International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

CLASSIFICATION OF MELANOMA FROM DERMOSCOPIC IMAGES USING DEEP LEARNING

Hima H A[1], Celine Mary Stuart[2]

[1]M. Tech Student, Dept. of Electronics and Communication, Government Engineering College, Kerala, India
[2]Associate Professor, Dept. of Electronics and Communication, Government Engineering College, Kerala, India

---

Abstract - *Malignant melanoma is one of the rapidly increasing and deadly diseases in the world. Early diagnosis is of great importance for treating the disease. Accurate observation of skin lesions is needed for melanoma detection. Dermoscopy is a non-invasive technique for observation of skin lesion. Manual observation of dermoscopic images for classification of lesion as benign or malignant can be inaccurate and subjective. Therefore computer aided diagnosis (CAD) plays a significant role for assisting in melanoma detection. Steps involved in traditional computer aided diagnosis of dermoscopic images involve a) Segmentation b) Feature extraction and c) Classification. Rather than using different tools for segmentation, feature extraction and classification, convolutional neural network (CNN) – a deep learning method can be used inorder to do automatic classification of skin lesion. In this work xception architecture of CNN is used as an end to end classification framework. This architecture is trained using three optimizers. Classification performance of the architecture with three optimizers is analysed using thresholding and ranking metrics. Using these metrics the optimal combination of xception architecture and optimizer for melanoma classification is identified.*

Key Words: **Melanoma, Xception, Optimizers, CNN, Classification**

1. INTRODUCTION

Skin cancer is the cancer that affects the skin. The exposure of epidermis (outermost skin layer) to UV radiation leads to damage in DNA. This result in an out of control growth of abnormal cells in the epidermis resulting in skin cancer. As time progress these abnormal cells begin to spread to other internal body parts. Skin cancers can be broadly divided into two: non melanoma skin cancer (NMSC) and melanoma. Basal cell skin cancer, squamous cell skin cancer along with some less common skin cancers come under NMSC. The most dangerous form of skin cancer is melanoma as shown in fig - 1. It results in 75% of deaths from skin cancer because of its high risk of spreading to other internal body parts. If diagnosed at early stages it can be treated easily without much spreading [1]. So early identification of malignant melanoma from dermoscopy images has great importance. The common symptoms of melanoma include mole with changes such as increase in size, colour, irregular edges, itchiness, skin breakdown. It also rarely occur on normally appearing skin.

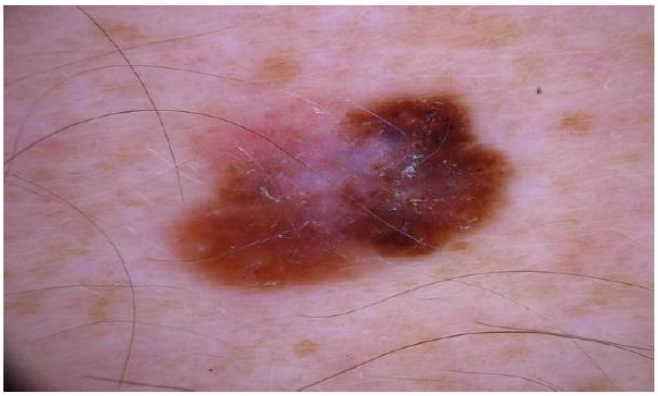

Fig -1: Melanoma affected skin

Clinicians observe skin lesions visually with the help of an instrument called dermatoscope. The main parts of a dermatoscope include a magnifier, a non - polarized light source, a transparent plate and a liquid medium between the instrument and the skin. This device allows skin lesion observation without being interfered by surface reflections. Based on the visual properties of the observed skin lesion different approaches used by clinicians to identify malignant melanoma from benign include ABCD rule [2], pattern analysis, Menzie's method [3] and the seven - point check list [4]. ABCD rule is the most commonly used method by clinicians where ABCD refers to asymmetry, border irregularity, non-uniform colour, diameter greater than 6 mm, and evolving size representing different clinical parameters. Some melanoma images are difficult to identify since they do not follow this rule and also high visual similarity between the benign and malignant lesions make human diagnosis tedious. Dermoscopy can increase the diagnosis sensitivity of melanoma detection by 10% – 27% if and only if the dermatologist is trained properly. Dermoscopy and its diagnostic effectiveness towards early melanoma detection has an inverse relation without proper training of the dermatologist. Also diagnosis made by human visual inspection is laborious, subjective, time consuming and may result in poor accuracy and reproducibility of disease diagnosis.

