

Mariame OUMOULYTE¹, Ali Omari ALAOU², Yousef FARHAOU²,
Ahmad El ALLAOU², Abdelkhalak BAHRI¹

¹ *Laboratory of Applied Sciences; Team: SDIC; National School of Applied Sciences
Al-Hoceima, AbdelmalekEsaadi University, Tétouan, Morocco*

² *L-STI, T-IDMS, FST Errachidia, Moulay Ismail University of Meknes, Morocco*

CONVOLUTIONAL NEURAL NETWORK-BASED SKIN CANCER CLASSIFICATION WITH TRANSFER LEARNING MODELS

Skin cancer is a medical condition characterized by abnormal growth of skin cells. This occurs when the DNA within these skin cells becomes damaged. In addition, it is a prevalent form of cancer that can result in fatalities if not identified in its early stages. A skin biopsy is a necessary step in determining the presence of skin cancer. However, this procedure requires time and expertise. In recent times, artificial intelligence and deep learning algorithms have exhibited superior performance compared with humans in visual tasks. This result can be attributed to improved processing capabilities and the availability of vast datasets. Automated classification driven by these advancements has the potential to facilitate the early identification of skin cancer. Traditional diagnostic methods might overlook certain cases, whereas artificial intelligence-powered approaches offer a broader perspective. Transfer learning is a widely used technique in deep learning, involving the use of pre-trained models. These models are extensively implemented in healthcare, especially in diagnosing and studying skin lesions. Similarly, convolutional neural networks (CNNs) have recently established themselves as highly robust autonomous feature extractors that can achieve excellent accuracy in skin cancer detection because of their high potential. The primary **goal** of this study was to build deep-learning models designed to perform binary classification of skin cancer into benign and malignant categories. The **tasks** to resolve are as follows: partitioning the database, allocating 80% of the images to the training set, assigning the remaining 20% to the test set, and applying a preprocessing procedure to the images, aiming to optimize their suitability for our analysis. This involved augmenting the dataset and resizing the images to align them with the specific requirements of each model used in our research; finally, building deep learning models to enable them to perform the classification task. The **methods** used are a CNNs model and two transfer learning models, i.e., Visual Geometry Group 16 (VGG16) and Visual Geometry Group 19 (VGG19). They are applied to dermoscopic images from the International Skin Image Collaboration Archive (ISIC) dataset to classify skin lesions into two classes and to conduct a comparative analysis. Our **results** indicated that the VGG16 model outperformed the others, achieving an accuracy of 87% and a loss of 38%. Additionally, the VGG16 model demonstrated the best recall, precision, and F1- score. Comparatively, the VGG16 and VGG19 models displayed superior performance in this classification task compared with the CNN model. **Conclusions.** The significance of this study stems from the fact that deep learning-based clinical decision support systems have proven to be highly beneficial, offering valuable recommendations to dermatologists during their diagnostic procedures.

Keywords: Deep learning; CNN; Transfer Learning; VGG19; VGG16; Skin cancer; Medical imaging.

1. Introduction

1.1. Motivation

Cancer is an ailment distinguished by the unrestrained splitting and proliferation of cells in organs or tissues, which can lead to their spread beyond the initial location [1]. Skin cancer is a dangerous and potentially deadly type of cancer [2 - 4]. It represents the most common type of cancer that threatens human beings. In the first instance, it is visually detected, and then by dermoscopic analysis, an early diagnosis makes it curable by almost 100% [5, 6].

Precisely diagnosing skin cancer poses a significant challenge for dermatologists, even when employing dermoscopy images, because of the initial

similarity in appearance among several types of skin cancer. Furthermore, even skilled dermatologists encounter limitations based on their education and experience in accurately diagnosing skin cancer. Their exposure is confined to a subset of potential skin cancer manifestations throughout their professional lifetime. Likewise, dermoscopy performed by less experienced dermatologists can lead to a decrease in the accuracy of skin cancer identification. Consequently, to solve the problems encountered by dermatologists there is an urgent necessity to create a swifter and more precise process for detecting and classifying skin lesions [7].

In recent years, researchers have invested substantial effort into crafting intelligent systems for applications in different fields such as object detection [8], emotion recognition [9] and healthcare [10].

Deep Convolutional Neural Networks (CNNs) have demonstrated their effectiveness in performing detection and classification tasks in medical image processing. Numerous studies have employed deep learning techniques for the classification of skin lesions, as documented in the existing literature.

Objectives

This study aims to develop automated tools that aid dermatologists in the precise diagnosis of skin lesions. To achieve this, we constructed classification models for skin lesions capable of predicting the class (benign or malignant) to which each dataset element belongs.

To accomplish the research objectives, the following tasks have been formulated:

1. Developing two pre-trained transfer learning models, VGG19 and VGG16, for feature extraction and classification, with the goal of accurately identifying the two categories of skin lesions.
2. Developing a customized CNN-based approach to achieve the classification process.
3. Training and testing the three deep neural networks mentioned above.
4. Conducting a comparative study aimed at evaluating the performance of various models and ultimately identifying the one that exhibits the most optimal results.

1.2. Content of the paper

The structure of the remaining sections in the paper is as follows: Materials and Methods are elaborated in Section 3: Section 3.1 introduces the used approach, Section 3.2 describes the dataset, Section 3.3 outlines the preprocessing procedures employed in the study, and Section 3.4 details the classification models used. The process of training and testing the models is described in Section 3.5. Furthermore, Section 3.6 delineates the methodology employed for performance evaluation. Section 4 outlines the findings and results of the conducted experiments, followed by the conclusion.

2. Literature Review

N. Nigar et al. [11] created a Convolutional Neural Network (CNN) model for the automated identification of six distinct skin conditions: actinic keratosis, benign keratosis, melanoma, basal cell carcinoma, insect bite, and skin acne. This model demonstrated an impressive accuracy rate of 97%, along with precision, recall, and F1-score, each reaching 91%. The primary objective of this model is to automatically distinguish between melanoma and non-melanoma skin cancer types.

S. M. Jaisakthi et al. [12] presented a transfer learning architecture based on EfficientNet. The evaluation of the system's performance was conducted using the Area under the curve ROC (AUC-ROC), yielding an impressive score of 0.9681. This achievement was realized through the optimal fine-tuning of EfficientNet-B6 using the Ranger optimizer.

S. Albawi et al. [13] designed and implemented a neural network-based method to predict skin cancer. The central focus of their work was refining the convolutional neural network (CNN) architecture and identifying optimal values for various CNN parameters. Their results demonstrated that the CNN approach achieved a notably elevated accuracy rate of 98.5%, surpassing the performance of other established methods.

A. Mahbod et al. [14] conducted a study to assess the performance of transfer learning with multi-scale and multi-network systems for classifying and detecting skin lesions. In addition, the authors examined the influence of dermoscopic image size on pre-trained CNNs using transfer learning.

