iJOE

International Journal of
Online and Biomedical Engineering

PAPER

Application of Convolutional Neural Networks in Skin Disease Prediction: Accuracy and Efficiency in Dermatological Image Analysis

Orlando Iparraguirre-Villanueva[1](✉), Michael Cabanillas-Carbonell[2]

[1]Facultad de Ingeniería, Universidad Tecnológica del Perú, Lima, Perú

[2]Facultad de Ingeniería, Universidad Privada del Norte, Lima, Perú

oiparraguirre@ieee.org

ABSTRACT

The use of dermatological images and convolutional neural networks (CNNs) to predict skin diseases is one of the most promising applications of data science to improve the diagnosis and treatment of skin diseases. The aim of this work was to achieve maximum accuracy and efficiency in skin disease prediction using dermatological images and CNN models. Based on dermatological images, the ability of five CNN models to predict skin diseases was evaluated. The ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet models were evaluated using the Kaggle HAM10000 (human against machine with 10000 training images) dataset. The metrics used were accuracy, recall, and F1 score. As a result, the study found that skin disease classification has variable performance. VGG-19 and DenseNet201 showed high values for accuracy, recall, and F1 score, with accuracy close to 98%. These models demonstrated an effective ability to identify and classify different types of skin diseases. In contrast, ResNet50 and Inception V3 obtained mixed results, while EfficientNet showed variable results in predicting skin diseases from dermatological images. Finally, the importance of choosing the right CNN model to predict skin diseases from dermatological images can be highlighted. VGG-19 and DenseNet201 performed well in classifying various skin diseases, which could be useful for developing dermatological diagnostic support systems.

KEYWORDS

prediction, disease, skin, imaging

1   INTRODUCTION

Dermatology is one of the many branches of medicine that specialize in the study, diagnosis, and treatment of skin diseases [1]. Skin diseases can range from mild to severe in terms of severity and origin [2], and can include conditions

such as dermatitis, acne, vascular lesions, melanoma, psoriasis, and skin cancer, among others [3]. Because of these diseases, patients may have a lower quality of life and even face a life-threatening situation in some cases [4]. Currently, the diagnosis of skin diseases depends largely on the physician's experience and clinical judgment [5], which can lead to diagnostic errors and delayed treatment [6], [7]. In addition, many parts of the world have limited access to dermatology specialists, which can further delay proper diagnosis and treatment. [8] The World Health Organization (WHO) claims that more than 900 million people worldwide suffer from skin diseases, making them one of the most common medical conditions among humans. It is estimated that more than 80% of all skin diseases are caused by five common conditions: psoriasis, atopic dermatitis, epidermolysis bullosa, melanoma, and vascular lesions [5]. For example, in countries of South and Central America, the number of deaths due to skin diseases in the year 2022 is Brazil, with 6,796, Mexico with 3,147, Argentina with 2,526, Colombia, 1,418, Chile with 803, Peru with 525, Ecuador with 311, Uruguay with 201, Nicaragua with 150, Guatemala with 140 and Costa Rica with 97 [9], as shown in Figure 1.

Fig. 1. Number of deaths due to skin diseases in South and Central American countries, 2022

In this context, computer science has made great progress in the field of dermatological image analysis with the use of innovative technologies such as convolutional neural networks (CNNs) [10], which have proven to be a very useful and practical tool in the prediction of skin diseases [11]. CNNs are artificial neural networks often used in image analysis. They have demonstrated high efficiency in pattern identification and image classification. [12], [13] In addition, CNNs can also play an important role in improving clinical outcomes, including treatment outcomes and patients' quality of life, by enabling earlier and more accurate detection of skin diseases [14], [15]. In addition, this technique could also be used in the training of physicians and in research into new treatments and therapies for skin diseases [16], and can be an effective tool in both areas [17]. More broadly, this study can also lead to the development of new techniques and technologies in medicine. The application of CNN in dermatological image analysis may inspire new research and developments in

artificial intelligence (AI) and deep learning [18], and can be applied in other fields of medicine and healthcare [19], [20].

The development of this work has the potential to make a valuable contribution and expand diagnostic options in dermatological diseases, facilitate access to health services in remote or resource-limited areas, and promote scientific and technological advances in medicine. Therefore, this work aimed to achieve maximum accuracy and efficiency in skin disease prediction with dermatological images using CNN models. It also seeks to demonstrate the importance of AI in interpreting the decisions made by CNN models. The model with the highest accuracy and efficiency can be used as a tool for early skin disease prediction and help improve the quality of medical care in clinical dermatology.

Health agencies such as WHO/PAHO, academics, and scientists have published research papers on skin diseases [21]. For example, in [22] they evaluated image recognition models to differentiate whether the models act differently between dermatological diseases. They used open-source melanoma and carcinoma images. According to the study, images with light skin colors had a higher sensitivity and positive predictive value, and they concluded that, for the AI models to perform equally well, more images of dermatological diseases need to be collected in people with a darker skin color than in people with a lighter skin color. Similarly, in [23], they developed a system to anticipate skin diseases accurately. Following a feature extraction approach is suggested for diagnosing skin diseases. As a result, they obtained an accuracy rate of 91.88% for the image set. Also, the authors in [24] developed a proposal to effectively classify five types of psoriasis using deep learning techniques, such as CNN and LSTM. Images of normal skin and psoriasis were used, and they were processed and segmented before feature extraction: color, texture, and shape. The classification models were trained with 80% of the images, and the results were 84.2% for CNN and 72.3% for LSTM.

Also, in [25] the authors evaluated the Tibot application's pre-accuracy with AI in identifying skin diseases compared to a dermatologist. They used photographs of lesions from 600 patients, and AI predicted three probability-based diagnoses compared with those of a dermatologist. The results showed a high application accuracy in the anticipated diagnosis, with a sensitivity and specificity of 97% and 98%, respectively. Finally, they concluded that the app has great potential for practical applicability. Similarly, the authors in [26] developed a computerized method for classifying skin diseases using MobileNet V2 and deep learning. The efficient and accurate model uses a gray-level co-occurrence matrix to assess disease progression. The method outperforms other state-of-the-art models with more than 85% accuracy. designed a mobile application to identify diseases in the early stage, which helps dermatologists diagnose skin conditions efficiently. Similarly, the authors [27] comprehensively analyzed the efficacy of CNNs to diagnose Lyme disease from images. They used 23 CNN architectures, such as VGG and ResNet50. The ResNet50 model achieved the best classification accuracy, with a 0.8442 ± 0.136 and a sensitivity of 0.8793 ± 0.147. The study demonstrated that some lightweight CNNs can be used for mobile Lyme disease pre-screening applications.

In addition, in [28], a deep learning and collective intelligence approach using CNNs was proposed to classify skin lesions based on deep learning techniques. Where they used CNNs such as AlexNet, GoogLeNet, among other models, then they analyzed the performance of each of the CNNs to evaluate their performance

and build a collective intelligence network system combined with the individual neural network. Furthermore, in [29], they built a system based on deep learning and trained it with a dataset representing the real context of a hospital in China. They used a dataset of 13 603 dermatologic images labeled by dermatologists and classified them into 14 classes. Based on the study results, it could be concluded that the approach achieved a high accuracy and sensitivity level of 94.8% when classifying the data. Finally, they concluded that the proposed framework can accurately classify the most common dermatoses. Also, in [30], they developed a two-level double hybrid two-hybrid CNN (2-HDCNNNN) function fusion approach for predicting malignant melanoma. Different classifiers used different models' classifiers, such as support vector machines and multilayer perceptron's. The results showed that the presented feature fusion approach achieved an accuracy of 0.9215, a specificity of 0.968, and a sensitivity of 0.8648 for diagnosing malignant melanoma. Also, in [31], they presented a framework for automated melanoma diagnosis, aiming to improve the performance of existing systems. In the present study, the team used the EfficientNetB5 model as the backbone of the networks to cluster pigmented lesions, using the EfficientNetB5 model to evaluate the dataset that included 8917 lesions. According to the results, the average (AUROC) for melanoma, 96% for nevus, and 97% for benign keratosis reached an average (AUROC) of 93%. Our results show that the framework provides an effective and explainable framework for making decisions.

