Non-dermatoscopic Image Analysis
for the Recognition of Malignant Skin Diseases
with Convolutional Neural Network
and Autoencoders

Ricardo Coronado[(✉)] [ID], Alexander Ocsa [ID], and Oscar Quispe [ID]

Universidad Nacional de San Agustín, Arequipa, Peru
{rcoronado,aocsa,oquispepo}@unsa.edu.pe

Abstract. Every year, people around the world are affected by different skin diseases or cancer. Nowadays, these can only be detected accurately by clinical analysis and skin biopsy. However, the diagnosis of this malignant disease does not ensure the survival of the patient, since many clinical cases are detected in the terminal phases. Only early diagnosis would increase the life expectancy of patients.

In this paper, we propose a method to recognition malignant skin diseases to identify malignant lesions in non-dermatoscopic images. For the method, we use Convolutional Neural Network and propose the use of autoencoders as another classification model that provides more information on the diagnosis. Experiments show that our proposal reaches up to 84.4% of accuracy in the well-known dataset of the ISIC-2016. In addition, we collect non-dermatoscopic images of skin lesions and developed a new dataset to demonstrate the advantage of our method.

Keywords: Skin cancer · Convolutional Neural Networks
Autoencoders · Image classification

1 Introduction

Malignant skin diseases take thousands of lives around of the world. For example, in 2016 in the United States, 83510 new cases of skin cancer have been diagnosed, from this, 13650 people have died [11]. The detection of this cancer is performed by clinical analysis, and the best clinical method used is ABCD [3,7]. This method analyzes the morphology of the lesion and its evolution. However, it requires a manual procedure and a high level of proficiency. As a solution for this problem, some researchers proposed computer-assisted methods, based on statistics, pattern recognition, machine learning, and deep learning, among others [2].

According to the state of the art, some researches achieves good results detecting malignant and benign lesions. However, this one could be insufficient in real scenarios due to correlations between diseases of different classes, it is common to find cases of benign lesions that become malignant over time. Find the

subclass of a sample could provide more information for a specialist to make a successful diagnosis. In addition, the datasets analyzed are made up of dermatoscopic images, such data samples are inaccessible to people who don't have dermatoscopes. On the other hand, we have made up a dataset of non-dermatoscopic images, these one are samples of skin lesions taken with a conventional camera.

Many methods were used for the skin detection, but currently, the best results have been obtained with the use of Convolutional Neural Network (CNN), as demonstrated in [1]. In this work, we use the CNN architecture VGG-19 as [6], but we propose the use of Autoencoders (AEs) instead of fully-connected networks. Further, we have tested this one on dataset with 3 class and 11 sub-class. The main contribution of our work is the use of AEs as method of classification, to identify the kind of skin disease which a sample belongs. The result of this, will classify the samples as benign, premalignant or malignant diseases.

This paper is organized as follow: Sect. 2 presents the concepts for the development of the proposal, specifically about CNN and AEs, Sect. 3 describes the datasets used, Sect. 4 shows the proposed method, Sect. 5 shows the experimental results. Finally, in Sect. 6, presents the conclusions of the paper.

2   Background

The methods for detecting skin diseases are based on feature extraction. There are two approaches, a clinical analysis method [3] based on the specialist's experience and a method computer-aided that uses Machine Learning for processing samples [6,9].

2.1   Convolutional Neural Network (CNN)

Originally a CNN requires a lot of training to obtain good results, depending on the complexity of the training data. To reduce the time required and improve the accuracy results, some works such as [4], use transfer learning to initialize the filters of the network. This helps the process of feature extraction made on convolutional layers. On the other hand, fully connected layers are restarted to fine-tune the CNN and set the number of classes.

2.2   Classification by Reconstruction

An autoencoder (AE) can be seen as neural network that tries to reconstruct the input data, these are known as a class of unsupervised learning algorithms [5]. Unlike supervised algorithms, not need labels or class information, also AEs have been used as a method to pre-train a network and initialize its weights. According to [8], this research introduce the use of AEs as a classification method.

3   Datasets

In this section, we present a new dataset of non-dermatoscopic images, built using different sources[1]. This dataset consists of 2360 unsegmented images of

---

[1] http://rikardocorp.me:8000/.

medium and high quality, divided into three main classes (benign, pre-malignant and malignant), each class is divided into subclasses as shown in (Table 1).

