Real-Time Recognition of Malignant Skin Lesions using Ensemble Modeling

V Kumar\(^1\) and T Choudhury\(^2\)*

\(^1\)Amity University, Noida, Uttar Pradesh, India
\(^2\)Department of Informatics, School of CS, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India

Received 31 July 2018; revised 19 December 2018; accepted 15 January 2019

Similarities between early malignant skin lesions and benign skin lesions make it challenging to correctly differentiate between the two. In this paper, a hybrid method for classification of skin lesions using an ensemble of deep predictive models (Feedforward Neural Network, Image Histogram Classification using KNN, LeNet-5 CNN architecture, VGG-11 CNN architecture) is proposed to correctly identify Malignant skin lesions. The image dataset contains multiple images of Benign (Nevus, Seborrheic Keratosis) and Malignant (Melanoma, Carcinoma) skin lesions supplied by the ISIC (International Skin Imaging Collaboration) Archive, an open source organization aiming for progressive studies in skin cancer detection. The challenge is to correctly segment the region of interest (skin lesion) in the initial stages, despite hairy images. We develop an ensemble of three models which are not correlated with each other (a pre-requisite for ensemble modeling). We also develop a MATLAB (64-bit R2016a, The Math Works) based Graphical User Interface (GUI) which provides real-time results for uploaded skin lesion images. The aim of the paper is to propose a method for early detection of melanoma, which is not far behind the time-consuming traditional biopsy procedures.

**Keywords:** VGG-11 Convolutional Neural Network, KNN Classification, MLP Classification, Melanoma

**Introduction**

Melanoma often resembles non-cancerous lesions in shape and color; the exception is that the former advances to other body parts if not detected early. The problem-in-hand has received much attention in the scientific and medical disciplines\(^{1,2,4,6}\), with many researchers focusing on the automatic detection of skin lesions. Such studies are based on several distinguishing features such as asymmetry, color, border and diameter (ABCD parameters). However, the precision of such studies still lacks behind the traditional biopsy procedures (a time-consuming procedure). VGG-11\(^5\) (along with two other auxiliary classifiers) produced positive results (93%), which is promising when compared to the existing state-of-the-art\(^7\). We create a new classifier (ensemble model) based on the results of three individual classifiers which predicts the results with better accuracies. We also compare our results with the existing state-of-the-art in the conclusion section.

**Materials and methods**

**ISIC archive**

Using the Girder Rest API provided by ISIC, we obtained around 2000 image sets of skin lesions (1500 Benign, 500 Malignant). In our image database, excessive hair was observed which concealed the skin lesion and caused numerous problems for obtaining the ROI (region of interest). We follow a procedure\(^7\) which removes hair from the gathered dataset using image morphological operations. The resulting images may have lower quality, but the results show improvement in obtaining the region of interest.

**Pre-processing hairy images**

Bottom Hat is a morphological operation used in image processing to identify small details (darker than surrounding) from the given images. The difference between the closing of an image and the input image leads to the bottom hat transformation. It results in extracting objects that are (a) smaller or thinner than the surroundings (b) are darker than the surrounding. After applying bottom hat operation, we invert the image set to highlight the hair pixels with a dark color on a white background (Figure 1A(V)). After inversion, a flood fill operation is performed to remove small black spots which leave us with long strands of hair. (Figure 1A(VI)) We then use Otsu segmentation to obtain a black & white image. At this stage, we have successfully segmented the hairy part of skin lesions. Since bottom hat dilates the width of hair pixels, we use erosion to augment the width of

\*Author for Correspondence
E-mail: tanupriya@ddn.upes.ac.in
hairs in the binary image. (Figure 1A(VIII)). The tricky part is replacing each identified dark pixel with the surrounding areas of the pixel. To perform this step, we dilate the original image. After dilation, the dark pixels of hairs are replaced with the surrounding pixels. However, this leads to degradation of the quality of the whole image. Hence, we only replace the segmented parts of the original images with the dilated version (using Figure 1A(IX.) as a mask), as shown in Figure 1A(X). In Equation 2, \( F \) is the final image, \( I \) is the input image, \( D \) is the dilated image and \( E \) is the eroded image. We apply Equation 1 for all the pixels \( p \) of original image to obtain the final image.

\[
F(p) = \begin{cases} 
D(p), & E(p) \text{ is black} \\
I(p), & E(p) \text{ is white} 
\end{cases} \quad \forall \ p \in I
\]

**ABCD features and multilayer perceptron**

The ABCD (Asymmetry, Border, Color, and Deviation) features of skin lesion are well known\(^3\). In this paper, we follow the ABCD rule. We extract circularity, luminosity, deviation, smoothness, and diameter from all images based on ABCD rule. Malignant skin lesions have varying shades, bigger diameter, indefinite borders and are asymmetrical in nature. On the other hand, benign moles have a smaller diameter, uniform and definite borders, constant color (brown) and symmetrical in nature. However, most of the early malignant moles have a resemblance to benign moles until their evolution. We calculate five parameters\(^3\) (circularity, smoothness, luminance, diameter, deviation) related to the ABCD parameters. Fig. 2 (B) shows the sample results for parameter measurement (with mean values). The

---

Fig. 1 — (A) shows the procedure followed for hair removal. The steps are ordered chronologically using Roman numerals. (B) shows sample results from the image set after hair removal steps.
results of feature extraction from the region of interest are summarized in Fig 2 (A). The scatterplots in the upper diagonal shows that none of the features share a strong correlation. The diagonal show stacked histogram for the features in Benign and Malignant cases. The lower diagonal is a kernel density estimate of the parameters using the Gaussian kernel. Based on the histograms and kernel density estimate, malignant and benign skin lesions show maximum differentiation using color deviation and diameter.

**Multilayer perceptron network**

We use a non-linear MLP classifier\(^9\)\(^{-11}\) with ReLu activation at three hidden layers and another at the output layer. We define two output nodes, one for Malignant lesions and another for Benign lesions. The output is a two-element vector (using one-hot encoding). We use Adam optimizer (it adapts the learning rate with training) for back-propagation with an initial learning rate of 0.003. We used categorical cross-entropy as the loss function and trained the network for 1750 epochs (batch size 128). The input to the neural network was a 6-element vector normalized from 0 to 1 (Diameter, Smoothness, Deviation, Circularity, Luminance, Age, Gender). The whole dataset (1500 Benign images, 500 Malignant images) was divided into a training set (60%), a validation set (15%) and testing set (25%) using random shuffle. The training accuracy and loss graph are shown in Fig 3 (A). Despite good accuracy, the precision value is still 73% (Table 2). The precision value is greatly improved using Ensemble Modelling.
Image histogram and KNN classification

We use image histograms and KNN classifier for differentiating between benign and melanoma image set. The advantage of using image histograms for classification instead of raw pixels is the reduced number of features. If a 32 px * 32 px with 3 channels (3072 features) image is converted to HSV histogram with each of the three histograms having 9 bins, the total number of features for the classification procedure reduces to 9 * 9 * 9 (729) features. That is a 76% decrease in the number of features with minimal loss in image information. One problem encountered during the above procedure was choosing the optimal hyperparameters (optimal K value, optimal bin combination). Different combinations of these two parameters produced different accuracies. To resolve this, we use the brute force method to find the correct combination. Bin size (5-20) and value of K (0-100) were chosen for the experiment. The accuracy obtained during the experiment increased for k = 1 to 4, remained constant from k = 5 to 25 and decreased sharply prior to that. Using the optimal hyperparameters deduced (9 bins and k = 4), we find the accuracies reported in Table 1 and Table 2. The accuracy using histogram classification is an overall improvement over raw pixel classification (32*32*3 features).

Convolutional neural network (VGG-11 Architecture)

Prior to using CNN for binary classification, normalization of image-set was performed to reduce computational time and remove background noise. We combine the normalized channels (using their mean and standard deviation) of each image to obtain normalized images (from 0 to 1).

The images (224 * 224 * 3) are fed into VGG-11 CNN5 architecture. While LeNet model could not achieve an accuracy greater than 82% in the test cases (before overfitting), the VGG-11 model displayed exceptional results (validation accuracy of 91% and testing accuracy of 88%). The kernel size for VGG is 3*3 and pool size is 2*2. Table 1 and Table 2 displays the comparison of LeNet and VGG model. The model was trained using stochastic gradient descent (SGD) with a learning rate of 0.06. The input size of images was 224 * 224 *3. The accuracy and loss graph for VGG-11 is shown in Fig 3 (B).

Results and Discussion

For Ensemble Modelling to be useful, it is required that the classifiers used are not correlated. We have used stacking⁸ for building the final model. First, we observe the accuracy, sensitivity, specificity of each model for a 10-fold cross-validation (Table 1). Based on the observations recorded in Table 1, we form a correlation chart of accuracies (for 10-fold cross-validation) for each pair of models as shown in Table 2. From the correlation chart, we see that the accuracies of LeNet and VGG-11 are correlated to each other. Furthermore, the accuracies of KNN (histogram) and KNN (raw) pixel are correlated. Since LeNet and KNN (raw) have low precision values, we drop these two models from Ensemble Model. To determine the optimal weights to find the best possible accuracy, we use a small neural network (3 input node and 1 output node) model with no

Fig. 3 — (A) MLP training and testing accuracy and loss (1750 epochs). (B) VGG-11 training accuracy and loss (85 epochs).
hidden layers and ReLu activation. In Equation 3, ML (MLP model) is input node 1, VG (VGG model) is input node 2 and KN (KNN model) is input node 3. The classifiers used for ensemble use disparate methodologies. ML uses ABCD features of Malignant and Benign images, KN focuses on spatial color information of images and VG extracts features by itself and classifies them. The final output (Opt) of ensemble modeling is produced according to Equation 2.

\[ \text{Opt} = \text{ReLu} (W1 \times ML + W2 \times VG + W3 \times KN + B) \]

After training the model on the same training dataset assigned to each of the classifiers, we find \( W1 = 0.27669 \), \( W2 = 0.52391 \), \( W3 = 0.15940 \) and \( B = -0.0384 \). The VGG model has the highest weight and the other two classifiers act as boosters. The accuracy of ensemble modeling is shown in Table 2. Using ensemble modeling, we achieve an unprecedented accuracy of 93% (96.2% in Benign cases and 86% in Malignant cases) which is much better than any of the individual classifiers. The confusion matrix of ensemble modeling shows better precision and recall values than any of the individual models (Table 2).

### Conclusions

In this paper, we presented an automated system for detecting the severity of skin lesions (cancerous or non-cancerous). Our proposed method is a collaboration between three different classification methods. Using deep procedures for classification, we are able to differentiate melanoma from benign lesions with 93% testing accuracy within seconds.
using MATLAB GUI (video). The stacked ensemble of the three classification methods produces good precision and recall rate, which is better than most of the existing automatic detection studies.

Tamanna Tabassum et al.\textsuperscript{2} achieved an accuracy of 89\% using their proposed method. However, the paper did not mention their region of interest segmentation technique in case of lesions covered in hair. Since we take care of hairy images in the pre-processing step, we performed better in feature extraction. Furthermore, we use a collaboration of multiple classifiers to surpass their accuracy of 89\%. We obtain an accuracy of 93\% and a precision of 88\% using the ensemble model. Balazs Harangi\textsuperscript{4} proposed an ensemble of deep convolutional networks for automatic skin lesion detection. However, the method proposed by them is computationally expensive (an ensemble of AlexNet, ResNet, VGG-16, Google LeNet). However, our proposed method is equally efficient and produces results within seconds.

Acknowledgement

We are indebted to Amity University, Noida for providing the support and necessary resources for completion of the project.

References

6 Lopez A R, Skin lesion classification from dermoscopic images using deep learning techniques, \textit{13th IASTED International Conference on Biomedical Engineering} (Innsbruck, Austria), 20-21 February 2017
8 Güneş F, Wolfinger R & Tan P Y, Stacked Ensemble Models for Improved Prediction Accuracy, SAS Conference Proceedings (Orlando, Florida), 2-5 April 2017