This work is organized as follows: Section 2 discusses the working method and the architecture of the CNN models for training, and the case study is also developed. Section 3 presents the results obtained from the training of the models. Section 4 discusses the results of previous studies related to the subject and compares them with the results obtained in this study. Finally, Section 5 presents the conclusions.

2　MATERIALS AND METHODS

This section briefly constructs the terminology of the CNN models used in this work. It then develops the proposed case study, which seeks to achieve maximum accuracy in skin disease prediction with dermatological imaging.

Convolutional neural network is one of the most specialized types of neural networks, designed to process structured data, such as audio signals and images, to detect patterns [32]. It is well known that these algorithms are very effective at detecting and extracting key features from this type of data and are used in a wide range of applications in the field of AI [33]. In this work, five CNN models are used to retrain the data, which are described below.

2.1　ResNet50

ResNet50 is a CNN that uses residual connections to solve the gradient fading problem by using residual connections [34]. Its modular architecture consists of several convolutional and clustering layers, activation layers, and finally, a final classification layer [35]. It is very effective in classifying images, achieving high accuracy even on large and complex datasets such as ImageNet [36], which has a lot of data to classify. Due to its scalable and adaptive design, it can be used to perform

other computer vision tasks, such as object detection and image segmentation [37]. Expressly, the building blocks are described in equation (1).

$$y = F(x, w_i) + x \qquad (1)$$

Equation (1) describes a residual mapping function to be learned on a group of layers, where x is the input y, y is the output of that group of layers, and $F(x, w_i)$ is the mathematical representation of the residual mapping.

2.2   Inception V3

Inception V3 is an advanced CNN model that uses parallel network modules to extract features of various scales and abstraction levels using the power of CNNs [38]. As presented in Figure 2, the model architecture is based on convolution filters of different sizes and depths, allowing it to extract detailed and accurate features from images. It also uses regularization techniques to improve the generalization of the model. Inception V3 has demonstrated an accuracy rate of over 95% on image classification and object detection tasks with large datasets such as ImageNet [39].
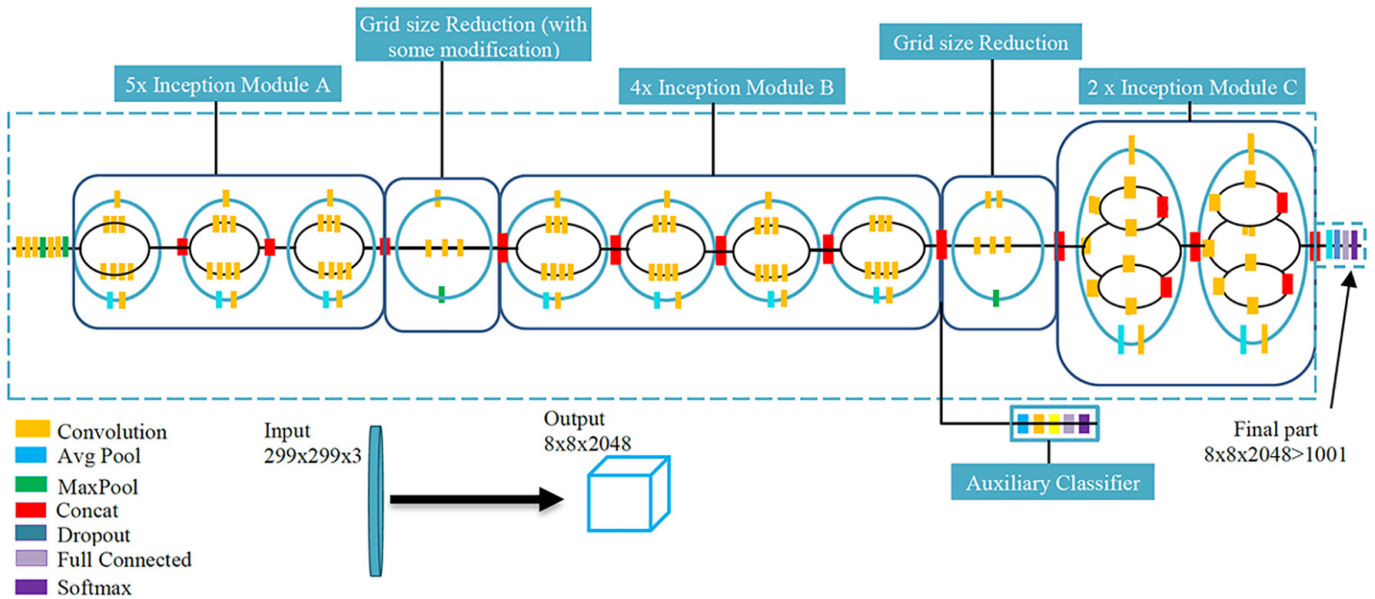

Fig. 2. Architecture of Inception V3

2.3   VGG-19

VGG-19 is a deep CNN model with 19 convolutional layers, followed by three fully connected layers, which are based on convolution and pooling [40]. Figure 3 shows its architecture, which is characterized by small convolutional filters with (3×3) layers and small pooling layers with a step length of (2×2). This allows the network to extract detailed features from the images, thus reducing its size [41]. When tested, the VGG-19 classifier has proven extremely effective in image classification tasks, achieving over 90% accuracy on the ImageNet dataset.
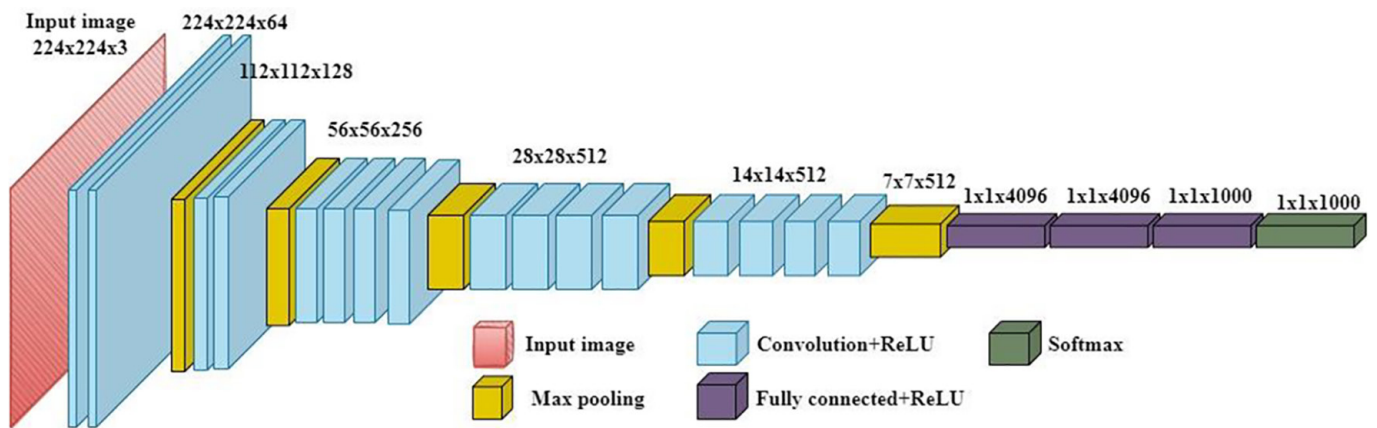
Fig. 3. Representation of the VGG-19 architecture