D. A. Rodrigues et al. [15] presented a unique method to classify skin lesions using deep learning, transfer learning, and IoT. Their proposal involved implementing deep learning and transfer learning in an IoT system, allowing clinicians to diagnose common skin lesions using CNNs as feature extractors. This study incorporated various pre-trained networks and machine-learning techniques.

K. M. Hosny et al. [16] proposed a skin lesion classification system using transfer learning and augmentation, specifically employing AlexNet architecture. The authors initialized the model's parameters with the settings of the original model and randomly initiated the replacement of the last three layer weights.

A. Singhal et al. [17] devised a transfer-learning skin lesion classification model employing four pre-trained networks: Inception v3, ResNet50, DenseNet201, and Inception ResNet v2. These networks were trained using a dataset comprising up to seven classes of skin lesions. The researchers conducted a study to evaluate the effectiveness of these models in classifying skin lesions.

G. Arora et al. [18] used fourteen transfer learning networks to classify seven types of skin lesions on unbalanced data. N. Kausar et al. [19] conducted a study on multiclass skin cancer classification using a set of fine-tuned deep-learning models. They proposed a model capable of accurately identifying the most prominent types of skin lesions. To achieve this, the authors employed transfer learning techniques, leveraging several pre-trained models and implementing class-weighted loss and augmentation methods during the classification stage, particularly using ResNet50.

A multiclass classification system for skin lesions by applying deep learning models was proposed by M. Tahir et al. [20] To categorize the four different skin diseases.

Al-Habib Islam et al. [21] presented a study on skin disease detection, where they employed a transfer learning approach called fine-tuned visual geometry group-19. They examined five transfer learning models VGG16, VGG19, MobileNetV2, InceptionV3, and MobileNet to determine the network that achieves the highest accuracy and the most effective transfer learning technique for the task of skin disease identification.

A. A. Nugroho et al. [22] developed a customized CNN model that achieved an accuracy of 78% when applied to the HAM10000 dataset. N. C. F. Codella et al. [23] utilized the ISIC 2017 dataset, which includes three categories of skin cancer, to employ machine learning methods for accurate melanoma prediction. However, the study encountered inaccurate results attributed to dataset bias and incomplete dermoscopic feature annotations.

In [24], a new approach was introduced that employs transfer learning and CNN to handle multi-resolution images captured by different sensors. The CNN was initially trained on a standard image dataset, and the learned weights were then transferred to other datasets with varying resolutions. Initially, the classification of skin cancer diseases was limited to two categories: benign and malignant.

3. Materials and Methods

3.1. The proposed approach

The evolution of machine learning techniques has had a profound impact on various sectors, and dermatology is no exception. The application of machine learning in dermatology has shown great promise in improving the accuracy and efficiency of skin disorder diagnosis and treatment. The research focus on medical imaging and computer vision has identified skin lesion classification and detection as pivotal areas, holding the potential to transform the early diagnosis and treatment of diverse skin disorders.

In this paper, a binary classification system for skin lesions based on convolutional neural network and transfer learning models is proposed using dermoscopic images. The models were trained on a dataset labeled with two classes: malignant and benign lesions. To enhance the original images, a set of preprocessing steps is applied to remove unnecessary information. The resulting improved images are then fed into the pattern recognition system. The classification process starts with the extraction of deep features from the preprocessed input image, followed by the utilization of

a dense layer to produce the final classification result. Six essential steps are used in the proposed approach: dataset collection, preprocessing, convolutional neural network, and VGG19 and VGG16 architectures for skin lesion classification, training and testing, classification, and performance evaluation. These steps are detailed in the following subsections. A general diagram of the proposed approach is illustrated in Figure 1, and a summary of the different layers in our models is shown in Figures 2 - 4.

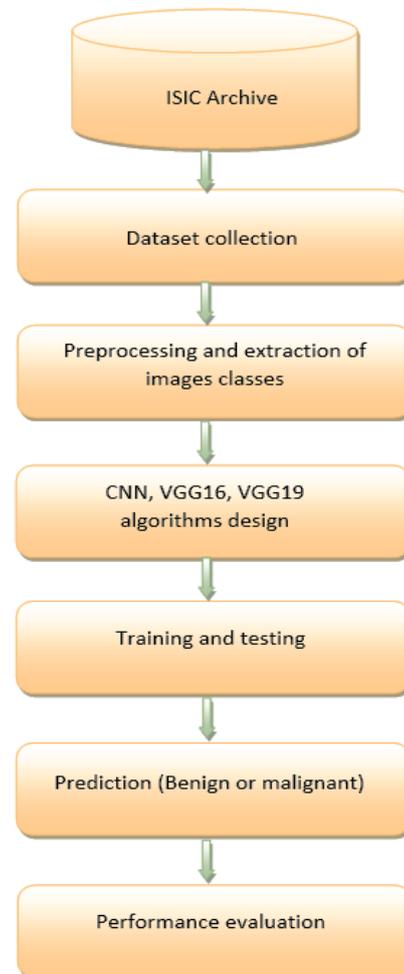


Fig. 1. Proposed approach architecture

3.2. Dataset

To conduct this study, we referred to the International Skin Image Collaboration Archive (ISIC) [25] from kaggle.com, which contains a dataset of various skin cancer types. This study focused on two categories selected from the dataset: benign and malignant moles. The benign type of skin cancer has 1800 samples in the dataset. However, the malignant type has 1497 samples, and the final dataset size is 3297. The images in the dataset were resized to (224x224x3) RGB and then labeled as "benign" and "malignant". Figures 5 and 6 show some example images from the dataset.

To ensure a comprehensive evaluation of our system's performance, we adopted a specific data distribution strategy. The database was divided, with 80% of the images allocated to the training set and the remaining 20% designated for the test set.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 64)	1792
max_pooling2d(MaxPooling2D)	(None, 112, 112, 64)	0
dropout (Dropout)	(None, 112, 112, 64)	0
conv2d_1 (Conv2D)	(None, 112, 112, 64)	36928
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0
dropout_1 (Dropout)	(None, 56, 56, 64)	0
flatten (Flatten)	(None, 200704)	0
dense (Dense)	(None, 128)	25690240
dense_1 (Dense)	(None, 2)	258

Fig. 2. Summary of the CNN model

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 512)	12845568
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 2)	514