This increases the necessity of computer aided diagnosis (CAD) systems that can handle all these issues and can

provide a second opinion which will be valuable for early diagnosis of disease. These systems also helps in the reduction of the large number of needless and costly biopsy procedures. For classification of melanoma using CAD systems machine learning or deep learning based techniques can be used. Recently deep learning techniques like convolutional neural network (CNN) have shown outstanding performance in many medical image analysis tasks [5]. Rather than using low level handcrafted features for melanoma recognition, CNN can be used for melanoma classification which has the ability to learn hierarchical features from raw dermoscopic images. Deep learning models are usually trained end to end and can directly predict the type of skin cancer resulting in a fully automatic system with minimum human interference. In this work xception architecture is used as an end to end classification system that classify the skin lesion as benign or malignant. The main objective of the work is to train the architecture using three optimizers namely stochastic gradient descent (SGD), root mean square propagation (RmsProp), adaptive moment estimation (Adam) and then to analyse the performance of the network for same classification task with three optimizers using evaluation metrics. Finally to identify the architecture – optimizer combination that gives best classification performance.

2. LITERATURE SURVEY

As the time of melanoma detection increases, the risk of spreading of melanoma to other organs through the lymph increases thereby increasing the mortality rate. This risk can be reduced with the help of CAD techniques. Classification of melanoma using CAD systems include machine learning or deep learning based techniques which differ in the steps involved for the task. Machine learning based melanoma classification involve three major stages: lesion segmentation, feature extraction and classification. The dermoscopic images must be pre - processed before segmentation. Basic pre - processing steps include variable lighting effect elimination [6], colour space conversion, appropriate colour channel selection [7], enhancement of the selected colour channel [8], contrast enhancement [9], normalizing colour variation caused by image acquisition, smoothing the image, hair removal [10] and removal of vignetting effect [11]. Accurate lesion segmentation requires a proper combination of pre - processing steps. Prior identification of type of noise or artifacts present in the image is necessary for proper selection of the pre – processing technique. But a generalisation of pre – processing techniques is not possible due to variation of artifacts in different dermoscopic images. After pre – processing only segmentation can be performed. Segmentation means separating the lesion (affected region) from the normal skin region. Segmentation methods can be broadly classified as thresholding, edge based and region based methods. Thresholding based segmentation involves clustering [12], histogram thresholding [13], adaptive thresholding. These methods achieve good results for dermoscopic images having bimodal image histogram that is there is a good contrast between the lesion and the skin. They fail when modes from the two regions overlap. Edge based approaches [14] where segmentation is done based on the zero-crossings of the laplacian of gaussian they perform poorly when there is a smooth transition between skin and lesion and also without well-defined boundaries resulting in leakage of contour through gaps in the edges. Region based methods include the multi scale region growing [15], the modified fuzzy c-means algorithm [16], the morphological flooding, a multi resolution markov random field algorithm and statistical region merging [17]. They result in over segmentation when the skin or lesion region is textured. From the segmented region features are extracted that can characterize the samples. Skin lesion features used for classification include morphological features, colour and texture based features. Asymmetry, border irregularity, eccentricity and diameter forms the morphological features [18]. But these features will not recognize some moles with malignancy at early stages, such as malignant melanoma with diameter smaller than 6mm. Global features such as colour and texture features can be calculated from the lesion area [19] but their values are corrupted by the presence of bubbles and other artifacts. This further affects the classifier performance. The final step is to use these features in a classifier to distinguish malignant melanoma from benign lesions. The most commonly used classifiers are support vector machines (SVMs), logistic regression, decision trees, ensemble learners [20], k-nearest neighbours (k-NN) [21], AdaBoost etc. The advantage with these classifiers is that they can be trained with lesser amount of data. But their robustness to classification is low. Handcrafted feature based diagnostic performance is found to be still unsatisfactory due to high intra - class and low inter - class variations in melanoma. Application of CNN in medical images include segmentation of a desired region or classification of diseases. CNN is used for segmentation of medical images such as x-ray, MRI, ventricle segmentation. CNN is used for diagnosis tasks such as tumour identification, tuberculosis diagnosis, lung cancer screening. There are several works that focus on end to end deep learning based melanoma classification such as deep learning ensembles [22], ResNet [23] etc. In this work xception architecture of CNN which is not pretrained is used for automatic melanoma classification.