2.4   DenseNet201

DenseNet201 is a deep CNN model characterized by dense connections between its layers. The input for each layer of DenseNet201 consists of the concatenation of the outputs of all the layers before it [42]. Information can now flow freely between layers, eliminating gradient fading and allowing information to flow freely between layers [43]. With DenseNet201, you can extract accurate and detailed features from images by using densely folded layer blocks as a method for processing images [44]. It has been shown that the model is very efficient in image classification tasks and can achieve more than 95% accuracy on the ImageNet dataset when used. Figure 4 shows the architecture of DenseNet201.
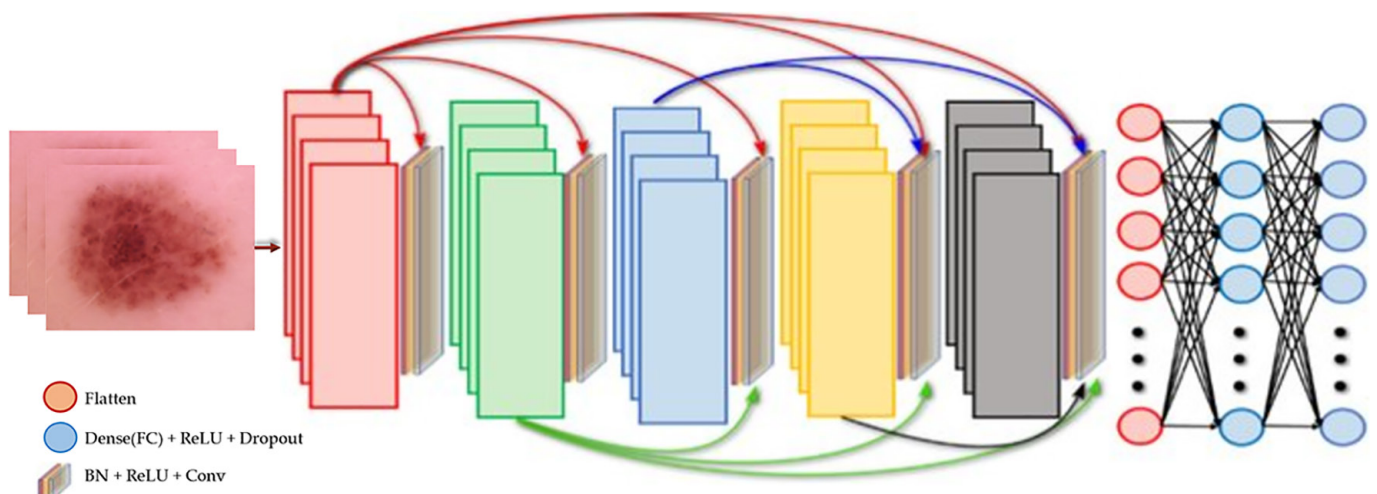

Fig. 4. Representation of the DenseNet201 architecture

2.5   EfficientNet

EfficientNet is a CNN model designed to achieve an optimal balance between accuracy and computational efficiency. It uses a composite scaling strategy to optimize the network architecture according to its size. This allows the model to adapt

to different image sizes and datasets efficiently [45]. The architecture of EfficientNet, as presented in Figure 5, is based on blocks of folded layers that allow accurate and detailed features to be extracted from images [46].
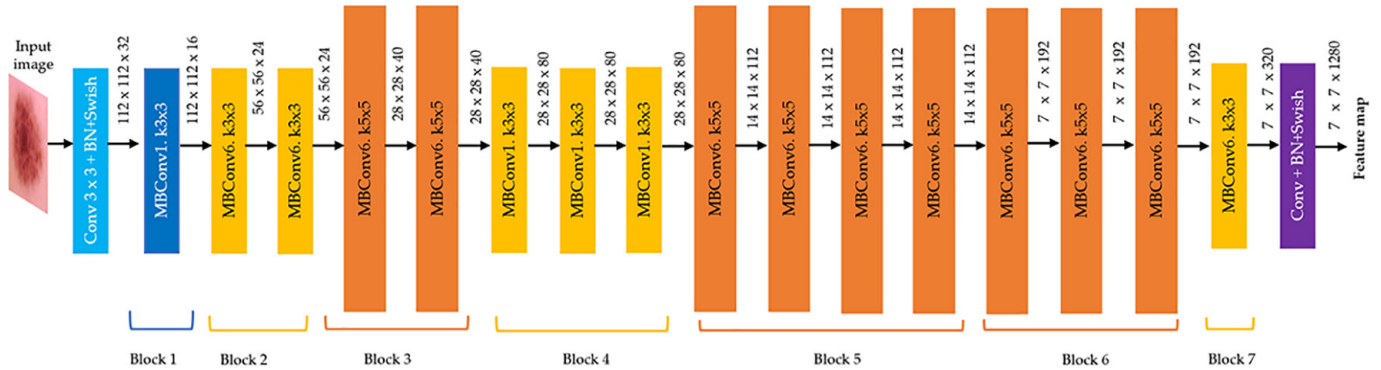

**Fig. 5.** Representation of the EfficientNet architecture

2.6 Case study

In this section, an explanation of the object of study is presented. An early and accurate prediction can help improve skin disease diagnosis and treatment. With the advancement of AI and deep learning, multiple techniques have been developed for disease classification using dermatological images. In this case study, CNNs are used, which is considered a powerful tool for skin disease classification from dermatological images. For this work, the HAM10000 (human against machine with 10000 training images) dataset is considered one of the largest dermatological image databases, comprising more than 10,000 high-resolution images of skin lesions. This work uses five CNNs (ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet) for the classification of seven types of skin diseases: melanocytic nevi (MN), melanoma (MEL), benign keratosis-like lesions (KBL), basal cell carcinoma (BCC), actinic keratoses (AK), vascular lesions (VASC), and dermatofibroma (DF). The objective is to achieve the highest accuracy in skin disease prediction with dermatological imaging and CNN models, evaluating the accuracy and efficiency of each model. The model that results in the highest accuracy and efficiency can be used as a tool for the early prediction of skin diseases and help improve the quality of medical care in clinical dermatology. The following is the development of each phase of applying CNNs in skin disease prediction.

Loading and cleaning of the data set. Initially, the PIL library was used to load the dataset (HAM10000), which provides wide compatibility with image formats and allows quick image access. After loading the images, the seven predictor variables and labels were assigned to each image (MN, MEL, BKL, BCC, AK, VASC, and DF), as shown in Table 1. Then, we verified that there would be no missing or null data and eliminated them if necessary. The next step is to resize the images because the original dimensions (450 * 600 * 3) take a long time to process with CNNs. This step is necessary because the images in the dataset have different resolutions and dimensions. The OpenCV library in Python was used to resize to a single size. In the next step, an exploratory analysis of the images is performed to understand the distribution of the classes and visualize the images.

Table 1. Characteristics of the data set

	Lesión	Image_id	dx_type	Age	Sex	Localization
0	HAM0000118	ISIC0027419	histo	75	man	scalp
1	HAM0000118	ISIC0025030	histo	65	man	scalp
2	HAM0002730	ISIC0026769	histo	80	man	scalp
3	HAM0002730	ISIC0025661	histo	75	female	scalp
4	HAM0001466	ISIC0031633	histo	80	man	ear
5	HAM0001466	ISIC0027860	histo	80	man	ear
6	HAM0002761	ISIC0029175	histo	70	female	face
7	HAM0002761	ISIC0029067	histo	70	male	face
8	HAM0005132	ISIC0025838	histo	68	female	back

