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

Aymane Elouadi, Oussama Nafi, Mohammed Ait Sidi Ahmed, Chaimaa Ouknider, Hamid Hrimech[1]
[1] ENSA, University Hassan First, Berrechid, Morocco

Abstract—Skin cancer is a significant global health concern, as cutaneous disorders affect a substantial percentage of individuals, ranging from 30% to 70% worldwide. However, diagnosing skin diseases poses considerable challenges due to various visual factors that influence the process, such as the intricate nature of skin texture, lesion localization, and the presence of hair. With more than 1500 identified skin disorders, encompassing infectious conditions, benign tumors, severe inflammatory diseases, and malignant tumors, the impact on individuals' quality of life is often profound.

This research paper proposes the utilization of several deep convolutional neural network (CNN) architectures to explore the potential of Deep Learning techniques trained on the extensive "DermNet" dataset for the accurate diagnosis of 23 different types of skin diseases. The performance of these architectures is thoroughly compared to identify the most effective approach. The findings reveal that DenseNet demonstrates superior performance in skin disease classification using the DermNet Dataset, achieving a remarkable Top-1 accuracy of 68.97% and a Top-5 accuracy of 89.05%. These results highlight the promising applications of Deep Learning in enhancing skin disease diagnosis and potentially improving patient outcomes in the future.

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

1 Introduction

Deep learning has emerged as a pivotal technology in the fields of machine learning and artificial intelligence, offering significant advancements. The concept of machine learning traces back to the mid-twentieth century when Alan Turing envisioned a "learning machine" capable of independent improvement in the 1950s. Over time, various machine learning techniques, including artificial neural networks (ANN), were developed, drawing inspiration from the functioning of biological neurons in the human brain. ANNs consist of interconnected artificial neurons, forming deeper networks with an increasing number of neurons.

The application of deep learning in the classification of diseases, particularly cancers, has become a significant challenge that researchers are actively pursuing. For instance, in a notable study, researchers developed a multi-level system based on an artificial neural network to detect eczema skin lesions. Skin diseases, encompassing various manifestations such as pimples, skin spots, redness, excessive sweating, growths, fungi, and infections, are prevalent among people.

Diagnosing skin diseases typically involves conducting a series of pathological laboratory tests to identify the specific condition. For skin cancer, melanoma, several

clinical diagnosis methods exist, including the ABCD rules, 7-point checklist, Menzies method, and the C.A.S.H algorithm. However, accurate diagnosis often requires expertise from dermatologists, and certain diseases pose challenges for precise identification.

In contrast to diagnoses reliant on subjective judgment and reproducibility challenges faced by human experts, computer-assisted diagnosis systems offer objectivity and reliability. These systems employ classification algorithms to extract essential features and achieve successful results.

Deep Convolutional Neural Networks (CNN) have gained popularity in recent years for feature learning and object classification. The use of high-performance GPUs facilitates large-scale dataset networking, improving CNN performance. Researchers worldwide have risen to the challenge of developing different CNN architectures to enhance performance. These networks have become increasingly deep, making visualizing the entire model extremely difficult. Many models have achieved success in competitions such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

This work focuses on developing a skin lesion classification method based on CNNs, utilizing various CNN architectures such as GoogleNet, InceptionV4, InceptionV3, NASNet-Large, MobileNetV3, InceptionResNetV2, VGG19, RestNet50, ResNext50, and DenseNet201. These architectures are compared, and the best-performing model is selected to generate a prediction model for skin disease data. The dataset used comprises skin disease images obtained from DermNet, a publicly available resource containing over 23,000 skin disease images representing 23 different types of skin diseases. Each top-level skin disease class includes a subset of bottom-level skin diseases.

The primary contribution of this study lies in conducting a comparative analysis of CNN models for the classification of skin diseases, covering a significant portion of the DermNet dataset's top-level taxonomy. This comparison aims to assist other researchers in selecting efficient CNN models for developing practical computer-aided systems for skin disease classification.

The remainder of this article is organized as follows: Section 2 provides an overview of relevant work related to computer-aided systems for skin lesion diagnosis. Section 3 outlines the methodology employed, including a brief description of public skin disease datasets, an introduction to deep learning, and popular architectures. Section 4 presents the preprocessing stage and experimental techniques utilized for model training. Furthermore, in section 5, the results obtained using the proposed approach are presented, and the performance of different CNN architectures is assessed through comparison.

2 Related work

Multiple researches aimed studying and using the performance of Deep Learning for dermatology and biological purpose, particularly for the accurate diagnosis of skin diseases. However, only few of them have focused on achieving a generalized classification of skin diseases, as most of them have restricted their studies to conditions like melanoma, a highly dangerous form of cancer [22]. In this section, we conduct a thorough examination of existing literature that explores the detection and classification of skin lesions using various Deep Learning techniques, with a specific emphasis on studies utilizing the DermNet database. We present a compilation of works on skin lesion classification employing Deep Learning methods 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

Based on previous research, Bajwa et al.'s study [23] has shown that Deep Learning (DL) has immense potential for classifying various skin diseases, surpassing human performance. They employed two techniques in their research, initially focusing on classifying 23 different types of skin lesions, achieving 80% accuracy and 98% AUC. In the second phase, they achieved 67% accuracy and 98% AUC when classifying 622 distinct sub-classes in the DermNet dataset.

In another study, Sah et al. [24] emphasized the role of image processing and image augmentation in improving the accuracy of skin lesion classification. They investigated the effectiveness of deep Convolutional Neural Network (CNN) models trained with the DermNet dataset, resulting in a favorable recognition rate.

Esteva et al. [25] demonstrated that deep neural networks (DNN) can match human performance in diagnosing skin cancer. They developed an attention-grabbing NN-based system that outperformed human dermatologists, achieving a top-1 accuracy of 60.0% and a top-3 accuracy of 80.3% in classification.

Haofu Liao proposed two studies in this field. In the first work [26], a global skin disease diagnostic system using deep CNN was suggested, achieving a top-1 accuracy

of 73.1% and a top-5 accuracy of 91.0%. In a subsequent study [27], Liao and colleagues demonstrated that utilizing skin lesion characteristics can aid in skin disease diagnosis due to the visual similarities among many diseases.
Kawahara et al. [28] utilized the MobileNet network, trained on the publicly available DermoFit library, to classify skin lesions into ten categories.

3 Methodology

Many efforts have been made to develop a computer vision system for skin disease classification. In the process, each image from the DermNet dataset undergoes a pre-processing step, where its size is adjusted according to the input requirements of each model. Additionally, various transformation techniques such as rotation, flipping, and zooming are applied to augment the dataset and increase the number of images. Subsequently, several models are proposed and trained using all the images in the training dataset. Finally, the models are evaluated on the test dataset to determine the most effective one. An overview of these steps is depicted in Figure 1..

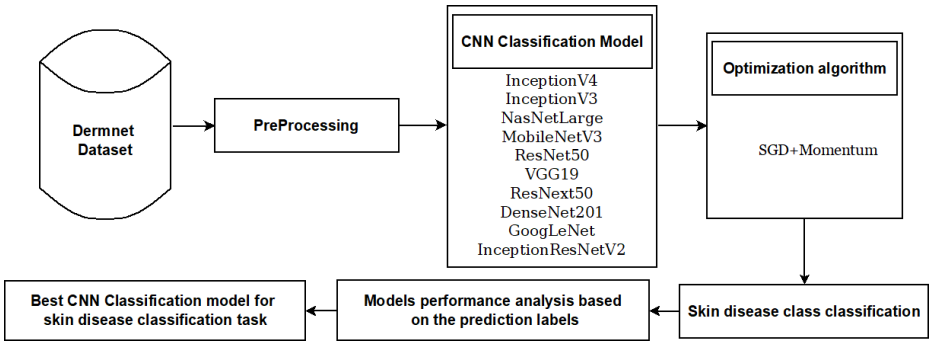

Fig. 1. Overview of the proposed skin diseases classification method

3.1 Dataset

DermNet dataset comprises over 23,000 dermoscopic images that encompass various types of skin diseases. These images are categorized based on a two-level taxonomy. Figure 2 showcases examples of lesion images extracted from the dataset, which encompasses 23 super-classes of skin diseases. Each image is accompanied by a diagnostic tag assigned by an expert in Dermatology Resources.

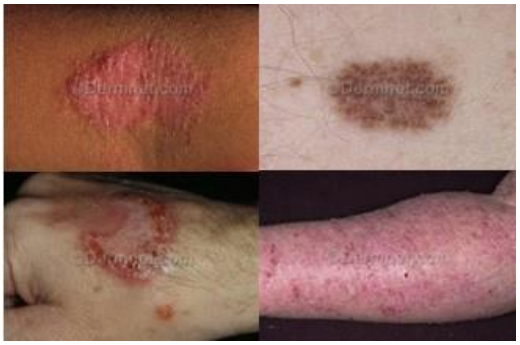

Fig. 2. Random samples of skin lesions from DermNet dataset

The DermNet dataset consists of diverse JPEG images depicting various skin diseases. During the dataset's quality control process, lower-quality images were removed. Following this step, the dataset was divided as follows: 80% of the samples, specifically 12,368 DermNet images, were allocated for training, 15% of the samples (3,085 images) were reserved for network validation, and the remaining 20% (4,002 images) were designated for the testing phase. Figure 3 visually represents the distribution of images across the 23 classes for the training, validation, and testing datasets.

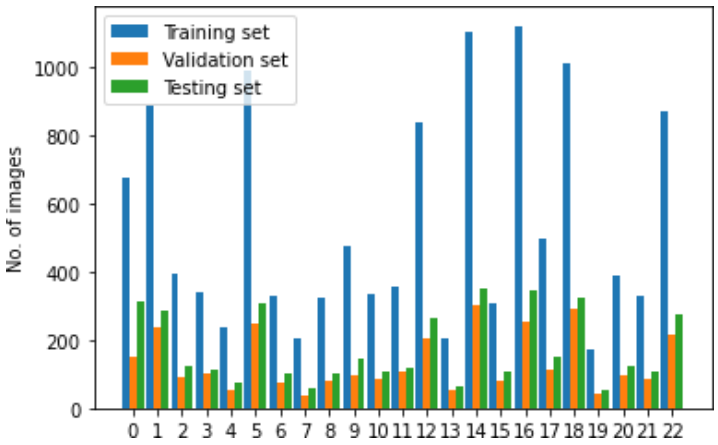

Fig. 3. Distribution of the available DermNet training and validation images

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

```
from tensorflow.python.client import device_lib

def get_available_devices():
    local_device_protos = device_lib.list_local_devices()
    return [x.name for x in local_device_protos]

print(get_available_devices())
```

```
['/device:CPU:0', '/device:GPU:0', '/device:GPU:1']
```

```
import tensorflow as tf
print(tf.__version__)
```

```
2.6.0
```

Figure 4 Usage of Tensorflow

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.

4 Experiments

Deep learning is a branch of machine learning that closely resembles the functioning of the human brain. It surpasses the traditional connection between data and algorithms by enabling machines to learn and improve through experience. Deep learning, drawing upon knowledge from neuroscience, mathematics, and technological advancements, is regarded as a remarkable revolution in the field of artificial intelligence. Recent advancements in computing power and the availability of extensive datasets have demonstrated the efficacy of deep learning algorithms, particularly in the medical domain. With the proliferation of smartphones, patients can now receive preliminary diagnoses of their health conditions, including skin diseases, thanks to the application of deep learning techniques.

4.1 Stratified cross validation

Implementing the concept of stratified sampling in cross-validation ensures the training and test sets have the same proportion of the feature of interest as in the original dataset. Doing this with the target variable ensures that the cross-validation result is a close approximation of generalization error.

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.

4.2 Data pre-processing

	lesion_id	image_id	dx	dx_type	age	sex	localization	dataset
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	vidir_modern
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	vidir_modern
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp	vidir_modern
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp	vidir_modern
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear	vidir_modern

Figure 5 : Data preprocessing

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

Data augmentation is a technique widely used in machine learning and deep learning to artificially increase the size and diversity of a dataset. By applying various transformations such as rotation, flipping, zooming, and adding noise to the existing data samples, data augmentation helps to create additional training examples. This approach proves particularly useful when working with limited datasets, as it enhances the model's ability to generalize and improves its robustness to variations in the input data. Data augmentation plays a crucial role in preventing overfitting and improving the overall performance and accuracy of machine learning models.

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

Image augmentation was performed using ImageDataGenerator API in Keras.

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 is a technique commonly employed in transfer learning, where a pretrained model is adapted to a specific task or dataset. Initially, a model is trained on a large dataset, typically from a related domain, to learn general features and patterns. Fine-tuning involves taking this pre-trained model and further training it on a smaller, task-specific dataset. By adjusting the model's parameters during this fine-

tuning process, the model can specialize and adapt its learned features to the specific nuances and characteristics of the target dataset.

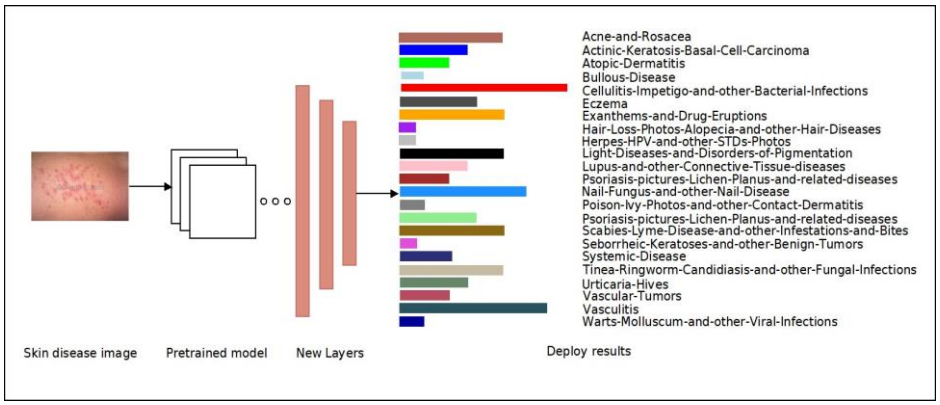

Fig. 6. Example of the finetuning of a pretrained model

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%. Finally, a SoftMax layer is set to produce the probability of each of the 23 output classes. The final deci- sion 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 important to evaluate a model, however they may be deceptive. In case of imbalanced dataset, 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. To overcome such imbalanced dataset scenarios, we can rely on other metrics which may be more useful and give more accurate insight. These metrics are precision, recall and AUC. These measures are defined as follows:

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

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

Precision or positive predictive value quantifies the proportion of the labels properly classified that are truly positive, as represented 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

5.1 Result and analysis

```
the_input [(None, 224, 224, 3)]
wavelet [(None, 112, 112, 12), (None, 56, 56, 12), (None, 28, 28, 12), (None, 14, 14, 12)]
adjust_channels_l3 (None, 28, 28, 64)
conv2d_11 (None, 28, 28, 64)
batch_normalization_9 (None, 28, 28, 64)
activation_9 (None, 28, 28, 64)
conv2d_12 (None, 28, 28, 64)
batch_normalization_10 (None, 28, 28, 64)
adjust_channels_l2 (None, 56, 56, 64)
adjust_channels_l1 (None, 112, 112, 64)
add_4 (None, 28, 28, 64)
conv2d_6 (None, 56, 56, 64)
conv2d_1 (None, 56, 56, 64)
activation_10 (None, 28, 28, 64)
batch_normalization_5 (None, 56, 56, 64)
batch_normalization_1 (None, 56, 56, 64)
conv2d_13 (None, 28, 28, 128)
adjust_channels_l4 (None, 14, 14, 64)
activation_5 (None, 56, 56, 64)
activation_1 (None, 56, 56, 64)
batch_normalization_11 (None, 28, 28, 128)
conv2d_21 (None, 14, 14, 64)
conv2d_7 (None, 56, 56, 64)
conv2d_2 (None, 56, 56, 64)
activation_11 (None, 28, 28, 128)
...
```

Figure 7 :Deep Learning Layers

The learning time and convergence of different models can vary based on their depth and number of parameters. To address this, we utilized the Early Stopping class provided by Keras. This technique monitors the validation accuracy value and stops the training if it does not improve for a predefined number of epochs. This helps prevent overfitting and optimizes the training process. By setting a higher number of epochs and allowing early stopping to intervene, the optimal epoch size can be determined automatically based on the model's performance on the validation set. The accuracy and loss for both the training and validation sets are illustrated in Figures 5 and 6.

```
skin_net = SkinNet(new_base, num_classes=7, activation='softmax')
skin_net.build_graph().summary()
```

```
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 224, 224, 3)]     0
_________________________________________________________________
Wavelet (Functional)         (None, 14, 14, 256)       5512256
_________________________________________________________________
GlobalBranchPooling (Sequent (None, 1024)              263424
_________________________________________________________________
Classifiers (Sequential)     (None, 7)                 7175
=================================================================
Total params: 5,782,855
Trainable params: 5,776,711
Non-trainable params: 6,144
_________________________________________________________________
```

Figure 8 Model 1 , Efficient-Net

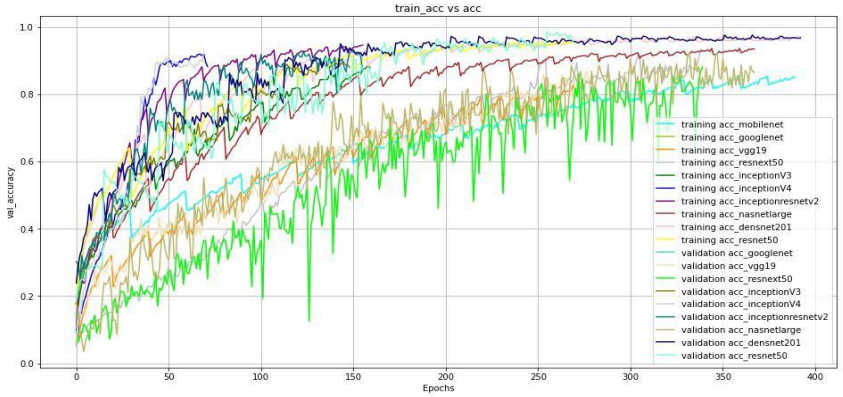

Fig. 9. Training and validation accuracy for the different models

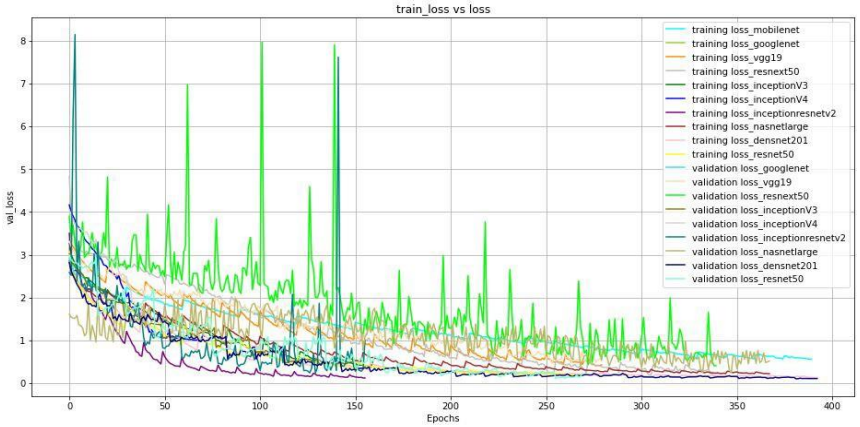

Fig. 10. Training and validation loss for the different models

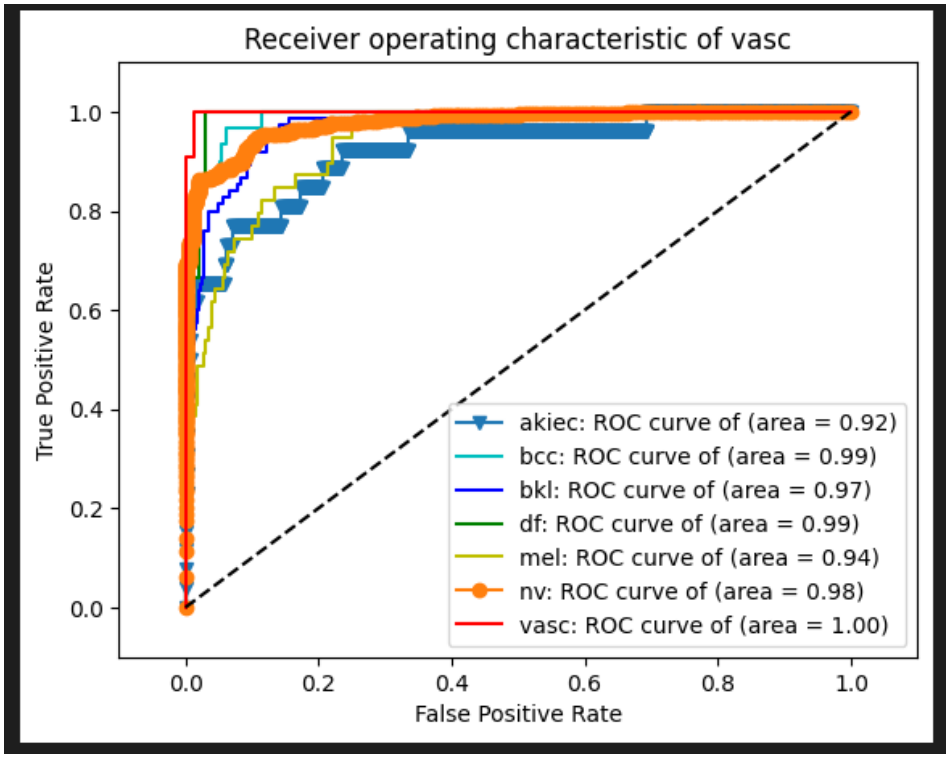

Figure 11 : Receiver operating characteristic of vasc

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 outperformed other architectures with significant mean differences in classification accuracies compared to InceptionV4, InceptionV3, MobilenetV3, ResNet50, VGG19, ResNext50, NASNet, GoogleNet, and InceptionResNet-v2. This could be attributed to the deep neural structure of DenseNet-201, allowing it to extract

distinct features and map complex patterns effectively, leading to more precise differentiation and diagnosis of classes. However, accuracy alone may not be the best metric for evaluating a model, and metrics like precision and recall provide valuable information. In terms of accuracy, NASNet-large achieved the highest result (65.94%), followed closely by ResNet50 (65.84%), GoogleNet (64.01%), InceptionV4 (63.47%), and InceptionV3 (63.47%). The recall rate followed a similar pattern. VGG19, ResNext50, and MobileNetV3 had insignificant accuracy rates, with MobileNetV3 being the lowest. DenseNet201 and InceptionResNetV2 showed acceptable accuracy results, but DenseNet201 was chosen due to its higher accuracy and lower number of parameters compared to InceptionResNetV2, making it computationally more efficient. Further evaluation and details can be obtained through metrics like Area Under the Receiver Operating Characteristic Curve (AUC). The performance of CNN models can vary depending on the data quantity and image complexity. Generally, a large dataset is required for the model to learn effectively and differentiate between classes. In this case, the dataset was not very large, particularly for certain classes, necessitating the use of transfer learning with pre-trained models like ImageNet. The specific characteristics and size of the dataset led to freezing the initial CNN layers and reusing the remaining parameters/weights, considering the computational cost. Reducing the number of frozen layers increases the computational cost of training.

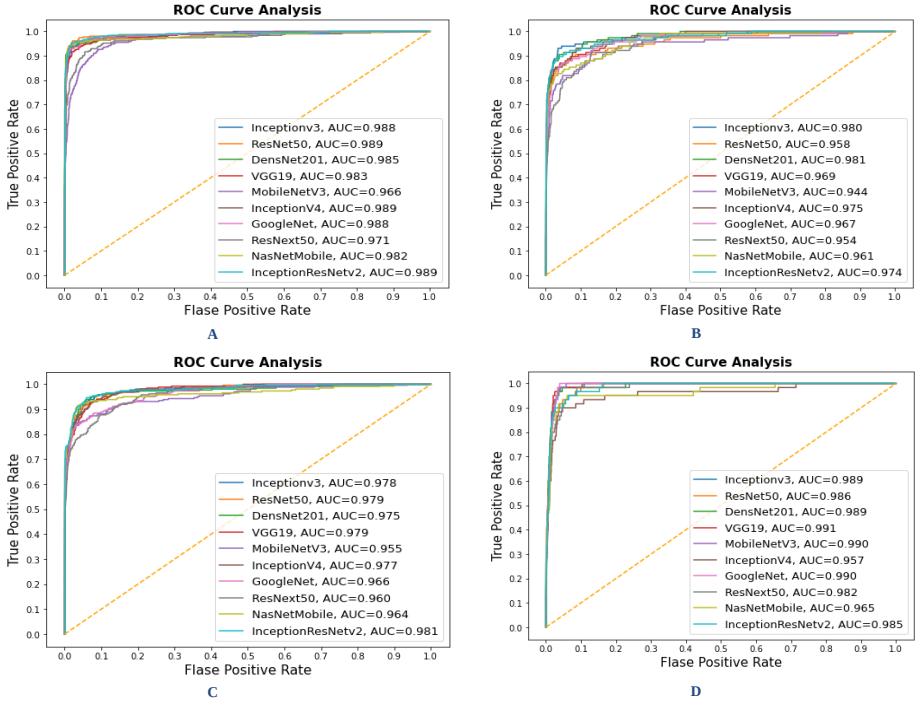

Fig. 12. ROC curve of the first class of skin lesions for the four classes predicted correctly (A: Acne and Rosacea, B: Psoriasis-pictures-Lichen-Planus-and-related-diseases, C: Nail Fungus and other Nail Disease, D: Hair-Loss-Photos-Alopecia-and-other-Hair-Diseases)

6 Conclusion

In this study, we trained and compared high-performance CNN architectures, including InceptionV4, InceptionV3, DenseNet-201, MobilenetV3, ResNet50, VGG19, ResNext50, NASNetLarge, GoogleNet, and InceptionResNetV2, using the DermNet dataset consisting of 19,434 images belonging to 23 classes of skin diseases. The objective was to select the most suitable model for developing a practical computer-aided diagnosis system in dermatology. Our findings revealed that DenseNet-201 performed the best among the deep learning neural networks (CNN architectures) for automatically diagnosing skin diseases using the DermNet dataset. This comparative study provides valuable insights into the behavior of CNNs in relation to a large number of classes, including the most common skin illnesses present in the DermNet dataset. The results can serve as a foundation for future research and the development of computer-assisted diagnostic systems for various dermatological conditions.

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

Chaimaa Ouknider is a Double Degree Student, an Information Systems and Big data engineering student, and a Bachelor of Business Administration, seeking to strengthen existing skills and develop new ones in data analysis and predictive modeling.

Mohammed Ait Sidi Ahmed is a

Oussama Nafi is a student specializing in Information Systems and Big Data Engineering, aiming to enhance proficiency in data analysis and predictive modeling and aspiring to bolster current skills and acquire new ones in the realm of data management and interpretation.

Aymane Elouadi is a

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.

Article submitted 2024-01-15.