Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection

[1] Ms. Prajakta P. Shirke, [2] Dr. Amit R. Gadekar

[1] Ph.D. Research Scholar, School of Computer Sciences and Engineering, School of Computer Sciences and Engineering, Sandip University, Nashik, Maharashtra, India

[2] Assistant Professor, School of Computer Sciences and Engineering, School of Computer Sciences and Engineering, Sandip University, Nashik, Maharashtra, India

*Corresponding author. Email: [1] prajakta.shirke@sandipuniversity.edu.in, [2] amit.gadekar@sandipuniversity.edu.in

Abstract— Melanoma, or skin cancer, is usually detected visually from dermoscopic pictures, which is a time-consuming and difficult job for the dermatologist. Existing systems either utilize classic machine learning models that concentrate on hand-picked acceptable features, or deep learning-based approaches that learn features from full pictures. Melanoma, or skin cancer, is usually detected visually from dermoscopic pictures, which is a time-consuming and difficult job for the dermatologist. Due to many artifacts such as poor contrast, diverse noise, presence of hair, fiber, and air bubbles, etc., such a visual examination with the naked eye for skin malignancies is tough and onerous. This paper provides a robust and automated system for Skin Lesion Classification (SLC), in which image augmentation, Deep Convolutional Neural Network (DCNN), and transfer learning are all combined. Our lesion classification experiments show that the suggested technique can effectively classify skin cancer with a high degree of accuracy, and that it can also identify skin lesions for melanoma detection.

Index Terms— Transfer Learning, Skin Lesion Classification, DCNN

I. INTRODUCTION

Skin tumors (lesion cancers) are the most frequent kind of cancer, characterized by abnormal cell proliferation that may spread to adjacent cells or other regions of the body. The majority of skin cancers are caused by exposure to UV radiation from the sun or by cosmetic devices like tanning beds. Melanoma, Squamous Cell Skin Cancer (SCC), and Basal Cell Skin Cancer are the three primary kinds of skin cancer (BCC) as illustrated in Figure 1. Melanoma is the most dangerous and prevalent kind of skin cancer, and it may form on the eyes or mouth, but SCC and BCC are less prevalent, do not spread to other locations, and are unlikely to cause mortality.

Figure 1: Sample Skin Cancer Images, Basal Cell Skin Cancer(A), Squamous Cell Skin Cancer(B), and Melanoma (C)

The formation of a big brown spot, spread by small and dark patches, changes in the color and size of moles on the body, or the descent of blood from them, as well as the formation of dark little & uneven borders lesion, commonly on the hands & feet, are the major signs of melanoma cancer. According to the American Cancer Society, there will be over a hundred thousand new melanomas cases in the United States in 2019, with 60% of men and 40% of women, and thousands of skin cancer patients expected to die from melanoma in 2019. "As a result, there is a critical need for a computer program that can deliver quick and accurate diagnosis for suspected skin cancer." Early detection of skin malignancies might improve survival rates and save medical costs & treatment for prospective patients. A computer-assisted application for identifying skin cancers may help to spot skin cancers earlier, before it's too late, and urge prospective patients to seek further pathology tests, medical guidance.

Skin cancer is nowadays diagnosed using a variety of imaging methods. Dermatologists most often utilize dermoscopic pictures, to study pigmented skin lesions. Because of the similarity between the lesions & healthy tissues, such a visual evaluation with the naked eye may induce a faulty-recognition. Dermatologists' hand inspections are sometimes arduous, time-consuming, and subjective, resulting in varying recognition outcomes. However, in order to overcome all of the aforementioned drawbacks and enhance the accuracy of skin cancer detection,

II. RELATED WORKS

Md. Kamrul Hasan et al., (2021) Despite the fact that automatic Skin Lesion Classification is an important step in computer-assisted diagnosis, it is a difficult process because