Exploratory data analysis. Exploratory data analysis (EDA) is a crucial part of understanding the data set better; in this case, EDA helps to understand the distribution of classes, the characteristics and patterns of the images, and the correlation between variables. In addition, EDA allows visualization of the data, as shown in Figure 6, which shows some dermatological images resized to (125 and 100) at a specific size for easier manipulation. Also, descriptive statistics and the identification of outliers, for which bar graphs are used, as depicted in Figure 6.
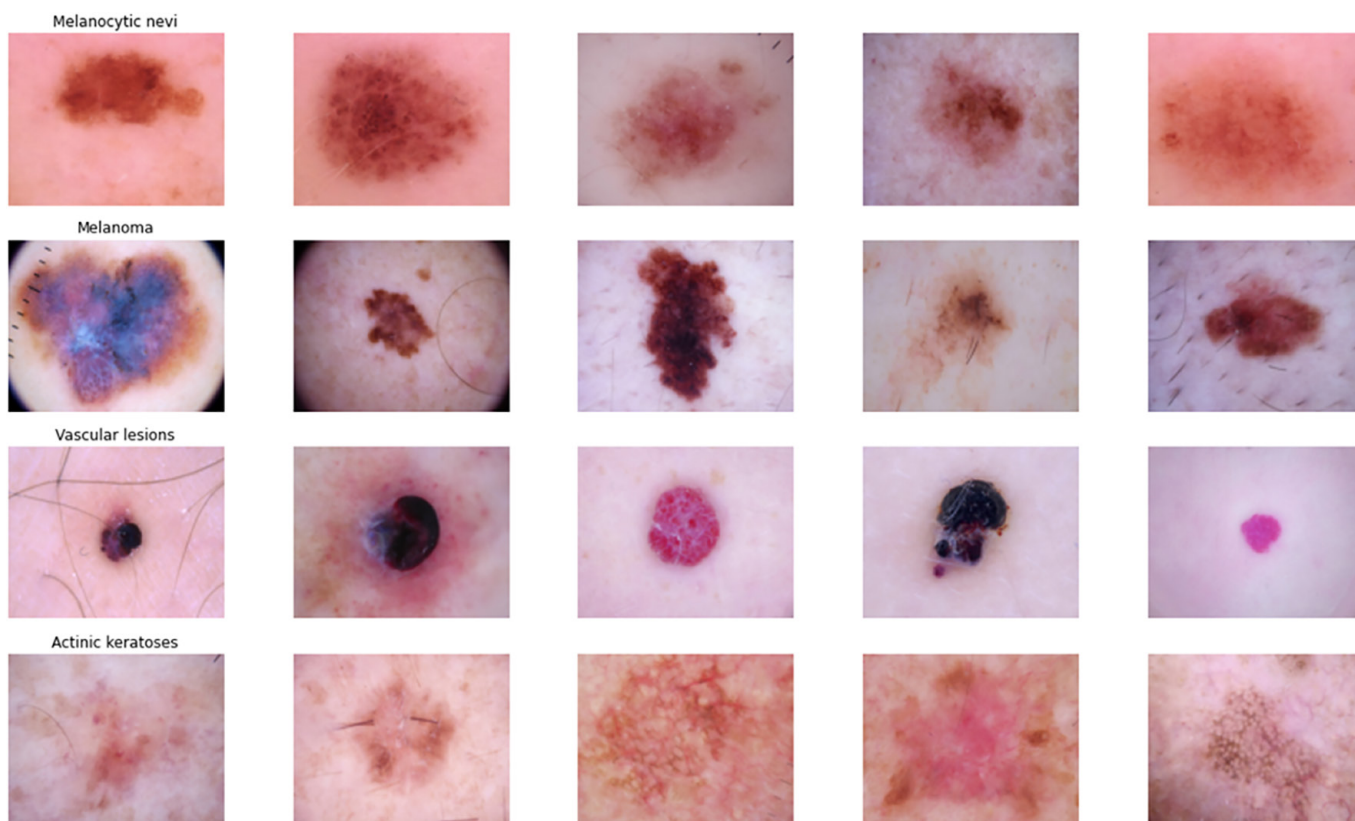
Fig. 6. Skin images resized to (125, 100)

Analyzing the data set, Figure 7 shows that people between 45 and 50 are more prone to skin diseases. In people under 10 years of age, skin diseases are minimal.

It is also observed that the probability of a skin disease increases as age increases. It is also observed that skin diseases are more frequent in men than in women. Skin diseases can occur anywhere on the body but are more visible on the "back" of the body and less on the extremities, fingers, or ears. The most frequent disease is MN, while the least frequent is DF. Also, the data set is classified by type of skin disease. For example, MN represents 69.9%, MEL 11.1%, BKL 11.0%, BCC 5.1%, AK 3.3%, VASC 1.4%, and DF 1.1%. The analysis also identified how the skin disease was discovered. For example, through histopathology, 53.3% were discovered; through follow-up examination, 37% were discovered; through expert consensus, 9% of the diseases were discovered; through confirmation by in vivo confocal microscopy, 0.7%.
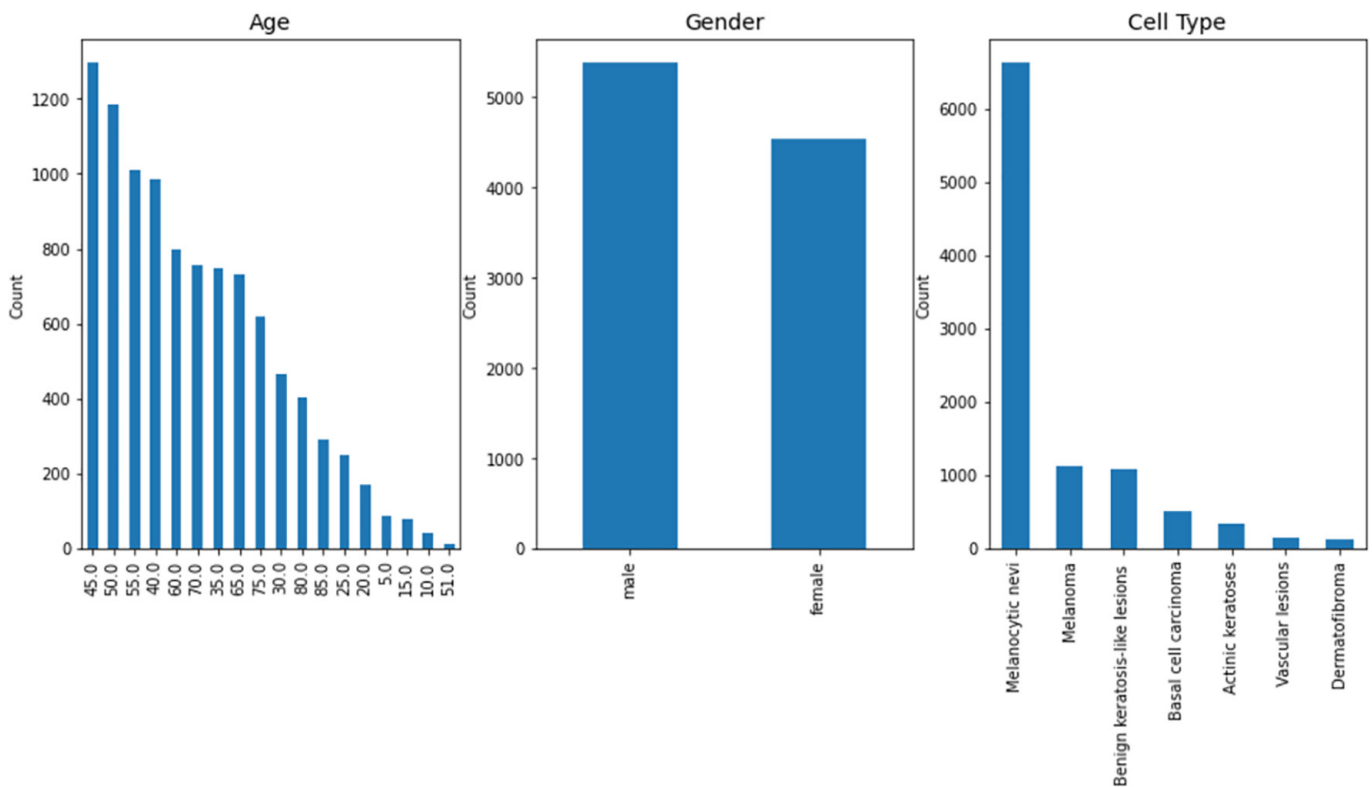

Fig. 7. Univariate analysis of the dataset (age, gender, and cell type)