3. PROPOSED WORK

To classify dermoscopic images into benign and malignant a recently developed deep learning architecture called xception is used. Classification of dermoscopic images are done automatically without applying lesion segmentation or complex image pre - processing. The proposed work involves analysing the performance of this architecture on selecting different optimizers, for selecting best optimizer for the network.

3.1 BLOCK DIAGRAM

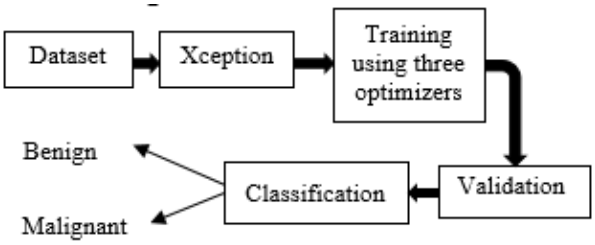

Fig -2: Block diagram

As seen in the fig -2 the input dataset used is the dermoscopic image dataset. It consists of separate sets of benign and malignant images. Pre - processing steps involve image resizing and normalization which is applied to all the images in the dataset. Here a two class problem is solved that is benign skin tumour (class 0) and malignant skin tumour (class 1). This labelling is done to the images in the dataset. Now the sets of benign and malignant images are shuffled together to avoid consecutive images having same label. Next is to split the shuffled dataset into training and validation set each containing a mix of benign and malignant images. The network architecture used is the xception architecture. Once the architecture is built next step is model fitting. Forward and backward pass of the training dataset through the network fits the model. This network is trained only for the dermoscopic image dataset and is not pretrained with any other dataset. Training of xception architecture is done using SGD, RmsProp and Adam optimizers. The classification performance of this network is analysed during validation. Predictions are made from the validation dataset and using these predictions model evaluation using evaluation metrics are done. The best one among the xception-optimizer combination can be used for melanoma classification.

3.2 XCEPTION NETWORK

Xception is a deep convolutional neural network involving depthwise separable convolution. In depthwise separable convolution, convolution process is divided into depthwise convolution and pointwise convolution. The idea behind this convolution is that the depth and spatial dimensions of a filter can be separated. Depth wise convolution and pointwise convolution uses two separate filters. For an RGB image having three channels depthwise convolution performs independent spatial convolution on each of these channels. Pointwise convolution projects the channels computed by depthwise convolution into a new channel space. In xception architecture depthwise convolution is followed by pointwise convolution. It has a linear stack of depthwise separable convolutional layers having residual connections. The feature extraction base is formed by 36 convolutional layers. These 36 layers are grouped into 14 modules which are arranged into entry flow, middle flow and exit flow blocks. Linear residual connections are present for each modules except for the first and last modules. Convolutional layer, batch normalization layer, an activation function, max pooling layer forms a module. Max pooling layer is not present in the first, last module and also in the middle flow block. The final output layer involves global average pooling and a softmax layer. Convolutional layers involves 3×3 depthwise convolutional filters and 1×1 pointwise convolutional filters. Batch normalization following the convolutional layer constantly corrects the activations to zero mean and unit standard deviation thereby speeding up the training process resulting in faster convergence. The activation function used here is a nonlinear activation function that is rectified linear unit (ReLU). ReLU rectifies vanishing gradient problem, it outputs zero for negative inputs and for any positive value it returns the same input value like a linear function. Max pooling calculates the maximum value in each patch of the feature map. It involves filters of size 3×3 with stride 2. Global average pooling (GAP) layer reduces each feature map to a single number by taking the average of all values. This layer minimizes overfitting by reducing the total number of parameters in the model. The output layer used for classification is the softmax classifier which uses the idea of softmax function. It outputs prediction probability in 0 – 1 range. The predicted output is then compared with the actual output and the change between the predicted output and the actual output is calculated as error. This error reduction is done by the optimizer during the back propagation process. The loss function chosen is binary cross entropy loss. When the probability of prediction diverges from the actual, loss increases. A model is perfect when it has zero loss. For binary classification problem where the number of classes M (benign, malignant) equals to 2, binary label y and predicted probability p, cross - entropy can be calculated as:

$$\text{Loss} = -(y \log p + (1-y) \log (1-p)) \quad (1)$$

3.2 OPTIMIZERS

Optimization is usually done to find the weights that minimizes the loss function. In neural networks loss reduction is done by changing the parameters of the neural network such as weights and learning rate using the optimizers. Different optimizers used for the proposed method include SGD, RmsProp and Adam. SGD optimizes an objective function iteratively and is regarded as a stochastic approximation of gradient descent optimization. It calculates the estimate of the gradient from a randomly selected subset of the data whereas in gradient descent this is calculated for the entire dataset. Each parameter update is computed over a mini-batch. It uses a constant learning rate and the data is randomly shuffled prior to each epoch of training. RmsProp is also a method of optimization in which learning rate is adapted for each of the parameter. It divides the learning rate using the running average of the magnitudes of recent gradients for a weight, restricts oscillations in vertical

direction. It is a very robust optimizer and can deal with stochastic objectives making it applicable to mini batch learning. Adam uses running averages of both the gradients and the second moments of the gradients. Using the estimates of first and second moments of the gradient this method computes individual adaptive learning rates for different parameters. The first moment is mean and second moment is uncentered variance. The update rule uses gradient invariant step size which is useful in areas with tiny gradients such as saddle points.

4. IMPLEMENTATION

The training is done on the ISIC (International Skin Imaging Collaboration) dataset. The dataset used consists of 782 benign images and 584 malignant images. This is a binary classification problem with labels 0 for benign or negative case and 1 for malignant or positive case. Before training the dataset is shuffled, then 1269 images are used for training and remaining 97 images are used for validation. The dataset consists of image dimensions ranging from 600×450 to 1504×1129. These images are resized to 128×128 for the purpose of giving it to the network. All these images are normalized by dividing the image pixels by 255 before giving to the network. The general hyper parameters for training include batch size and epoch. The training is done with 25 epochs and a batch size of 8. The learning rate is initialised to 0.001. The programming is done using python language on Jupyter notebook platform.

4.1 EVALUATION METRICS

To evaluate the classification performance of the xception architecture trained using different optimizers the evaluation metrics are used. Evaluation metrics used are thresholding metrics and ranking metrics. Thresholding metrics include accuracy, sensitivity, specificity, precision and negative predict value (NPV). These metrics are evaluated at the image level. The values used for finding these metrics are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). TP are cases in which classifier predicts a positive case as positive. TN are cases in which classifier predicts a negative case as negative. FP are cases in which classifier predicts a negative case as positive and is also known as a type I error. FN are cases in which classifier predicts a positive case as negative, it is also known as type II error. Accuracy is the proportion of correct predictions (both true positive and true negative) among total predictions. Sensitivity is the proportion of positive data points that are correctly considered as positive among all positive cases. Specificity is the proportion of negative data points that are correctly considered as negative among all negative cases. Precision or positive predict value is the number of correct positive results divided by the number of positive results predicted by the classifier. Negative predict value is the number of correct negative results divided by the number of negative results predicted by the classifier.

The ranking metrics used is the receiver operating characteristic (ROC) curve. An ROC curve is a basic tool for analysing the classifier. It is a plot between sensitivity (true positive rate) and specificity (false positive rate). It shows the ability of a classifier to separate the classes. The region of ROC curve is divided into three. The diagonal line shows equal weightage for both classes (benign, malignant). The region below the diagonal shows bad prediction. The region above the diagonal is the best region and for a good classifier the curve is close to the top left corner.

