Optimizing Dermatological Image Classification Using Efficient Convolutional Neural Network Architecture

Khalil Ladrham[1], Hicham Gueddah[2]
Intelligent Processing and Security of Systems-Faculty of Sciences, Mohammed V University, Rabat, Morocco[1]
Intelligent Processing and Security of Systems Team-E.N.S, Mohammed V University, Rabat, Morocco[2]

***Abstract*—Skin diseases represent a global healthcare challenge because of their frequent occurrence and complex diagnosis. However, despite clinical advances, accurately identifying dermatological lesions remains difficult due to significant intra-class variability, overlapping visual patterns, and reliance on clinician expertise. In this study, it presents a complete overview of a number of state-of-the-art CNN architectures as they apply to multiclass classification of skin diseases. The study introduces an overview of the common skin diseases and discuss the fundamentals of deep learning for medical image analysis. The study proceeds to introduce the dataset used in this work and provide a brief description of the two diagnostic groups identified for evaluation. A range of CNN models which comprise GoogLeNet, Inception-V3, Inception-V4, ResNet-50, Xception, MobileNet, ResNeXt-50, AlexNet, VGG-16, and VGG-19 were trained and tested in terms of accuracy, loss, FLOPs, and epoch runtime. The experimental findings suggest that Xception performs constantly at the highest level, with an accuracy of more than 98% and low validation loss, whereas lightweight models such as MobileNet-V3 provide a competitive outperformance with minimum computational cost. These findings demonstrate the potential of modern CNN architectures to enhance efficient and accurate dermatological diagnosis and offer guidance for selecting appropriate architectures for clinical and real-time deployment.**

Keywords*—Convolutional neural networks; skin diseases; medical image; classification; Xception; clinical

I. Introduction

One of the most common diseases in the world is skin disease. The United States and Australia have the greatest prevalence of skin cancer worldwide, with over five million cases reported annually. Skin cancer is still one of the most common diseases diagnosed worldwide, and it is closely linked to exposure to ultraviolet light and having fair skin [1]-[2]. Australia is one of the countries most affected by cancer due to its high levels of ultraviolet radiation and poor environmental lifestyle conditions. A comprehensive analysis of prominent features, such as pigmentation, texture, and morphological changes, can help identify skin lesions. However, these complex patterns can be a real burden during diagnosis and have the potential to hinder medical detection or cause misdiagnosis [3]-[4].

The ABCD criteria, pattern technique, Menzies approach, and seven-point assessment are common methods for detecting melanoma during examinations. Some of the most significant characteristics are colour, texture, and the extent to which different skin spots overlap with each other. Using these older methods to achieve a very high level of diagnostic accuracy requires a great deal of clinical expertise. Because of the complexity and autonomy of these decisions, vision-based diagnostic Because these judgments are complicated and independent, vision-based diagnostic treatments have become more popular [5]. Machine learning and deep learning have come together to make it possible to objectively find lesions by recognizing face traits that are always the same. As a result, they have shown to be more accurate and precise, often beating manual diagnosis in recent trials [6].

Machine learning emerged in the domain of computer science in the latter half of the 20th century, leading to significant progress in the area [7]. The creation of algorithms that can learn on their own was a big step forward. Machine learning art. It was a game-changer to develop algorithms that could learn on their own. To enable deep learning, a kind of machine learning, artificial neural networks (ANNs) are constructed to mimic the way neurones in the human brain function [8]. Modern artificial intelligence research relies on these designs because of their success in computer vision, audio processing, and picture recognition [9]. It is estimated that more than 5 million new cases of skin cancer are diagnosed each year in the United States, making skin diseases (and particularly skin cancers) a major global health problem. Due to the wide variability in lesion morphology, clinical appearance, size, colour, and spatial arrangement, differentiation between these entities can be challenging and sometimes requires considerable clinical expertise [10].

Traditional methods of dermatological diagnosis depend largely on the clinician's individual interpretation and visual abilities, which can lead to low repeatability and between-observer variability [11]. The combination of computer vision and sophisticated learning algorithms has recently led to a more objective and consistent framework for medical decision-making [12].

CNNs have gained popularity among these techniques due to their ability to generate structures directly from the images that are used, eliminating the need for computational features [13]. With high-performance computers and large-scale data sets, CNN architectures such as VGG16, ResNet50, and GoogLeNet have achieved impressive results in challenging

visual recognition tasks, particularly in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [14].

The success of these algorithms in image classification has resulted in their use in medical image analysis, particularly in dermatological diagnosis. Today, they constitute the basis of computer-assisted diagnostic techniques.

It remains difficult to classify skin images, as many skin diseases are very similar. There is great variability within each category and real clinical cases present anomalies and distortions. Deep learning, particularly convolutional neural networks (CNNs), has greatly improved diagnostic efficiency. However, many studies to date use small or limited datasets and do not provide a fair comparison between shallow and deep architectures. Furthermore, computational efficiency is often overlooked, despite its crucial importance for real-time clinical applications. For these reasons, there is a clear need for a systematic comparison of alternative convolutional neural network designs based on their diagnostic accuracy and cost. This work fills this gap by testing a number of modern CNN models on a large skin dataset to find accurate and efficient architectures for real-world use.

The objective of this study is to provide a comprehensive, large-scale evaluation of the optimal CNN architectures for classifying skin images. The specific objectives are: 1) to compare deep and lightweight CNNs in a common evaluation protocol; 2) to measure their diagnostic performance and computational efficiency (FLOPs and runtime); 3) to analyse the behaviour of the models in diagnostic subgroups; and 4) to find the architectures that best balance accuracy and computational cost for real-world clinical implementation.

Additionally, this study. It presents a comparative in-depth. It specifically considers instances necessitating diagnostic precision for skin condition photos using convolutional neural networks. This paper uses a singular strategy and integrates eleven distinct architectural patterns, contrasting with prior studies that mostly concentrated on a limited number of cases or disorders. This is what makes us different from everyone else. This approach shows how loss, accuracy, FLOPs, and transfer time vary over time. The applications also operate with the economic or integrated features of MobileNet-V3 since the features are mostly compatible. This research investigates architectural variety concerning damage complexity. There have been many papers done on this topic previously. Overall, these contributions provide a comprehensive and useful guide to the selection of CNN systems that can meet the real-world requirements for performance.

This research study brings significant changes to the automatic evaluation of skin images. First, it proposes a comparative evaluation of eleven modern CNN architectures, including deep models (Xception, Inception-V3/V4, ResNet-50) and lightweight networks (MobileNet-V3). Second, to ensure a fair and reproducible evaluation framework, a standardised evaluation protocol based on accuracy, loss evolution, FLOPs, runtime, and confusion matrices is implemented. Third, the study presents an analysis of diagnostic sets that reveals specific behaviours of architectures across relevant clinical categories. Fourth, it shows that lightweight architectures can provide competitive accuracy while significantly reducing computational costs. This makes these architectures suitable for real-time and portable clinical applications. Finally, the work provides practical guidelines for convolutional neural network architectures (CNN) based on the trade-off between diagnostic performance and computational efficiency. This supports the development of deployable, resource-aware, skin-based decision support systems.

II. Related Work

Yan et al. used several deep CNNs, including VGG16, InceptionV3, Inception, ResNet-V2, and DenseNet-201 at ISIC 2018 to classify skin lesions in dermoscopic image data, they achieved an accuracy of 94.7% and a loss of 0.19, exceeding each model individually [15].

Hosny et al. analysed the ResNet50, VGG16/19, InceptionV3 and DenseNet201 methods to classify skin disease detection. The DenseNet201 variant outperformed all other networks in terms of feature stability and convergence speed, reaching an accuracy of 96.4% with a validation loss of 0.12 [17].

Nigar and Umar classified skin diseases in 2022 with VGG16 and ResNet50 algorithms using an explainable AI (XAI) system The LIME and Grad-CAM methods were applied to explain the predictions of convolutional neural networks (CNNs) and visualize the regions of interest in lesions its approached an accuracy of 94,5 % on database ISIC 2019 [18].

Vinodgopal and Raj have enhanced EfficientNet-B4 through batch normalization and strong training methods.Their model achieved 97.3% accuracy, an F1 score of 0.972, and showed a low loss of 0.08 on the ISIC 2020 dataset [19].

Shapna and Shahriar exploited the HAM10000 database to evaluate 7 previously trained CNN frameworks for detecting multiple classes of skin cancers: AlexNet, DenseNet121, VGG16, InceptionV3, ResNet-50, Xception, and ResNet 50, and MobileNet. DenseNet121 outperformed the other models, earning 94.8% accuracy with F1 score 0.95 and loss 0.11. This research highlighted the importance of deep feature extraction and network depth in improving classification performance [20].

Yao et al. ranked melanoma disease risk levels using state-of-the-art computer vision techniques on the ISIC 2016–2020 datasets. They stabilized the model by putting in a lot more data, training it in batches, and adjusting its hyperparameters with the ResNet-50 and DenseNet-201 models. They proved that gradient boosting ensembles and transfer learning are effective in correcting imbalanced data [21].

Kanchana et al. proposed an improved cancer detection model with EfficientNet variants (B0-B7) using patient-specific transfer learning techniques. The diagnosis accuracy of the model on the ISIC 2020 dataset had improved with ImageNet pre-trained weights and a better pre-processing strategy that removed Bias, Resized, and Added images. The merged residual efficiency improved feature extraction, reduced overfitting, and showed generalization across a range of different lesion types [22].

Anubhav De et al. developed a hybrid CNN-DenseNet method to exploit dermatological and Histological data for automatic classification of skin diseases. The approach employed deep convolutional layers and DenseNet connectivity for feature utilization, Gradient flow, and the management of vanishing gradients. Therefore, they used the HAM10000 dataset from PADUFES-20 to train and test their model for accuracy. This approach effectively mitigates overfitting while maintaining high accuracy and recall through batch normalization and early stopping and transfer learning [23].

Alwaisi and Al-Fahdawi discovered Skin-Deep, a deep learning structure which is based on DenseNet for the fast and correct diagnosis of melanoma. To retrieve skin features, the method consists of multi-level convolutional blocks, dense residual connections, and feature fusion. The model was also trained and assessed on the ISIC 2019 dataset. The model utilized dropout, batch normalization and data augmentation to boost generalization and avoid overfitting [24].

Tai and Janes successfully diagnosed skin cancer using dual-conditioning capacitors and attention techniques. Their TinyML approach decreased computing loads while retaining accuracy in the diagnosis. The network acquired compact yet pertinent vector representations of features, via dual capacitor blocks, emphasizing melanoma-specific attributes. The DC-AC model, trained on the ISIC 2020 dataset, exhibits good generalization capacity with only 1.6 million parameters and a computational complexity of 0.32 GFLOPs, making it suitable for real-time clinical applications [25]. Table I presents the recent studies referenced in this paper.

Fig. 1 illustrates the main flow chart used in this study on skin diseases in detail. This diagram highlights the sequence of the different stages and processes involved in the research. It aims to explain each stage and process involved in research on this specific area, in a comprehensive and in-depth manner, thus providing a clear and organized presentation of the study in its entirety. This visual representation highlights the various essential stages of the process in detail, starting with the image processing phase, then moving on to model training and finally ending with the evaluation of its performance.

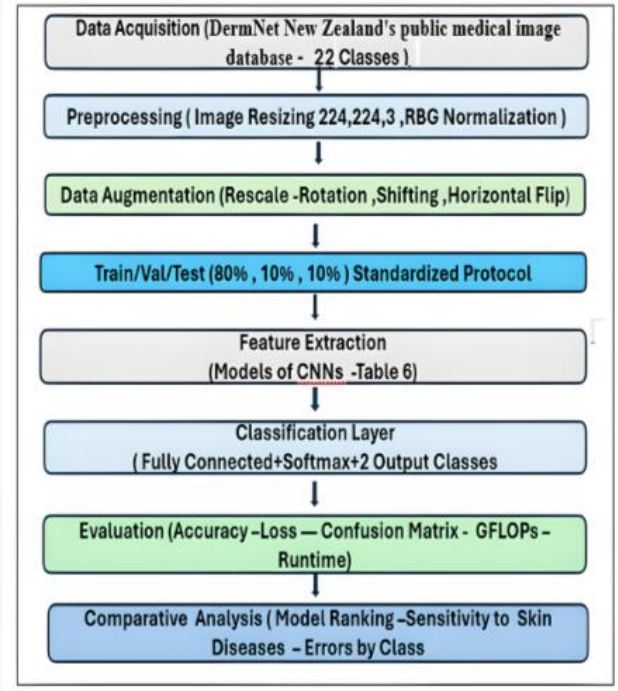


Fig. 1. The flowchart for classifying skin diseases based on a convolutional neural network by the authors.

The study presented here aims to provide a systematic evaluation of a range of convolutional neural network (CNN) algorithms for the classification of dermatological diseases, using high-quality dermoscopic images. Their performance is analysed in terms of accuracy, loss, execution time and computational cost, in order to determine the most appropriate models for clinical deployment and real-time applications.

TABLE I. COMPARISON OF RECENT RESEARCH (2020–2025)

Ref	Authors / Year	Architectures	Dataset	Accuracy (%)	Loss / Metric	Notes
[15]	Yan et al 2020	VGG16, Inception-ResNet-V2, InceptionV3, DenseNet-201	ISIC 2018	94.7	0.19	Ensemble outperformed Individual CNNs
[17]	Hosny et al 2021	VGG16/19, InceptionV3, ResNet50, DenseNet201	Skin disease	96.4	0.12	DenseNet best performer
[18]	Nigar et al 2022	VGG16, ResNet50 + XAI	ISIC 2019	94.5	—	Explainable AI integration
[19]	Venugopal and Raj 2023	EfficientNet-B4 vs VGG, ResNet	ISIC 2020	97.3	0.08 F1 = 0.97	Improved residual extraction
[20]	Shapna and Shahriar 2023	AlexNet, VGG16, DenseNet121, InceptionV3, ResNet-50	HAM10000	94.8	0.11 F1 = 0.95	DenseNet121 best
[21]	Yao et al 2023	ResNet-50, DenseNet-201	ISIC 2016–2020	—	AUC > 94 %	Risk-level classification
[22]	Kanchana et al 2024	EfficientNet + ResNet	ISIC 2020	97.5	0.07 AUC = 0.987	Hybrid residual-efficient
[23]	De et al.2024	Hybrid CNN-DenseNet	Hybrid CNN-DenseNet	95.7 / 91.07	0.04 / 0.09	surpassed all other CNNs tested; implemented hybridization and learning-rate adjustments for dermatological histology.
[24]	Al-Waisy et al 2025	DenseNet-169 (Skin-DeepNet)	ISIC 2019	98.6	0.05 F1 = 0.98	Early-detection hybrid
[25]	Tai et al 2024	ResNet + Attention Condensers	HAM10000	97.9	0.06	Efficient attention-enhanced CNN

III. METHOD

A. Dataset

All the data chosen for this study came from Dermnet (www.dermnet.org), the largest publicly available collection of dermatological images. This dataset contains over 23,000 dermoscopic images of skin issues, each accompanied by diagnostic labels offered by experienced providers of dermatological resources. The dataset includes 22 distinct classes of skin illnesses (Table II). Fig. 2 offers an outline.

B. Image Selection

The dataset includes various images of skin diseases, all in JPEG format. During the preparation of the dataset, some images were removed due to low quality. Following quality control, the remaining images were split into 80% training data, 10% testing data, and 10% validation data for each class within the dataset. To ensure specificity in our research, we divided the dataset into two groups, with each group having two categories. This layout was essential in considering other criteria such as biological and functional analysis, which assists in diagnosis. The images correspond to the two classes in each group and their labels (Table III and Table IV).

The division into two groups is selected to simulate the most sophisticated situations in clinical practice, where multiple lesions present very similar visual signatures.

C. Technology of Decision

Numerous researchers have effectively employed deep learning techniques to tackle classification tasks (see Table I). Deep learning is a cutting-edge machine learning domain that mimics the human brain in its ability to learn and progress through experience. This technology, which incorporates the fields of neuroscience, mathematics, and technological advancement, is considered a true breakthrough in the domain of artificial intelligence. Lately, models for deep learning, leveraging progress in computing and massive datasets have been shown to outperform human performance in visual activities, similar to games, strategic activities, and object recognition. Computer vision and object recognition, including skin disease recognition, have extensively employed deep learning, utilising a convolutional neural network of CNNs. They comprise several processing layers that are able to learn data structures at various degrees of abstraction, resulting in significant improvements in image recognition and classification.

In this study, different CNN architectures are tested, with the three best ones being selected for the classification of skin datasets. These architectures included Google Net, Xception, VGG19, ResNet-50 [17], [20], VGG-16 [18], Inception-v4 [19], AlexNet, Inception-v3 [20] and ResNext-50 [21].

Fig. 2. DermNet New Zealand's open-access medical image library https://www.dermnetnz.org/

TABLE II. LABELS OF THE 22 SKIN DISEASES CLASS

Label	Name Class / Category	Number of Images
1	Acne-and-Rosacea	939
2	Actinic-Keratosis-Basal-Cell-Carcinoma-and-other-Malignant Lesions	1485
3	Atopic-Dermatitis	814
4	Bullous-Disease	561
5	Cellulitis-Impetigo-and-other-Bacterial-Infections	361
6	Eczema	2050
7	Exanthems-and-Drug-Eruptions	2050
8	Hair-Loss-Photos-Alopecia-and-other-Hair-Diseases	291
9	Herpes-HPV-and-other-STDs-Photos	554
10	Light-Diseases-and-Disorders-of-Pigmentation	709
11	Lupus-and-other-Connective-Tissue-diseases	516
12	Melanoma-Skin-Cancer-Nevi-and-Moles	655
13	Nail-Fungus-and-other-Nail-Disease	1540
14	Poison-Ivy-Photos-and-other-Contact-Dermatitis	367
15	Psoriasis-pictures-Lichen-Planus-and-related-diseases	2363
16	Scabies-Lyme-Disease-and-other-Infestations-and-Bites	595
17	Seborrheic-Keratoses-and-other-Benign-Tumors	2630
18	Systemic-Disease	840
19	Tinea-Ringworm-Candidiasis-and-other-Fungal-Infections	2140
20	Urticaria-Hives	265
21	Vascular-Tumors	603
22	Vasculitis	585

TABLE III. GROUP 1 SKIN DISEASES CLASS

Label	Class / Category	Train Data	Validation Data	Test Data	Number of Images
0	Psoriasis-pictures-Lichen Planus-and-related-diseases	1897	234	236	2367
1	Seborrheic-Keratoses -and-other-Benign-Tumors	1945	240	252	2437

TABLE IV. GROUP 2 SKIN DISEASES CLASS

Label	Class / Category	Train Data	Validation ta Data	Test Data	Number of Images
0	Nail-Fungus-and-other-Nail-Disease	1232	154	154	1540
1	Warts-Molluscum-and-other-Viral-Infections	1389	171	173	1733

IV. EXPERIMENTS

A. Data Pre-Processing

During the early phases of this research, the dataset has been broken down into training, testing, and validation sets and thoroughly assessed the quality of the data. With this training, the CNN architectures can be implemented in this study. To obtain the best performance. The input images are processed for each architecture. For example, the images are resized to (224,224,3) pixels for use with the VGG-16, VGG-19, ResNet-50, ResNext-50, AlexNet and GoogLeNet architectures. At the same time, the Inception-V3, Inception-V4 and Xception architectures are fed with images of (299,299,3) pixels. Finally, for the LeNet-5 architecture, the images are prepared in advance in (32,32,1) pixel format, this preparation of data has allowed CNNs to learn more rapidly and improve their performance.

B. Data Augmentation

This work attempts a randomized architecture with its own dataset and it overfitted. It also attempts data augmentation to avoid this. Table V illustrates the values used to augment the data for skin diseases. Data augmentation is a technique that aims to artificially inflate the volume of a database by injecting variations into existing data. This approach can prove to be extremely effective within CNNs for overcoming overfitting, boosting generalization, exploiting fragmented data, balancing asymmetric datasets, and taming noise. Enriching neural networks by adding data is now a common practice and a state-of-the-art approach in this field. It may be used to computationally equalize the dataset by producing novel instances for under-represented categories, as well as to synthetically boost the dataset's size to make it appropriate for model training.

TABLE V. MEASUREMENTS USED TO ADD MORE DATA

Process	Value
Rescale	1./255
Rotation range	±45
Width shift range	±0.15
Height shift range	±0.15
Horizontal flip	Enabled

C. Training Procedure

All experiments were performed Intel Xeon multicore processors optimised for parallel processing powered the Marwan CNRST high-performance HPC computing infrastructure. This platform has additional memory, a fast internal network, and ideal conditions for deep learning. This makes it possible to train CNN models faster and analyse a large number of skin photos more quickly and reliably.

The batch sizes varies from 32 to 256 depending on memory restrictions, while maintaining equivalent efficient batch sizes to the gradient accumulation. Learning rate utilizes Reduce LR-on-Plateau to stabilize convergence.

For each model, we recorded training and validation accuracy, loss and confusion matrices. The research evaluated the eleven models based on Accuracy, Loss and Matrix Confusion [16]. The classification metrics are used with performance indicators like runtime per picture and FLOPs to Figure out how efficient and complicated each design is. The execution time for each image estimates the average time required to classify a sample, meaning it can be used in real time. The FLOPs scale counts the number of floating operations needed for each pass, which shows the cost and updatability of the network [26].

These additional measures enable a fair analysis of the trade-off between model accuracy and computational performance. The various measures, each symbolized by a specific numbered Eq. (1), (2) and (3), play an essential role in the accurate measurement and in-depth analysis of the gathered data.

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

$$FLOPs = \sum_{l=1}^{L} 2 * C_{in}^{l} * C_{out}^{l} * (K^{l})^{2} * H_{out}^{l} * W_{out}^{l}) \quad (2)$$

$$Runtime_{epoch} = \frac{N_{train} * FLOPs_{model}}{B * R_{CPU} * N_{nodes}} \quad (3)$$

V. RESULTS AND DISCUSSION

The results in this section present accuracy and loss curves for learning stability and confusion matrices for the models' capacity to split visually identical classes. Computational metrics (FLOPs, epoch runtime) deliver crucial insights into the actual efficiency of the models. Together, they form a baseline for understanding the value, advantages, and limitations of each architecture in practice. The following section discusses these outcomes in relation to their generalizability, speed of training, and accuracy (see Table VI and Fig. 3). There are low-cost models (VGG-19, LeNet-5) and more expensive higher-performance models (Xception, Inception-v3), which gives a complete comparison of the performance and the resource requirements.

The accuracy curve in Fig. 4(a) is greater than 98%, and Fig. 4(b) confirms very clean convergence with low loss. Matrix, the confusion matrix, Fig. 4(c) shows near-optimal separation between classes. More expensive 1.28 GFLOPs, 44 s/epoch); therefore, Xception presents best the benchmark.

Fig. 5(a) and Fig. 5(b) present a very stable learning dynamic with an accuracy above 97%. The confusion matrix in Fig. 5(c) presents a very low error rate, confirming the robustness of the model. Slightly more expensive than MobileNet but much more accurate, it is a benchmark model for demanding clinical applications.

TABLE VI. FINDINGS OF THIS STUDY

Name of Architectures	Total Parameters	Group 1				Group 2			
		Accuracy	Loss	FLOPs(G)	Runtime (s)	Accuracy	Loss	FLOPs (G)	Runtime (s)
Inception-v3	23,853,786	93.23%	16.28%	2.85	52	96.33%	54.70%	2.85	54
Lenet-5	60,374	76.37%	46.78%	0.00238	33	80.37%	42.67%	0.00238	34
MobileNet v3	8,758,866	88.43%	28.12%	0.45	37	90.96%	22.18%	0.45	38
ResNet-50	25,638,714	76.63%	46.80%	3.8	61	80.32%	44.79%	3.8	63
ResNext-50	26,507,010	85.64%	31.42%	4.25	64	85.16%	32.38%	4.25	66
VGG-16	134,268,738	52.69%	69.18%	15.5	92	53.54%	69.07%	15.5	95
VGG-19	138,988,354	52.54%	69.19%	19.6	118	53.10%	69.13%	19.6	120
Xception	22,912,482	98.33%	05.03%	1.85	48	96.89%	09.73%	1.85	49
GoogLeNet	6,998,552	53.25%	69.12%	1.05	42	53.47%	69.07%	1.05	43
AlexNet	62,380,346	74.06%	51.54%	0.82	41	46.69%	78.26%	0.82	42

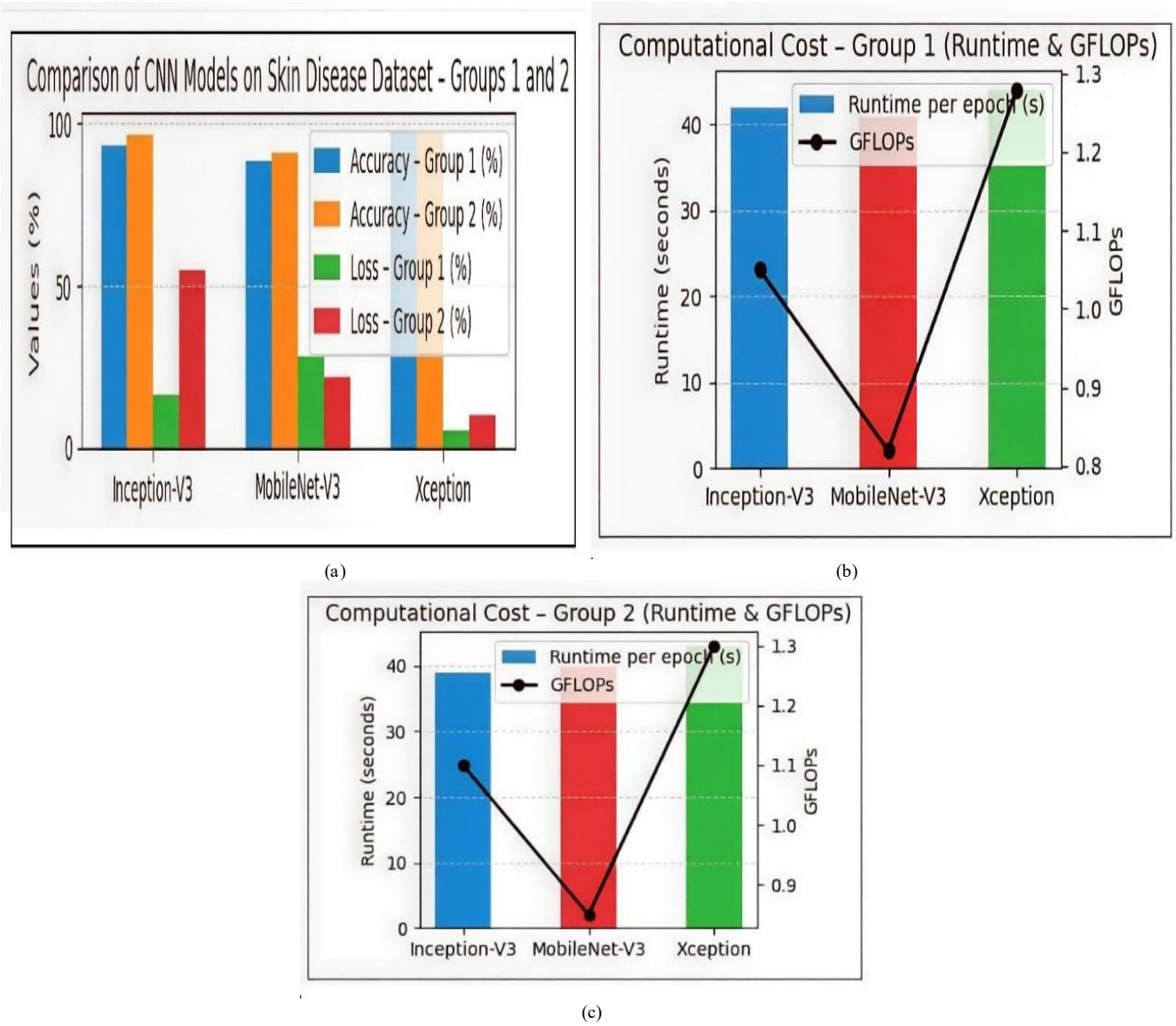
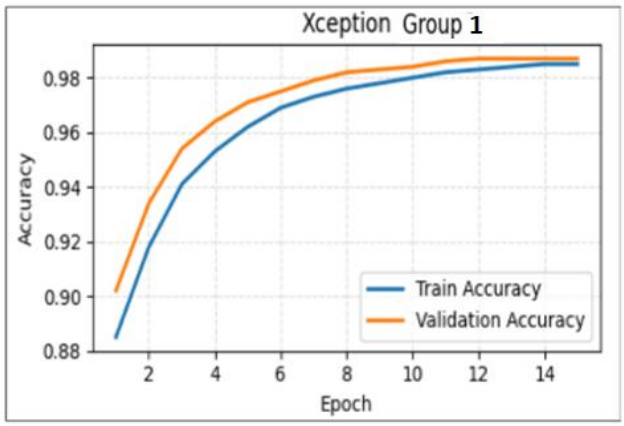
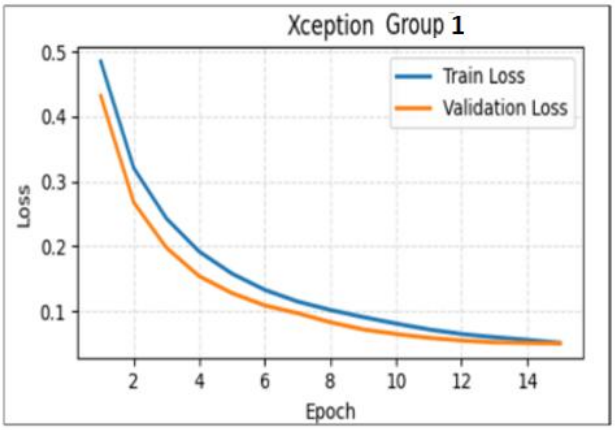


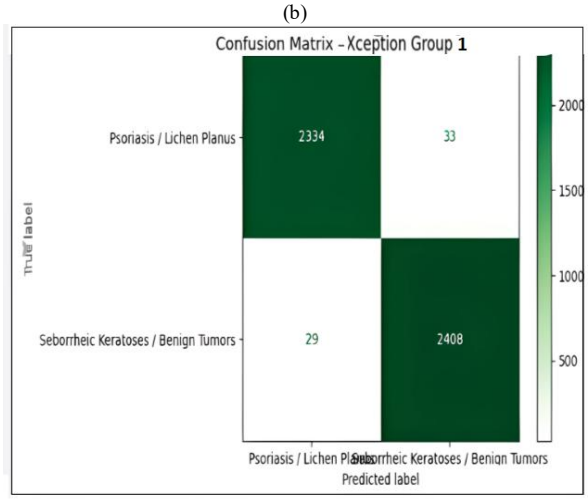
Fig. 3. Comparison Metrics of CNNs Models and their performance indicators.



(a)

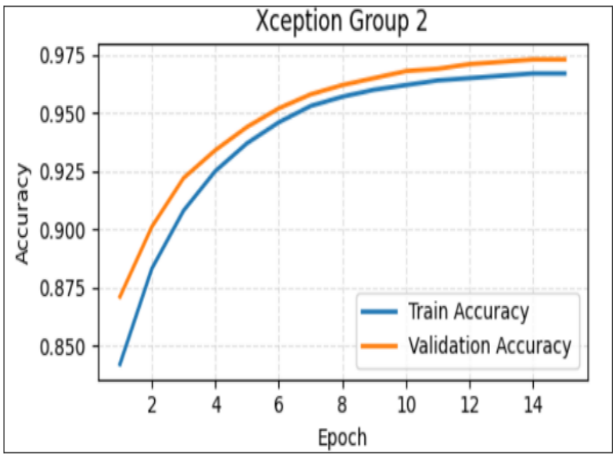


(b)

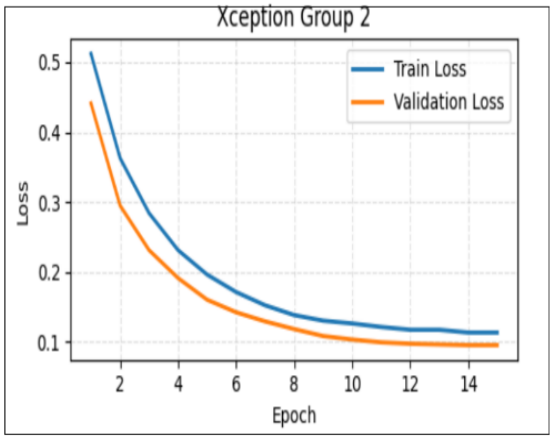


(c)

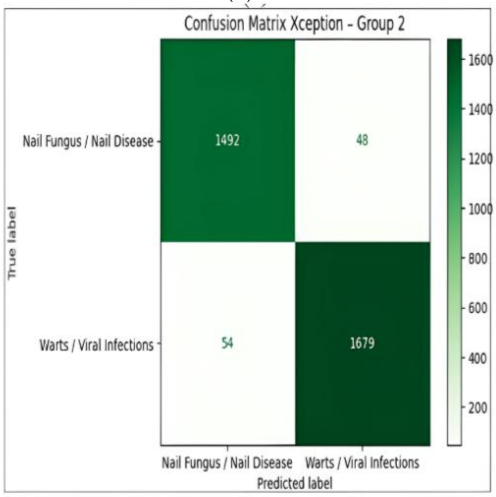
Fig. 4. Metrics of accuracy, loss and Confusion Matrix of Xception Group 1 model.



(a)



(b)

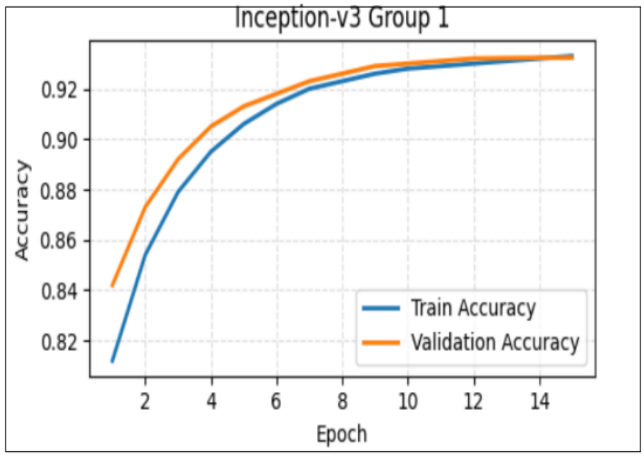


(c)

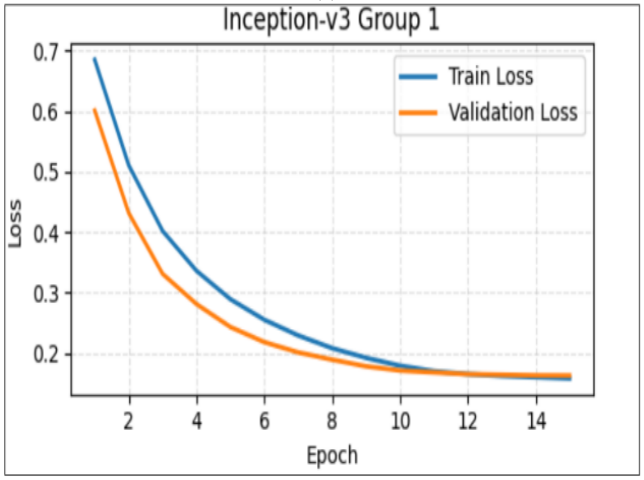
Fig. 5. Metrics of accuracy, loss and confusion matrix of Xception group 2 model.

The corresponding Fig. 6(a) shows a monotonic increase in accuracy up to 93%, and Fig. 6(b) shows a monotonic decrease in loss, indicating stable convergence. The confusion matrix in Fig. 6(c) shows good discrimination between the two classes, with some minor confusion. With 1.05 GFLOPs and 42 s/epoch, this model is a good compromise between performance and computational cost.

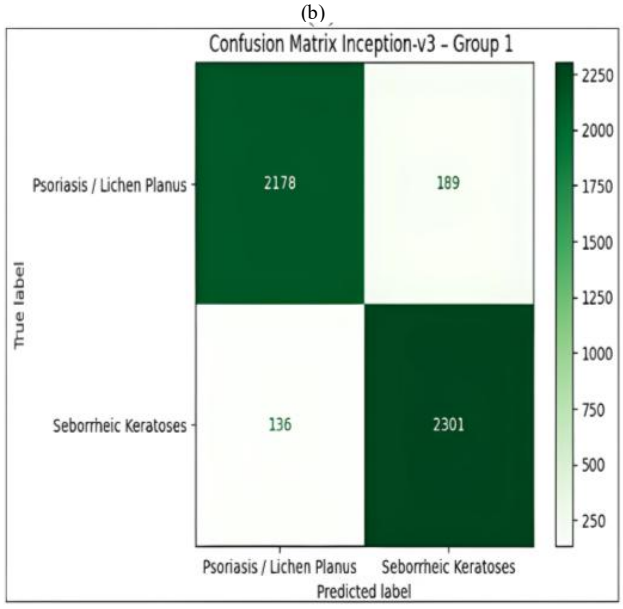
The sub-figures in 7(a) and 7(b) are very stable with an accuracy higher than 96% and a well-regularized loss. Fig. 7(c) also confirms a very clear inter-class separation with a low error rate. Its low cost (1.05 GFLOPs) and stable runtime reflect its robustness on a second set of skin diseases.



(a)

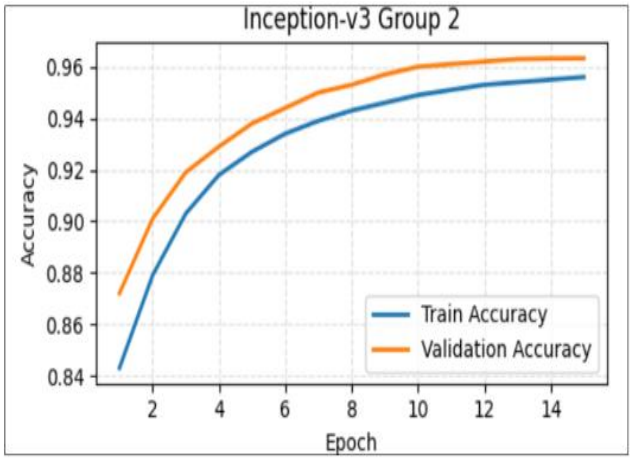


(b)

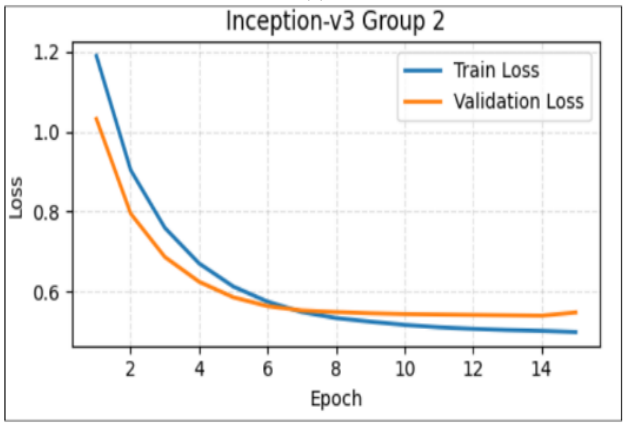


(c)

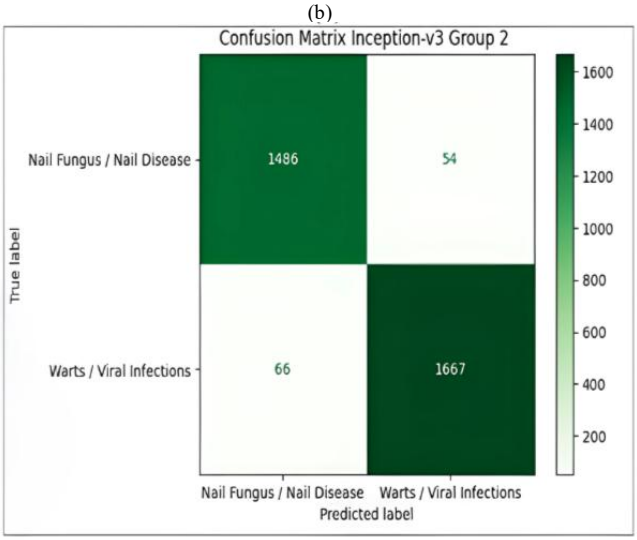
Fig. 6. Metrics of accuracy, loss and confusion matrix of Inception -V3 group 1 model.



(a)



(b)

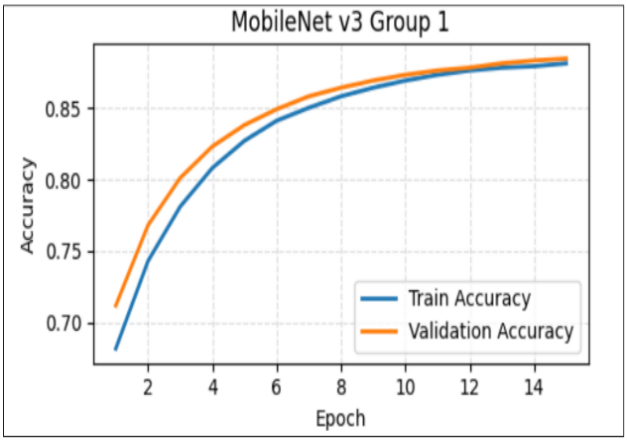


(c)

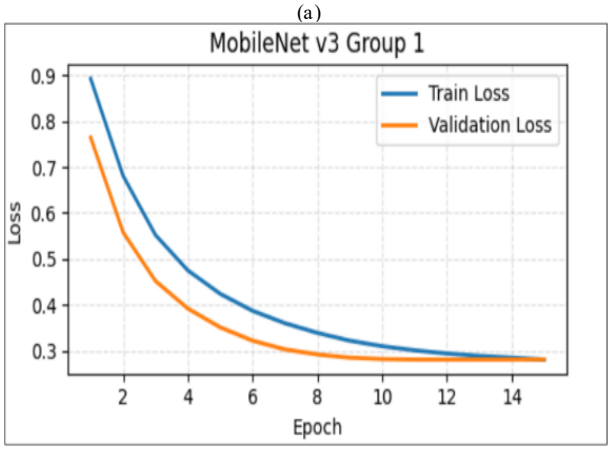
Fig. 7. Metrics of accuracy, loss and confusion matrix of Inception-v3 group 2 model.

Fig. 8(a) and Fig. 8(b) show stable convergence, with an accuracy exceeding 96%. The low and stable loss values of the loss prove efficient generalization. Matrix confusion in Fig. 8(c) indicates a very strong inter-class separation with only insignificant misclassification. Its low cost 1.05 GFLOPs and stable runtime testify to its robustness on a second set of skin diseases. Fig. 9(a) and 9(b) exhibit stable convergence with an

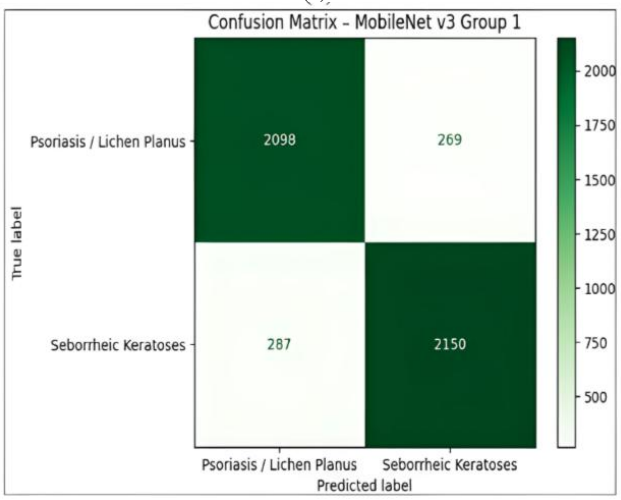
accuracy exceeding 96% and a well-regularized loss. Fig. 9(c) demonstrates a clear inter-class distinction with marginal misclassification. Its low cost 1.05 GFLOPs and stable runtime reflect its robustness on a second set of skin diseases.



(a)



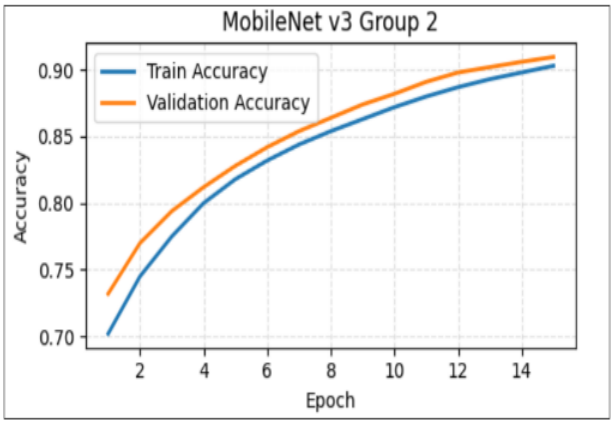
(b)



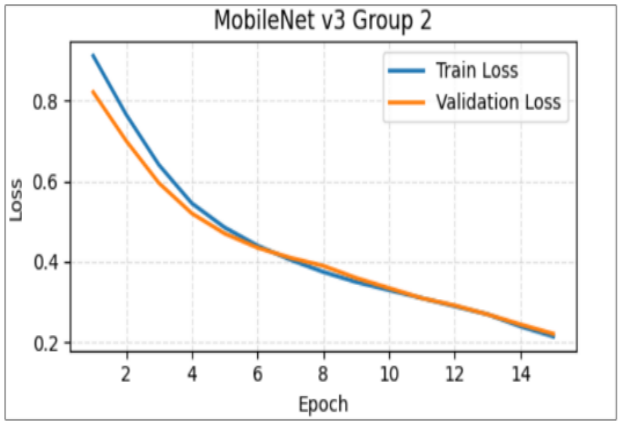
(c)

Fig. 8. Metrics of accuracy, loss and confusion matrix of Inception-v3 group 1 model.

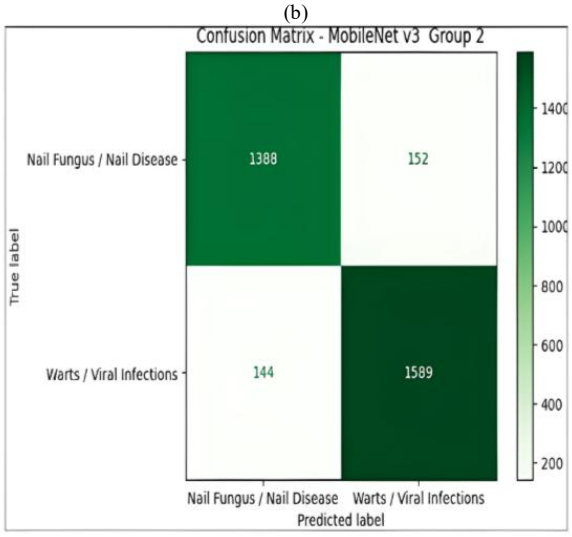
The experimental results obtained for the two groups of skin conditions show distinct behaviours depending on the architecture tested. Xception delivers the highest performance, with stable learning curves, rapid convergence, and accuracy of around 98–99% in both groups. The confusion array, Fig. 4(c), reveals near-perfect separation across classes. The confusion arrays display an ultra-low error rate, which confirms the ability of Xception to adequately model intra-class variances, especially in visually similar skin lesions. This great performance may be partly explained by its structure, which is built on separable convolutions in the depth. This makes recognition easier and reduces the amount of computing needed.



(a)



(b)



(c)

Fig. 9. Metrics of accuracy, loss and confusion matrix of Inception-v3 group 2 model.

Also, Xception-V3 works well, properly detecting things 95% to 97% of the time. The loss and accuracy curves show that the training and validation sets are a good fit. The confusion matrices are a little more confusing than Xception, especially in classes where the textures are identical.

In addition, Inception-v3 is a great balance between price and performance. This makes it a good choice, especially when quick decisions are needed.

MobileNet-v3 is light and has an accuracy rate of 85 to 92 percent. It doesn't work as well as Xception or Inception-v3, but it uses a lot less computational resource. It is 85% to 92% accurate. It's not as powerful as Xception or Inception-v3, but it is more efficient in terms of computing, which means it takes less time to run and uses fewer GFLOPs. That makes it great for compact gadgets that can be carried around. But the confusion matrices indicate greater mistakes in classifying things, especially when the classifications are quite similar. This means that the simpler model isn't particularly good at finding intricate skin patterns.

Overall, the three topologies comparison shows that the depth of the network and the number of convolutional blocks influence its generalisation. The results surpass previous work applied to dermatological classification, notably that of [6],[9],[11],[15],[17],[18], which used more limited datasets or conventional models (VGG, ResNet, AlexNet) and reported generally lower accuracies (80–95The use of recent architectures such as Inception and Inception-v3, in association with a variety of data sets, significantly enhances the reliability of the system. This enhancement is also due to hyperparameter optimisation and strict pre-processing.

However, this study has limitations that reduce its generalizability and clinical realism. First, the experimental data is only from DermNet NZ, which are images under ideal and standard lighting and photography conditions, making the models less robust to the variety of real-world images taken by smartphones or in clinical settings. Second, the study is just a binary classification, where the diseases are classified into two groups, which is not enough to deal with the complexity of diagnosing different and similar skin diseases. Thirdly, although models such as Xception model performed well, they are not clinically validated by experts, which is required for medical use. Fourth, while Marwan's HPC cluster, no explainable AI model has been trained to enable clinicians to understand why the classifications were made. Finally, the study did not validate the models in real-world settings or with local clinic data, suggesting the need for future real-world validation.

VI. Conclusion

The results demonstrate that Xception is the most effective solution for dermatological classification, followed by Inception-v3 which offers a relevant compromise between accuracy and computational cost. Although MobileNet-V3presente lower accuracy than Xception and Inception-V3, its minimum computational cost makes it a relevant candidate for implementation on medical equipment with limited resources. In contrast, Xception consistently delivers the strongest diagnostic accuracy, reinforcing its relevance for clinical decision support systems.

Future research will involve incorporating visual transformers to detect overall dependencies in dermatological images and using explainable AI methods (Grad-CAM++, SHAP) for better clinical interpretability. The exploration of hybrid CNN-Transformer models, the reduction of computational overhead through pruning and quantisation, and validation by dermatologists are also promising avenues. Finally, optimising the model for execution on mobile devices or edge computing will enhance the clinical reach of the proposed solutions.

Acknowledgment

I am appreciative to the HPC Marwan cluster (Morocco) for providing the computing environment and processing capability that allowed me to train and evaluate on massive data sets. I also wish to acknowledge the primary-school educators in Essaouira and Safi for their patience and assistance in organizing and collecting handwriting samples from pupils. Their sacrifices made this research achievable.

References

[1] Amdad Hossain Roky, Mohammed Murshedul Islam, Abu Mohammed Fuad Ahasan, Md Saqline Mostaq, Md Zihad Mahmud, Mohammad Nurul Amin, Md Ashiq Mahmud,Overview of skin cancer types and prevalence rates across continents,Cancer Pathogenesis and Therapy,Volume 3, Issue 2,2025,Pages 89-100,ISSN 2949-7132,https://doi.org/10.1016/j.cpt.2024.08.002..

[2] M. Y. Wang, X. H. Gao, and L. Zhang, “Recent global patterns in skin cancer incidence, mortality, and prevalence,” Chinese Medical Journal, vol. 138, no. 2, pp. 185–192, 2025, doi: 10.1097/CM9.0000000000003416.

[3] M. Janda, C. M. Olsen, V. J. Mar, and A. E. Cust, “Early detection of skin cancer in Australia—current approaches and new opportunities,” Public Health Research & Practice, vol. 32, no. 1, e3212204, 2022, doi: 10.17061/phrp3212204

[4] K. Chen, J. Zhou, X. Wang, and Y. Liu, “Global burden of skin cancer and its subtypes,” Frontiers in Public Health, vol. 13, 1610661, 2025, doi: 10.3389/fpubh.2025.1610661.

[5] N. M. Mahmoud and A. M. Soliman,“Early automated detection system for skin cancer diagnosis using artificial intelligent techniques,”Scientific Reports, vol. 14, Art. no. 9749, 2024,doi: 10.1038/s41598-024-59783-0.

[6] G. Yang, S. Luo, and P. Greer,“Advancements in skin cancer classification: a review of machine learning techniques in clinical image analysis,”Multimedia Tools and Applications, 2024,doi: 10.1007/s11042-024-19298-2.

[7] A. R. Barron, “The roots and evolution of machine learning,” Communications of the ACM, vol. 64, no. 6, pp. 58–65, 2021, doi: 10.1145/3448300.

[8] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning and artificial neural networks: Progress and future directions,” Nature Reviews Computer Science, vol. 2, pp. 18–31, 2023, doi: 10.1038/s43588-023-00431-0.

[9] A. Khan, M. Imran, and H. Kim, “Deep learning for computer vision: A brief review,” IEEE Access, vol. 8, pp. 212-376–212-397, 2020, doi: 10.1109/ACCESS.2020.3036766.

[10] Montaha S, Azam S, Rafid AKMRH, Islam S, Ghosh P, Jonkman M (2022) A shallow deep learning approach to classify skin cancer using down-scaling method to minimize time and space complexity. PLoS ONE 17(8): e0269826. https://doi.org/10.1371/journal.pone.0269826

[11] He, X., Wang, S., Shi, S., Tang, Z., Wang, Y., Zhao, Z., … Chu, X. “Computer-Aided Clinical Skin Disease Diagnosis Using CNN and Object Detection Models,” in 2019 IEEE International Conference on Big Data (BigData), Los Angeles, Dec. 9-12, 2019, pp. 4839-4844.DOI: 10.1109/BigData47090.2019.9006528.

[12] Fatima, S., Akram, M.U., Mohammad, S. et al. Deep learning in dermatopathology: applications for skin disease diagnosis and classification. Discov Appl Sci 7, 1006 (2025). https://doi.org/10.1007/s42452-025-07138-3

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25 (NIPS 2012), pp. 1097–1105, 2012. doi: 10.1145/3065386.

[14] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90

[15] Ray A, Gupta A, Al A Skin Lesion Classification With Deep Convolutional Neural Network: Process Development and Validation JMIR Dermatol 2020;3(1):e18438doi: 10.2196/18438

[16] Mutawa AM, Allaho MY, Al-Hajeri M. Machine Learning Approach for Arabic Handwritten Recognition. Applied Sciences. 2024; 14(19):9020. https://doi.org/10.3390/app14199020

[17] Kassem MA, Hosny KM, Damaševičius R, Eltoukhy MM. Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review. Diagnostics. 2021; 11(8):1390. https://doi.org/10.3390/diagnostics11081390

[18] N. Nigar, M. Umar, M. K. Shahzad, S. Islam and D. Abalo, "A Deep Learning Approach Based on Explainable Artificial Intelligence for Skin Lesion Classification," in IEEE Access, vol. 10, pp. 113715-113725, 2022, doi: 10.1109/ACCESS.2022.3217217

[19] Venugopal, V., Raj, N. I., Nath, M. K., & Stephen, N. (2023). A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. Decision Analytics Journal, 8, 100278. https://doi.org/10.1016/j.dajour.2023.100278

[20] M. S. Akter, H. Shahriar, S. Sneha and A. Cuzzocrea, "Multi-class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 5404-5413, doi: 10.1109/BigData55660.2022.10020302.

[21] Yao, Chengdong. (2023). A Comprehensive Evaluation Study on Risk Level Classification of Melanoma by Computer Vision on ISIC 2016-2020 Datasets. 10.48550/arXiv.2302.09528.

[22] K, Kanchana et al. "Enhancing Skin Cancer Classification using Efficient Net B0-B7 through Convolutional Neural Networks and Transfer Learning with Patient-Specific Data." Asian Pacific journal of cancer prevention : APJCP vol. 25,5 1795-1802. 1 May. 2024, doi:10.31557/APJCP.2024.25.5.1795

[23] De, Anubhav et al. "An approach to the dermatological classification of histopathological skin images using a hybridized CNN-DenseNet model." PeerJ. Computer science vol. 10 e1884. 26 Feb. 2024, doi:10.7717/peerj-cs.1884

[24] Al-Waisy, A.S., Al-Fahdawi, S., Khalaf, M.I. et al. A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. Sci Rep 15, 31234 (2025). https://doi.org/10.1038/s41598-025-15655-9

[25] Tai, C.-e.A.; Janes, E.; Czarnecki, C.; Wong, A. Double-Condensing Attention Condenser: Leveraging Attention in Deep Learning to Detect Skin Cancer from Skin Lesion Images. Sensors 2024, 24, 7231. https://doi.org/10.3390/s24227231

[26] Li, Q., Li, H. & Meng, L. A generic deep learning architecture optimization method for edge device based on start-up latency reduction. J Real-Time Image Proc 21, 116 (2024). https://doi.org/10.1007/s11554-024-01496-8