Bivariate analysis allows for a statistical analysis of the data set to determine the relationship between the variables. For example, in which areas of the body there are more lesions per type of disease, as shown in Figure 8. Infection in the body's lower extremities is more visible in women. Some unknown regions also show infections, which are visible in both men and women. Also, acral surfaces show the least cases of infection, and only in males. The face is the area most affected by benign keratotic lesions. Also, melanocytic nevi infect more body parts (except the face). The age group from zero to 75 years is the most affected by melanocytic nevi. On the other hand, people aged 80–90 years are more affected by benign keratotic lesions. The most common disease affecting the scalp is benign keratotic lesions. Melanocytic nevi are a disease that mostly affects people in the back area. As evidenced in Figures 7 and 8, EDA helps to identify problems in low-quality images and class imbalance that may affect the accuracy of the models. Also, it allows for selecting relevant features and discarding irrelevant or redundant ones, which can help improve model efficiency.
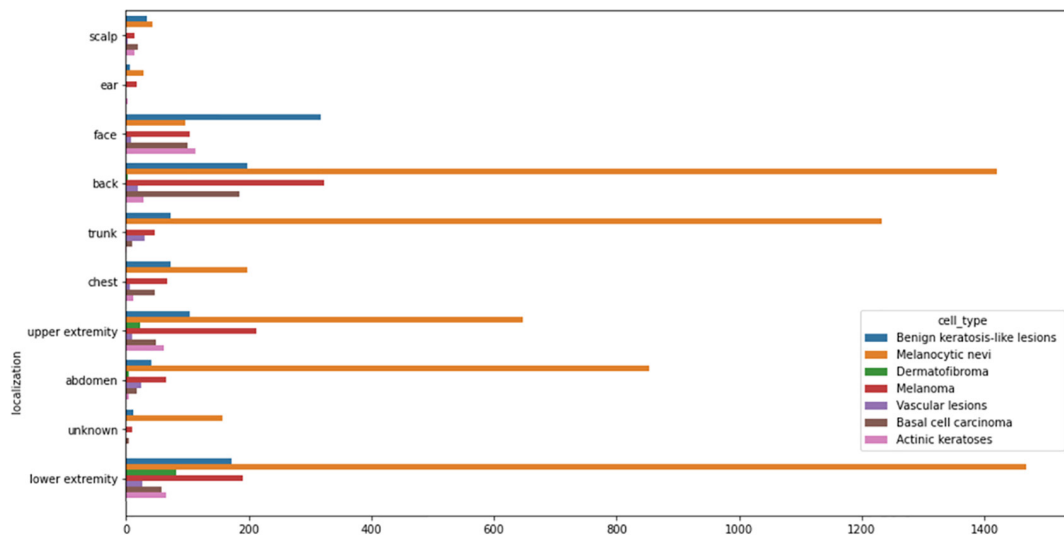
Fig. 8. Localization of skin disease vs. cell type

Processing and standardization. It is very important to remember that before feeding the images to the CNN models, it is necessary to preprocess them. For this purpose, normalization is used as a technique to standardize the pixel values of the images and improve the convergence of the model to scale the numerical values of the variables so that they are in a common and comparable range, as seen in Table 2. The data set must be divided into training, validation, and testing to proceed with the next step. As part of the training process, we are going to create a training set for fitting the parameters of the model, we are going to create a validation set for adjusting the hyperparameters and preventing overfitting, and we are going to use the test set for the final evaluation of the model. In this case, we will use a ratio of 60%, 20%, and 20% for the training, validation, and test sets. Now, Python libraries such as Keras are used to implement CNN models. This work uses five retrained models: ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet. These algorithms have been demonstrating high performance in image classification tasks. Next, the hyperparameters are tuned using cross-validation. Cross-validation is a technique to evaluate model performance and avoid overfitting.

Table 2. Scaling of variable number values

array ([[	
[0.75686275, 0.55686275, 0.63137255],	[0.81176471, 0.60784314, 0.62352941],
[0.76078431, 0.56078431, 0.64705882],	[0.79215686, 0.56078431, 0.56862745],
[0.76862745, 0.57254902, 0.64705882],	[0.78823529, 0.57647059, 0.58039216]
],[	
[0.75294118, 0.55294118, 0.61176471],	[0.81568627, 0.61960784, 0.63137255],
[0.78039216, 0.58039216, 0.64705882],	[0.80784314, 0.61176471, 0.61960784],
[0.76470588, 0.56470588, 0.63529412],	[0.80392157, 0.61960784, 0.63137255]
],[	
[0.75294118, 0.54117647, 0.58823529],	[0.83137255, 0.63921569, 0.65882353],
[0.76470588, 0.56862745, 0.62352941],	[0.81568627, 0.62745098, 0.63921569],
[0.78431373, 0.58039216, 0.64313725],	[0.81960784, 0.63921569, 0.66666667],
	]])

Training and evaluation of models. In this phase, we proceed with the training and validation of the models. To do this, we first proceeded with the training, validation, and test dataset division. In this paper, the training set is used as the basis for training the model (Inception V3, ResNet50, DenseNet201, EfficientNet), the validation set is used as the basis for tuning the hyperparameters and performing the cross-validation, and the test set is used as the basis for testing the model. Model performance. The next step consists of creating the data generators responsible for loading the data into the batches during training. The loss function was minimized using optimization algorithms such as stochastic gradient descent, and the model weights were adjusted based on the results of this phase. The model is then compiled, specifically the loss function, the optimization algorithm, and the evaluation metrics. Finally, the models are trained using the training data. There will be several model iterations in which the data batches will be passed to the model, the loss will be calculated, and the model weights will be updated, as shown in Table 3.

Table 3. Training example of the Inception V3 model

Layer (Type)	Param#	Connected to
input_1 (Input Layer)	0	
Layer_conv2d (Conv2D)	4507392	mixed10 [0][0]
average_pooling2d (AvgPool)	0	conv2d [0][0]
Layer_max_pooling2d (MaxPooling2D)	0	average_pooling2d [0][0]
Layer_flatten (Flatten)	0	max_pooling2d [0][0]
Layer_dense (Dense)	1052672	flatten [0][0]
Layer_dropout (Dropout)	0	dense [0][0]
Layer_dense_1 (Dense)	16781312	dropout [0][0]
Layer_dropout_1 (Dropout)	0	dense_1 [0][0]
Layer_dense_2 (Dense)	28679	dropout_1 [0][0]
SoftMax (SoftMax)	0	dense_2 [0][0]
Total: 21, 861, 055		
Trainable: 6, 270, 663		
Non-trainable: 15, 590, 392		

Table 3 shows the output of the Inception V3 model with the additional layers, such as the convolution layer, average pooling, and max pooling; these have different output shapes (none, 6, 6, 6, 256), (none, 3, 3, 3, 256), and (none, 1, 1, 256), respectively. After the flatten layer, the fully connected layer is applied with the output (none, 4096), and the last dense layer with shape (none, 7) represents the seven output classes, and a Softmax activation is applied to obtain the probabilities.

Regarding validation, it should be noted that, after each iteration, the performance of the models in the validation set is evaluated; this involves passing the validation data through the trained model and calculating the evaluation metrics, such as accuracy or error. In addition, at this stage, it is very important to adjust the

model's hyperparameters, such as the learning rate, the batch size, or the number of epochs, to improve the performance of the models. After completing all the training epochs and adjusting the hyperparameters, we evaluate the final model performance on the test set, as presented in Figure 9. This allows us to estimate the model performance with previously unseen data.
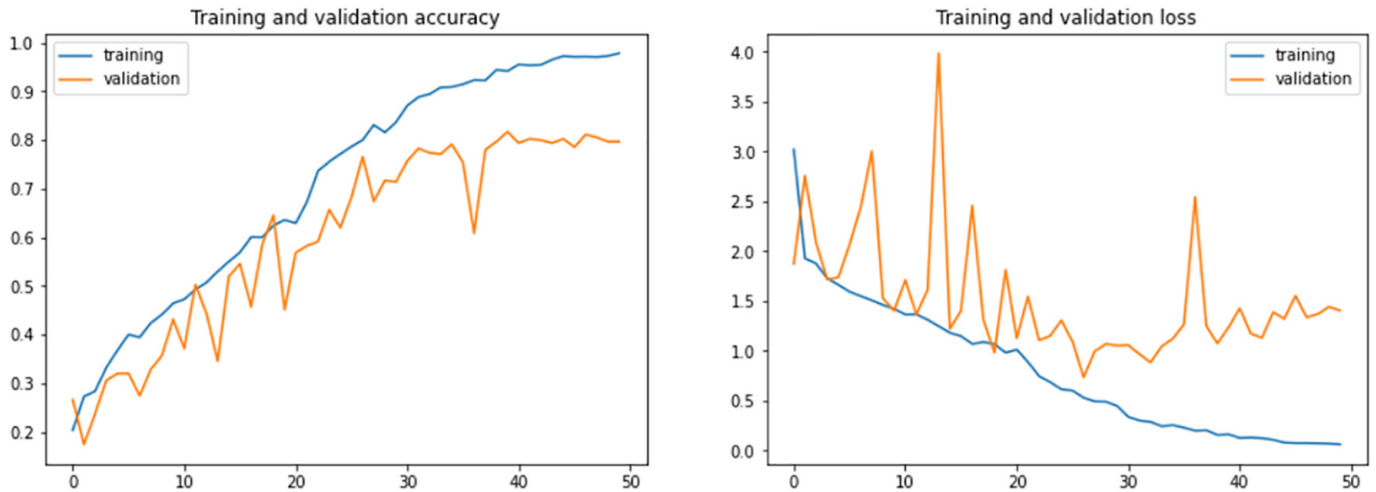

Fig. 9. Inception V3 model performance during training and validation

Figure 9 shows how the accuracy of the Inception V3 model improves as it iterates through the epochs. For example, the training accuracy curve shows an upward trend and reaches a high accuracy near 100% at the end of the epochs. This indicates that the model is learning the training data patterns well and improving its performance as it trains more. It is also seen that the validation accuracy curve follows a similar trend to that of in training up to epoch 35, then has a decrease and goes up again and stays at 80% accuracy.

On the other hand, training and loss validation show how the model loss changes as training is performed. The training curve shows a downward trend and decreases as the epochs progress; this indicates that the model is increasingly adjusting to the training data and improving its ability to minimize error. While the validation loss curve shows that it does not have a continuous downward trend, it is observed that it has ups and downs; this reflects that throughout the epochs, the model is not reducing the error in the data optimally.

3   RESULTS

This section presents the training results of the models (ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet), for which the HAM10000 dataset obtained from Kaggle was used. Training was performed for each model, considering each type of skin disease. Metrics such as accuracy, recall, and F1 score were used. The training results of the models are presented in Table 4.

Table 4. Model training results

	Precision	Recall	F1-Score	Support
ResNet50				
AK	0.7045	0.6793	0.7678	50
BCC	0.8509	0.8515	0.8429	60
BKL	0.7152	0.7512	0.9036	50
DF	0.6107	0.8341	0.7892	55
MN	0.9234	0.6027	0.8174	48
MEL	0.7661	0.8534	0.8665	35
Vascular skin lesions	0.8003	0.8432	0.8351	52
Accuracy			0.7824	350
Macro	0.7724	0.7668	0.8284	350
Weighted	0.7676	0.7686	0.8109	350
Inception V3				
AK	0.6852	0.7400	0.7115	50
BCC	0.8333	0.6667	0.7407	60
BKL	0.6909	0.7600	0.7238	50
DF	0.9818	0.9818	0.9818	55
MN	0.7750	0.6458	0.7045	48
MEL	0.6222	0.8000	0.7000	35
Vascular skin lesions	0.9623	0.9808	0.9714	52
Accuracy			0.7971	350
Macro	0.7930	0.7964	0.7906	350
Weighted	0.8052	0.7971	0.7973	350
VGG-19				
AK	0.8273	0.9598	0.8943	50
BCC	0.9452	0.8706	0.8971	60
BKL	0.9264	0.9684	0.9267	50
DF	0.8168	0.9379	0.8614	55
MN	0.8387	0.8222	0.8719	48
MEL	0.8935	0.8757	0.8773	35
Vascular skin lesions	0.8719	0.9051	0.9275	52
Accuracy			0.9823	350
Macro	0.8773	0.9001	0.8885	350
Weighted	0.8961	0.8991	0.8899	350

(Continued)

Table 4. Model training results *(Continued)*

	Precision	Recall	F1-Score	Support
DenseNet201				
AK	0.8985	0.9624	0.8934	50
BCC	0.9001	0.9359	0.9581	60
BKL	0.9493	0.9049	0.9325	50
DF	0.9338	0.9101	0.8867	55
MN	0.9304	0.8575	0.8693	48
MEL	0.8957	0.8827	0.9256	35
Vascular skin lesions	0.8786	0.9322	0.8579	52
Accuracy			0.9834	350
Macro	0.9092	0.9141	0.9038	350
Weighted	0.9093	0.9139	0.8965	350
EfficientNet				
AK	0.7643	0.8897	0.7813	50
BCC	0.8721	0.8496	0.8602	60
BKL	0.9067	0.7143	0.6905	50
DF	0.7589	0.8362	0.9351	55
MN	0.8142	0.7338	0.7317	48
MEL	0.6825	0.9269	0.8128	35
Vascular skin lesions	0.9263	0.7874	0.7549	52
Accuracy			0.6973	350
Macro	0.8101	0.8093	0.7914	350
Weighted	0.8005	0.7984	0.7858	350

Table 4 shows that the ResNet50 model showed variable performance for each class, with precision, recall, and F1-score values varying between types. The results are accurate at approximately 78.24%, indicating that the samples were correctly classified. 81.09% average F1-score was achieved. The Inception V3 model also shows variable performance for each class, with differing precision, recall, and F1 score values. The accuracy obtained is approximately 79.71%, indicating that the samples were classified correctly. As for the weighted average F1 score, a value of approximately 79.73% is obtained. The VGG-19 model performs well in terms of accuracy, recall, and F1 score for all classes.

The accuracy obtained is approximately 98.23%, indicating that the samples were correctly classified. In terms of the weighted average F1-score, a value of approximately 88.99% is obtained. Also, the DenseNet201 model performs well in terms of accuracy, recall, and F1-score for all classes. The accuracy obtained is approximately 98.34%, indicating that the samples were correctly classified. In terms of the weighted average F1-score, a value of approximately 89.65% is obtained. Similarly, the EfficientNet model shows variable performance in terms of accuracy, recall, and F1 score for each class. The accuracy obtained is approximately 69.73%, indicating that the samples were correctly classified.

In general terms, CNN models, such as ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet, are used to predict skin diseases. In this specific case in dermatological image processing, the VGG-19 and DenseNet201 models obtain the best results. Both models show high precision, recall, and F1-score values and high accuracy. DenseNet201 obtains a slight advantage between the two in the weighted average F1-score, with a value of 89.65%. They are closely followed by ResNet50. These models may be useful in skin disease diagnosis and classification applications. Therefore, it could be concluded that DenseNet201 performed better in dermatological image processing compared to the other mentioned models.

4 DISCUSSION

This work evaluated the performance of several CNN models (ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet) in skin disease prediction using dermatological image analysis. The objective was to achieve the highest accuracy in skin disease prediction with dermatological images. The results reveal significant differences in the performance of the different models evaluated. In general terms, it is observed that the VGG-19, DenseNet201, and EfficientNet models achieved superior accuracy and efficiency compared to ResNet50 and Inception V3.

The VGG-19 model demonstrated high accuracy and recall in all disease categories, indicating a robust ability to identify and pre-accurately classify different types of skin diseases. In addition, VGG-19 achieved a high F1 score in most categories, indicating a balance between accuracy and recall. This result is related to the results of [30], where they developed a CNN function approach to predict malignant melanoma, reaching an accuracy level of 92.15%, like the result obtained in this work with the VGG-19 model that reached an accuracy of 98.23%. DenseNet201 also showed outstanding accuracy with 98.34%, obtaining high F1-Score scores in most disease categories. This suggests that DenseNet201 can capture more complex relationships and obtain a deeper representation of dermatological image features. The results obtained in [28] proposed a dermatological image classification system using the DenseNet201 model, achieving significant results in model performance for the specific case. On the other hand, the EfficientNet model showed mixed results, with varying accuracy and recall in the different disease categories. Although EfficientNet achieved high accuracy in melanoma detection, its recall was relatively low, reaching only 69.73%. This indicates that, although EfficientNet can be efficient in classifying some diseases, it could have difficulties detecting certain cases of melanoma. This result differs from that achieved in [22], where they worked with melanoma images, and the results showed that images with light skin color obtained higher sensitivity and positive predictive value, concluding that more dermatological images of people with dark skin color are necessary since the results are very balanced. The ResNet50 and Inception V3 models obtained more modest results than those evaluated. Although they achieved reasonable accuracy in several categories, their recall and F1 scores were lower compared to VGG-19 and DenseNet201; this does not mean that the models are deficient; the performance of the models is subject to different criteria and data volumes, such as the case, for example, in the works [25] and [27] analyzed the efficacy of these CNN models to diagnose Lyme disease from dermatological images, obtaining an accuracy of 84.42%, which resembles that obtained in this work, a modest result for some types of skin disease. It is important to note that, although these models achieved promising results, there are still challenges to face in skin disease prediction using image analysis. For example, the early

and accurate detection of melanoma remains an area for improvement, especially in the case of EfficientNet.