Fig. 3. Summary of the VGG19 model

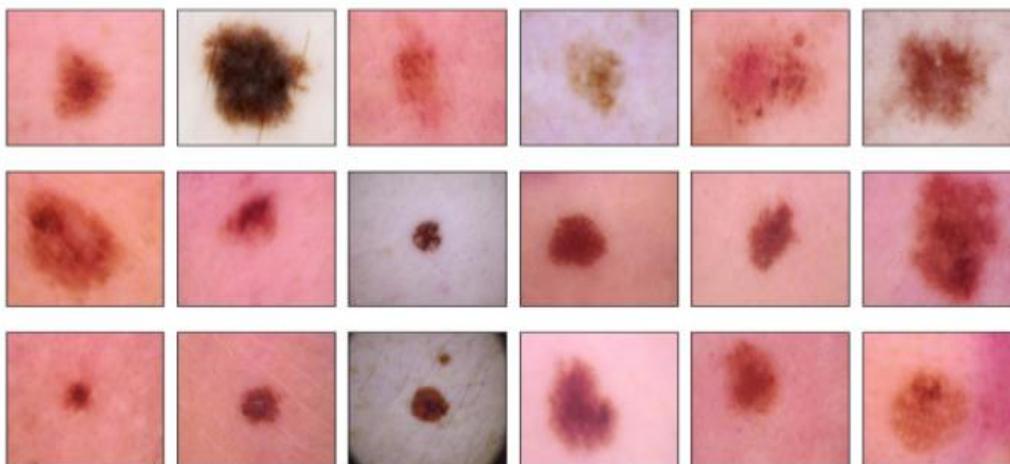


Fig. 5. Example of benign images in the dataset [25]

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 512)	12845568
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 2)	514

Fig. 4. Summary of the VGG16 model

3.3. Preprocessing

We implemented a preprocessing image process using the included pre-processing function of the Keras ImageDataGenerator to prepare the data. This involved:

- Resizing the images to suit the distinct specifications of each model employed in our study. This careful preparation allowed us to obtain accurate and meaningful results for detecting and classifying skin lesions;

- Normalization of data. Data normalization is the process of structuring a database to minimize data redundancy, ensure data integrity, and eliminate undesirable characteristics. Several established techniques for normalization exist, including methods

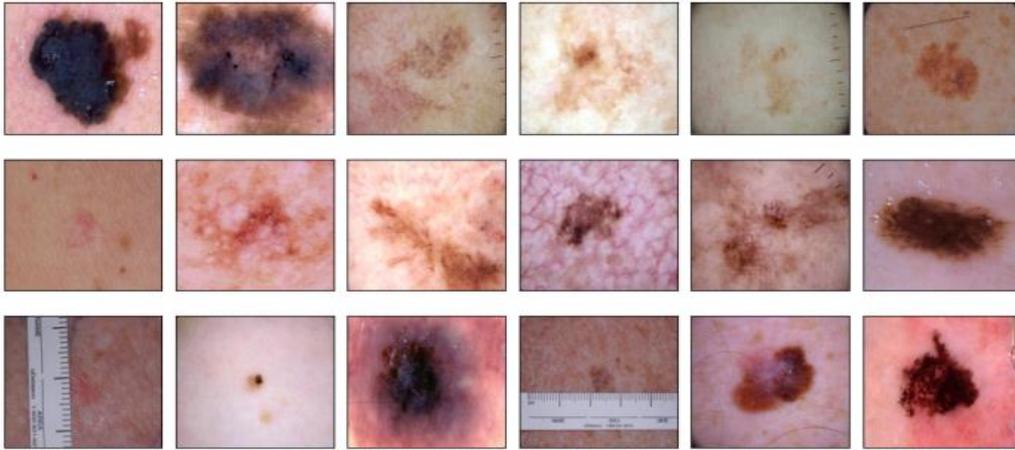


Fig. 6. Example of malignant images in the dataset [25]

such as min-max normalization, z-score normalization, and decimal scaling normalization. In our study, we normalized the dataset by dividing it by 255, which corresponds to the gray scale value of an image;

- Augmentation of data. To effectively train our models. The data augmentation technique holds considerable importance in the training process. This method prevents distortion and maintains the inherent coherence of the input and output data. Furthermore, this process is performed in real-time during the training phase, contributing to an enhanced model output and addressing the challenge of overfitting. Multiple options for image augmentation are available, including choosing values from a range of sizes such as shear range, zoom range, rotation range, and horizontal flip. The settings for image augmentation employed in our experiment are outlined in Table 1.

Table 1

Images augmentation setting

Augmentation setting	Range
Shear range	0.2
Zoom range	0.2
Width shift range	0.2
Height shift range	0.2
Rotation range	40
Horizontal flip	True

3.4. Classification models

3.4.1. Transfer Learning model structure

Transfer Learning is a machine learning technique that leverages an existing model to address distinct yet interconnected problems. In essence, it involves harnessing the knowledge gained from one task to enhance generalization in another. This is achieved by employing the pre-trained weights or model architecture

of an existing model to address our specific problem. Transfer learning offers several advantages, including accelerated training time, improved performance in many instances, and decreased demand for an extensive dataset. In this study, VGG16 and VGG19 known as VGGNet CNN architecture developed by the Visual Geometry Group of the University of Oxford, were used, trained and tested to determine the type of skin lesions.

3.4.1.1. VGG19 Convolutional neural network architecture

The VGG19 network (Figure7) [26] has 19 layers, including 16 convolutional layers, 3 fully connected layers, 5 MaxPooling layers, and 1 SoftMax layer. It accepts an input image of size 224×224 into the network. The network uses a kernel of size 3×3 with a stride of 1 and space padding. Max pooling is performed in step 2. The two layers are fully connected, and the last layer has been eliminated and reacquired with a softmax layer that classifies the two types of skin lesions. In this work, we add a preprocessing layer in front of VGG19, freeze the existing layers, and finally create a final fully connected layer.

3.4.1.2. VGG16 Convolutional neural network architecture

The VGG16 network, as illustrated in Figure 8 [27], consists of 13 convolutional layers, 5 max-pooling layers, and 3 dense layers; however, it includes only 16 weight layers (layers with learnable parameters). VGG16 accepts 224×224 images with 3 RGB channels. What is special about VGG16 is that rather than focusing on several parameters, it focuses on convolutional layers of 3×3 step 1 filters but uses the same padding and max pooling layers as 2×2 step 2 filters.

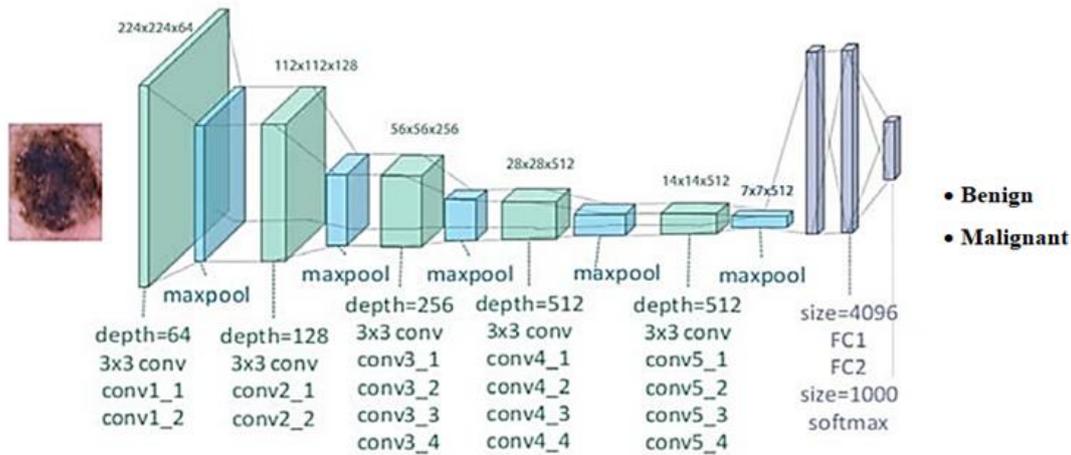


Fig. 7. VGG19 model architecture [26]

