]$ which refer to the classes ['akiec', 'bcc', 'bkl', 'df', 'mel', 'nv', 'vasc'] respectively. However, during the training additional weight is added for different classes for different networks to make each network has more sensitivity to a specific class. For example, 6.0 is set as a weight for the first class (AKIEC) while training the VGG 16 Network and due to that, it is obvious that this network shows the best results for the first class. Figure 5 displays the confusion matrix and AUC-ROC for the best networks, which will be used.

As ResNet152, VGG16, DensNet169, and Xception achieved the best accuracy among the other networks, cross-validation technique with 5 folds was applied in order to ensure that these models are unbiased to any class. The results of cross-validation are shown in Tab. 3, where a slight varying in the accuracy was detected. Therefore, the training and testing data could be considered as unbiased.

E. Fusing models

After training all networks individually using the same data and the same procedure, ResNet152, VGG16, DensNet169, and Xception were chosen, in order to ensemble them due to their good behaviour at classifying different classes. Thus, each network will be superior to a particular class. Then, the trained networks are merged by calculating the average of the outputs of the last layer for each network instead of taking only one output value from each network, then selecting the highest probability. Indeed, this strategy could exploit the ability of each network to distinguish between all classes. This could be done by taking the

Lesion/ network	Inception V3	DenseNet 121	ResNet 50	DenseNet 169	ResNet 152	Xception	VGG 16
AKIEC	51	57	55	63	61	60	67
BCC	73	81	72	81	76	76	80
BKL	71	73	72	75	72	76	74
DF	69	44	68	65	79	71	51
MEL	64	66	67	69	71	70	68
NV	93	93	93	93	93	94	93
VASC	82	90	90	84	91	84	83
accuracy	85	85	85	86	86	87	86

Tab. 2: F1-score and accuracy of different kinds of networks on the holdout test set after fine-tuning.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	AVG
VGG16	87.620	86.352	85.540	86.961	85.439	86.3824
Xception	87.048	86.555	86.251	87.570	88.026	87.09
DenseNet 169	86.859	86.352	85.773	86.193	86.316	86.2986
ResNet 152	85.701	85.954	86.715	85.904	86.309	86.1166

Tab. 3: Accuracy of the models using cross-validation technique.

outputs from the last layer of each network after applying a linear activation function instead of using Softmax, and then, all outputs will be added to each other to compute the average of each class individually. Thus, the model will give seven values, a value for each class, and the result will be the maximum value of them. This combines the different behaviour of the networks on the classes to give even better results that none of the models could achieve individually. As shown in Tab. 4 that after ensembling, the network achieved F1-score better than all networks for four different classes. For instance, the model achieved 88% for the BCC class, where the other networks (DenseNet 169, ResNet152, Xception and VGG16) achieved (81,76,76,80)% F1-score respectively. Similarly, the proposed model achieved (81-77-96)% F1-score for (BKL, MEL, NV) classes where the best values achieved by individual models were (76-71-94)% respectively.

Comparing the AUC for the proposed model and its individual networks, from Figs. 4 and 5, using the holdout dataset it is noticeable that the proposed model has a similar behaviour to its parents in distinguishing between classes precisely with a slight improvement. Furthermore, the confusion matrices show a significant improvement in the accuracy of the proposed

model compared to its parents. Whereas the accuracy is increased from (87%, 86%) to 90%.

F. Computational Cost

The computational cost of the model is an important parameter to define the quality of the model. It detects the training time of each network and the execution time in the online mode of the model. In many cases, the purpose of the models is to work online so that it should be fast and suitable for the device that the model will work on it. Indeed the computational cost of the proposed model is high due to using four different architecture for deep neural networks which means the number of the parameters in the model will be very high and the execution time will be high too. To decrease the cost global average pooling was used in each network, in addition to using a small number of neurons in the Dense layers of each network 512 instead of 4096 or 2048 like the original networks. Therefore, the number of parameters in the proposed model is 109,817,668 which takes around 8 seconds to load all the weights of the model using google collabs configurations.