5. RESULTS

Once the network is trained using the three optimizers its classification performance is evaluated using the validation dataset. The thresholding metrics obtained is shown in table-1. From the thresholding metrics it is clear that xception trained with Adam optimizer obtained the highest accuracy of 0.88 for melanoma classification. The ROC curve obtained for three optimizers is shown fig -3, fig -4 and fig -5. From the three graphs it is clear that xception trained with Adam optimizer is the best candidate for melanoma classification.

Table -1: Comparison of architecture performance for different optimizers

Metrics	Xception+ SGD	Xception + RmsProp	Xception+ Adam
Accuracy	0.80	0.86	**0.88**
Sensitivity	0.84	0.78	**0.89**
Specificity	0.82	0.77	**0.86**
Precision	0.86	0.93	0.89
NPV	0.80	0.77	**0.86**

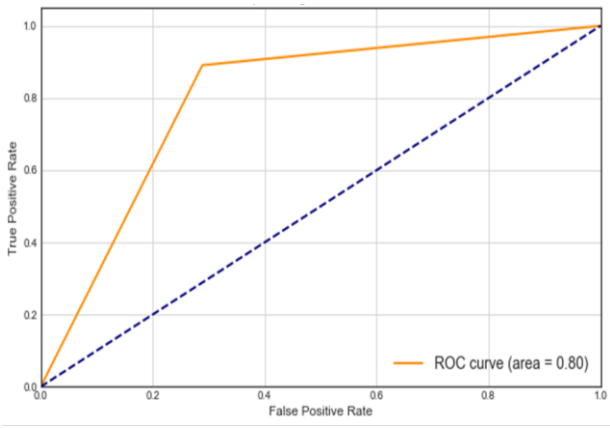

Fig -3: ROC curve for xception trained with SGD

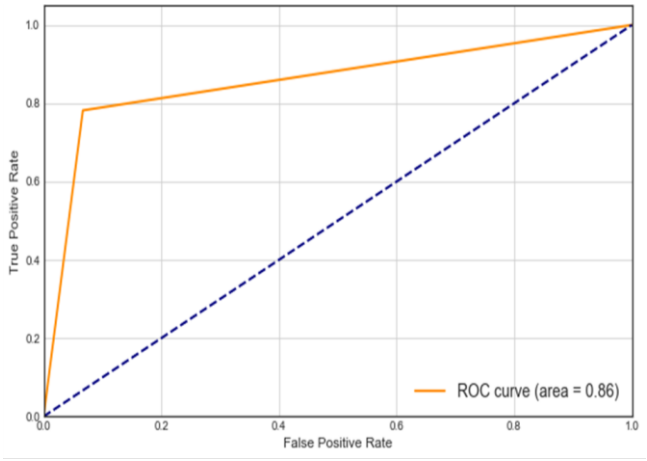

Fig -4: ROC curve for xception trained with RmsProp

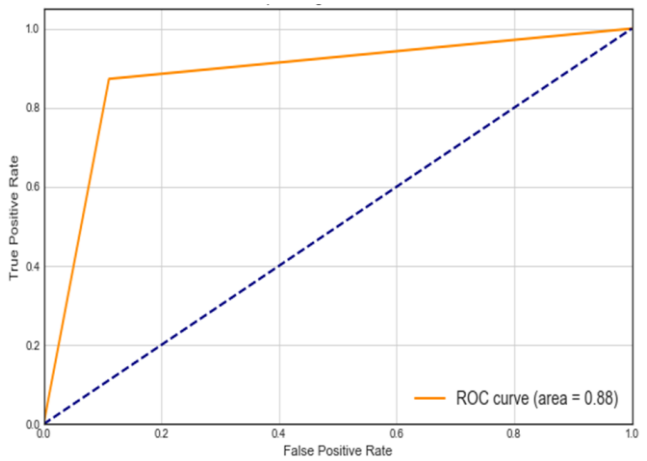

Fig -5: ROC curve for xception trained with Adam

This classifier is tested for 108 cases of which 55 are malignant and 53 are benign. Among the 55 malignant cases, for 48 cases malignancy is correctly predicted. This shows a correct malignancy prediction of 0.87. Table -2 shows the resulting confusion matrix.