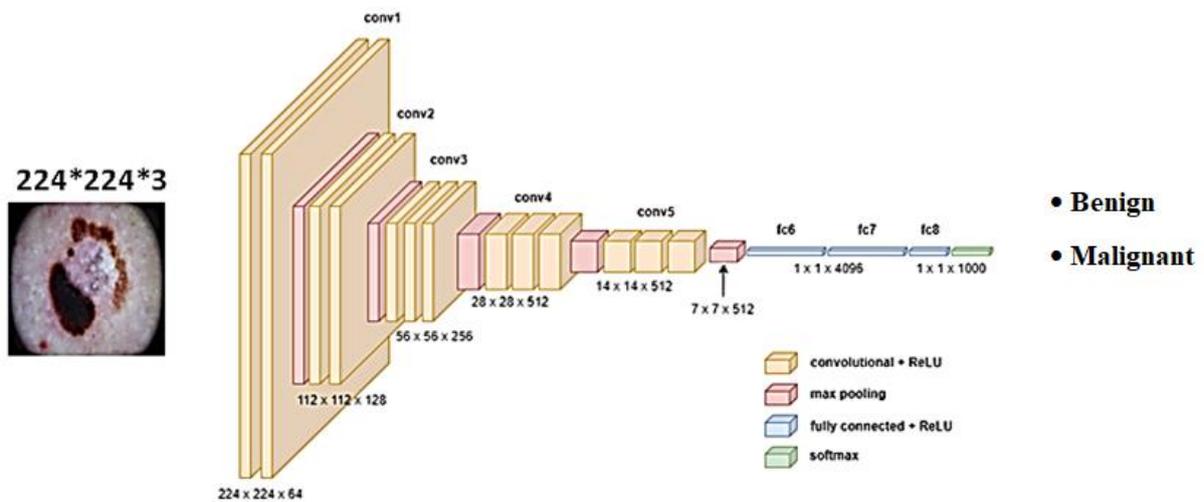


Fig. 8. VGG16 model architecture [27]

3.4.2. Convolutional neural network model

A CNN represents a network architecture within the realm of deep learning, designed to learn directly from data. CNNs are especially effective for identifying patterns within images, facilitating the recognition of objects, classes, and categories.

A CNN can comprise dozens or even hundreds of layers, each dedicated to discerning distinct features within an image. Filters are employed to process every training image across various resolutions, and the outcomes of these convolved images serve as inputs for subsequent layers. The filters initiate as rudimentary attributes such as brightness and edges, gradually evolving to encompass intricate features that are characteristic identifiers of objects.

The network presented in Figure 9 [28] is structured into 3 layers. The first layer serves as the input layer responsible for forming and training the dataset. It collects the data, assigns weights to the hidden layers, and allows the neurons in the hidden layers to separate

features from the data, ultimately deriving a model. This model forms the foundation for the generative layers, which are responsible for selecting the appropriate classes. The network employs binary classification to make the final decision. In the given example, class 0 represents a benign tumor, whereas class 1 indicates a malignant tumor. This system’s implementation revolves around the CNN.

3.5. Training and testing

To implement the CNN, VGG19, and VGG16 models, we used Keras, which is a deep-learning API for Python [29]. All methods were implemented using Keras. In this study, a dataset containing 2637 training images and 660 test images was used. The training spanned 50 epochs, utilizing a batch size of 64 and a learning rate of 0.01 for the models. Adam [30] was the optimization function used for weight updating. The loss function chosen for this model was "binary_crossentropy". To combat overfitting, a dropout rate of 0.5 was used for the fully connected layers.

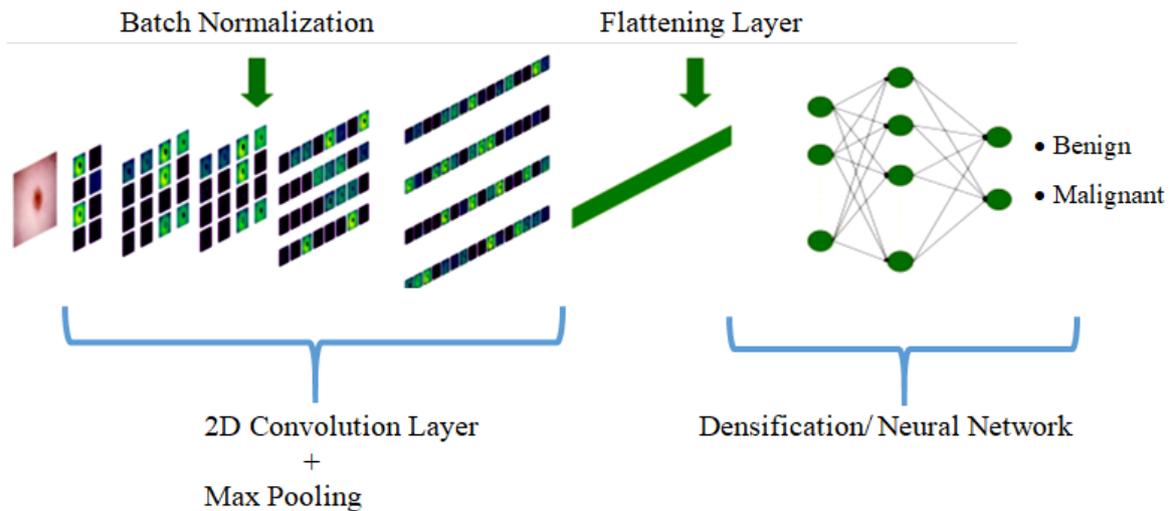


Fig. 9. CNN model [28]

After training for 50 epochs, the highest-performance model was selected on the basis of the evaluation metrics. The model parameters were then used for the test images to evaluate the overall network performance on previously unseen data.

Nevertheless, in the context of examining the number of parameters in the pre-trained models, VGG16 and VGG19, they were initially very high. However, through the use of transfer learning and the process of freezing some layers at the start of the network, the number of parameters decreased significantly. As a result, the required parameters for training were drastically reduced from 27,692,098 to 12,977,410 in the case of VGG16 and from 33,001,794 to 12,977,410 in the case of VGG19. Table 2 displays the training parameters for each model.