to inconsistencies in texture, color, indistinct borders, and forms. In paper, present Dermoscopic Expert, an automated and robust system for dermoscopic SLC (DermoExpert). Preprocessing, a (hybrid-CNN), and transfer learning make up the DermoExpert. The proposed hybrid-CNN classifier comprises of three independent feature extractors that are combined to provide better-depth feature maps of the corresponding lesion using the same input pictures. The various fully linked layers are then ensembled to produce a final prediction probability from those unique and fused feature maps. We apply lesion segmentation, augmentation, and class rebalancing in the preprocessing. We also used geometric and intensity-based augmentation, as well as class rebalancing. We also employ information from a pre-trained model, to develop a general classifier, despite the fact that only tiny datasets are employed. Finally, we build and build a web application for automated lesion identification using the weights of our DermoExpert. We tested our DermoExpert on the ISIC-2016, ISIC-2017, and ISIC-2018 datasets, and our DermoExpert scored 0.96, 0.95, and 0.97, respectively, in the area under the receiver operating characteristic curve (AUC). In terms of AUC, the experimental findings surpass the current state-of-the-art by 10.0 percent and 2.0 percent for the ISIC-2016 and ISIC-2017 datasets, respectively. For the ISIC-2018 dataset, the DermoExpert also excels by 3.0 percent in terms of balanced accuracy. Our ISIC-2018 dataset source code and segmented masks will be made openly accessible for the research community to enhance. [1]

Zheng et al., (2020) Breast cancer is one of the most serious illnesses, and it is the second leading cause of mortality among women. When malignant, cancerous tumors form in the breast cells, breast cancer develops. Self-tests and periodic clinical checks aid in early detection and, as a result, greatly enhance the odds of survival. Breast cancer categorization is a medical strategy that poses a significant challenge to academics and scientists. In the categorization of cancer data, neural networks have lately become a prominent method. The DLA-EABA has been theoretically suggested in this research for breast cancer diagnosis using sophisticated computational approaches. Tumor classification techniques based on transfers are being actively developed utilizing deep convolutional neural networks, in addition to classic computer vision techniques (CNNs). This research begins by looking at how CNN-based transfer learning may be used to describe breast masses for diagnostic, predictive, or prognostic purposes across a variety of imaging modalities, including digital breast tomosynthesis, and mammography, MRI, ultrasound. Several convolutional layers, LSTM, and Max-pooling layers are included in the deep learning framework. A fully connected layer and a softmax layer have categorization and error estimates built in. This work focuses on finding the best way by integrating various machine learning methodologies with methods for choosing and extracting features, as well as assessing their output using classification and segmentation algorithms. When compared

to other current systems, the testing findings demonstrate that the high accuracy level of 97.2 percent, sensitivity of 98.3 percent, and specificity of 96.5 percent. [2]

Gulati et al., (2019) Because it may readily spread to other parts of the body, malignant melanoma is the most serious kind of skin cancer. In paper, a CAD system is constructed using the deep learning idea, which has recently gained popularity for its ability to provide greater accuracies. Convolutional neural networks, the most common deep neural network, are used. AlexNet and VGG16, two pre-trained networks, were employed in two distinct ways like Transfer learning and feature extractor. The transfer learning-based method has been shown to provide efficient outcomes for both CNNs, with AlexNet with transfer learning achieving 95% accuracy and AlexNet as a feature extractor achieving 90% accuracy. The same is true for VGG16, which has an accuracy of 97.5 percent in the first scenario and 95 percent in the second. Finally, all strategies are compared, and it is clear that VGG16 with transfer learning outperforms them all, with accuracy, sensitivity, and specificity of 97.5, 100, and 96.87 percent, respectively. Aside from that, the sensitivity attained for both transfer learning-based designs is 100%, implying that all melanoma cases are accurately detected. [4]

Adegun et al., (2019) Melanoma is the most dangerous kind of skin cancer. However, distinguishing melanoma lesions from non-melanoma lesions has proven difficult. In this research, we present a DL-based system for automated melanoma lesion identification and segmentation that addresses these constraints. For effective learning and feature extraction, an upgraded encoder-decoder network with encoder and decoder sub-networks linked by a series of skip routes is presented, bringing the semantic level of the encoder feature maps closer to that of the decoder feature maps. For pixel-wise categorization of melanoma lesions, the system uses a multi-stage and multi-scale technique, as well as a softmax classifier. We developed a novel approach called Lesion-classifier that divides skin lesions into melanoma and non-melanoma based on the findings of pixel-wise classification. Our tests on two well-known public benchmark skin lesion datasets, the International Symposium on Biomedical Imaging (ISBI)2017 and the Hospital Pedro Hispano (PH2), show that our technique outperforms several state-of-the-art approaches. On the ISIC 2017 dataset, we obtained accuracy and dice coefficient of 95% and 92 percent, respectively, while on the PH2 datasets, we obtained accuracy and dice coefficient of 95% and 93 percent. [5]