Finally, this study highlights the importance of carefully selecting the CNN model for dermatological image-based skin disease prediction. Among the models trained in this context, we have VGG-19, DenseNet201, and EfficientNet, which showed the best balance between accuracy and recall.

5 CONCLUSIONS

In this work, we evaluated the performance of five CNN models in skin disease prediction based on dermatological image analysis. The models used were ResNet50, Inception V3, VGG-19, DenseNet201, and EfficientNet, and were trained and tested with the Kaggle HAM10000 dataset. Metrics such as accuracy, recall, and F1-score were used to evaluate each model's performance.

The results showed that the models classified skin diseases differently. The ResNet50 model obtained mixed results regarding accuracy, recall, and F1 score for each class. The accuracy obtained was 78.24%, indicating that the samples were classified correctly. The Inception V3 model also showed variable performance, with 79.71% accuracy. On the other hand, the VGG-19 and DenseNet201 models showed high performance in all classes, with high values for accuracy, recall, and F1-score. The accuracy obtained for VGG-19 was 98.23%, while for DenseNet201 it was 98.34%. These results indicate that about 98% of the samples were correctly classified by both models. These findings suggest that the VGG-19 and DenseNet201 models are more effective in dermatological image processing than the other models evaluated. These models could be useful in skin disease diagnosis and classification applications. However, it is important to note that skin disease prediction using image analysis still faces challenges. Early and accurate melanoma detection remains an area for improvement, especially with EfficientNet. Furthermore, selecting the appropriate model depends on the context and volume of data available.

For future research, combining multiple CNN architectures could be explored to leverage each individual's strengths and improve the accuracy and efficiency of skin disease prediction. This work provides promising evidence for applying CNNs in skin disease prediction from dermatological images. The VGG-19, DenseNet201, and EfficientNet models showed outstanding performance in terms of pre-accuracy and efficiency, supporting their usefulness in clinical practice. However, further research is required to address limitations and improve these networks' performance in real clinical settings.

6 REFERENCES

[1] A. M. Dessie, S. F. Feleke, S. G. Workie, T. G. Abebe, Y. M. Chanie, and A. K. Yalew, "Prevalence of skin disease and its associated factors among primary schoolchildren: A cross-sectional study from a Northern Ethiopian town," *Clin. Cosmet. Investig. Dermatol.*, vol. 15, pp. 791–801, 2022. https://doi.org/10.2147/CCID.S361051

[2] N. Prasitpuriprecha, S. Santaweesuk, P. Boonkert, and P. Chamnan, "Prevalence and DALYs of skin diseases in Ubonratchathani based on real-world national healthcare service data," *Scientific Reports*, vol. 12, 2022. https://doi.org/10.1038/s41598-022-20237-0

[3] Y. Xue *et al.*, "Global burden of bacterial skin diseases: A systematic analysis combined with sociodemographic index, 1990–2019," *Front. Med.*, vol. 9, 2022. https://doi.org/10.3389/fmed.2022.861115

[4] K. J. Pulsipher, M. D. Szeto, C. W. Rundle, C. L. Presley, M. R. Laughter, and R. P. Dellavalle, "Global burden of skin disease representation in the literature: Bibliometric analysis," *JMIR Dermatol.*, vol. 4, no. 2, p. e29282, 2021. https://doi.org/10.2196/29282

[5] R. J. Hay *et al.*, "The global burden of skin disease in 2010: An analysis of the prevalence and impact of skin conditions," *Journal of Investigative Dermatology*, vol. 134, no. 6, pp. 1527–1534, 2014. https://doi.org/10.1038/jid.2013.446

[6] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: Diagnosis to detection," *Multimed. Tools Appl.*, pp. 1–39, 2022. https://doi.org/10.1007/S11042-022-14305-W/FIGURES/3

[7] O. Iparraguirre-Villanueva, A. Epifanía-Huerta, C. Torres-Ceclén, J. Ruiz-Alvarado, and M. Cabanillas-Carbonell, "Breast cancer prediction using machine learning models," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023. https://doi.org/10.14569/IJACSA.2023.0140272

[8] W. Abbes *et al.*, "Fuzzy decision ontology for melanoma diagnosis using KNN classifier," *Multimed. Tools Appl.*, vol. 80, pp. 25517–25538, 2021. https://doi.org/10.1007/s11042-021-10858-4

[9] WHO Mortality Database, "Skin diseases," World Health Organization (WHO), 2022. [Online]. Available: https://platform.who.int/mortality/themes/theme-details/topics/topic-details/MDB/skin-diseases [Accessed: May 1, 2023].

[10] Y. Hong, G. Zhang, B. Wei, J. Cong, Y. Xu, and K. Zhang, "Weakly supervised semantic segmentation for skin cancer via CNN superpixel region response," *Multimed. Tools Appl.*, vol. 82, pp. 6829–6847, 2023. https://doi.org/10.1007/s11042-022-13606-4

[11] M. M. Mijwil, "Skin cancer disease images classification using deep learning solutions," *Multimed. Tools Appl.*, vol. 80, pp. 26255–26271, 2021. https://doi.org/10.1007/s11042-021-10952-7

[12] T. Diwan, R. Shukla, E. Ghuse, and J. V. Tembhurne, "Model hybridization & learning rate annealing for skin cancer detection," *Multimed. Tools Appl.*, vol. 82, pp. 2369–2392, 2023. https://doi.org/10.1007/s11042-022-12633-5

[13] R. Lohith, N. N. Govinda, K. Pruthvi, V. Janhavi, and H. L. Gururaj, "Facial skin disease detection using image processing," *International Journal of Bioinformatics and Intelligent Computing*, vol. 2, no. 1, pp. 1–11, 2023. https://doi.org/10.61797/ijbic.v2i1.207

[14] S. Inthiyaz *et al.*, "Skin disease detection using deep learning," *Advances in Engineering Software*, vol. 175, p. 103361, 2023. https://doi.org/10.1016/j.advengsoft.2022.103361

[15] M. Ahammed, M. Al Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthcare Analytics*, vol. 2, p. 100122, 2022. https://doi.org/10.1016/j.health.2022.100122

[16] M. C. Schielein *et al.*, "Outlier detection in dermatology: Performance of different convolutional neural networks for binary classification of inflammatory skin diseases," *Journal of the European Academy of Dermatology and Venereology*, vol. 37, no. 5, pp. 1071–1079, 2023. https://doi.org/10.1111/jdv.18853

[17] P. R. Kshirsagar, H. Manoharan, S. Shitharth, A. M. Alshareef, N. Albishry, and P. K. Balachandran, "Deep learning approaches for prognosis of automated skin disease," *Life*, vol. 12, no. 3, p. 426, 2022. https://doi.org/10.3390/life12030426

[18] A. Kalaivani and S. Karpagavalli, "Deep neural network optimization for skin disease classification forecast analysis," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2022, pp. 708–713. https://doi.org/10.1109/ICACCS54159.2022.9785004

[19] A. Jain *et al.,* "Multi-type skin diseases classification using OP-DNN based feature extraction approach," *Multimed. Tools Appl.*, vol. 81, pp. 6451–6476, 2022. https://doi.org/10.1007/s11042-021-11823-x

[20] S. Hossain, S. Umer, R. K. Rout, and M. Tanveer, "Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling," *Appl. Soft Comput.*, vol. 134, p. 109997, 2023. https://doi.org/10.1016/j.asoc.2023.109997

[21] I. A. Bratchenko, L. A. Bratchenko, Y. A. Khristoforova, A. A. Moryatov, S. V. Kozlov, and V. P. Zakharov, "Classification of skin cancer using convolutional neural networks analysis of Raman spectra," *Comput. Methods Programs Biomed.*, vol. 219, p. 106755, 2022. https://doi.org/10.1016/j.cmpb.2022.106755

[22] P. Aggarwal, "Performance of artificial intelligence imaging models in detecting dermatological manifestations in higher Fitzpatrick skin color classifications," *JMIR Dermatol.*, vol. 4, no. 2, p. e31697, 2021. https://doi.org/10.2196/31697

[23] A. Kalaivani, S. Karpagavalli, and K. Gulati, "Expert automated system for prediction of multi-type dermatology sicknesses using deep neural network feature extraction approach," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 3s, pp. 170–178, 2023. Accessed: Apr. 24, 2023. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/2557

[24] S. F. Aijaz, S. J. Khan, F. Azim, C. S. Shakeel, and U. Hassan, "Deep learning application for effective classification of different types of Psoriasis," *J. Healthc. Eng.*, vol. 2022, no. 1, pp. 1–12, 2022. https://doi.org/10.1155/2022/7541583

[25] S. S. Marri, A. C. Inamadar, A. B. Janagond, and W. Albadri, "Analyzing the predictability of an artificial intelligence app (Tibot) in the diagnosis of dermatological conditions: A cross-sectional study," *JMIR Dermatol.*, vol. 6, p. e45529, 2023. https://doi.org/10.2196/45529

[26] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with mobilenet v2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021. https://doi.org/10.3390/s21082852

[27] S. I. Hossain *et al.,* "Exploring convolutional neural networks with transfer learning for diagnosing Lyme disease from skin lesion images," *Comput. Methods Programs in Biomed.*, vol. 215, p. 106624, 2022. https://doi.org/10.1016/j.cmpb.2022.106624

[28] D. Popescu, M. El-Khatib, and L. Ichim, "Skin lesion classification using collective intelligence of multiple neural networks," *Sensors*, vol. 22, no. 12, p. 4399, 2022. https://doi.org/10.3390/s22124399

[29] C. Y. Zhu *et al.,* "A deep learning based framework for diagnosing multiple skin diseases in a clinical environment," *Front. Med.*, vol. 8, 2021. https://doi.org/10.3389/fmed.2021.626369

[30] Y. Nancy Jane, S. K. Charanya, M. Amsaprabhaa, P. Jayashanker, and H. K. Nehemiah, "2-HDCNN: A two-tier hybrid dual convolution neural network feature fusion approach for diagnosing malignant melanoma," *Comput. Biol. Med.*, vol. 152, p. 106333, 2023. https://doi.org/10.1016/j.compbiomed.2022.106333

[31] A. C. Foahom Gouabou *et al.,* "Computer aided diagnosis of melanoma using deep neural networks and game theory: Application on dermoscopic images of skin lesions," *Int. J. Mol. Sci.*, vol. 23, no. 22, p. 13838, 2022. https://doi.org/10.3390/ijms232213838

[32] R. Suresh Kumar, B. Nagaraj, P. Manimegalai, and P. Ajay, "Dual feature extraction based convolutional neural network classifier for magnetic resonance imaging tumor detection using U-Net and three-dimensional convolutional neural network," *Computers and Electrical Engineering*, vol. 101, p. 108010, 2022. https://doi.org/10.1016/j.compeleceng.2022.108010

[33] O. Iparraguirre-Villanueva *et al.*, "Text prediction recurrent neural networks using long short-term memory-dropout," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 3, pp. 1758–1768, 2023. https://doi.org/10.11591/ijeecs.v29.i3.pp1758-1768

[34] L. Nanni *et al.*, "Feature transforms for image data augmentation," *Neural Comput. Appl.*, vol. 34, pp. 22345–22356, 2022. https://doi.org/10.1007/s00521-022-07645-z

[35] L. Pan *et al.*, "MFDNN: Multi-channel feature deep neural network algorithm to identify COVID19 chest X-ray images," *Health Inf. Sci. Syst.,* vol. 10, 2022. https://doi.org/10.1007/s13755-022-00174-y

[36] R. Aloo *et al.*, "Ensemble method using real images, metadata and synthetic images for control of class imbalance in classification," *Artificial Life and Robotics*, vol. 27, pp. 796–803, 2022. https://doi.org/10.1007/s10015-022-00781-8

[37] H. Kim, H. Lee, and S. H. Ahn, "Systematic deep transfer learning method based on a small image dataset for spaghetti-shape defect monitoring of fused deposition modeling," *J. Manuf. Syst.*, vol. 65, pp. 439–451, 2022. https://doi.org/10.1016/j.jmsy.2022.10.009

[38] X. Wang *et al.*, "A recognition method of ancient architectures based on the improved Inception V3 model," *Symmetry*, vol. 14, no. 12, p. 2679, 2022. https://doi.org/10.3390/sym14122679

[39] M. A. S. Al Husaini *et al.*, "Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4," *Neural Comput. Appl.*, vol. 34, pp. 333–348, 2022. https://doi.org/10.1007/s00521-021-06372-1

[40] Z. Cao, J. Huang, X. He, and Z. Zong, "BND-VGG-19: A deep learning algorithm for COVID-19 identification utilizing X-ray images," *Knowl. Based. Syst.*, vol. 258, p. 110040, 2022. https://doi.org/10.1016/j.knosys.2022.110040

[41] A. K. Pradhan *et al.*, "A COVID-19 X-ray image classification model based on an enhanced convolutional neural network and hill climbing algorithms," *Multimed. Tools. Appl.*, vol. 82, pp. 14219–14237, 2022. https://doi.org/10.1007/s11042-022-13826-8

[42] O. Iparraguirre-Villanueva *et al.*, "Disease identification in crop plants based on convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, 2023. https://doi.org/10.14569/IJACSA.2023.0140360

[43] L. S. Chow, G. S. Tang, M. I. Solihin, N. M. Gowdh, N. Ramli, and K. Rahmat, "Quantitative and qualitative analysis of 18 deep convolutional neural network (CNN) models with transfer learning to diagnose COVID-19 on chest X-Ray (CXR) images," *SN Comput. Sci.*, vol. 4, 2023. https://doi.org/10.1007/s42979-022-01545-8

[44] M. Effati and G. Nejat, "A performance study of CNN architectures for the autonomous detection of COVID-19 symptoms using cough and breathing," *Computers*, vol. 12, no. 2, p. 44, 2023. https://doi.org/10.3390/computers12020044

[45] P. Kaur *et al.*, "Recognition of leaf disease using hybrid convolutional neural network by applying feature reduction," *Sensors*, vol. 22, no. 2, p. 575, 2022. https://doi.org/10.3390/s22020575

[46] A. Garg, S. Salehi, M. La Rocca, R. Garner, and D. Duncan, "Efficient and visualizable convolutional neural networks for COVID-19 classification using Chest CT," *Expert Syst. Appl.*, vol. 195, p. 116540, 2022. https://doi.org/10.1016/j.eswa.2022.116540

7 AUTHORS

Orlando Iparraguirre-Villanueva ⓘ ＳＣ Systems Engineer with a master's degree in information technology management, and a PhD in Systems Engineering from Universidad Nacional Federico Villarreal—Peru. ITIL® Foundation Certificate in IT Service, Specialization in Business Continuity Management, Scrum Fundamentals

Certification (SFC). National and international speaker/panelist (Panama, Colombia, Ecuador, Venezuela, México). Specialist in software development, IoT, business intelligence, free software, augmented reality, machine learning, text mining, and virtual environments (E-mail: oiparraguirre@ieee.org).

Michael Cabanillas-Carbonell ⓘ ᴤᴄ is an engineer with a Masters in Systems Engineering, pursuing a PhD in Systems Engineering and Telecommunications at the Polytechnic University of Madrid. Conference Chair of the Engineering International Research Conference IEEE Peru EIRCON. Research professor and international lecturer specializing in software development, artificial intelligence, machine learning, business intelligence, and augmented reality. He has authored more than 100 scientific articles indexed in IEEE Xplore, Scopus, and WoS (E-mail: mcabanillas@ieee.org).