Table 2

Training parameters of the networks

Model	Total parameters	Trainable parameters	Non-trainable parameters
VGG19	33,001,794	12,977,410	20,024,384
VGG16	27,692,098	12,977,410	14,714,688
CNN	25,729,218	25,729,218	0

3.6. Performance evaluation

We validated the models' performance by evaluating metrics such as Recall, Precision, and F1-score. see equation 1, 2, and 3:

– Recall, the fraction of true positives that are correctly identified;

– Precision, is the fraction of retrieved instances that are relevant;

– F1-score, is the weighted average of Precision and Recall.

$$\text{Recall/Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

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

$$\text{F1 - score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Where:

– TP (True Positives). These instances refer to cases in which the model accurately predicted the positive class when the actual class was indeed positive;

– FN (False Negatives). These occurrences correspond to situations in which the model forecasted the negative class, but the actual class was positive. In simpler terms, the model failed to identify a positive case;

– FP (False Positives). These scenarios arise when the model predicts a positive class, but the actual class is negative. In such instances, the model mistakenly assigns a negative case as positive.

4. Results and discussion

4.1. Accuracy and loss

During the training process of our three proposed models, the loss function for both the training and validation data consistently decreased as the number of epochs increased. Simultaneously, the accuracy of the models increases. The accuracy and loss values for each model are presented in Table 3. The pre-trained VGG16 model attained a peak accuracy of 87 % with a loss of 0.38. In contrast, the VGG19 architecture achieved the best loss value of 0.39, but a slightly lower accuracy of 84.37%. The CNN model, on the other hand, attained an

accuracy of 74% with a loss of 0.52. The results clearly show that the VGG16 model significantly outperforms the other two models, resulting in better values in terms of accuracy, precision, recall, and F1 score. The CNN classifier performs the poorest among its counterpart algorithms.

Figures 10 and 11 display, respectively, the graph of accuracy and loss of training and validation on the skin cancer dataset for the pre-trained models VGG19, VGG16, and CNN. According to the graphs, it appears that VGG19 exhibits significant overfitting of the provided data throughout each training epoch, with high training and testing loss observed at the initial epoch. In contrast, for the VGG16 model, there is a minimal difference in the results between the training and testing accuracy, suggesting that the network is not overfitting. Additionally, the loss and accuracy stabilize between epochs 40 and 50, indicating a stable network performance. The CNN model consistently demonstrated an increase in accuracy with each iteration without exhibiting signs of network overfitting. Simultaneously, the system loss continues to decrease progressively with each iteration.

In summary, the VGG16 model has demonstrated its reliability and robustness, enabling the attainment of high accuracy even with limited data. It effectively mitigates overfitting, as corroborated by [31] and [32].

Table 3
Accuracy and Loss of proposed VGG19, VGG16, and CNN models(%)

Model	Accuracy	Loss
VGG19	84.37	39
VGG16	87	38
CNN	74	52

4.2. Confusion matrix

The confusion matrix is employed to evaluate the performance of classifier models displaying the number of correct and incorrect predictions based on each class in the test dataset (Benign, Malignant). Figure 12 illustrates the confusion matrix of the three proposed models:

– For the VGG19 model, out of 360 benign images, 302 images were predicted correctly and 58 images were predicted incorrectly. On the other hand, we found that out of 300 malicious images, 252 were correctly predicted and 48 were incorrectly predicted.

– For the VGG16 model, we found that for 360 benign images, 313 images were correctly predicted, whereas 47 images were incorrectly predicted. On the other hand, we found that for 300 malignant images,

261 images were correctly predicted and 39 images were incorrectly predicted;

– For custom CNN, we note that out of 360 benign images, 280 images were predicted correctly while 80 images had a wrong prediction. On the other hand, we notice that on 300 malignant images, 234 images were predicted correctly and 66 images had an incorrect prediction.

Based on the information provided earlier, it can be concluded that the VGG16 model demonstrates higher performance.

4.3. Recall, Precision, and F1-score

Tables 4 - 6, show precision, recall, and F1-score metrics results for the three proposed models. The results indicate that the VGG16 transfer learning model outperformed the other two models in terms of precision, recall and F1-score for both benign and malignant classes. Specifically, it achieved 88% precision, 87% recall, and 87% F1-score for the benign class, and 85% precision, 87% recall, and 86% F1-score for the malignant class. Following closely, the VGG19 model achieved the second-best results, while the CNN model was ranked last.

Table 4

VGG19 model

Class	Precision	Recall	F1-score	Support
Benign	0.86	0.83	0.84	360
Malignant	0.81	0.84	0.82	300

Table 5

VGG16 model

Class	Precision	Recall	F1-score	Support
Benign	0.88	0.87	0.87	360
Malignant	0.85	0.87	0.86	300

Table 6

CNN model

Class	Precision	Recall	F1-score	Support
Benign	0.80	0.77	0.78	360
Malignant	0.78	0.74	0.75	300

5. Conclusions

The realm of computerized diagnosis through image processing has garnered considerable interest in recent times. Accessibility of affordable software and