Table 1. Skin diseases dataset

Category	Skin disease	Class	Sub-class	Train	Test	Total
Bening	Melanocytic Nevus	0	0	135	55	190
	Dermatofibroma	0	1	80	20	100
	Lentigo Solar	0	2	45	8	53
	Seborrheic Keratosis	0	3	249	75	324
Pre Malignant	Atypical Nevi	1	4	51	17	68
	Actinic Keratosis	1	5	60	20	80
	Keratoacanthoma	1	6	136	50	186
Malignant	Bowen	2	7	54	7	61
	Squamous Cell Carcinoma	2	8	49	11	60
	Melanoma	2	9	193	66	259
	Basal Cell Carcinoma	2	10	117	56	173

The sub-classes considered in this dataset were selected because they are the most common, lethal and easily confused by other lesions of less severity. This was done with the help of a specialist in dermatology and oncology, a professional at the National Institute of Neoplasm Diseases of Peru (INEN[2]). Finally, we pick 1554 samples for the final dataset, this was split in training (1169) and testing (385) samples.

In addition, we used 4 different datasets to validate our proposal; MNIST[3], CIFAR-10[4], SVHN[5] and the ISBI Challenge 2016 Dataset[6].

4 Proposed Method

Our proposed method is based on the use of Convolutional Neural Network (CNN) and Autoencoders (AE). For the evaluations, we measure the accuracy, precision, recall and fβ metric.

4.1 Data Preprocessing

Using a semi-automated process (Fig. 1), we segment the images using thresholding techniques [9], available in the Python library sklearn-image[7], we fine-tune

---

[2] http://www.inen.sld.pe/.
[3] http://yann.lecun.com/exdb/mnist/.
[4] https://www.cs.toronto.edu/~kriz/cifar.html.
[5] http://ufldl.stanford.edu/housenumbers/.
[6] https://isic-archive.com/.
[7] http://scikit-image.org/.

segmentation using a hand-craft tool available at github[8]. Then, we generate the images dataset at 224×224 pixels dimension.

ooo Limit of the lesion [Area A²]   ▬ Thresholding   ▬ Total area (zoom) [Area (A(1+x²))]
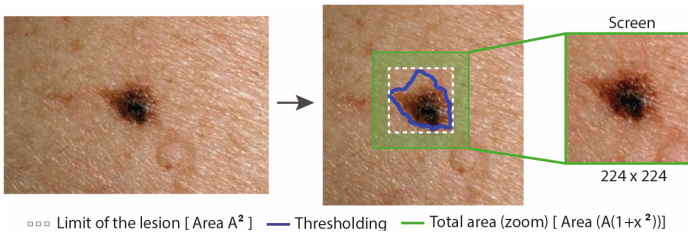
Fig. 1. Image segmentation process

Then, we generate synthetic data to increase the number of samples. For this, through the clinical analysis and specialist assistance (INEN), we pick the best samples of each subclass and performed rotations in 0, 90, 180 and 270 degrees to increase the training data by 33%. The testing data is immutable.

4.2   Features Extractor and Classifier

We use a VGG19 network architecture [6], which consists of 19 layers, as we can see in Fig. 2, this scheme conforms the feature extractor that we use. According to [10], to obtain the most general-purpose representation for learning is used the output of the last convolutional layer of the CNN. The original classifier is modified so that our network can classify three types of classes. Thus, the network weights are pre-trained on Imagenet[9] and the CNN was fine-tuned to the target dataset by transfer learning.

4.3   Clasification with Autoencoders

Before using AE as a classification method, we have to save the feature vectors of the first fully connected layer of the network, as shown in the Fig. 2, this one is the new representation of each image which consists of 4096 values. Then, we can train our first AE [5] which we will call global-AE. The global-AE is trained with the entire training dataset until the reconstruction error is minimized. Additionally, we generate n-autoencoders (n-AEs) which are cloned from global-AE, where n is the number of classes in the dataset (See Fig. 3).

Then, in the training phase, each training sample will only feed the AE associated with its class. What happens here is that each AE will be to specialize in reconstructing the data of its own class.

---

[8] https://github.com/rikardocorp/skin-diseases.
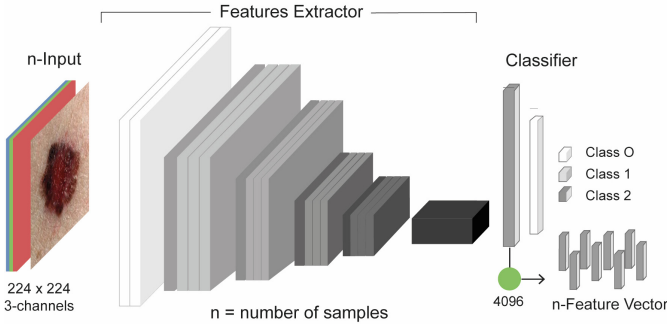[9] http://www.image-net.org.

Fig. 2. Convolutional Neural Network VGG-19

For the test phase, each sample is tested by the n-AEs, to generate a reconstruction error vector by sample, as we see in the Fig. 3. Finally, we get the minimum reconstruction error for each vector to know the class to which the sample belongs.
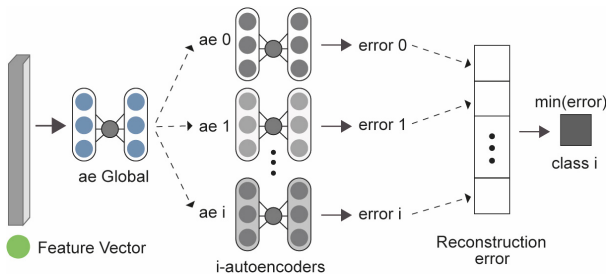

Fig. 3. Each AE consists of 5 layers (1 to feed, 2 for enconding and 2 for decoding)

5   Results

To validate the classification model with AEs, We train the CNN network with different datasets (MNIST, CIFAR-10, SVHN and ISBI) to get the accuracy classification and the feature vectors as described in Sect. 4.2. These feature vectors feed our method of classification with AE named CNN-AE described in Sect. 4.3. We setting our CNN with a minibatch of 30 samples, and learning rate between $<10^{-3}, 10^{-5}>$. In Table 2, we can see the results obtained by CNN and CNN-AE. Here we can observe that CNN-AE is comparable to CNN. If we focus only on the accuracy indicator, we get results that are in general slightly worse.

In addition, we compared our results with the results of the ISBI Contest Dataset. It is available on its ISBI-2016 webpage[10]. Table 3 shows these results.

---

[10] https://challenge.kitware.com/#phase/56fc2763cad3a54f8bb80e51.

Table 2. Comparison of CNN and CNN-AE classification models

	CIFAR-10		MNIST		SVHN		ISBI	
	CNN	CNN-AE	CNN	CNN-AE	CNN	CNN-AE	CNN	CNN-AE
Accuracy	0.910	0.896	0.990	0.978	0.909	0.860	0.846	0.844
Precision	0.911	0.895	0.990	0,977	0.903	0.846	0.759	0.762
Recall	0.910	0.896	0.990	0.977	0.904	0.855	0.759	0.702
$F\beta$	0.910	0.895	0.990	0.977	0.903	0.847	0.759	0.744

The winner of the contest has an Accuracy 1% higher than our model, while our Average Precision is slightly higher. However, according to the sensitivity metric, our model is better identifying True Positive, equivalent to cases of skin cancer.

Table 3. Proposed method (CNN, CNN-AE) vs the winner of ISBI contest

	Accuracy	Area under Roc	Avg. precision	Sensitivity	Specificity
1er Place	0.855	0.783	0.624	0.547	0.931
*CNN	0.846	0.758	0.651	0.613	0.904
*CNN-AE	0.844	0.702	0.610	0.466	0.937

Finally, we perform a comprehensive evaluation for the dataset we presented in this work, which is conformed by 3 classes and 11 sub-classes. First, we train the VGG19 CNN network to classify (VGG19 with 3 classes) which we will call CNN-3. In Fig. 4 shows confusion matrix for CNN-3 network.
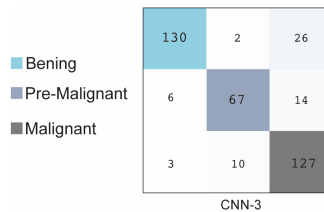
Fig. 4. Confusion Matrix for CNN-3 network for our own dataset

Know only the skin lesion class is not enough for an adequate diagnosis. It is important to know the sub-class (kind of disease) that is being detected. To achieve this, we use the CNN-3 network and use it as a feature extractor Sect. 4.2. Moreover, we performed the training VGG19 CNN network with AE to classify (11 sub-classes), which we will call CNN-AE-11 described in Sect. 4.3.
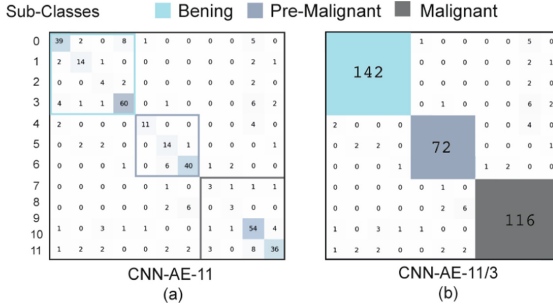
Fig. 5. Confusion Matrix for CNN-AE-11 and CNN-AE-11/3 for our own dataset

In Fig. 5(a), we see the confusion matrix of CNN-AE-11 with a accuracy (0.722) lowest that CNN-3 (0.841). However, if we analyze the results of CNN-AE-11, we can see that some samples were wrongly classified as sub-class, but were correctly classified as class, according to Table 1. Therefore, we group the results of CNN-AE-11 by class hits (benign, pre-malignant and malignant) and we will call CNN-AE-11/3, as we can see in the Fig. 5(b). Now, we can deduce that the accuracy of CNN-AE-11/3 is 85.71%, improving the accuracy of CNN-3 (84.15%). The results obtained in this test are shown in the Table 4.

Table 4. Evaluation for the own dataset, CNN-3, CNN-AE-11, and CNN-AE-11/3.

	Own dataset			
	Accuracy	Precision	Recall	FBetal
CNN-3	0.8415	0.8479	0.8333	0.8425
CNN-AE-11/3	**0.8571**	*0.9141*	*0.9217*	*0.9179*
CNN-11	0.1948	0.0177	0.0909	0.0211
CNN-AE-11	**0.7220**	*0.6424*	*0.6380*	*0.6365*

6   Conclusions

– This work has reached the first place of the ISBI-2016 Contest 3. Moreover, according to the sensitivity metric, our model is better identifying True Positive, equivalent to cases of skin cancer. So, our model is better due to the fact that there is a greater risk for sick people who are classified as healthy.
– Classification using autoencoders is a novel method for the malignant diseases diagnosis. It has shown comparable results as demonstrated in Table 2, even with unbalanced datasets. This feature is important for this kind of research, since there is no availability of large datasets with images, to build datasets.

– Finally, the detection of malignant diseases requires the analysis of all the information that we can obtain from a diagnostic method; even the errors provide information, patterns, and behavior, which resemble the clinical diagnosis that is performed by specialists.

References

1. Fornaciali, M., Carvalho, M., Bittencourt, F.V., de Avila, S.E.F., Valle, E.: Towards automated melanoma screening: Proper computer vision & reliable results. CoRR, abs/1604.04024 (2016)
2. Jafari, M.H., Nasr-Esfahani, E., Karimi, N., Reza Soroushmehr, S.M., Samavi, S., Najarian, K.: Extraction of skin lesions from non-dermoscopic images using deep learning. ArXiv e-prints, September 2016
3. Whited, J.D., Grichnik, J.M.: Does this patient have a mole or a melanoma? JAMA **279**(9), 696–701 (1998)
4. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1397–1400, April 2016
5. Le, Q.V., Google Brain and Google Inc.: A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks (2015)
6. Menegola, A., Fornaciali, M., Pires, R., Bittencourt, F.V., de Avila, S.E.F., Valle, E.: Knowledge transfer for melanoma screening with deep learning. CoRR, abs/1703.07479 (2017)
7. Nachbar, F., Stolz, W., Merkle, T., Cognetta, A.B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., Plewig, G.: The ABCD rule of dermatoscopy. J. Am. Acad. Dermatol. **30**(4), 551–559 (1994)
8. Othman, E., Bazi, Y., Alajlan, N., Alhichri, H., Melgani, F.: Using convolutional features and a sparse autoencoder for land-use scene classification. Int. J. Remote Sens. **37**(10), 2149–2167 (2016)
9. Premaladha, J., Ravichandran, K.S.: Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. J. Med. Syst. **40**(4), 1–12 (2016)
10. Sablayrolles, A., Douze, M., Jégou, H., Usunier, N.: How should we evaluate supervised hashing? CoRR, abs/1609.06753 (2016)
11. Siegel, R., Miller, K., Jemal, A.: Cancer statistics. CA Cancer J. Clin. **66**(1), 7–30 (2016)