Table -2: Confusion matrix

	Actual : 1	Actual : 0
Predicted : 1	TP : 48	FP : 5
Predicted : 0	FN : 7	TN : 48

Using this architecture prediction of new dermoscopic images that have not been used for training and validation is done. Depending on the type of the image the network predicts it as either benign or malignant. Examples of such predicted images are shown in fig -6 which shows robust prediction for similar dermoscopic images eventhough they differ in their type.

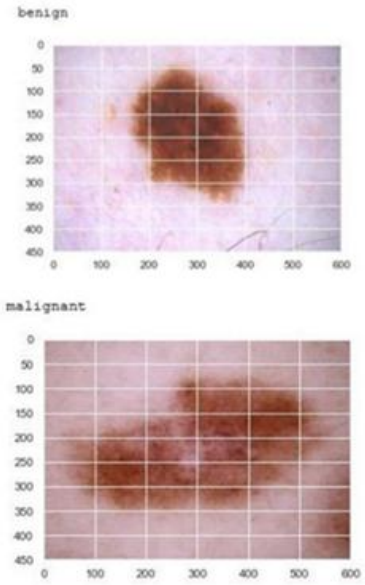

Fig -6: Predictions made using the network

6. CONCLUSION

In this work, a melanoma classification technique using CNN is studied. A recently developed deep learning framework xception is used to classify dermoscopic images into benign and malignant. The xception architecture is trained using three optimizers for the ISIC data set. Melanoma classification is performed without applying complex image pre – processing, lesion segmentation and feature extraction. Thresholding and ranking metrics evaluates the classifier performance for three optimizers. From the evaluation metrics it is clear that the xception architecture trained with Adam optimizer outperforms the network trained with other two optimizers for lesion classification task. In future this method can be combined with an android application. This aids in taking the skin lesion image through a mobile camera and then diagnosing the skin lesion using the app supporting in telemedicine. Also the possibilities of this network can be exploited to other medical diagnostic tasks.

REFERENCES

[1] A. R. A. Ali and T. M. Deserno, "A Systematic Review of Automated Melanoma Detection in Dermatoscopic Images and its Ground Truth Data," Proc. SPIE, vol. 8318, Feb. 2012, pp.454–462.

[2] Naheed R, Abbasi, Helen M. Shaw, Darrell S. Rigel, Robert J. Friedman, William H. McCarthy, Iman Osman, Alfred W. Kopf and David Polsky, "Early Diagnosis of Cutaneous Melanoma Revisiting the ABCD Criteria," American Medical Association, vol. 292, Dec. 2004, pp.2771-2776.

[3] S. W. Menzies, "Frequency and Morphologic Characteristics of Invasive Melanomas Lacking Specific

Surface Microscopic Features," Arch. Dermatol., vol. 132, Oct. 1996, pp. 1178–1182.

[4] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco and M. Delfino, "Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions: Comp. of the ABCD rule of Dermatoscopy and a New 7-Point Checklist Based on Pattern Analysis," Archives of Dermatology , vol. 134, Dec. 1998, pp. 1563-1570.

[5] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu and M. S. Lew, "Deep Learning for Visual Understanding: A Review," Neurocomputing, vol. 187, Apr. 2016, pp. 27-48.

[6] J. Glaister, R. Amelard, A. Wong and D. Clausi, "MSIM: Multistage Illumination Modeling of Dermatological Photographs for Illumination Corrected Skin Lesion Analysis," IEEE Transactions on Biomedical Engineering, vol. 60, Feb. 2013, pp. 1873-1883.

[7] G. Schaefer, M. I. Rajab, M. E. Celebi and H. Iyatomi, "Colour and Contrast Enhancement for Improved Skin Lesion Segmentation," Computerized Medical Imaging and Graphics, vol. 35, March 2011, pp. 99-104.