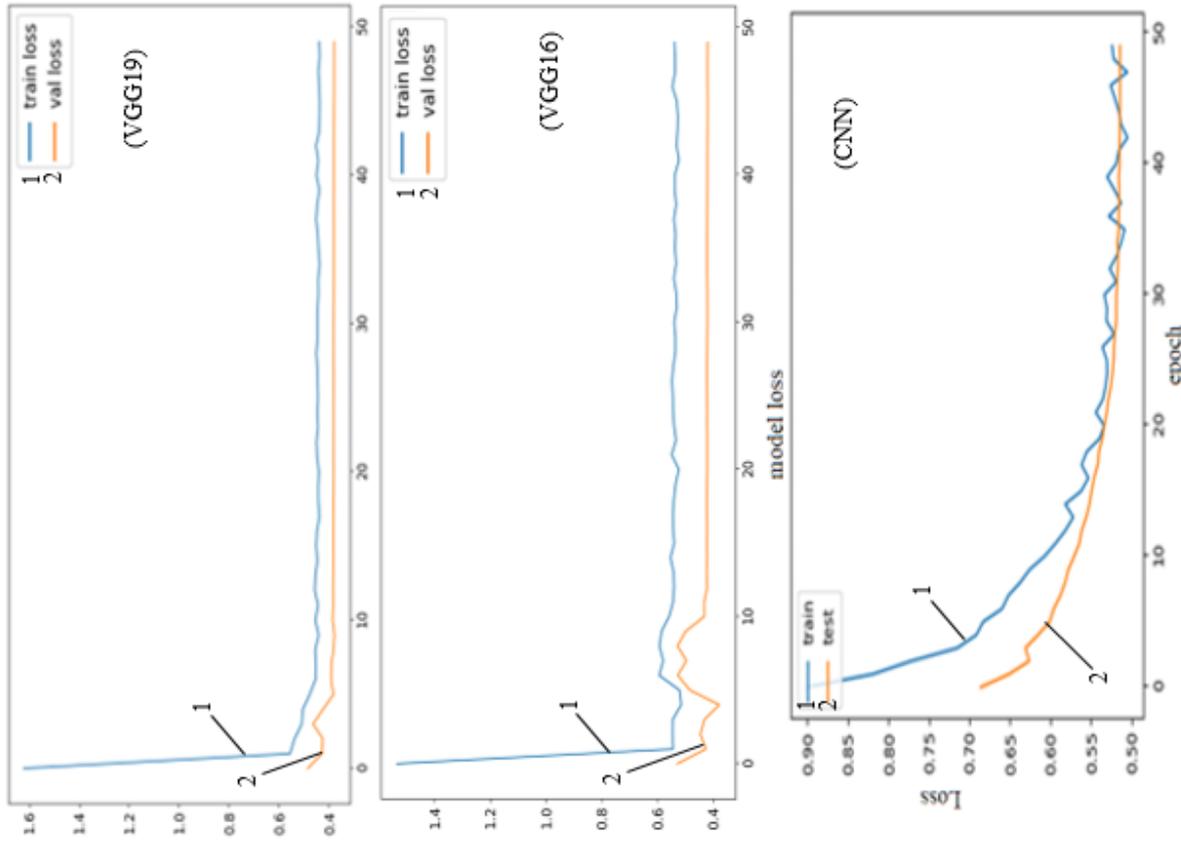


Fig. 11. Loss curves for proposed models

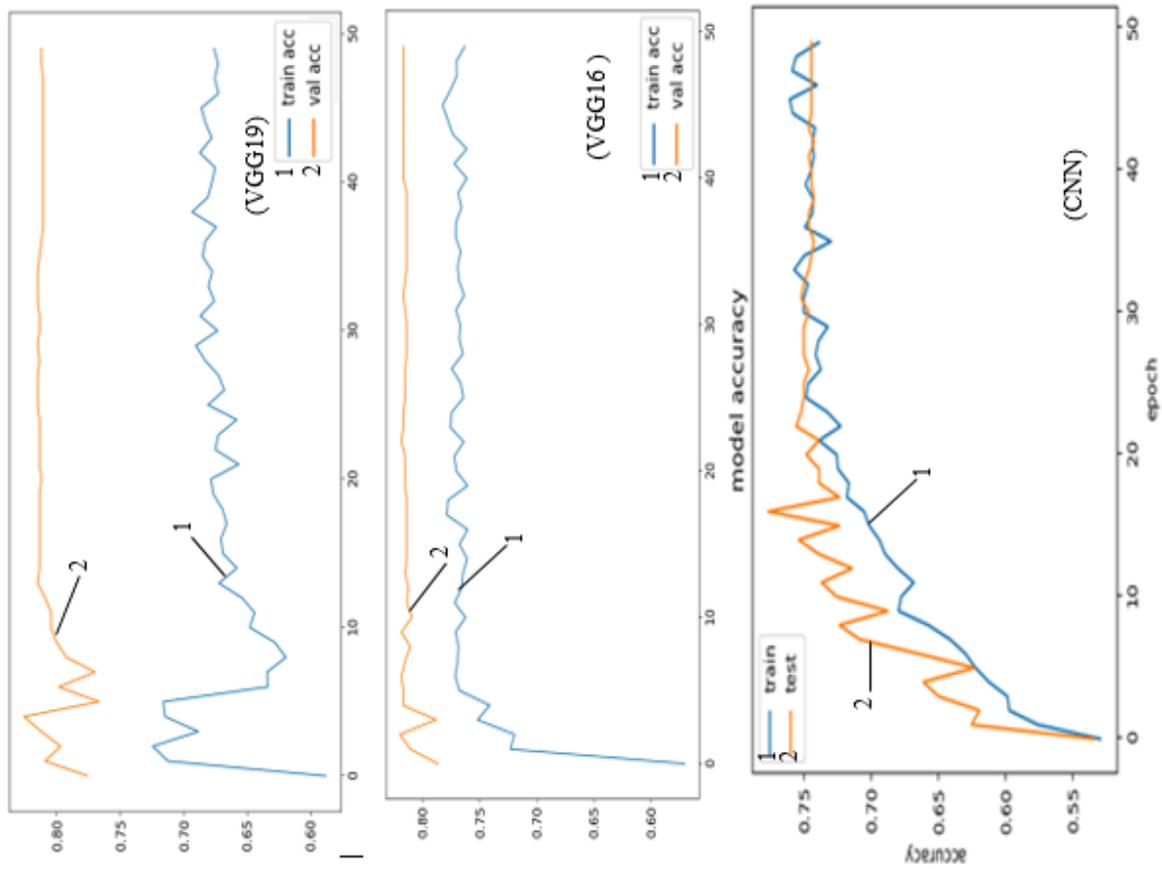


Fig. 10. Accuracy curves depicting the performance of our proposed models

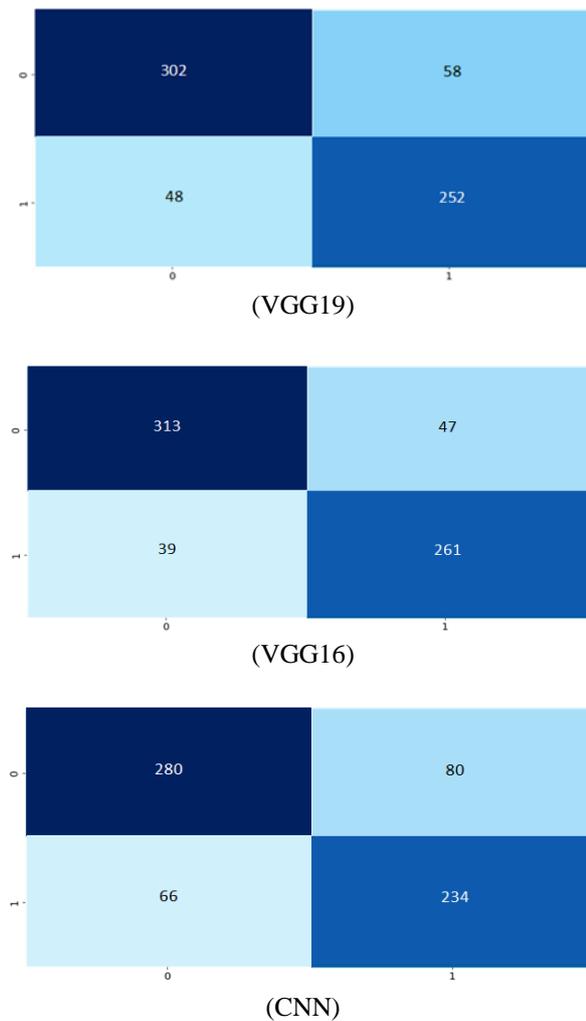


Fig. 12. The confusion matrix for our three proposed models

hardware has opened up new possibilities for image analysis applications. Automated diagnostic systems have become prevalent and are tailored for specific medical image analyses and disease diagnosis. Notably, in recent years, deep neural networks and other variants of deep learning algorithms have captured significant attention, achieving remarkable classification results in the field of computer vision. The primary goal behind developing this efficient system is to fulfill a real-time clinical diagnostic project by diagnosing skin lesions.

The proposed approach comprises several stages, including data normalization, data augmentation, feature extraction using deep learning models (CNN, VGG19, VGG16), and classification. These models were trained and tested using 3297 skin lesion images sourced from the International Skin Image Collaboration archive. According to the training and testing results, we deduced that the VGG16 model achieved good performance for the benign and malignant classes with an accuracy of 87% and a loss rate of 0.38. The VGG16 model provid-

ed the highest performance among the others with 88% precision, 87% recall, and 87% F1-score for the benign class and 85% precision, 87% recall, and 86% F1-score for the malignant class.

In future studies, we are considering improving these models so that they can classify several categories of skin cancer (Actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, etc).

Contributions of authors: Conceptualization, methodology – **Mariame Oumoulyte, Ahmad El Allaoui**; formulation of tasks, analysis – **Mariame Oumoulyte, Abdelkhalak Bahri, Ali Omari Alaoui**; development of model, software, verification – **Mariame Oumoulyte**; analysis of results, visualization – **Mariame Oumoulyte, Ahmad El Allaoui**; writing – original draft preparation, writing – review and editing – **Mariame Oumoulyte, Ali Omari Alaoui, Yousef Farhaoui**.

All authors have read and approved the published version of this manuscript.

References

1. Akyel, C., & Arici, N. Cilt Kanserinde Kılı Temizliği ve Lezyon Bölütlemesinde Yeni Bir Yaklaşım[A New Approach to Hair Noise Cleansing and Lesion Segmentation in Images of Skin Cancer]. *Politeknik Dergisi*, 2020, vol. 23, no. 3, pp. 821-828. DOI: 10.2339/politeknik.645395. (in Turkish)
2. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. M., Alraddadi, M. O., & Mahnashi, M. H. Skin Cancer Detection: A Review Using Deep Learning Techniques. *International Journal of Environmental Research and Public Health*, 2021, vol. 18, no. 10, article no. 5479. DOI: 10.3390/ijerph18105479.
3. *Better Health Channel: Melanoma*. Available at <https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/melanoma#what-is-melanoma>. (accessed 22.06.23).
4. Naqvi, M., Gilani, S. Q., Syed, T., Marques, O., & Kim, H. C. Skin Cancer Detection Using Deep Learning – A Review. *Diagnostics*, 2023, vol. 13, no. 11, article no. 1911. DOI: 10.3390/diagnostics13111911.
5. Chuchu, N., Dinnes, J., Takwoingi, Y., Matin, R. N., Bayliss, S. E., Davenport, C., Moreau, J. F., Bassett, O., Godfrey, K., O'Sullivan, C., Walter, F. M., Motley, R., Deeks, J. J., & Williams, H. C. Tele dermatology for diagnosing skin cancer in adults. *Cochrane Database of Systematic Reviews*, 2018, iss. 12, article no. CD013193. DOI: 10.1002/14651858.CD013193.
6. Maglogiannis, I., & Doukas, C. N. Overview of Advanced Computer Vision Systems for Skin Lesions Characterization. *IEEE Transactions on Information*

Technology in Biomedicine, 2009, vol. 13, no. 5, pp. 721–733. DOI: 10.1109/TITB.2009.2017529.

7. Ali, M. S., Miah, M. S., Haque, J., Rahman, M. M., & Islam, M. K. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Machine Learning with Applications*, 2021, vol. 5, article no. 100036. DOI: 10.1016/j.mlwa.2021.100036.

8. Moskalenko, V. V., Moskalenko, A. S., Korobov, A. H., Zarets'kyy, M. O., & Semashko, V. A. Model' ta alhorytm navchannya systemy detektuvannya malorozmirnykh ob'yektiv dlya malohabarynykh bezpilotnykh lital'nykh aparativ [A model and learning algorithm of small-sized object detection system for compact drones]. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2018, no. 4, pp. 41–52. DOI: 10.32620/reks.2018.4.04. (in Ukrainian)

9. Hazra, S. K., Ema, R. R., Galib, S. M., Kabir, S., & Adnan, N. Emotion recognition of human speech using deep learning method and MFCC features. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2022, no. 4, pp. 161–172. DOI: 10.32620/reks.2022.4.13.

10. Krivtsov, S., Meniailov, I., Bazilevych, K., & Chumachenko, D. Predictive model of COVID-19 epidemic process based on neural network. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2022, no. 4, pp. 7–18. DOI: 10.32620/reks.2022.4.01.

11. Nigar, N., Wajid, A., Islam, S., & Shahzad, M. K. Skin cancer classification: a deep learning approach. *Pakistan Journal of Science*, 2023, vol. 75, no. 02, pp. 210–218. DOI: 10.57041/pjs.v75i02.851.

12. Jaisakthi, S. M., Mirunalini, P., Chandrabose, A., & Rajagopal, A. Classification of skin cancer from dermoscopic images using deep neural network architectures. *Multimed Tools Appl*, 2023, vol. 82, pp. 15763–15778. DOI: 10.1007/s11042-022-13847-3

13. Albawi, S., Arif, M. H., & Waleed, J. Skin cancer classification dermatologist-level based on deep learning model. *Acta Scientiarum Technology*, 2023, vol. 45, pp. e61531–e61531. DOI: 10.4025/actascitechnol.v45i1.61531.

14. Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., & Ellinger, I. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Computer Methods and Programs in Biomedicine*, 2020, vol. 193, article no. 105475. DOI: 10.1016/j.cmpb.2020.105475.

15. Rodrigues, D. A., Ivo, R. F., Satapathy, S. C., Wang, S., Hemanth, J., & Filho, P. P. R. A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system. *Pattern Recognition Letters*, 2020, vol. 136, pp. 8–15. DOI: 10.1016/j.patrec.2020.05.019.

16. Hosny, K. M., Kassem, M. A., & Foad, M. M. Classification of skin lesions using transfer learning

and augmentation with Alex-net. *PLOS ONE*, 2019, vol. 14 no. 5, article no. e0217293. 17 p. DOI: 10.1371/journal.pone.0217293.

17. Singhal, A., Shukla, R., Kankar, P. K., Dubey, S., Singh, S., & Pachori, R. B. Comparing the capabilities of transfer learning models to detect skin lesion in humans. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, IMECHE*, 2020, vol. 234, no. 10, pp. 1083–1093. DOI: 10.1177/0954411920939829.

18. Arora, G., Dubey, A. K., Jaffery, Z. A., & Rocha, A. A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data. *Neural Computing and Applications*, 2023, vol. 35, no. 11, pp. 7989–8015. DOI: 10.1007/s00521-022-06922-1.

19. Kausar, N., Hameed, A., Sattar, M., Ashraf, R., Imran, A. S., Abidin, M. Z. Ul., & Ali, A. Multiclass Skin Cancer Classification Using Ensemble of Fine-Tuned Deep Learning Models. *Applied Sciences*, 2021, vol. 11, iss. 22, article no. 10593. DOI: 10.3390/app112210593.