Lesion/ network	DenseNet 169	ResNet 152	Xception	VGG 16	Proposed Model
AKIEC	63	61	60	<u>67</u>	64
BCC	81	76	76	80	<u>88</u>
BKL	75	72	76	74	<u>81</u>
DF	65	<u>79</u>	71	51	71
MEL	69	71	70	68	<u>77</u>
NV	93	93	94	93	<u>96</u>
VASC	84	<u>91</u>	84	83	86
accuracy	86	86	87	86	<u>90</u>

Tab. 4: Comparison between the ensembling model with the individual models for both F1-score and accuracy.

5. Results

The normalized confusion matrix and AUC-ROC are shown in Fig. 6. Moreover, the final model achieved 98.85% for top-3 accuracy, 96.61% for top-2 accuracy, 90.02% for top-1 accuracy, 98.4% for AUC, 90.34% for precision, and 90.00% for recall. We have involved the state-of-the-art approach for a comprehensive comparison with the proposed model, which is evaluated with the same metrics for the same purpose. Table 5 displays the results of previous works in addition to the top-5 ranked models in ICIS competition for 2019, which is an annual competition created to support the effort of researchers and developers worldwide to find the solution to this complicated problem. As shown in Tab. 5, the proposed model achieved better performance than any other model in term of accuracy, precision, recall and AUC. Although some of these studies used the same number of networks such as [17, 23]. Furthermore, using shallower networks like in [22] or using less number of networks like in [19, 22] could be better in term of complexity and computational cost but it is significantly worst than the proposed model. Moreover, one network could be more efficient for computers with low performance. So that many models were implemented using only one model, even by exploiting predefined networks like efficientNetB4 [24] for this task or by building customized networks [25]; however, they still lack in the accuracy. In addition, the proposed model performs better than other models which are used to classify less number of lesions (two [22] or three [23]), or even for the models where different loss function was used for training [24].

On the other hand, the proposed ensembling technique was better than other techniques like simple average [19], voting technique [23] or using SVM for fusing networks [22]. In addition to models that applied transfer learning without retraining the whole layers where they preserve the original weight of networks. AUC and Recall results are much higher due to the powerful training, ensembling techniques and the power of the weighted categorical cross-entropy loss in treating imbalanced data.

Therefore, the proposed model can be helpful even if it is used to give multiple diagnosis choices to the dermatologists with high accuracy and keep the last decision for them.

6. Conclusions

Convolutional Neural Network applications for skin lesion detection are promising. Therefore, this study investigated the power of the proposed ensembling approach, where four kinds of CNN architectures are trained individually using transfer learning and weighted categorical-cross-entropy loss function to rebalance the training process. After that, the proposed ensembling technique was used to fuse the last layer from every network to raise the accuracy significantly and overcome their weaknesses, even for classes which lack in the training samples.

Averaging the last layer of each network, instead of taking the output of each one and averaging them, was able to increase the accuracy of the models from 86-87% to 90%. The proposed model was able to achieve high accu-

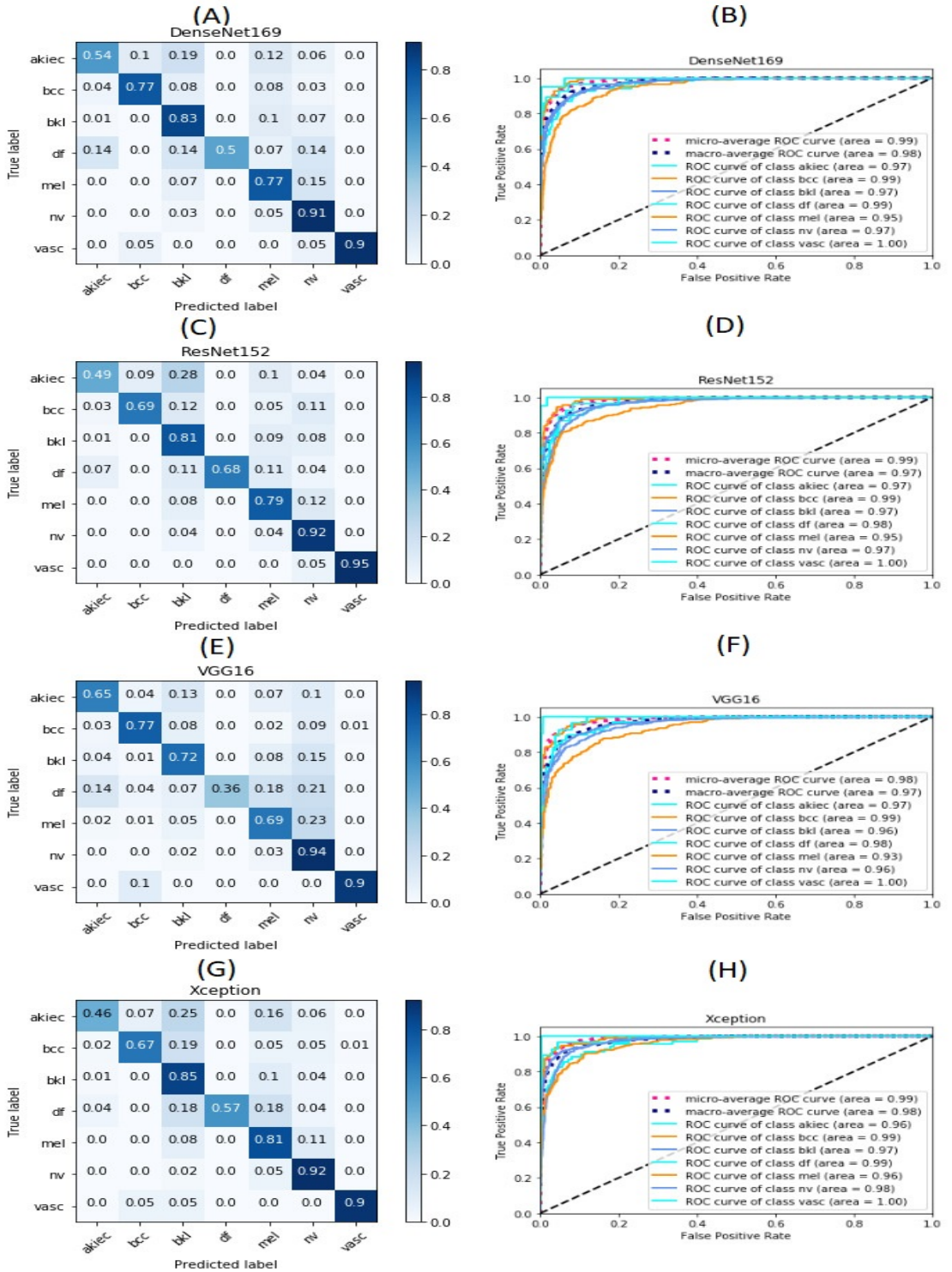


Fig. 5: The behavior of the selected networks, Confusion matrix on the left and AUC-ROC on the right. DenseNet169 (A-B), ResNet152 (C-D), VGG16 (E-F), and Xception (G-H).

paper	Networks	Precision	Recall	Accuracy	AUC
[17]	PNASNet-5-Large	-	-	76	-
	InceptionResNetV2			70	
	SENet154			74	
	InceptionV4			67	
	Ensemble			73	
[21]	Top-5 ranked models in ISIC 2019 Competition.			88.6	
				88.0	
				85.4	
				85.4	
				84.8	
[19]	ResNet50 + Inception V3	86.2	79.6	89.9	-
[22]	AlexNet + VGG16 + ResNet18	-	-	-	90.69
[23]	AlexNet + VGG16 + ResNet18 + Inception	-	-	86.6	89.1
[24]	EfficientNetB4 with CLF	89.0	86.13	89.97	-
[25]	Customized model	-	-	77.98	-
Our study	Proposed Ensemble Model	90.34	90.00	90.02	99

Tab. 5: Comparison between the proposed model and the previous works.

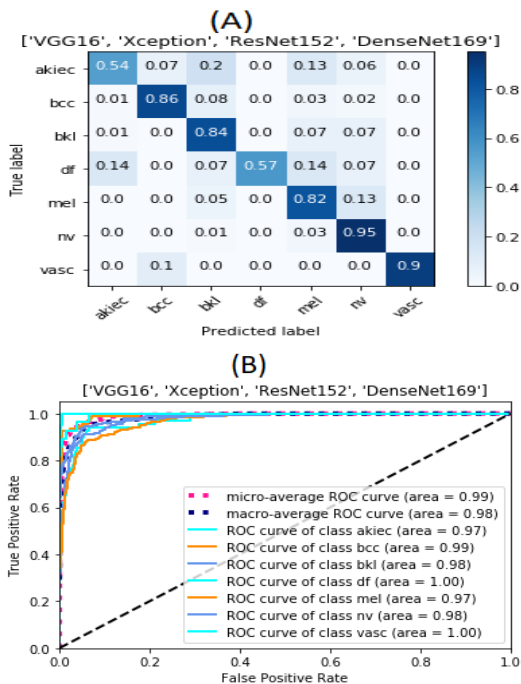


Fig. 6: The proposed model results (A) confusion matrix and (B) ROC curve..

racy, although no other kinds of segmentation, de-noising or colour-space conversion steps were added to the classification pipeline. Indeed the proposed model needs large memory because of

using four different architectures. However, this can be ignored regarding its good performance.

Finding useful methods for segmentation and de-noising will give better results due to removing harmful signals in addition to reduce overfitting. Furthermore, the number of networks could be extended to include additional CNNs since choosing shallower networks could reduce overfitting and decrease the required memory for the final model as well.

References

- [1] Melanoma Skin Cancer Statistics. (2019). <https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html>
- [2] Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., Cabo, H., ... & Lallas, A. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *The lancet oncology*, 20(7), 938-947.
- [3] Lee, T., Ng, V., Gallagher, R., Coldman, A., & McLean, D. (1997). *Dullrazor®: A*

- software approach to hair removal from images. *Computers in biology and medicine*, 27(6), 533-543.
- [4] Razmjoooy, N., Mousavi, B. S., Soleymani, F., & Khotbesara, M. H. (2013). A computer-aided diagnosis system for malignant melanomas. *Neural Computing and Applications*, 23(7-8), 2059-2071.
- [5] Messadi, M., Bessaid, A., & Taleb-Ahmed, A. (2009). Extraction of specific parameters for skin tumour classification. *Journal of medical engineering & technology*, 33(4), 288-295.
- [6] Yang, T., Chen, Y., Lu, J., & Fan, Z. (2019). Sampling with level set for pigmented skin lesion segmentation. *Signal, Image and Video Processing*, 13(4), 813-821.
- [7] Elgamal, M. (2013). Automatic skin cancer images classification. *IJACSA International Journal of Advanced Computer Science and Applications*, 4(3), 287-294.
- [8] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., & Ellinge, I. (2019, May). Skin lesion classification using hybrid deep neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1229-1233). IEEE.
- [9] Ünver, H. M., & Ayan, E. (2019). Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics*, 9(3), 72.
- [10] Narayana, C. H. V., Reddy, E. S., & Prasad, M. S. (2012). A new method for gray level image thresholding using spatial correlation features and ultrafuzzy measure. *Global Journal of Computer Science and Technology Graphics & Vision*, 12(15).
- [11] Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., & Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, 6(1), 014006.
- [12] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
- [13] Chen, E. Z., Dong, X., Wu, J., Jiang, H., Li, X., & Rong, R. (2018). Lesion Attributes Segmentation for Melanoma Detection with Deep Learning. *bioRxiv*, 381855.
- [14] Shimizu, K., Iyatomi, H., Celebi, M. E., Norton, K. A., & Tanaka, M. (2014). Four-class classification of skin lesions with task decomposition strategy. *IEEE transactions on biomedical engineering*, 62(1), 274-283.
- [15] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., & Ellinge, I. (2019, May). Skin lesion classification using hybrid deep neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1229-1233). IEEE.
- [16] Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics*, 86, 25-32.
- [17] Milton, M. A. A. (2019). Automated skin lesion classification using ensemble of deep neural networks in isic 2018: Skin lesion analysis towards melanoma detection challenge. *arXiv preprint arXiv:1901.10802*.
- [18] Premaladha, J., & Ravichandran, K. S. (2016). Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. *Journal of medical systems*, 40(4), 96.
- [19] Shahin, A. H., Kamal, A., & Elattar, M. A. (2018, December). Deep ensemble learning for skin lesion classification from dermoscopic images. In *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 150-153). IEEE.
- [20] Tschandl, P. Harvard Dataverse <https://doi.org/10.7910/DVN/DBW86T> (2018)

- [21] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & 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.
- [22] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., & Ellinge, I. (2019, May). Skin lesion classification using hybrid deep neural networks. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1229-1233). IEEE.
- [23] Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics*, 86, 25-32.
- [24] Pham, T. C., Doucet, A., Luong, C. M., Tran, C. T., & Hoang, V. D. (2020). Improving Skin-Disease Classification Based on Customized Loss Function Combined With Balanced Mini-Batch Logic and Real-Time Image Augmentation. *IEEE Access*, 8, 150725-150737.
- [25] Kanani, P., & Padole, M. (2019). Deep Learning to Detect Skin Cancer using Google Colab. *International Journal of Engineering and Advanced Technology Regular Issue*, 8(6), 2176-2183.

About Authors

Tajjour SALWAN His M. Tech. Degree in 2020 in Information and Communication Technology from Marwadi University, India. His main research area is to apply Machine Learning in Solar Energy. Presently, he is a faculty member and Research Scholar in Center of Excellence in Energy Science and Technology at Shoolini University.

Patel SHOBHIT K. has done his Ph.D. on Electronics & communication engineering at Charotar University of Science and Technology, Changa, India. He has completed his post-doctoral research in electrical and computer

engineering at the University of Nebraska, Lincoln, USA. He is currently working in the area of Photonics using AI, IoT, Optics and communication networks. He has published several research papers in high impact SCI journals. He has also filed 7 Indian patents on different novel research done by him. He received DST international travel grant in the year 2014 to present a paper in IEEE APS-URSI symposium at Memphis, USA. He also received DST International Travel Grant to present a paper in PIERS Symposium, NTU, Singapore. He is currently working on many graphene-based project and has received funding from SERB, DST, for his research. He has been honoured with awards for the achievements in area of the research field.

Parmar CHANDRASINH has received a master's degree in Electronics and Communication specialized in Communication System Engineering from Gujarat University, Ahmedabad. He has academic experience of over two decades and guided many projects in the field of sensors and IoT related to solved real-time problems. He is guiding master's students for the use of machine learning in the field of disease detection, defect detection and sports result prediction. He has been awarded by IUCEE for his project-based learning approach which he has practiced for more than fifty IoT based project guidance for graduate and master degree students. He is currently guiding a few funded projects in IoT and Machine learning related problems. His key interest is design thinking, sensors and IoT and Machine learning.

Vigneswaran DHASARATHAN His major research includes RF and Photonics for artificial intelligence applications, sensing application and higher speed data compatible networks. He is also working for biosensor, nonlinear optical sensor, metamaterial, super-conductor properties and silicon photonics. He has published 60 papers in refereed national/international journals and more than 15 papers in conferences.