[8] M. E. Celebi, H. Iyatomi and G. Schaefer, "Contrast Enhancement in Dermoscopy Images by Maximizing a Histogram Bimodality Measure," in 6th IEEE International Conference on Image Processing (ICIP), IEEE, Nov. 2009, pp. 2601-2604.

[9] Q. Abbas, I. F. Garcia, M. E. Celebi, W. Ahmad and Q. Mushtaq, "A Perceptually Oriented Method for Contrast Enhancement and Segmentation of Dermoscopy Images," Skin Research and Technology, vol. 19, Aug. 2012, pp. 490-497.

[10] Q. Abbas, I. F. Garcia, M. E. Celebi and W. Ahmad, "A Feature Preserving Hair Removal Algorithm for Dermoscopy Images," Skin Research and Technology, vol. 19, Feb. 2013, pp. 27-36.

[11] F. Y. Xie, Y. Lu, A. Bovik, Z. Jiang and R. Meng, "Application-Driven No-Reference Quality Assessment for Dermoscopy Images with Multiple Distortions," IEEE Transactions on Biomedical Engineering, vol. 63, June 2016, pp.1248-1256 .

[12] M. Mete, S. Kockara and K. Aydin, "Fast Density-based Lesion Detection in Dermoscopy Images," Computerized Medical Imaging and Graphics, vol. 35, Sep. 2010, pp. 128-136.

[13] R. Garnavi, M. Aldeen and M. E. Celebi, G. Varigos and S. Finch, "Border Detection in Dermoscopy Images Using Hybrid Thresholding on Optimized Color Channels," Computerized Medical Imaging and Graphics, vol. 35, Mar. 2011, pp. 105-115.

[14] Q. Abbas, M. E. Celebi, Irene Fondon Garcia and M. Rashid, "Lesion Border Detection in Dermoscopy Images using Dynamic Programming," Skin Research and Technology, vol. 17, Feb. 2011, pp. 91-100.

[15] B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker and E. Hva-tum, "Automatic Lesion Boundary Detection in Dermoscopy Images using Gradient Vector Flow Snakes," Skin Res. & Technol., vol. 11, Feb. 2005, pp. 17–26.

[16] D. H. Chung and G. Sapiro, "Segmenting Skin Lesions with Partial-Differential Equations Based Image Processing Algorithms," IEEE Trans. Med. Imag., vol. 19, July. 2000, pp. 763–767.

[17] M. Celebi, Y. Aslandogan and P. Bergstresser, "Unsupervised Border Detection of Skin Lesion Images," in Int. Conf. Information Technology: Coding and Computing (ITCC 2005), vol. 2, May 2005, pp. 123–128.

[18] M. Celebi, H. Kingravi and J. Lee, "Fast and Accurate Border Detection in Dermoscopy Images using Statistical Region Merging," in Proc. SPIE Medical Imaging, vol. 6512, March 2007.

[19] Marghoob, Ashfaq. A and Alon Scope. "The Complexity of Diagnosing Melanoma," Journal of Investigative Dermatology, vol. 129, Jan. 2009, pp. 11-13.

[20] Ebtihal Almansour and M. Arfan Jaffar, "Classification of Dermoscopic Skin Cancer Images Using Color and Hybrid Texture Features", IJCSNS, vol.16, April 2016, pp. 135-139.

[21] G. Schaefer, B. Krawczyk, M. E. Celebi and H. Iyatomi, "An Ensemble Classification Approach for Melanoma Diagnosis," Memetic Computing, vol. 6, Oct. 2014, pp. 233-240.

[22] Codella. N. C, Nguyen. Q. B, Pankanti. S, Gutman. D, Helba. B, Halpern. A and Smith. J. R, "Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images", IBM Journal, vol. 61, July 2017.

[23] Zhen Yu , Xudong Jiang , Feng Zhou , Jing Qin , Dong Ni, Siping Chen, Baiying Lei, and Tianfu Wang, "Melanoma Recognition in Dermoscopic Images via Aggregated Deep Convolutional Features", IEEE Transactions On Biomedical Engineering, vol. 66, April 2019, pp. 1006-1016.