20. Tahir, M., Naeem, A., Malik, H., Tanveer, J., Naqvi, R. A., & Lee, S.-W. DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images. *Cancers*, 2023, vol. 15, no. 7, article no. 2179. DOI: 10.3390/cancers15072179.

21. Islam, M. Al-H., Shahriyar, S. M., Alam, M. J., Rahman, M., & Sarker, M. R. K. R. Skin disease detection employing transfer learning approach a fine-tune visual geometry group-19. *Indonesian Journal of Electrical Engineering and Computer Science*, 2023, vol. 31, no. 1, pp. 321–328. DOI:10.11591/ijeecs.v31.i1.pp321-328.

22. Nugroho, A. A., Slamet, I., & Sugiyanto. Skin cancer identification system of HAM10000 skin cancer dataset using convolutional neural network. *AIP Conference Proceedings*, 2019, vol. 2202, no. 1, article no. 020039. DOI: 10.1063/1.5141652.

23. Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kallou, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). *Proceeding of 2018 IEEE 15th International Symposium on Biomedical Imaging*, 2018, pp. 168–172. DOI: 10.1109/ISBI.2018.8363547.

24. Farooq, A., Jia, X., Hu, J., & Zhou, J. Knowledge Transfer via Convolution Neural Networks for Multi-Resolution Lawn Weed Classification. *Proceeding of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote*

Sensing, 2019, pp. 01-05. DOI: 10.1109/WHISPERS.2019.8920832.

25. *The International Skin Imaging Collaboration*. Available at: <https://www.isic-archive.com/> (accessed 10.05.2023).

26. Zheng, Y., Yang, C., & Merkulov, A. Breast cancer screening using convolutional neural network and follow-up digital mammography. *Computational Imaging III*, SPIE, 2018, vol. 10669, article no. 1066905. DOI:10.1117/12.2304564.

27. Leung, K. How to Easily Draw Neural Network Architecture Diagrams. *Towards data science*. Available at: <https://towardsdatascience.com/how-to-easily-draw-neural-network-architecture-diagrams-a6b6138ed875> (accessed 10.05.2023).

28. Oumoulyte, M., El Allaoui, A., Farhaoui, Y., Amounas, F., & Qaraai, Y. Deep Learning Algorithms for Skin Cancer Classification. *Artificial Intelligence and Smart Environment. ICAISE 2022. Lecture Notes in*

Networks and Systems, Springer, Cham, 2022, vol. 635, pp. 345-351. DOI: 10.1007/978-3-031-26254-8_49.

29. *The Keras ecosystem*. Available at: <https://keras.io/> (accessed 10.05.2023).

30. *Optimizer that implements the Adam algorithm*. Available at: <http://keras.io/api/optimizers/adam/> (accessed 10.05.2023).

31. Ibrahim, A., Elbasheir, M., Badawi, S., Mohammed, A., & Alalmin, A. Skin Cancer Classification Using Transfer Learning by VGG16 Architecture (Case Study on Kaggle Dataset). *Journal of Intelligent Learning Systems and Applications*, 2023, vol. 15, no. 3, pp. 67-75. DOI: 10.4236/jilsa.2023.153005.

32. Arshed, M. A., Mumtaz, S., Ibrahim, M., Ahmed, S., Tahir, M., & Shafi, M. Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network Based Pre-Trained Models. *Information*, 2023, vol. 14, iss. 7. DOI: 10.3390/info14070415.

Received 22.08.2023, Accepted 20.11.2023

КЛАСИФІКАЦІЯ РАКУ ШКІРИ НА ОСНОВІ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ З МОДЕЛЯМИ ТРАНСФЕРНОГО НАВЧАННЯ

Маріам Умуліт, Алі Омарі Алауї, Юсеф Фархауї,
Ахмад Ель Аллауї, Абдельхалак Бахрі

Рак шкіри – це захворювання, що характеризується атипичним розростанням клітин шкіри. Це відбувається, коли ДНК у клітинах шкіри пошкоджується. Крім того, це поширена форма раку, яка може призвести до летального результату, якщо її не виявити на ранніх стадіях. Біопсія шкіри є необхідним кроком у визначенні наявності раку шкіри, однак ця процедура вимагає часу і досвіду. Останнім часом штучний інтелект і алгоритми глибокого навчання демонструють вищу продуктивність порівняно з людиною у виконанні візуальних завдань. Це можна пояснити покращеними можливостями обробки та наявністю великих наборів даних. Автоматизована класифікація, заснована на цих досягненнях, має потенціал для полегшення ранньої діагностики раку шкіри. Традиційні методи діагностики можуть пропустити певні випадки, тоді як підходи на основі штучного інтелекту пропонують ширшу перспективу. Трансферне навчання виділяється як широко використовувана техніка в глибокому навчанні, що передбачає використання попередньо навчених моделей. Ці моделі широко застосовуються в охороні здоров'я, особливо в діагностиці та вивченні уражень шкіри. Аналогічно, згорткові нейронні мережі (CNN) нещодавно зарекомендували себе як високонадійні автономні екстрактори ознак, які завдяки своєму високому потенціалу можуть досягти відмінної точності у виявленні раку шкіри. Основною метою цього дослідження є побудова моделей глибокого навчання, що не мають знаків, для виконання бінарної класифікації раку шкіри на доброякісні та злоякісні категорії. Завдання, які необхідно вирішити: розбиття бази даних, виділення 80% зображень у навчальну вибірку, решта 20% - у тестову вибірку; застосування процедури попередньої обробки зображень з метою оптимізації їхньої придатності для аналізу. Це включало розширення набору даних і зміну розміру зображень для приведення їх у відповідність до специфічних вимог кожної моделі, використаної в нашому дослідженні; нарешті, створення моделей глибокого навчання для того, щоб вони могли виконувати завдання класифікації. Використані методи – це модель CNN і дві моделі навчання з переносом, такі як Visual Geometry Group 16 (VGG16) і Visual Geometry Group 19 (VGG19), вони були застосовані до дерматологічних зображень з набору даних Міжнародного архіву зображень шкіри (ISIC) для класифікації уражень шкіри на два класи і проведення порівняльного аналізу. Наші результати показали, що модель VGG16 перевершила інші, досягнувши точності 87% і втрат 38%. Крім того, модель VGG16 продемонструвала найкращі показники пригадування, точності та F1. Для порівняння, моделі VGG16 та VGG19 продемонстрували кращі результати в цьому завданні класифікації порівняно з моделлю CNN. Висновки. Важливість цього дослідження впливає з того, що сис-

теми підтримки прийняття клінічних рішень на основі глибокого навчання виявилися дуже корисними, пропонуючи цінні рекомендації дерматологам під час діагностичних процедур.

Ключові слова: CNN; Глибоке навчання; Навчання з перенесенням; VGG19; VGG16; Рак шкіри; