2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Melanoma Skin Cancer Detection Using Recent Deep Learning Models*

Takfarines Guergueb, Moulay A. Akhloufi, *Senior Member IEEE*

***Abstract*— Melanoma is considered as one of the world's deadly cancers. This type of skin cancer will spread to other areas of the body if not detected at an early stage. Convolutional Neural Network (CNN) based classifiers are currently considered one of the most effective melanoma detection techniques. This study presents the use of recent deep CNN approaches to detect melanoma skin cancer and investigate suspicious lesions. Tests were conducted using a set of more than 36,000 images extracted from multiple datasets. The obtained results show that the best performing deep learning approach achieves high scores with an accuracy and Area Under Curve (AUC) above 99%.**

I. INTRODUCTION

Skin cancer is regarded as the most prevalent disease all over the globe. There are 5.4 million cases of skin cancer in the United States each year [1]. The global statistics regarding skin cancer are equally alarming as well. Recent studies show that the diagnosis of new melanoma cases has increased by 53% on an annual basis from 2008 to 2018 [2], with the rate of mortality of skin cancer expected to increase over the next few years. The rate of survival from skin cancer is less than 14%, if it is diagnosed in later stages. On the other hand, the rate of survival is about 97% if the skin cancer is detected at an early stage [3].

A trained dermatologist is found to typically adopt a sequence of steps, beginning with naked eye examination of suspicious lesions, accompanied by dermoscopy and biopsy. This will take time and may progress to later stages of the disease for the patient. Moreover, based on the ability of the clinician, a successful diagnosis is subjective. The most experimented dermatologists has been shown to have a accuracy of less than 80% when diagnosing skin cancer [4]. In public healthcare, there are not enough trained dermatologists available worldwide. The most hopeful sign of making recovery possible is the early detection of cancer. Early diagnosis of cancer produces a 10-year rate of survival [5]. Early detection of skin cancer is a difficult task because skin lesions are similar to each other and it is hard to determine if the lesion is malignant or benign. For identifying the lesion category, extensive analysis is required. Traditionally, a dermatoscope is used for studying the lesion. However, dermatoscope is an expensive instrument and only dermatologists have this instrument.

Extensive testing solutions have been established to rapidly detect skin cancer and address any of the above problems by designing computer image processing algorithms. Most of these algorithmic approaches were parametric, which implied that they needed a constant distribution of data. As it is not feasible to monitor the nature and variability of the data, these approaches may be ineffective to diagnose the disease correctly. Non-parametric approaches, though, are not based on the restriction that the data is in the usual distribution form. This study aims at detecting skin cancer melanoma using a deep learning approach.

This paper presents a comparative study to evaluate the best solution for detecting melanoma skin cancer. We use image data collected from the Society for Imaging Informatics in Medicine and the International Skin Imaging Collaboration (SIIM-ISIC) 2020 [6], a competition to identify melanoma in images of skin lesions. Additionally, we added external images from ISIC's archive [7], ISIC 2019 [8], ISIC 2018 [9], [10] and ISIC 2017 [11].

II. RELATED WORK

The recent development in the application of deep learning techniques has led important achievements in the detection of skin cancer. Traditional techniques for the detection and diagnosis of skin cancer use visual inspection and manual screening. However, the traditional process of manual screening is error prone, complex and time-consuming [12]. The skin lesion images have a very complex nature and it is not possible to accurately detect the skin cancer.

To address the issue, Aldwgeri *et al.* [13], proposed a transfer learning and CNN based classification model to classify skin disease. The models were pretrained on DenseNet121, InceptionV3, Xception, ResNet50 and VGG-Net. Finally, an ensemble approach is evaluated by combining the previous models. The results show an accuracy of 97%, a sensitivity of 80% and a specificity of 98.1% . For improving the accuracy, Majtner *et al.* [14], developed a technique based on the combination of linear discriminant analysis (LDA) and the features extracted from a pretrained AlexNet model . The results showed and improved accuracy of 85.8%, a sensitivity of 52%, a specificity of 97.4% and an AUC of 80.5%.

Khalid *et al.* [15], proposed a pretrained transfer learning and deep learning network for classification of skin cancer. The AlexNet architecture was used and a transfer learning was applied to identify the skin lesions. The results show an accuracy of 98.33%, a sensitivity of 98.93% and a specificity of 97.73%.

Bisla *et al.* [16], proposed a deep learning based automated system designed for the detection of melanoma. The ResNet50 architecture was used and a fine-tuning was

*This research was enabled in part by support provided by the New Brunswick Health Research Foundation (NBHRF).

T. Guergueb and M.A. Akhloufi are with the Perception, Robotics, and Intelligent Machines Research Group (PRIME), Dept of Computer Science, Université de Moncton, Moncton, NB, Canada {etg7520, moulay.akhloufi}@umoncton.ca

applied. The data augmentation and purification technique were used to address the issue of limited datasets. The paper achieved an AUC of 88% for melanoma classification.

Jayapriya *et al.* [17], proposed Fully Convolutional Networks (FCNs) framework for detecting the melanoma lesions. FCN is a deep learning model based on a convolution network that utilizes the pixels of the images for the prediction of melanoma disease. The paper incorporated VGG16 and GoogLeNet architectures for enhancing the accuracy of segmentation and a support vector machine for the classification after extracting the features from a segmented lesion by using deep residual networks and hand-crafted features. The proposed framework achieved an accuracy of 88.92%, a sensitivity of 69.33% and a specificity of 93.75% .

Zhang *et al.* [18], proposed a CNN based model to detect the melanoma by analysis of skin lesion images using the EfficientNetB6. The proposed technique achieved a AUC-ROC score of 91%.

III. PROPOSED APPROACH

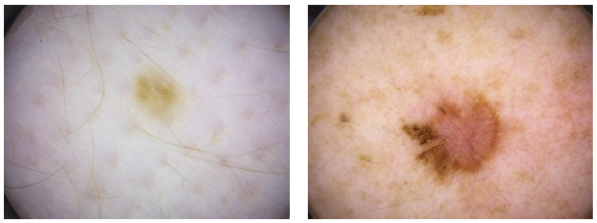
A. Image datasets

The data used in this research is collected from the Kaggle SIIM-ISIC 2020 competition [5]. The challenge was proposed in collaboration with the Society for Imaging Informatics in Medicine (SIIM) and the International Skin Imaging Collaboration (ISIC).

In this competition, the general objective is to identify melanoma in images of skin lesions. In particular, determine which are likely to represent a melanoma using images from the same patient. The dataset consists of both metadata and images in the DICOM format commonly used for medical imaging. The images have been resized to 1024x1024 and are available as TFRecord format.

This challenge corresponds to a binary classification (benign and malignant). The set of data is highly imbalanced, 33126 images of which 584 are malignant images. To fix this issue, we added 580 malignant images from ISIC's archive [7], 1182 malignant images from ISIC 2019 [8], and 1614 malignant images from ISIC 2018 and ISIC 2017 [9]–[11].

The new dataset contains a total of 36,502 images, where 32542 are benign images and 3960 are malignant images. Figure 1 shows example images of benign and malignant melanoma disease.



(a) (b)

Fig. 1: Sample images from the skin cancer dataset (a): Benign; (b): Malignant.

B. Data augmentation

Data augmentation is a technique that increases the diversity of the datasets for training models without collecting new data. We used a translation of +- 10, rotations between -5 and +5 with a 1 degree increment. Figure 2 shows the output after applying the data augmentation technique on a single input image.

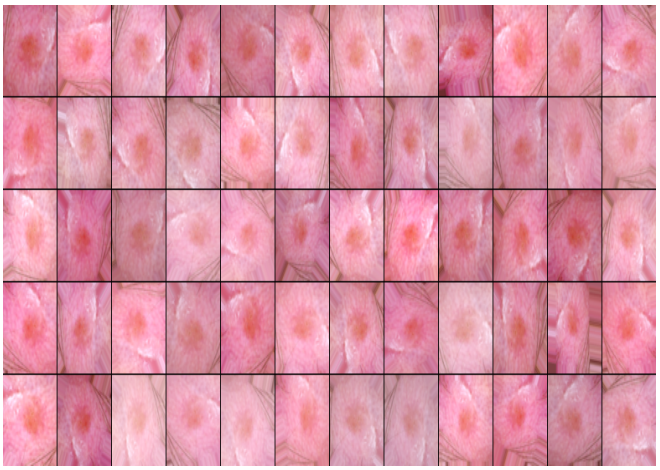


Fig. 2: The image showing the data augmentation applied on a single input image. The first image is the original image and other images are transformed images after applying the data augmentation technique.

C. Learning rate scheduler

The learning rate scheduler is used to reduce the learning rate (LR) during the training process of deep neural networks. We used 1cyclic policy learning rate scheduler due to its ability to use higher learning rates and achieve faster convergence (Super-Convergence) [19]. The 1cyclic policy is the combination of linearly increasing the learning rate with an option of decreasing the momentum during the first half cycle, afterwards the learning rate is decreased gradually is the second half cycle while increasing the momentum. The cycle starts with a minimum and maximum learning rate boundaries with a step-size. The cycle is based on two steps where one step increases linearly from minimum to maximum and second step decreases linearly. Finally, the learning rate is sharply reduced at the end of the cycle.

The 1cyclic policy is based on cyclic learning rates (CLR) with a slight modification, instead of several cycles, only one cycle will be used during all the training process. Mathematically the proposed formula for CLR is as follows:

$$\eta_t = (\eta_{min} + (\eta_{max} - \eta_{min})(max(0, 1 - x)) \quad (1)$$

where x is defined as:

$$x = \left| (\frac{iterations}{stepsize} - 2(cycle) + 1) \right| \quad (2)$$

and cycle can be calculated as:

$$cycle = floor\left(1 + \frac{iterations}{2(stepsize)}\right) \quad (3)$$

where η_{min}, η_{max} are the boundaries of the learning rate, iterations represents the number of completed mini-batches and stepsize defines one half of a cycle length. The term (1-x) should always be positive [19].

The proposed model sets a warm up period of 35 epochs and after which the learning rate starts decreasing to a minimum LR of 0.00001. To keep the same LR for the next period of training we have another variable called LR sustain. In our work we set LR sustain to 0, meaning that the LR changes during all the training process. The decay rate is fixed to 0.8 as it can optimally vary the learning rate for the optimizer functions [19]. In our experience we set the initial LR to 0.00001 and the maximum LR to 0.00040 [19].

D. Deep learning models

We tested various pre-trained models on different architectures such as VGG [20], DenseNet [21], ResNet [22], EfficientNet [23], MobileNet [24], NASNet [25], Inception-ResNetV2 [26], Xception [27] and Inception [28] to select the optimal model for classifying the images into malignant or benign [29]. details can be seen in table I.

TABLE I: Model architectures

Model	Sub models
VGG	VGG16, VGG19
ResNet	ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, ResNet152V2.
EfficientNet	EfficientNetB5, EfficientNetB6, EfficientNetB7
DenseNet	DenseNet121, DenseNet169, DenseNet201
Inception	InceptionResNetV2, InceptionV3, Xception
MobileNet	MobileNetV2, NASNetLarge

The hyperparameters of the models are fine-tuned to improve the performance. The classification layer of each model functioning as the output is replaced with two layers; Global Average Pooling layer (GAP) and Sigmoid layer. The Global Average Pooling layer (GAP) is added to improve the accuracy and reduce the overfitting and Sigmoid layer predicts the output in the form of probabilities for melanoma detection.

IV. EXPERIMENTAL RESULTS

A. Training and test datasets

The proposed models are trained and tested on a new dataset of 36502 images created from multiple public datasets such as SIIM-ISIC 2020 [6], ISIC's archive [7], ISIC 2019 [8], ISIC 2018 [9], [10] and ISIC 2017 [11]. The images are resized to 512 × 512 pixels. The dataset is randomly divided into train with 28719 samples and test with 7300 samples. To overcome the problem of overfitting the data augmentation was applied to the training set. The Rectified Adam is used as optimizer and we train the models with 1cyclic policy. The models are trained for 100 epochs using a batch size of 64, even at 25 epochs the model achieved good results.

B. Performance evaluation

The proposed models are evaluated using the following metrics: Area Under Curve (AUC), Sensitivity (SN), Specificity (SP), and Accuracy. Table II shows the results of different DL models for the detection of melanoma. It can be clearly seen in table II that EfficientNetB7 outperforms other models with an ACC of 99.33%, a SN of 98.78% and a SP of 99.38%. The proposed model is trained with datasets that have heterogeneous data samples and are being generated from multiple datasets. Moreover, the proposed 1cyclic policy converged the model faster with higher learning rates and as a result, the proposed model achieved the highest AUC, sensitivity and accuracy compared to the existing models proposed in the past for the detection of melanoma.

Table III shows the comparison of the most efficient of the proposed models with state-of-the-art models. Our proposed model achieved the highest AUC, sensitivity and accuracy compared to the most recently published work for the detection of melanoma.

V. CONCLUSION

The paper presented the comparative analysis of deep learning (DL) models for the detection of melanoma. We trained and tested our models in a new large dataset built by combining multiple public datasets. In our study, we compared 20 different DL models. The results show that EfficientnetB7 outperforms all other models achieving an AUC of 99.01%.

In future work, we will further explore the techniques to tackle the imbalance in the available datasets. Besides we will work on the multi-classification features of skin lesions such as basal cell carcinoma, vascular lesions, benign keratosis, dermatofibroma, and actinic keratosis.

REFERENCES

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer Statistics for Hispanics/latinos, 2012," *CA: a cancer journal for clinicians*, vol. 62, no. 5, pp. 283–298, 2012.

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level Classification of Skin Cancer with Deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[4] R. Marks, "Epidemiology of melanoma: Clinical dermatology. review article," *Clinical and Experimental Dermatology: Clinical dermatology*, vol. 25, no. 6, pp. 459–463, 2000.

[5] P. D. Sonali Patil and C. J. Sankirtan Bhatt, "Skin cancer detection and classification," in *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI), Langkawi*. IEEE, 2017, pp. 1–6.

[6] SIIM-ISIC, "SIIM-ISIC Melanoma Classification," https://www.kaggle.com/c/siim-isic-melanoma-classification/data, online; accessed February 2021.

TABLE II: Results of deep learning models for melanoma detection

Methodes	**ACC**	**SN**	**SP**	**AUC**
VGG16	96.01	85.01	97.01	91.01
VGG19	96.12	77.11	97.73	87.42
ResNet50	99.01	96.14	99.25	97.70
ResNet50V2	98.62	96.14	98.83	97.49
ResNet101	99.13	97.57	99.27	98.42
ResNet101V2	98.94	96.69	99.14	97.92
ResNet152	99.01	97.90	99.01	98.50
ResNet152V2	98.84	96.14	99.07	97.61
InceptionResNetV2	99.18	97.35	99.33	98.34
InceptionV3	99.01	97.57	99.19	98.38
Xception	99.21	98.01	99.31	98.66
MobileNet	98.94	97.68	99.01	98.37
MobileNetV2	98.84	93.28	99.31	96.30
NASNetLarge	96.92	63.80	**99.73**	81.77
EfficientNetB5	99.19	98.34	99.26	98.80
EfficientNetB6	99.19	98.56	99.24	98.90
EfficientNetB7	99.33	**98.78**	99.38	**99.01**
DenseNet121	99.23	98.45	99.29	98.87
DenseNet169	99.21	98.45	99.28	98.87
DenseNet201	**99.36**	97.79	99.49	98.64

TABLE III: Comparison with past work

Author	**ACC**	**SN**	**SP**	**AUC**
Aldwgeri *et al.* [16]	97.01	80.01	98.10	-
Majtner *et al.* [18]	85.80	52.01	97.40	80.50
Khalid *et al.* [18]	98.33	98.93	97.73	-
Bisla *et al.*	-	-	-	88.01
Jayapriya *et al.*[20]	88.92	69.33	93.75	-
Zhang *et al.* [22]	91.01	-	-	-
Proposed model	**99.33**	**98.78**	**99.38**	**99.01**

[7] ISICs, "International Skin Imaging Collaboration ISICs Archive," https://challenge.isic-archive.com/landing, online; accessed February 2021.

[8] ISIC 2019, "International Skin Imaging Collaboration ISIC 2019," https://challenge.isic-archive.com/landing/2019, online; accessed February 2021.

[9] V. R. Noel Codella, M. E. C. Philipp Tschandl, D. G. Stephen Dusza, A. K. Brian Helba, M. M. Konstantinos Liopyris, and A. H. Harald Kittler, "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)," *https://arxiv.org/abs/1902.03368*, 2018.

[10] P. Tschandl and H. Rosendahl, C.and Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." *doi:10.1038/sdata.2018.161 (2018).*, 2018.

[11] G. D. Codella N, H. B. Celebi ME, D. S. Marchetti MA, L. K. Kalloo A, K. H. Mishra N, and H. A., "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)." in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 168–172.

[12] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE transactions on medical imaging*, vol. 36, no. 4, pp. 994–1004, 2016.

[13] A. Aldwgeri and N. F. Abubacker, "Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images," in *International Visual Informatics Conference*. Springer, 2019, pp. 214–226.

[14] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Optimised deep learning features for improved melanoma detection," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11 883–11 903, 2019.

[15] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*. IEEE, 2018, pp. 90–93.

[16] D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polsky, "Towards automated melanoma detection with deep learning: Data purification and augmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 2720–2728.

[17] K. Jayapriya and I. J. Jacob, "Hybrid fully convolutional networks-based skin lesion segmentation and melanoma detection using deep feature," *International Journal of Imaging Systems and Technology*, vol. 30, no. 2, pp. 348–357, 2020.

[18] R. Zhang, "Melanoma detection using convolutional neural network," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*. IEEE, 2021, pp. 75–78.

[19] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay," *arXiv preprint arXiv:1803.09820*, 2018.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[21] Z. L. Gao Huang and L. van der Maaten, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016.

[22] X. Z. Kaiming He and J. S. Shaoqing Ren, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[23] M. Tan and Q. V. Le, "Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019.

[24] M. Z. Andrew G. Howard, D. K. Bo Chen, T. W. Weijun Wang, and H. A. Marco Andreetto, "Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.

[25] V. V. Barret Zoph and Q. V. L. Jonathon Shlens, "Learning transferable architectures for scalable image recognition," *CoRR*, vol. abs/1707.07012, 2017.

[26] S. I. Christian Szegedy and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 201.

[27] F. Chollet, "Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016.

[28] W. L. Christian Szegedy, P. S. Yangqing Jia, D. A. Scott Reed, V. V. Dumitru Erhan, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

[29] C. Zhang, P. Benz, D. M. Argaw, S. Lee, J. Kim, F. Rameau, J.-C. Bazin, and I. S. Kweon, "Resnet or densenet? introducing dense shortcuts to resnet," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 3550–3559.