Moldovan, Dorin. (2019) The classification of skin cancer pictures is a key scientific problem since the population is predicted to grow dramatically in the next years, resulting in an increase in the number of individuals who may get skin cancer at some point in their lives. Some kinds of skin cancer may be effectively treated if found early on, therefore studying skin cancer photos utilizing the newest technology breakthroughs may provide better results than

using older approaches. This article describes a two-step approach for classifying skin cancer photos that is based on transfer learning and deep learning and consists of two parts. The classification models were created in Python using the PyTorch machine learning toolkit, and the Human Against Machine with 10000 training pictures (HAM10000) dataset was utilized as experimental support for testing and verifying the transfer learning-based strategy. In the first phase, the prediction model for testing data is 85% accurate, and in the second step, the prediction model for testing data is 75% accurate. [6]

III. RESEARCH METHODOLOGY

In this work, we use the Jupiter Anaconda Navigator tool to execute CNN. Also, a data set based on skin infected augmented photos was collected, and all of the data was compared using a simulation tool. The numerous layers kmeans and augmented layers are employed in the simulation tool, which is programmed in Python. The Analysis based on materials and methods in Literature are given as follows illustrated in table 1:

Table 1: ISIC-2017 dataset - Data distribution

S. No	Class Types	Description	Train	Validation	Test
01	Melanoma (Mel)	Malignant skin tumor obtained from melanocytic	369	29	115
02	Seborrheic Keratosis (SK)	Benign skin tumor obtained from keratinocytes	256	44	91
03	Nevus (Nev)	Benign skin tumor obtained from melanocytic	1375	77	394
Total Images			2000	150	600

- **Proposed Framework**

As illustrated in Figure 2, the suggested framework uses an end-to-end CNN architecture to automate feature extraction & classification of the skin lesion for melanoma detection, with image augmentation and normalization being critical and important aspects of the proposed framework. Since 2012, deep CNNs have been frequently employed in both medical and natural picture categorization. It often outperforms human knowledge. A CNN was trained on more than 1.0 million frontal-view chest X-rays in CheXNet, and it has ability to get superior recognition results than the four experts on average.
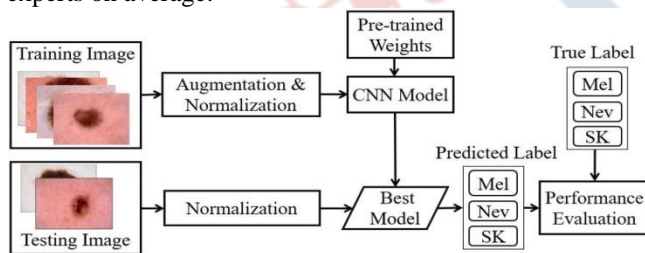

Figure 2: Proposed framework for an automatic SLC towards melanoma recognition

In the paper, the CNN model is gives in Figure 3, which has 13 convolutional layers in the 5 convolutional blocks. Each block ends up with a
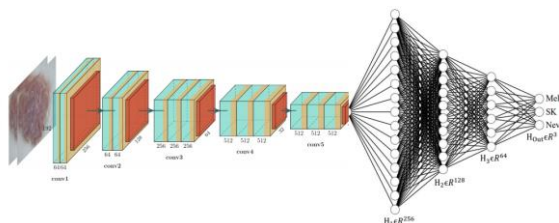

Figure 3: The CNN network for the SLC, where Hm ∈ Rn is the mth hidden layer in n-dimensional space. The output layer, HOut lies in 3-dimensional (Mel, SK, and Nev) space.

The max-pooling layer reduces the amount of connections between convolutional layers, lowering computational complexity and speeding up the generalization of CNN models by decreasing overfitting. The output of the convolutional layers is sent into the Fully Connected (FC) layers, which consist of three FC layers and one output layer. Due to state-of-the-art performance for image classification, global average pooling (GAP) was employed between convolutional layers and FC layers instead of the typical flatten layer The height weight depth dimensional tensor was reduced to 11 depth in GAP, and each height width feature map was converted to a single integer by averaging the height width values. However, multiple layers of the feature extractor were frozen to identify the ideal amount of layers to be frozen by maximizing the AUC for the SLC for fine-tuning the convolutional layers.

Transfer learning may be used to overcome the lack of training pictures, particularly if annotation is time-consuming and expensive. The feature extractor kernels (2-D convolutional layers) were initialized using the previously learned weights from ImageNet, as illustrated in Figure 3. A glorot uniform distribution was used to initialize the kernels in FC layers. The Glorot distribution, also known as the Xavier distribution, has a mean of 0 and a standard deviation of $\sqrt{(2)/(F_{in} + F_{out})}$, where F_{in} and F_{out} are the number of input and output units in the weight tensor, respectively.

In the suggested pipeline, random rotation (0 90), width shift (10%), height shift (10%), random shearing (20%), and horizontal & zooming (20%), vertical flipping were employed as picture augmentation, with the outside pixels being filled using a reflection approach. Before being sent into the CNN network, the photos were also uploaded to [0 1]. The Aspect Ratio (AR) distribution of the ISIC-2017 photos shows that AR is 3:4. As a result, we used nearest-neighbor interpolation to shrink the input photos to 192-256 pixels. The loss function in our framework was

categorical cross-entropy, which was tuned to maximize the average accuracy of lesion categorization.

Without the AMSGrad variation, the loss function was optimized using the Adam optimizer with initial Learning Rate (LR), exponential decay rates (1, 2) of LR = 0.0001, 1 = 0.9, and 2 = 0.999, respectively. If validation loss does not improve after 5 epochs, the initial learning rate is cut by 20.0 percent. The proposed pipeline was trained on a machine with a batch size of 8.0, as indicated in paragraph 2.1.

- **Evaluation Criterion**

The proposed pipeline was assessed using the False Positive (FP), True Positive (TP), False Negative (FN), and True Negative (TN) confusion matrix (TN). The precision, recall, and F1-score were also utilized, where the recall measures the percentage of properly identified positive patients from all positive recognition and the precision measures the type-II error (the sample having target syndromes but incorrectly fails to be denied). The F1-score is the harmonic mean of recall and accuracy, indicating the tradeoff. The AUC also assesses how well lesion predictions are prioritized rather than their absolute value.

IV. RESULTS AND DISCUSSION

The findings of the ISIC-2017 test dataset (see Table 1) for the SLC are reported in this section. On the same dataset, state-of-the-art approaches are compared to the suggested framework at the conclusion of this section.

Table 2 shows a classification report that visualizes recall, F1-score, precision, and support scores on a per-class basis. A categorization report is a more detailed report.

Table 2: SLC Classification report, where weights of the classes were calculated from assisted samples for averaging the metrics.

Class	Precision	Recall	F1-score	Support
Mel	0.53	0.63	0.57	117
Nev	0.87	0.77	0.81	393
SK	0.55	0.74	0.63	90
Weighted Average	0.76	0.73	0.74	600

Intuition for the classifier's quantitative assessment, which may also reveal the classifier's flaws in a specific class of a multi-class issue. Table 2 shows that the properly categorizing samples of Mel, Nev, and SK are 63.0 percent, 77.0 percent, and 74.0 percent, respectively, while the type-II error (false-negative rate) is 38.0 percent, 24.0 percent, and 27.0 percent, respectively. The support-weighted recall of 74.0 percent implies that 27.0 percent of samples have target symptoms, however the suggested framework incorrectly fails to reject this. Only 24.0 percent of samples are incorrectly categorized among all identified true classes, according to the support-weighted precision of 77.0 percent.

Although the unequal class distribution was employed for training, the F1-score suggests that the suggested framework for the SLC of the ISIC-2017 has improved accuracy and recall. The confusion matrix in Table 3 contains further specifics on the planned SLC's class-wise study. Table 3 shows the confusion matrix, which shows the number of accurate and inaccurate predictions made by the suggested framework with a count value.

Table 3: Test results Confusion matrix, where each column and row represent the instances in a predicted & actual class.

Actual	Predicted		
	Mel	Nev	SK
Mel	70	28	19
Nev	57	300	36
SK	9	15	68

Table 3 reveals that Mel, Nev, and SK classes are categorized as (70, 28, 19), (57, 300, 36), & (9, 15, 68) accordingly among 117, 393, 92 samples of Mel, Nev, and SK classes. There are 117 Mel samples classed as the Nev, one of which is FN. Table 1 reveals that in the training dataset, there are more samples in the Nev class, indicating that the biased classifier is biased towards Nev. The findings, as given in Table 3, also suggest that a large percentage of Mel and SK samples are expected to be Nev. Table 4 shows some of the misclassified photos produced by the proposed model, as well as the difficulty in building right classifications. Table 4 shows that images with a true class of Mel are categorized as SK & Nev, respectively, with confidence probabilities of 0.934 and 0.88. Identically, with a degree of confidence possibly, all other photos in Table 4 are incorrectly identified. If we look at those photographs closely, they seem to be Mels, particularly in the fifth row, since the texture of the lesion region is more complicated. The dermatologist, on the other hand, says it's Nev. The lack of variety in the training samples, inter-class variety, and intra-class similarity are likely reasons for the proposed framework's incorrect categorization.

The ROC and precision-recall curves are shown in Figure 4 (a) and Figure 4 (b), respectively. It is observed from Figure 4 (a) that for a 10.0 % FP rate, the TP rates for the SLC is approximately 62.0 %. The corresponding macro-average AUC from the ROC curve is 0.873, which specify that for any given random sample, the probability of correct recognition as Mel, Nev, and SK is as highest as 87.3 %. The precision-recall curve, as shown in Figure 4 (b), shows the tradeoff between precision and recall for different thresholds, where the high area under the curve represents recall and precision with high values. High-rise scores for both express that the proposed pipeline is a blessing with correct results as well returning a majority (High recall) of all positive results. The macro-average precision from the proposed pipeline is 80.6 %,. which shows that the proposed pipeline is suitable for the SLC for melanoma recognition.

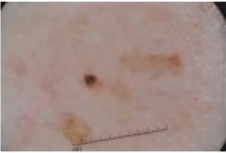**Table 4:** Some examples of mistakenly classified images, from the proposed frame- work, with confidence probability.
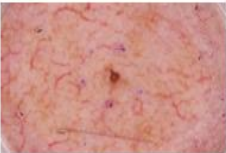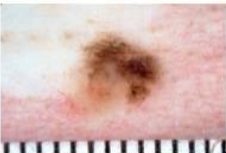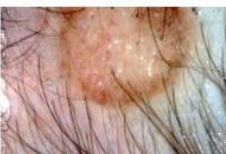
Example Images	Predicted Class with Confidence
	Actual Class: Mel Predicted Class: SK Confidence Probability: 0.934
	Actual Class: Mel Predicted Class: Nev Confidence Probability: 0.880
	Actual Class: SK Predicted Class: Nev Confidence Probability: 0.994
	Actual Class: SK Predicted Class: Mel Confidence Probability: 0.876
	Actual Class: Nev Predicted Class: Mel Confidence Probability: 0.972
	Actual Class: Nev Predicted Class: SK Confidence Probability: 0.941

Table 5 represents state-of-the-art comparative analysis of proposed pipeline with latest works, where AlexNet and ResNet-101 were implemented for the SLC. The proposed framework shows the best classification of the skin lesion, as shown in Table 5. Our pipeline conctruct the best results concerning
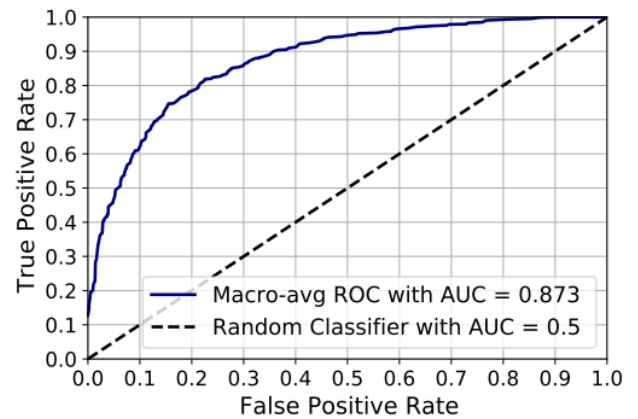
(a)
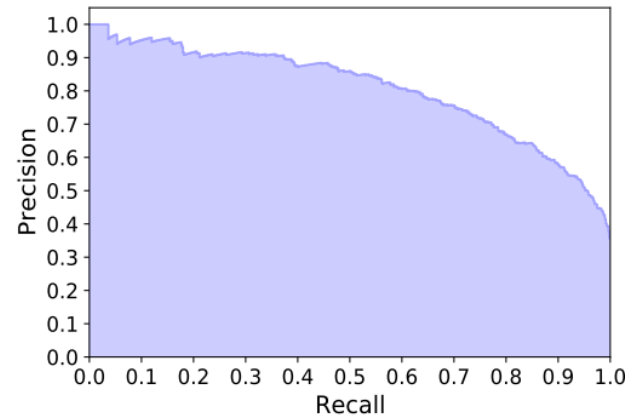
(b)

Figure 4: (a) ROC curve to summarize the trade-off between the TP rate & FP rate, and (b) Precision-recall curve to summarize the trade-off between the TP rate & positive predictive value.

By a margin of 12.0% and 5.0 percent, respectively, recall and accuracy outperformed state-of-the-art works. Our technique outperforms the approach by 39.0 percent in recall and 5.0 percent in accuracy, despite the fact that the AUC is the same in both techniques.

V. CONCLUSION

The review of DCNN model achieves a better classification rate when estimate to other transfer learning models. The capability of the proposed method is to classify benign and malicious skin lesions by replacing the output activation layer with sigmoid for the binary classification. Several detailed experiments have been used to verify the proposed framework's potential. Utilizing transfer learning, earlier taught weights, and geometric augmentation, the limitation of using fewer manually annotated dermoscopic pictures to develop a generic framework was addressed. Additional tweaking of the hyper-parameters and associated augmentations may result in improved SLC identification results. In the future, ''the suggested methodology will be evaluated on various datasets of dermoscopic pictures for the

SLC." The suggested pipeline will be used for recognition in additional domains to ensure its flexibility and universality.

REFERENCES

[1] Md. Kamrul Hasan, Md. Toufick E Elahi, Md. Ashraful Alam (2021) "DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation" medRxiv 2021.02.02.21251038; doi: https://doi.org/10.1101/2021.02.02.21251038

[2] Zheng, Jing & Lin, Denan & Gao, Zhongjun & Wang, Shuang & He, Mingjie & Fan, Jipeng. (2020). Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2993536.

[3] Harangi, B.; Baran, A.; Hajdu, A. Assisted deep learning framework for multi-class skin lesion classification considering a binaryclassification support.Biomed. Signal Process. Control2020,62, 102041

[4] Gulati, Savy & Bhogal, Rosepreet. (2019). Detection of Malignant Melanoma Using Deep Learning. 10.1007/978-981-13-9939-8_28.

[5] Adegun, Adekanmi & Viriri, Serestina. (2019). Deep Learning-Based System for Automatic Melanoma Detection. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2962812.

[6] Moldovan, Dorin. (2019). Transfer Learning Based Method for Two-Step Skin Cancer Images Classification. 10.1109/EHB47216.2019.8970067.

[7] Nida, Nudrat & Irtaza, Aun & Javed, Ali & Yousaf, Muhammad Haroon & Mahmood, Muhammad. (2019). Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. International Journal of Medical Informatics. 124. 10.1016/j.ijmedinf.2019.01.005.

[8] Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images ofcommon pigmented skin lesions.Sci. Data2018,5.

[9] Farooq, Muhammad & Azhar, Muhammad & Raza, Rana. (2016). Automatic Lesion Detection System (ALDS) for Skin Cancer Classification Using SVM and Neural Classifiers. 301-308. 10.1109/BIBE.2016.53.

[10] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper withconvolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA,USA, 7–15 June 2015.

[11] Bruno, D.O.T., et al., LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues. Expert Systems with Applications, 2016. 55: p. 329-340.

[12] Abbasi, N.R., et al., Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. Jama, 2004. 292(22): p. 2771-2776.

[13] Johr, R.H., Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist. Clinics in dermatology, 2002. 20(3): p. 240-247.

[14] Kittler, H., et al., Morphologic changes of pigmented skin lesions: a useful extension of the ABCD rule for dermatoscopy. Journal of the American Academy of Dermatology, 1999. 40(4):p. 558-562.

[15] Almansour, E. and M.A. Jaffar, Classification of Dermoscopic skin cancer images using color and hybrid texture features. IJCSNS Int J Comput Sci Netw Secur, 2016. 16(4): p. 135-9.

[16] Xie, F., et al., Melanoma classification on dermoscopy images using a neural network ensemble model. IEEE transactions on medical imaging, 2016. 36(3): p. 849-858.