Comparative Study of Multiple CNN Models for Classification of 23 Skin Diseases

https://doi.org/10.3991/ijoe.v18i11.32517

Amina Aboulmira[1](✉), Hamid Hrimech[1], Mohamed Lachgar[2]
[1] ENSA, University Hassan First, Berrechid, Morocco
[2] ENSA, University Chouaib Doukkali, El Jadida, Morocco
a.aboulmira@uhp.ac.ma

Abstract—Cutaneous disorders are one of the most common burdens worldwide, that affects 30% to 70% of individuals. Despite its prevalence, skin disease diagnosis is highly difficult due to several influencing visual clues, such as the complexities of skin texture, the location of the lesion, and presence of hair. Over 1500 identified skin disorders, ranging from infectious disorders and benign tumors to severe inflammatory diseases and malignant tumors, that often have a major effect on the quality of life. In this paper, several deep CNN architectures are proposed, exploring the potential of Deep Learning trained on "DermNet" dataset for the diagnosis of 23 type of skin diseases. These architectures are compared in order to choose the most performed one. Our approach shows that DenseNet was the most performed one for the skin disease classification using DermNet Dataset with a Top-1 accuracy of 68.97% and Top-5 accuracy of 89.05%.

Keywords—skin lesion, classification, DermNet, deep learning, convolutional neural networks

1 Introduction

Deep learning has quickly become one of the most important machine learning and artificial intelligence technologies. The notion of machine learning extends back to the mid-twentieth century. Alan Turing, a British mathematician, envisaged a computer capable of learning, a "learning machine" in the 1950s [1]. Over the following decades, various machine learning techniques were developed to create algorithms that could learn and improve independently. Among these techniques, there are artificial neural networks (ANN). Their concept is inspired by the functioning of biological neurons in the human brain [2]. They consist of an interconnection of several artificial neurons connected to each other. The more neurons we have, the "deeper" the network. Deep learning's applicability for diseases or cancers classification is a challenge, and a better choice for good results, that many researchers are now using. For example, in [3] researchers created a multi-level system based on an artificial neural network for detecting eczema skin lesions. Skin diseases are one of the most seen infections among people, there are characterized as changes or disorders of the skin that

testify to a problem related to the skin, which can manifest itself in different forms (pimples, skin spots, redness, excessive sweating, growths, fungi, infections. of the skin, discoloration of the skin) [4].To diagnose a skin disease, a series of pathological laboratory tests are performed to be able to identify the correct disease. For example, it exists four major clinical diagnosis methods to study skin cancer, melanoma: ABCD rules [5], 7-point checklist [6], Menzies method[7] and the C.A.S.H algorithm [8]. To have an accurate diagnosis with these methods, a degree of expertise from the dermatologist is necessary, especially for certain diseases that can often be poorly diagnosed [9]. Contrary to a diagnosis made by human experts, which depends a lot on subjective judgement and difficulty to reproduce, a computer-assisted diagnosis system is more objective and reliable, using one of the classification algorithms to extract the important features and get successful results.

In recent years, Deep Convolutional Neural Networks (CNN) has become very popular for feature learning and object classification [10]. The use of high-performance GPUs allows networking of a large-scale dataset for better performance. With CNN, many researchers around the world are taking up the challenge of developing different CNN architecture to improve CNN performance [11]. These networks have become so deep that the whole model become so extremely difficult to visualize. Most of these models have won competitions like the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). In this work, a skin lesion classification method based on convolutional neural networks (CNNs) is developed with different CNN architectures (GoogleNe t[12], InceptionV4 [13], InceptionV3[14], NASNet-Large [15], MobileNetV3 [16], InceptionResNetV2 [13], VGG19 [17], RestNet50[18], ResNext50 [19], DenseNet201[20]). The models are compared and the best of them is choose, to generate a model of prediction for skin disease data. For the dataset, skin disease images are obtained from DermNet [21], a publicly available dataset of resource containing over 23000 skin disease images, with 23 different types of skin diseases Each top-level skin disease class comprises a subset of the bottom-level skin diseases. In general, this study will design a classification model for skin diseases based on a comparison of different CNN pretrained models in aims to choose the CNN model that fits best with DermNet dataset.

The main contribution of this work is that we performed a comparative study of the CNN models for the classification of skin diseases including all the classes in the top-level taxonomy of the DermNet dataset, which encompasses a large part of the diseases skin. This comparison aims to help other researchers to choose the efficient CNN model for building a practical computer aided system for skin disease classification.

The remainder of the article is organized as follows: Section 2 summarizes some of the relevant work based on computer-aided systems for the diagnosis of skin lesions. Section 3 exposes the methodology followed to achieve the purpose of classification skin lesion with CNN, a brief description of the public skin disease datasets, and the conception of deep learning and popular architectures. In section 4, we present the preprocessing stage and the experimental techniques for the training of the models. Next in section 5, we present results obtained from using the proposed approach and assess the performance of the different CNN architectures, compared to each other.

2 Related work

Many studies have tried to exploit the performance of Deep Learning in favor of dermatology, especially for the diagnosis of skin diseases. However, few studies that were interest to a universal classification of skin disease, most of them restrict the problem to certain skin diseases, such as melanoma, which is a deadly cancer [22]. In this section, we review the various existing work in the literature for the purpose of detecting and classifying skin lesion by exploiting the different techniques of Deep Learning, by focusing more on the research that has been developed with the DermNet database. Works on skin lesion classification with deep learning methods are listed in Table 1.

Table 1. References of skin disease classification with deep learning

Ref	Authors	Year	End Point	Dataset	Result	Model
[23]	Bajwa et al.	2020	Classification of 23 diseases	DermNet database	Accuracy: 80% AUC: 98%	Fine-tuned DenseNet-161, SE-ResNeXt-101 and NASNet
[24]	Sah et al.	2019	Classification of 10 different classes of skin disease	DermNet database	Accuracy: 76.3%	Fine-tuned of pretrained VGG16 model
[25]	Esteva et al.	2017	Robust system of classification skin cancers	Clinical images dataset	Accuracy: 55.4 ± 1. %	Pretrained Inception v3
[26]	Haofu Liao et al.	2016	Classification of 38 disease-targeted and lesion-targeted	AtlasDerm Derma DermIS Dermnet DermQuest	Top-1 accuracy: 27.6% Top-5 accuracy: 57.9%	AlexNet model trained from scratch and using fine-tuning
[27]	Kawahara et al.	2016	Classification of skin lesions into ten categories	Dermofit	Accuracy: 81.8%	Pretrained AlexNet
[28]	Haofu Liao	2015	Classification of 23 diseases	Dermnet OLE	Top-1 accuracy: 73.1% Top-5 accuracy: 91.0%	Deep convolutional neural network (CNN) using VGG19

Looking at previous works, the research by Bajwa et al [23] proved that DL have enormous potential to classify a vast array of skin diseases challenging the human performance, they used two techniques in their researcher, they first focus on classification of 23 skin classes lesions and achieved 80% accuracy and 98% AUC. In second phase, they achieved 67% accuracy and 98% AUC in classification of 622 distinct sub-classes in DermNet dataset. Sah et al. in [24], highlighted the role of image processing part and image augmentation to improve the accuracy of skin lesion classification. They investigated the ability of deep CNN models trained with DermNet dataset to achieve a good recognition rate. Esteva et al. in [25], proved that deep neural networks (DNN) rival human performance, by building a strong system that attracted a big attention, they developed a NN-based system that can diagnose cancer

from skin image, reaching the level of human dermatologists. They achieved 60.0% top-1 accuracy and 80.3% top-3 accuracy classification, outperforming human specialists' performance. Haofu Liao proposed two works. In the first, he suggested a global skin disease diagnostic system using deep convolutional neural network (CNN) [26], achieving 73.1% Top-1 accuracy and 91.0% Top-5 accuracy. In 2016, he proposed a second work with other researchers [27], when they proved that using skin lesion characteristics facilitate skin disease diagnosis as many diseases are so similar in the visual aspects. Kawahara et al in [28] used the MobileNet network trained on a public library DermoFit and classified the skin lesions into ten categories.

3 Methodology

Many steps have been taken in aims to develop a computer vision-based system for skin disease classification. Each image in the DermNet dataset goes through the preprocessing step where it sizes changes according to the input of each model, and goes through different transformation phases (rotation, flip, zoom, etc.) to increase the number of images in the dataset. Several models are proposed afterwards and are trained on all the images of the training dataset, to finally be able to evaluate them on the test dataset in order to choose the most efficient model. The overview of the different steps is shown in Figure 1.

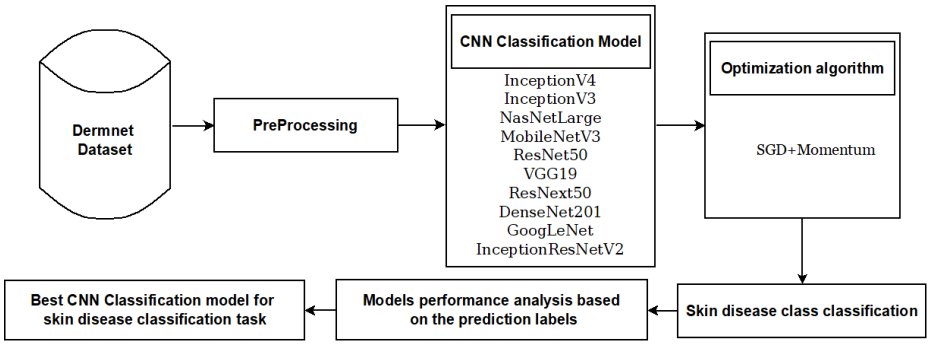

Fig. 1. Overview of the proposed skin diseases classification method

3.1 Dataset

DermNet dataset includes a collection of more than 23,000 dermoscopic images belonging to different skin diseases. The images are organized according to a two-level taxonomy. Figure 2 illustrates sample lesion images from the dataset, it has 23 super-classes of skin diseases. Each image has a diagnostic tag provided by an expert in Dermatology Resources.

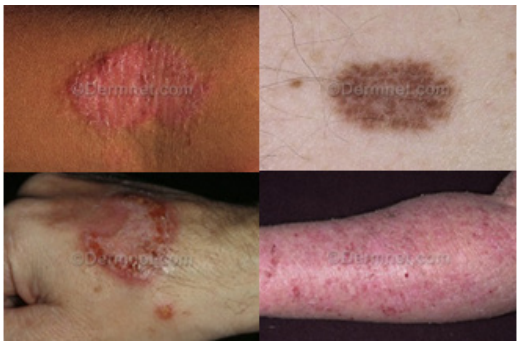

Fig. 2. Random samples of skin lesions from DermNet dataset

DermNet dataset contains different images of skin diseases, in JPEG extension, several images with lower quality were deleted in the checking of the dataset. After quality control step, the dataset was divided into 80% of the samples i.e., 12368 DermNet images are used for training and 15% of the samples i.e., 3085 images are used for the validation of the network, and 20% of the samples i.e., 4002 images are used for the testing phase. Figure 3 show the distribution of images in each of the 23 classes for training, validation, and testing data.

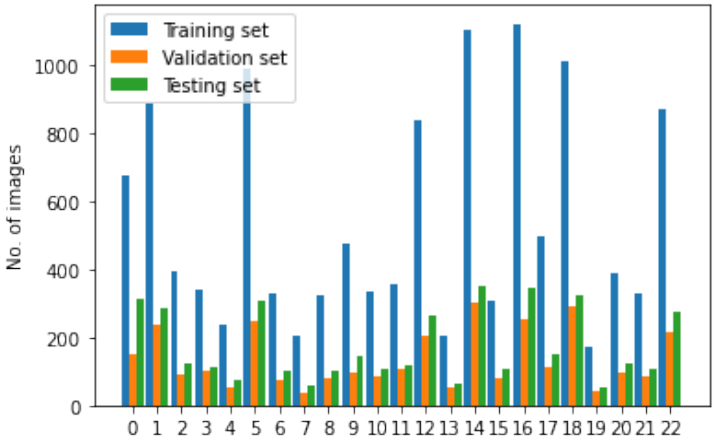

Fig. 3. Distribution of the available DermNet training and validation images

Table 2 shows the number of DermNet images in each of the 23 classes, each category name represents a class of skin disease, and each label is associated with its category.

Table 2. Labels of the 23 skin diseases class

Label	Class Name
0	Acne-and-Rosacea
1	Actinic-Keratosis-Basal-Cell-Carcinoma
2	Atopic-Dermatitis
3	Bullous-Disease
4	Cellulitis-Impetigo-and-other-Bacterial-Infections
5	Eczema
6	Exanthems-and-Drug-Eruptions
7	Hair-Loss-Photos-Alopecia-and-other-Hair-Diseases
8	Herpes-HPV-and-other-STDs-Photos
9	Light-Diseases-and-Disorders-of-Pigmentation
10	Lupus-and-other-Connective-Tissue-diseases
11	Psoriasis-pictures-Lichen-Planus-and-related-diseases
12	Nail-Fungus-and-other-Nail-Disease
13	Poison-Ivy-Photos-and-other-Contact-Dermatitis
14	Psoriasis-pictures-Lichen-Planus-and-related-diseases
15	Scabies-Lyme-Disease-and-other-Infestations-and-Bites
16	Seborrheic-Keratoses-and-other-Benign-Tumors
17	Systemic-Disease
18	Tinea-Ringworm-Candidiasis-and-other-Fungal-Infections
19	Urticaria-Hives
20	Vascular-Tumors
21	Vasculitis
22	Warts-Molluscum-and-other-Viral-Infections

3.2 Experimental setup

Keras library was used to implement the networks. As previously indicated, the pretrained models were trained to perform the 23-classification task using transfer learning, and then they were fine-tuned using a categorical cross-entropy cost function and stochastic gradient descent (SGD) with a small learning rate of 1e-4 and momentum value set to 0.9. Indeed, because of its efficiency, SGD has become one of the most used optimization methods. Combined with momentum, it usually converges faster, the momentum helps to follow prevalent descent directions and dampens oscillation caused by the variance, which accelerates gradient vectors in the appropriate directions and leads to faster convergence [29].

4 Experiments

Deep learning is a field of machine learning. It is closer to the human brain; it goes further than the connection between data and algorithms since it allows the machine to learn and progress through its experience. Deep Learning, which uses a combina-

tion of knowledge in neuroscience, mathematics and technological advances, is today hailed as a real revolution in the field of artificial intelligence. It has recently been shown that deep learning algorithms, based on advances in computing and very large datasets, can be well served in the medical field, especially with the development of smartphones where the patient can nowadays have a primary diagnosis of his health condition as the case of skin diseases [30].

4.1 Stratified cross validation

As we can see from Figure 4, the DermNet dataset is highly imbalanced. This may lead to a misclassification of the minority class relative to the dominant class. To deal with the negative effect of imbalanced data, we have proposed the use of stratified cross validation. The Stratified k-Fold CV technique is a useful technique in situations where we have an unbalanced set. This is a variation of the classic k-Fold CV approach, which is based on dividing the dataset into k parts, each with approximately the same percentage of samples from each class as the full set. The average of the values obtained in each division is the performance metric given by the k-fold cross validation. This technique is computationally expensive, but it does not waste a lot of data (unlike defining an arbitrary validation set) and has huge advantages in problems such as inverse inference or in situations where we have a dataset with small number of samples [31].

4.2 Data pre-processing

In the previous part of this paper, dataset was divided in 3 parts (training, testing, validation). A check control is carried out before that, to delete images with lower quality. The data entry for all the CNN architectures is prepared, a resizing images step is required: pre-processing of (224,224,3)-pixel-sized input images for DenseNet201, VGG19, GoogLeNet, ResNext50, ResNet50, InceptionResNetV2 and InceptionV4, and pre-processing of (299,299,3)-pixel-sized input images for InceptionV3, and NasNetLarge architecture. Labels in Table 2 are used for generating the training and Test labels. The original pretrained model must be modified to suit our needs, the final fully connected output layer must perform a multiple classification of 23 classes not 1000 classes. Typically, images are standardized by subtracting the averaged pixel over the training set to center pixel values around zero.

4.3 Data augmentation

In general, deep learning models perform better when they have access to more data, since they will have more information to extract which gives them the possibility of learning more. Certainly, in some cases we cannot access to a large dataset, one of the alternatives is to do some transformations on the images of the dataset (like rotation, flip, change brightness...) in way to have more images from the same base set. Table 3 shows the different transformations made to the input images.

Table 3. Values used in data augmentation

Process	Value
Rescale	1./255
Rotation_range	45
Width_shift_range	0.15
Height_shift_range	0.15
Horizontal_flip	True
Zoom_range	0.2

To achieve a good performance, image augmentation was performed using ImageDataGenerator API in Keras. Image Augmentation is used to artificially generate training images from the dataset images, performing various transformation methods.

4.4 Transfer learning

Due to limited data and the necessity of high computation power, transfer learning can be used to train a deep CNN efficiently, so that, besides using the same architecture of the pretrained model, we can let the model learns new tasks based on parameters learned by the previous training on ImageNet Dataset, instead of training the network from randomly initialized parameters.

4.5 Fine tuning

Fine tuning consists of adjusting a pre-trained model by refining its weights when training on a new set, to make it more adaptable and efficient on this set. This technique is applied to the pre-trained models to make them more relevant for the skin disease classification task. The last layers of the pretrained networks are configured for 1000 classes. We removed the last layer and replaced it with a new classifier. The classifier's role is to classify input images based on the activations it gets from the feature extraction phase of every CNN, Figure 4 shows the fine-tuning process.

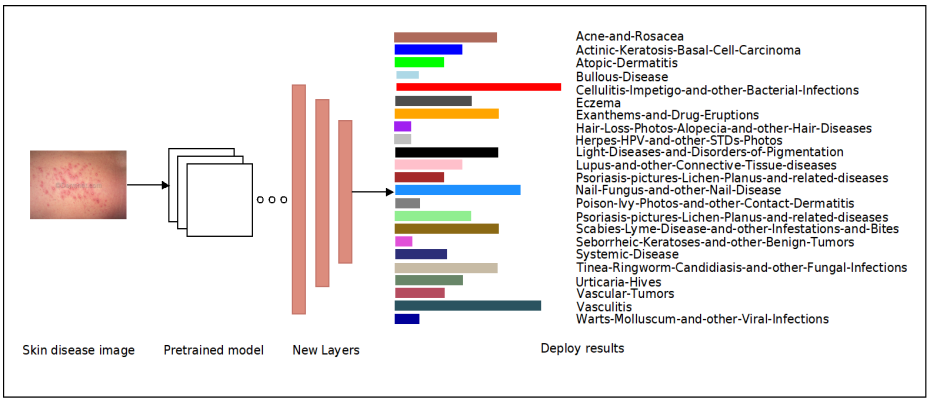

Fig. 4. Example of the finetuning of a pretrained model

We first add a global average pooling 2D layer, then a fully connected layer and a dropout layer were included. Adding dropout layers can improve neural networks by reducing overfitting by using a value between 20%–50% [32]. Finally, a SoftMax layer is set to produce the probability of each of the 23 output classes. The final decision of the disease class is picked from the class with the highest probability. Figure 4 shows an example of the fine-tuning process where the final layers was replaced by our classification task.

4.6 Performance measures

Accuracy and error rate measures are very important to evaluate a model, however they may be deceptive in some situation as these are data dependent. In case of imbalanced dataset where the number of majority samples is too large compared to the minority samples, the classifier gets biased towards the majority samples. We can see in such cases that we have a high accuracy value and very low error rate. These results project as if the classifier is an ideal one which is not the real scenario. Indeed, even if these two metrics indicate that the classifier is efficient, the minority classes are not well classified. There are several alternatives to overcome such imbalanced dataset scenarios, we can rely on other metrics which may prove more useful and give more accurate insight into the performance of the classifier. These metrics are precision, recall and AUC. These measures are defined as follows:

Accuracy (ACC) quantifies the proportion of the labels that were correctly classified divided by all predictions that were made on the test set, which is formally expressed as Equation 1:

$$Accuracy = \frac{(TN+TP)}{(TN+FP+FN+TP)} \tag{1}$$

Precision also known as or positive predictive value quantifies the proportion of the labels properly classified that are truly positive, as represented formally in Equation 2:

$$Precision = \frac{TP}{(FP+TP)} \tag{2}$$

Recall quantifies the proportion of misclassified labels that are truly positive, as represented formally in Equation 3:

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{3}$$

5 Results and discussion

The experiments were conducted on a Windows10 machine with the following hardware configuration: Intel Core™ i7-10750H CPU @ 2.6GHz processor with Install memory (RAM): 16.0GB and GeForce RTX 2060 GPU with 6GB GDDR5 memory.

5.1 Result and analysis

Depending on the depth and number of parameters of each model, the learning time and the number of epochs required to converge may differ from one model to another. For this, we used the Early stopping class provided by Keras, which is a technique that will stop the training if the validation accuracy value does not change for a predefined number of epochs, which avoid which prevents the model from starting the overfitting. Early stopping can help optimize the epoch's size by setting a high number of epochs and letting early stopping automatically stop the training once the model no longer performs on the validation set. Figure 5 and 6, shows the accuracy and loss for training and validation set.

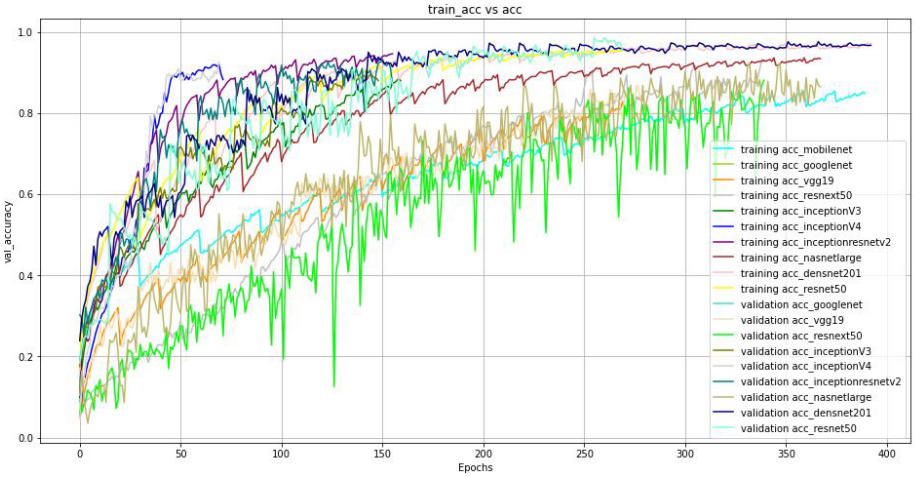

Fig. 5. Training and validation accuracy for the different models

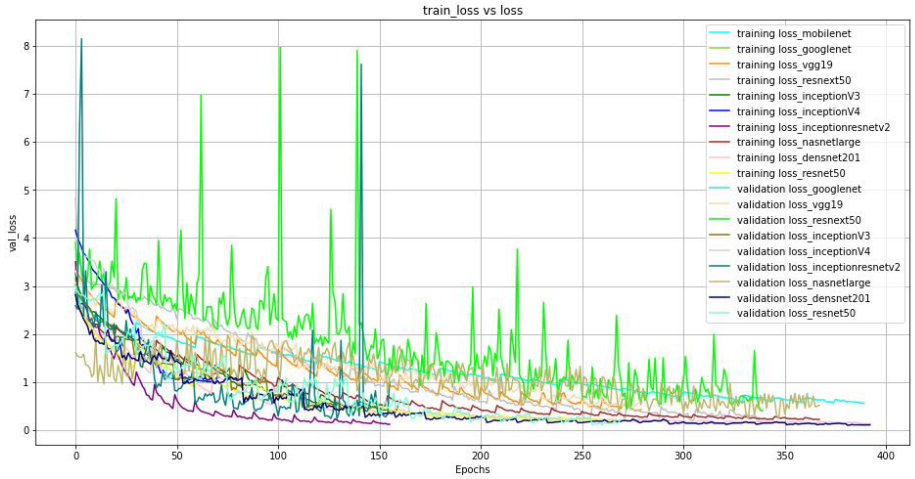

Fig. 6. Training and validation loss for the different models

While many of the models present similar results for the given dataset, we can notice that some of them stand out. DenseNet-201 outperformed all other models, it achieved significantly higher performance with a small difference compared with the InceptionResNetV2. The results in Table 4 summarizes the classification report on testing dataset for the different networks.

Table 4. Results of implementation CNNs

Name of Architecture	Total params	Top-1 accuracy	Top-5 accuracy	Precision	Recall
InceptionV4	41,979,287	63.47%	88.25%	64.2%	63.5%
InceptionV3	22,068,023	62.94%	87.45%	50.1 %	44.8%
DenseNet201	18,570,839	68.97%	89.05%	67.3%	67.2%
MobileNetV3	3,363,031	48.67%	83.18%	51.2%	48.7%
ResNet50	23,852,951	65.84%	88.38%	66.2%	65.8%
VGG19	20,090,564	59.82%	87.90%	58.2%	56.3%
ResNext50	26,679,063	53.72%	82.33%	54.2%	53.7%
NASNetLarge	85,436,009	65.94%	87.88%	66.2%	65.8%
GoogLeNet	6,147,879	64.01%	89.23%	64.5%	64.0%
InceptionResnetV2	54,378,231	67.99 %	89.45%	68.6%	68.0%

(*) We used the weighted average for the precision and recall rate

Densenet performed better than all architectures with mean differences in classification accuracies of 5.5%, 6.03%, 20.3%, 3.13%, 9.15%, 15.25%, 3.03%, 4.96 and 0.98% when compared to InceptionV4, InceptionV3, MobilenetV3, ResNet50, VGG19, ResNext50, NASNet, GoogleNet and InceptionResNet-v2, respectively. The reason may be due to the depth of DenseNet-201 as it has the deepest neural structure compared to other deep neural networks in use, which allows it to extract more distinct features and map more complex patterns, which helps to differentiate the classes and have a more precise diagnosis. However, the accuracy metric is not always the best metric to evaluate a model; precision and recall can give us more information. Ideally, for our model, we would like to avoid all situations where the patient has a disease, but the model misdiagnoses him, which means we need to focus more on having a high recall.

ResNet50 which introduced the concept of the repetitive residual blocks, NASNet-large which use a reinforcement learning search method to optimize the architecture configurations, GoogleNet, InceptionV3 and InceptionV4 who uses inception modules that are composed of multiple filters of different sizes over simple convolution layers, all these models showed a similar performance. Starting with the accuracy, NASNet-large had the higher result 65.94% followed closely by ResNet50 with 65.84%, GoogleNet with 64.01%, InceptionV4 by 63.47% and the lower accuracy with InceptionV3 which obtained 63.47%. In the same way, the order of the recall rate was equal to the accuracy, where the lower percentage was obtained by InceptionV3 with 44.8% and the highest by NASNet and ResNet50 with 65.8%. VGG19 and ResNext50 had insignificant accuracy rate and MobileNetV3 was the lowest.

In one hand, DenseNet201 and InceptionResNetV2 achieved both the most acceptable accuracy results compared to other models. InceptionResNetV2, on the other hand,

while having a greater number of learnable parameters than DenseNet-201, was unable to outperform DenseNet201 in terms of accuracy. This, makes the InceptionResNetV2 train slower with more computation cost, as compared to DenseNet-201. Hence, after this evaluation, only architecture that have higher top-1 accuracy score and don't have a high total of parameters is kept: DenseNet201.

For more evaluation of the results and more details, we can refer either in the Area Under the Receiver Operating Characteristic Curve (AUC). This measure ensures that the classifier is evaluated at many different thresholds, i.e., the distance cutting points above which a sample is considered anomalous. Figure 7 shows the ROC curve and Area under these ROC curves for the 4 best classes predicted correctly, that is plotted with true positive rate against the false positive rate.

Depending on each issue, the amount of data and the complexity of the images, CNNs models perform differently in each situation. In general, a large amount of data is required for the model to learn well and extract the characteristics of classes and to be able to differentiate between them. The dataset we have is not very large, especially for some classes, which prompted us to use transfer learning to take advantage of ImageNet's weights. However, the specificity of each dataset (target dataset and source dataset), as well as the size of the dataset led us to simply freeze the first CNN levels of the pre-trained models and recycle the remaining parameters / weights. It is intuitive that the more the number of frozen layers is reduced, the more the computational cost of the formation is considerable [33].

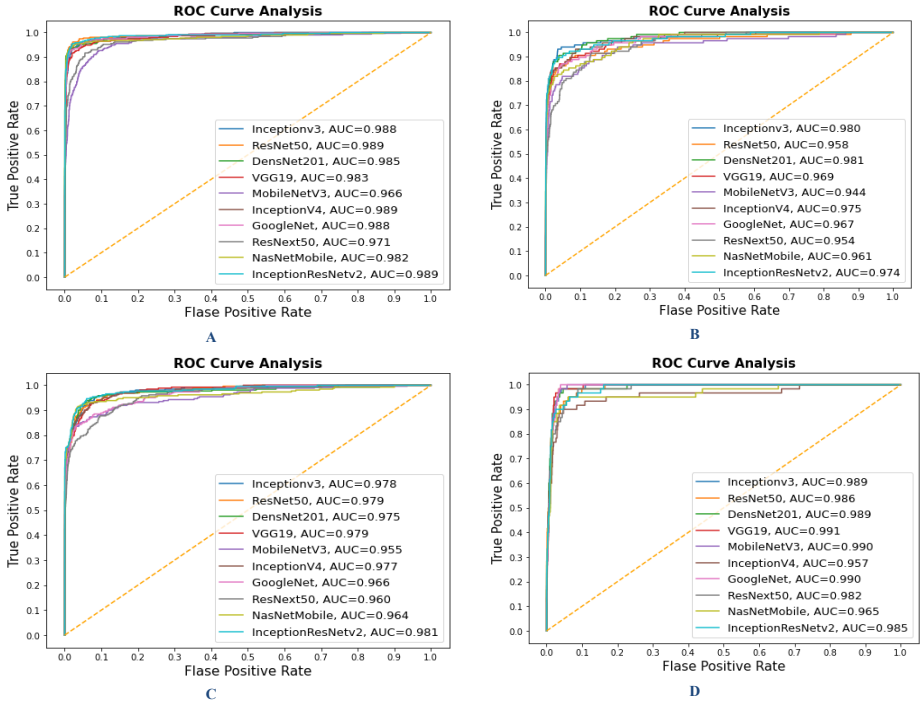

Fig. 7. ROC curve of the first class of skin lesions for the four classes predicted correctly (A: Acne and Rosacea, B: Psoriasis-pictures-Lichen-Planus-and-related-diseases, C: Nail Fungus and other Nail Disease, D: Hair-Loss-Photos-Alopecia-and-other-Hair-Diseases)

6 Conclusion

In this work, high-performance CNN architectures such as InceptionV4, InceptionV3, DenseNet-201, MobilenetV3, ResNet50, VGG19, ResNext50, NASNetLarge, GoogleNet and InceptionResNetV2 were trained using DermNet dataset which contain 19434 images of skin diseases belonging to 23 classes.

These architectures are used to classify skin diseases, in order to develop a comparison between them, which helps to select the best model to build a practical computer-aided diagnosis system that can be used in the dermatology field. DenseNet-201 was the best deep learning neural network (CNN architecture) for automatic diagnosis of skin diseases using the DermNet dataset. This study helped to select the best CNN model for the automatic diagnosis of skin diseases using DermNet Dataset. The main purpose of this work is to provide comparative study of CNN behavior toward the dermnet dataset, which has been little treated previously with this large number of classes including most common skin illnesses. Thus, this can be used as a basis for future research to develop a computer-assisted diagnostic system for the most common dermatological conditions.

While our approach learns features end to end, another promising way might be fusing CNN models to build a robust deep learning architecture, another approach is to use handcrafted features to our advantage, as diagnostic markers typically tracked by dermatologists.

In a vision, and for further development of this study, we aim to use segmentation to the skin region with the classification algorithm (K-means), which is a Machine Learning algorithm, to minimize prediction errors. Another approach is to collect more clinical data to improve the model's performances and test other type of dataset, for a bit of knowledge, is what it to give a better prediction score for all types of datasets.

7 References

[1] S. Muggleton, *Alan Turing and the development of Artificial Intelligence*.

[2] E. Grossi and M. Buscema, "Introduction to artificial neural networks," *Eur. J. Gastroenterol. Hepatol.*, vol. 19, pp. 1046–54, Jan. 2008. https://doi.org/10.1097/MEG.0b013e3282f198a0

[3] L. Guzman, R. Maglaque, V. Torres, S. Zapido, and M. I. Cordel, Design and Evaluation of a Multi-model, Multi-level Artificial Neural Network for Eczema Skin Lesion Detection. 2015. https://doi.org/10.1109/AIMS.2015.17

[4] "Skin Disorders: Pictures, Causes, Symptoms, Treatments, and Prevention." https://www.healthline.com/health/skin-disorders (accessed Feb. 23, 2021).

[5] F. Nachbar *et al.*, "The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions," *J. Am. Acad. Dermatol.*, vol. 30, no. 4, pp. 551–559, Apr. 1994. https://doi.org/10.1016/S0190-9622(94)70061-3

[6] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Arch. Dermatol.*, vol. 134, no. 12, pp. 1563–1570, Dec. 1998. https://doi.org/10.1001/archderm.134.12.1563

[7] S. W. Menzies, C. Ingvar, K. A. Crotty, and W. H. McCarthy, "Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features," *Arch. Dermatol.*, vol. 132, no. 10, pp. 1178–1182, Oct. 1996.

[8] J. S. Henning *et al.*, "The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy," *J. Am. Acad. Dermatol.*, vol. 56, no. 1, pp. 45–52, Jan. 2007. https://doi.org/10.1016/j.jaad.2006.09.003

[9] A. Sit, "Patient Buzz: Misdiagnosed Skin Conditions," *Next Steps in Dermatology*, Dec. 23, 2020. https://nextstepsinderm.com/derm-topics/patient-buzz/patient-buzz-misdiagnosed-skin-conditions/ (accessed Aug. 14, 2021).

[10] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, *Deep convolutional neural networks for LVCSR," ICASSP*. 2013.

[11] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020. https://doi.org/10.1007/s10462-020-09825-6

[12] C. Szegedy *et al.*, "Going Deeper with Convolutions," *ArXiv14094842 Cs*, Sep. 2014, Accessed: Feb. 23, 2021. [Online]. Available: http://arxiv.org/abs/1409.4842

[13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *ArXiv160207261 Cs*, Aug. 2016, Accessed: Jun. 29, 2021. [Online]. Available: http://arxiv.org/abs/1602.07261

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *ArXiv151200567 Cs*, Dec. 2015, Accessed: Feb. 23, 2021. [Online]. Available: http://arxiv.org/abs/1512.00567

[15] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," *ArXiv170707012 Cs Stat*, Apr. 2018, Accessed: Jul. 26, 2021. [Online]. Available: http://arxiv.org/abs/1707.07012

[16] A. Howard *et al.*, "Searching for MobileNetV3," *ArXiv190502244 Cs*, Nov. 2019, Accessed: Feb. 23, 2021. [Online]. Available: http://arxiv.org/abs/1905.02244

[17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ArXiv14091556 Cs*, Apr. 2015, Accessed: Feb. 23, 2021. [Online]. Available: http://arxiv.org/abs/1409.1556

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

[19] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *ArXiv161105431 Cs*, Apr. 2017, Accessed: Aug. 04, 2021. [Online]. Available: http://arxiv.org/abs/1611.05431

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243

[21] Dermnet, "Homepage," *Dermatology Education*. http://www.dermnet.com/ (accessed Feb. 23, 2021).

[22] A. Naeem, S. Farooq, A. Khelifi, and A. Abid, "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," *IEEE Access*, vol. PP, pp. 1–1, Jun. 2020. https://doi.org/10.1109/ACCESS.2020.3001507

[23] M. N. Bajwa *et al.*, "Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks," *Appl. Sci.*, vol. 10, p. 2488, Apr. 2020. https://doi.org/10.3390/app10072488

[24] A. Sah, S. Bhusal, S. Amatya, M. Mainali, and S. Shakya, *Dermatological Diseases Classification using Image Processing and Deep Neural Network*. 2019, p. 386. https://doi.org/10.1109/ICCCIS48478.2019.8974487

[25] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, Art. no. 7639, Feb. 2017. https://doi.org/10.1038/nature21056

[26] Haofu Liao, Yuncheng Li, and Jiebo Luo, "Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Dec. 2016, pp. 355–360. https://doi.org/10.1109/ICPR.2016.7899659

[27] J. Kawahara, A. Bentaieb, and G. Hamarneh, *Deep features to classify skin lesions*. 2016, p. 1400. https://doi.org/10.1109/ISBI.2016.7493528

[28] H. Liao, "A Deep Learning Approach to Universal Skin Disease Classification," 2015. /paper/A-Deep-Learning-Approach-to-Universal-Skin-Disease-Liao/af34fc0aebff011b56ede8f46ca0787cfb1324ac (accessed Jan. 13, 2021).

[29] S. Ruder, "An overview of gradient descent optimization algorithms," *ArXiv160904747 Cs*, Jun. 2017, Accessed: Jul. 25, 2021. [Online]. Available: http://arxiv.org/abs/1609.04747

[30] L. Rosado, M. J. Vasconcelos, and M. Ferreira, "A Mobile-Based Prototype for Skin Lesion Analysis: Towards a Patient-Oriented Design Approach," *Int. J. Online Biomed. Eng. IJOE*, vol. 9, pp. 27–29, Dec. 2013. https://doi.org/10.1109/ISBI.2016.7493528

[31] S. N. M. Safuan, M. R. M. Tomari, and W. N. W. Zakaria, "Cross Validation Analysis of Convolutional Neural Network Variants with Various White Blood Cells Datasets for the Classification Task," *Int. J. Online Biomed. Eng. IJOE*, vol. 18, no. 02, Art. no. 02, Feb. 2022. https://doi.org/10.3991/ijoe.v18i02.27321

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.

[33] J. Kunze, L. Kirsch, I. Kurenkov, A. Krug, J. Johannsmeier, and S. Stober, "Transfer Learning for Speech Recognition on a Budget," in *Proceedings of the 2nd Workshop on Representation Learning for NLP*, Vancouver, Canada, Aug. 2017, pp. 168–177. https://doi.org/10.18653/v1/W17-2620

8 Authors

Amina Aboulmira has received a master's degree in computer science from Hassan Premier University of Science and Technology in 2021. Currently she is pursuing her Ph.D. program at 'Systems Analysis and Modeling and Decision Support' laboratory in National School of Applied Sciences. Her research interest includes computer vision, image processing, deep learning, and machine learning.

Hamid Hrimech is an assistant Professor in the Computer Science and Mathematics Department of the National School of Applied Sciences.at Hassan 1st University. He has a Ph.D. from Arts et Métiers ParisTech (France) in computer science. His research interests include machine learning and driving simulation.

Mohammed Lachgar has received his Ph.D. in Computer Science at the Cadi Ayyad University in 2017, He is a professor in Computer Science at the National School of Applied Sciences, Chouaib Doukkali University, El Jadida, Morocco. His research interests are in the areas of automation tools development in embedded software, software modeling & design, metamodel design, model transformation, model verification & validation method, machine learning and deep learning.

Article submitted 2022-05-17. Resubmitted 2022-06-14. Final acceptance 2022-06-14. Final version published as submitted by the authors.