

Skin lesion classification using convolutional neural networks based on Multi-Features Extraction

Benyahia Samia¹, Meftah Boudjela²[0000-0001-5772-1878], and Olivier Lézoray³[0000-0003-0540-543X]

¹ Department of Computer Science, Faculty of Exact Sciences, University of Mascara, Mascara, Algeria

² LRSBG Laboratory, University of Mascara, Mascara, Algeria

³ Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

Abstract. In the recent era, deep learning has become a crucial technique for the detection of various forms of skin lesions. Indeed, Convolutional neural networks (CNN) have become the state-of-the-art choice for feature extraction. In this paper, we investigate the efficiency of three state-of-the-art pre-trained convolutional neural networks (CNN) architectures as feature extractors along with four machine learning classifiers to perform the classification of skin lesions on the PH2 dataset. In this research, we find out that a DenseNet201 combined with Cubic SVM achieved the best results in accuracy: 99% and 95% for 2 and 3 classes, respectively. The results also show that the suggested method is competitive with other approaches on the PH2 dataset.

Keywords: Feature extraction · Classification · Skin lesion · Convolutional Neural Networks.

1 Introduction

Cancer is considered one of the leading causes of death in the world. It is estimated that the number of people diagnosed with cancer will double in the next few decades. Fortunately, some cancers have a high chance of cure through early detection, appropriate treatment, and care in the early stages. Skin lesions are an abnormal change in the tissue either on the surface of the skin or under the skin. Skin lesion usually grows in an irregular way beyond their usual boundaries as compared to the surrounding tissue. It is primarily caused by excessive exposure to ultraviolet radiation. Skin lesions can be classified into two categories: benign skin tumors such as nevus, or malignant tumors such as melanoma that is the least common but the most harmful form of skin cancer. Until the last few years, computer-aided diagnosis was used for the early detection of skin cancer from dermoscopy images. It was based on handcrafted features extraction, such as statistical pixel-level features, shape features, texture features, and relational features from the images in order to train classical machine learning models for

distinguishing the lesion from the surrounding healthy skin. Currently, such a computer-aided diagnosis is still a challenging task. Most of researchers are now investigating deep learning techniques. It is expected that they will in the next future be able to reach a performance similar to those of dermatologists, directly from dermoscopic images. Indeed, in recent years there have been many successful applications of machine learning techniques (ANN, SVM, KNN) as well as deep learning approaches and in particular of Convolutional Neural Networks (CNNs) such as AlexNet, VGGNet, ResNet, DenseNet, GoogleNet, Inception, EfficientNets. Several contributions have been proposed recently in terms of new models that have made significant improvements in the detection and the classifications of skin lesions. Hopefully, such new techniques will help to improve patient survival rates.

In this paper, we investigate the efficiency of various commonly pre-trained convolutional neural networks architectures as feature extractors with various machine learning classifiers to perform the classification for dermoscopic images from the PH2 dataset.

The rest of the paper is structured as follows. Section 2 presents some studies on the application of CNN for melanoma diagnosis on the PH2 dataset. Section 3 describes the dataset used, the various pretrained CNN used as feature extractor, the machine learning classifiers, and the evaluation metrics. Section 4 describes experiments and discusses of the obtained results. Last Section provides the conclusion of this study.

2 Related works

Several automated recognition methods for skin lesion images have been proposed in the last decade. In particular, recent classification approaches of skin lesions have been dominated by CNN approaches. In this section, we review some studies on the application of CNN for melanoma diagnosis on the PH2 dataset.

Ozkan et al. [1] proposed a study on skin lesions classification based on dermoscopic images to classify images of the PH2 datasets into three classes: normal, abnormal, and melanoma. They used four different machine learning classifiers: ANN, SVM, KNN, and Decision Tree. The achieved accuracies were 92.50%, 89.50%, 82.00%, and 90.00% for ANN, SVM, KNN, and DT, respectively.

Ghasem et al. [2] proposed two-hybrid approaches to combine four heterogeneous classifiers KNN, SVM, ENN, and MLP. Their first approach was the Structure-Based on Stacking (SBS), while the other was the Hierarchical Structure Based on Stacking (HSBS). The authors have considered a preprocessing step, a segmentation step for analyzing the lesion area and have applied different feature extraction methods based on the shape, color, and texture. Finally, a classification step is used either SBS or HSBS by combining the different classifiers. The evaluation was based on the PH2 dermoscopic images with different selected features. The achieved accuracy was 96.7% for the HSBS method and 98.5% for SBS.

Ann et al. [3] used a deep learning approach to classify skin lesions in dermoscopic images. First, the authors preprocess the images to remove unwanted artifacts, such as hairs, using morphological operators and an inpainting algorithm. They classify the images using a CNN AlexNet architecture, and finally, tested the classifier using both preprocessed and unprocessed images from the PH2 dataset. The obtained accuracy for the two classes was 93%, while the accuracy for the three classes was 67.5% with preprocessed images.

Singh et al. [4] proposed an approach for the skin lesions classification using a segmentation step by a thresholding method and the ABCD rule for feature extraction. Finally, a SVM classifier is used to make a decision. The proposed system achieved an accuracy of 92.5% on the PH2 data set.

Filali et al. [5] proposed a skin lesion classification system based on a fusion of handcrafted features (shape, skeleton, color, and texture) and features extracted from deep learning architectures (VggNet16, ResNet18, AlexNet, and GoogLeNet). They used a Support Vector Machine (SVM) classifier to make a final decision. Their approach achieved an accuracy of 98% on the PH2 dataset.

Sanket et al. [6] proposed a skin lesion diagnosis system. The authors begin with a processing step using median and Wiener filters to remove noise, followed by a segmentation step using a watershed and morphological filters. After that, a feature extracting step is performed using the Grey Level Co-occurrence Matrix (GLCM), color and geometrical features. Finally, the classification is performed using KNN, SVM, and an ensemble method. They achieved an accuracy of 92% on the PH2 dataset.

Khalid et al. [7] proposed a skin lesions classification framework based on fine-tuning a pre-trained deep learning network AlexNet. They replaced the last layer with a softmax to classify the lesion on three different classes from the PH2 dataset and achieved an accuracy of 98.61%.

3 Methodology and Materials

The process we propose is made up of four successive parts. The first component is the input of dermoscopic images. The second is the feature extraction part with three commonly pre-trained convolutional neural networks (CNN) architectures. From the obtained features, two databases are created: learning and test databases. The third component of the system is the learning step by a set of classifiers where each one is performed individually. Finally, the last is the validation step from the test dataset.

In addition, as shown in Figure 1, there are two kinds of extracted features with the skin lesions dataset:

1. Extracted features using the original dataset.
2. Extracted features with an augmented dataset using data augmentation techniques (will be detailed later).

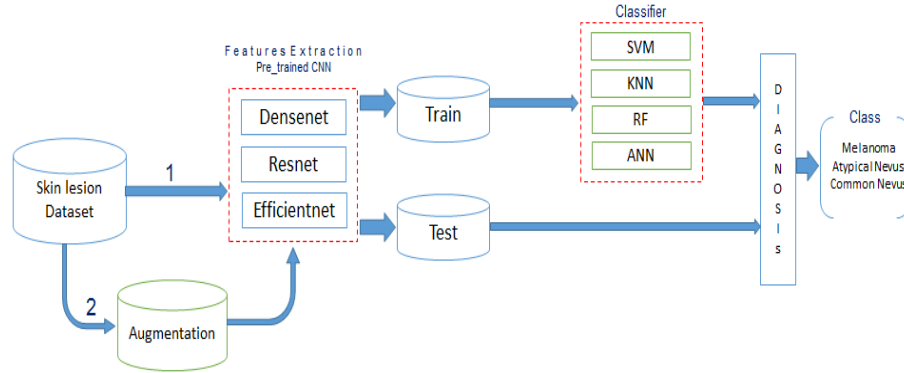


Fig. 1. Proposed method's flowchart.

3.1 PH2 dataset

The Portuguese dermatological service of the Pedro Hispano Hospital and the University of Porto have collaborated to establish the PH2 database [16]. There are 200 dermoscopic images in the PH2 database, with 80 common nevus, 80 atypical nevus, and 40 melanomas. These are 8-bit RGB color images with a 768 x 560 pixels. The images can be obtained online from Hospital Pedro Hispano (<https://www.fc.up.pt/addi/ph2%20database.html>) and some are shown in Figure 2.



Fig. 2. Example of PH2 skin images.

3.2 Convolutional Neural Networks

The convolutional neural networks (CNN) are a special type of neural networks, that possesses several convolutional layers to extract learned features directly from images in a hierarchical manner, and they can be used for classification tasks. Recently, CNN have been used to improve the performance in many different applications with various architectures such as LeNet, AlexNet, VGGNet, ResNet, DenseNet, GoogLeNet among others. Such CNN architectures can be used as pre-trained architectures that were initially trained on large datasets. In

this section, we present a brief overview of such pre-trained CNNs that can be used as extractors of features for skin lesion images.

ResNet ResNet [13] introduces residual blocks that include skip-connection between layers where each layer feeds into the next layer and directly into the layers about 23 hops away. ResNet architecture contains a 3x3 convolution filter, global average pooling, and max-pooling layers, residual blocks, batch normalization layers followed by a fully connected layer, and softmax for the classification. ResNet architecture has several variations with a different number of layers such as ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152.

DenseNet DenseNet architecture [14] uses shortcut connections composed of a dense block linked by transition layers. Each dense block contains convolution layers where each layer is connected to all preceding other layers. All the feature maps from previous layers are passed to all subsequent layers of the same block. DenseNet architecture contains a convolution filter, global average pooling, max-pooling layers, transition layers, dense blocks followed by fully connected layers, and softmax for the classification. DenseNet architecture proposes several variations with a different number of layers such as DensNet-121, DensNet-169, DensNet-201, and DensNet-246.

EfficientNet EfficientNet architecture [15] is based on the principle of scaling the different dimensions such as depth, width, and image resolution of the network at the same time uniformly by using a fixed compound coefficient. EfficientNet is considered as a family of eight different CNN models: EfficientNet-B0 to B7. The EfficientNet-B0 represents the baseline version with an input size of 224×224 , and it is based on the inverted bottleneck residual and squeeze-and-excitation blocks.

3.3 Classifiers

Accurate automated classification of a skin lesion in its early stages saves effort, time, and human life. The purpose of our study will be to find the most accurate classifier from several machine learning classifiers for skin lesion classification. This section provides a summary of four machine learning classifiers that will be employed for the classification of skin lesions: Artificial neural network, Support Vector Machines, K-Nearest Neighbor, and Random Forest.

Artificial Neural Network ANN are described as compound systems made up of at least two layers of neurons, an input layer, and an output layer, usually including hidden layers [9]. Each layer contains a large number of artificial neurons that constitute an interconnected network in a weighted way. Each neuron in the network receives digital information as signals from neighboring neurons, and each of these values is assigned a particular "weight" representative of the strength of the connection.

Support Vector Machines SVMs algorithm plots the feature's value at a point in a high-dimensional space [10]. Then, it performs classification by finding the hyperplane, that determines a an hyperplane that can classify the data and differentiates the space into two zones. The set of points near the hyperplane is referred to as the Support Vectors.

K-Nearest Neighbor KNN is a nonparametric method associated with only one parameter that represents the number K of nearest neighbors and a training data set [11]. The principle of the KNN model consists of choosing the K data points closest to the point under study in order to predict its class.

Random Forest RF [12] consists in constructing and training multitude of decision trees in parallel on slightly different data subsets with random variables.

3.4 Evaluation metrics

In order to evaluate the performance of our classifiers and to compare the different results of different scenarios, we used four evaluation metrics. These measures are accuracy, sensitivity, specificity, and precision.

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \quad (1)$$

$$Sensitivity = \frac{T_P}{T_P + F_N} \quad (2)$$

$$Specificity = \frac{T_N}{T_N + F_P} \quad (3)$$

$$Precision = \frac{T_P}{T_P + F_P} \quad (4)$$

Where T_P , T_N , F_P , F_N , and refer to true positive, true negative, false positive and false negative respectively.

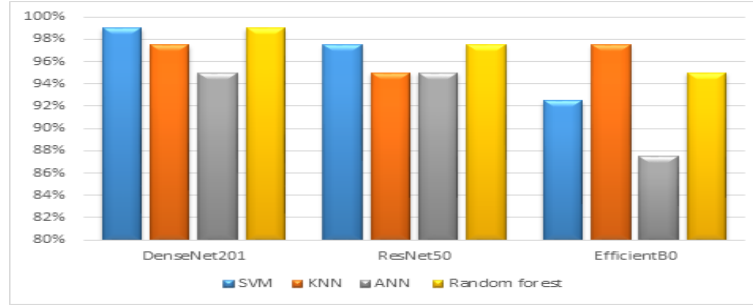
4 Experimental results and discussion

In our study for a melanoma diagnosis, two scenarios of experiments are proposed using the PH2 dataset:

- Classification of skin lesions into two types of lesions: melanoma or non-melanoma
- Classification of skin lesions into three types: melanoma, atypical nevus, or common nevus

Table 1. Summary of results for melanoma or non-melanoma

	DenseNet201				ResNet50				EfficientB0			
	Acc.%	Sens.%	Spec.%	Prec.%	Acc.%	Sens.%	Spec.%	Prec.%	Acc.%	Sens.%	Spec.%	Prec.%
SVM	99	99	99	99	97.5	83.33	83.33	98.68	92.5	65.31	65.31	72.36
KNN	97.5	98.64	98.64	87.5	95	81.98	91.98	91.98	97.5	98.64	98.64	87.5
ANN	95	96.96	96.96	88.88	95	97.36	97.36	75.00	87.50	78.12	78.12	81.16
RF	99	99	99	99	97.5	83.33	83.33	98.68	95	97.29	97.29	80

**Fig. 3.** Accuracy according to DenseNet201, ResNet50, and EfficientB0 with four classifiers for melanoma or non-melanoma .

For both scenarios DenseNet201, ResNet50, EfficientB0 pre-trained CNN models are used as feature extractors, and ANN, SVM, KNN, RF classifiers used for classification. The latter are compared one to the other on the PH2 dataset. As previously mentioned, the PH2 dataset contains a total of 200 images that we divided into two parts.

- 80% of the dataset (160 images) is used for training
- 20% of the dataset (40 images) is used for testing the effectiveness of the models.

For the second scenario concerning the classification into three classes, a data augmentation techniques are included with the following augmentations: rotations, zooming, shearing, flips (top-bottom, left right), skew-left-right, contrast enhancement. All experiments were performed with the same training dataset and tested with the same test set.

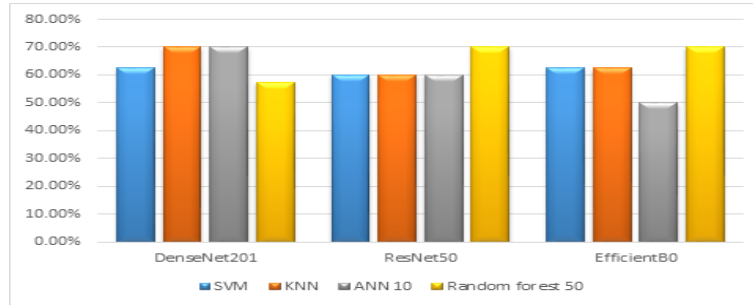
Table 1 and Figure 3 depict the performance results obtained for classifying PH2 dataset in two classes: melanoma or non-melanoma with DenseNet201, ResNet50, EfficientB0 CNN extractor and ANN, SVM, KNN, RF classifiers.

Table 2 and Figure 4 depict the performance results obtained for classifying PH2 dataset in 3 classes: melanoma, atypical nevus or common nevus without augmentation with DenseNet201, ResNet50, EfficientB0 CNN extractor and ANN, SVM, KNN, RF classifiers.

Table 3 and Figure 5 depicts the performance results obtained for classifying PH2 dataset in 3 classes: melanoma, atypical nevus or common nevus with aug-

Table 2. Summary of results for melanoma, atypical nevus or common nevus without augmentation

	DenseNet201				ResNet50				EfficientB0			
	Acc.%	Sens.%	Spec.%	Prec.%	Acc.%	Sens.%	Spec.%	Prec.%	Acc.%	Sens.%	Spec.%	Prec.%
SVM	62.50	62.77	78.38	73.19	60.00	60.20	77.16	71.44	62.50	40.09	76.82	60.45
KNN	70.00	70.37	81.48	75.91	60.00	57.06	75.59	61.11	62.50	41.14	76.20	62.19
ANN	70	72.80	83.33	72.87	60.00	28.57	74.04	34.43	50.00	30.51	72.17	57.44
RF	57.50	58.69	75.30	69.59	70.00	44.75	82.09	76.50	70.00	47.78	82.38	67.35

**Fig. 4.** Accuracy according to DenseNet201, ResNet50, and EfficientB0 with four classifiers for melanoma, atypical nevus or common nevus without augmentation.**Table 3.** Summary of results for melanoma, atypical nevus or common nevus with augmentation.

	DenseNet201				ResNet50				EfficientB0			
	Acc.%	Sens.%	Spec.%	Prec.%	Acc.%	Sens.%	Spec.%	Prec.%	Acc.%	Sens.%	Spec.%	Prec.%
SVM	95	96.67	97.22	96.29	87.50	91.25	92.77	91.02	90.00	92.91	94.16	92.59
KNN	87.50	91.25	93.51	86.31	82.50	80	90.37	77.14	82.50	87.50	89.72	87.30
ANN	92.50	87.77	93.88	92.50	95.00	96.29	96.00	96.66	85.00	88.88	90.90	89.16
RF	85.00	88.75	91.57	83.57	82.50	87.50	90.92	79.39	82.50	87.50	90.04	82.47

mentation with DenseNet201, ResNet50, EfficientB0 CNN extractor and ANN, SVM, KNN, RF classifiers.

Through the evaluation and results obtained from Table 1 to Table 3, we find that the DenseNet201 model achieved the best results as compared to other CNN models. On the other side, the cubic SVM classifier achieved the best results as compared to others classifiers ANN, KNN, RF. Through Figure 3 to Figure 5, we notice that data augmentation considerably improves the results of classification for three classes. From Table 1, Table 2 and Table 3, the statistics show that SVM classifiers perform the highest accuracy of 99% for melanoma and non-melanoma detection while 95% for melanoma, atypical nevus, or common nevus with augmentation combined with DenseNet architecture.

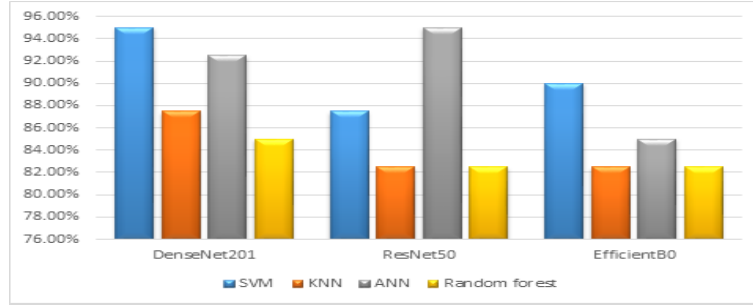


Fig. 5. Accuracy according to DenseNet201, ResNet50, and EfficientB0 with four classifiers for melanoma, atypical nevus or common nevus with augmentation.

Table 4. Results of the proposed approach compared to various approaches for PH2 dataset

Authors	No. of classes	Method	Accuracy
Ghasem et al. [2]	2	SBS	98.50%
Ghasem et al. [2]	2	HSBS	96.70%
Ann et al. [3]	2	AlexNet	93.00%
Filali et al. [5]	2	SVM	98.00%
Sanket et al. [6]	2	SVM	92.00%
Proposed work	2	DenseNet201+SVM	99.00%

Table 5. Results of the proposed approach compared to various approaches for PH2 dataset

Authors	No. of classes	Method	Accuracy
Ozkan et al. [1]	3	MLP	92.50%
Ann et al. [3]	3	AlexNet	67.50%
Singh et al. [4]	3	SVM	92.50%
Khalid et al. [7]	3	AlexNet	98.61%
Proposed work	3	DenseNet201+SVM	95.00%

Table 4 and Table 5 depict the comparison of the proposed approach with the state-of-the-art models in terms of accuracy for the PH2 dataset. As it can be seen our proposed approach is competitive with the state-of-the-art.

5 Conclusion

In order to take advantage of deep learning models' ability to extract features from skin lesion images, several experiments were conducted with convolutional neural networks (CNN) architectures combined with various classifiers. In this work, we have proposed to use DenseNet201, ResNet50, EfficientB0 pre-trained CNN architectures as feature extractors and the Artificial neural network (ANN), Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random Forest

(RF) as classifiers to evaluate the classification of skin lesions from PH2 datasets with 2 or 3 classes. The results found show that the DenseNet201 model combined with the SVM classifier gives the better score for two classes with an accuracy of 99%, moreover for three classes with 95%, which is competitive with the actual state-of-the-art.

Acknowledgement

This work was completed as part of the Hubert Curien Partnership (PHC) TAS-SILI cooperation program between France and Algeria under the project code 19MDU212.

References

1. Ozkan, I., Koklu, M.: Skin Lesion Classification using Machine Learning Algorithms. *International Journal of Intelligent Systems and Applications in Engineering* **5**, 285–289 (2017)
2. Ghasem Shakourian, G., Kordy, H. M., Ebrahimi, F.: A hierarchical structure based on Stacking approach for skin lesion classification. *Expert Syst. Appl.* **145**, 113–127 (2020)
3. Salido, Julie Ann A., Ruiz, C. R.: Using Deep Learning to Detect Melanoma in Dermoscopy Images. *International Journal of Machine Learning and Computing* **8**(1), 61–68 (2018)
4. Singh, L., Janghel, R. R., Sahu, S.: Designing a Retrieval-Based Diagnostic Aid using Effective Features to Classify Skin Lesion in Dermoscopic Images. *Procedia Computer Science* **167** : 2172–2180 (2020)
5. Filali, Y., El Khoukhi, H., Sabri, M., Aarab, A.: Efficient fusion of handcrafted and pre-trained CNNs features to classify melanoma skin cancer. *Multimedia Tools and Applications*, 1–20 (2020)
6. Sanket, K., Chandra, J.: Skin Cancer Classification using Machine Learning for Dermoscopy Image **1457** (2019)
7. Khalid, M. H., Kassem, M. A., Foad, M. M.: Skin Cancer Classification using Deep Learning and Transfer Learning. 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), 90–93 (2018)
8. Satheesha, T. Y., Satyanarayana, D., Prasad, M., Kashyap D. Dhruve: Melanoma Is Skin Deep: A 3D Reconstruction Technique for Computerized Dermoscopic Skin Lesion Classification. *IEEE Journal of Translational Engineering in Health and Medicine* **5** 1–17 (2017)
9. Livingstone, D. J.: *Artificial Neural Networks: Methods and Applications*, Humana Press (2011)
10. Vapnik, V.: *Statistical learning theory* (1998)
11. Larose, Daniel T., Larose, Chantal D.: *Discovering Knowledge in Data: an Introduction to Data Mining*, Wiley (2014)
12. Breiman, L.: Random Forests, *Machine Learning* **4** 5–32 (2001)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (2016)

14. Huang, G., Liu, Z., Weinberger, K. Q.:Densely Connected Convolutional Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261–2269 (2017)
15. Tan, M., Le, Q.V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, ArXiv (2019)
16. Mendonan, T. and Ferreira, P. and Marques, J. and Maral, A. and Rozeira, J.: PH2 - A dermoscopic image database for research and benchmarking, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 5437–5440 (2013)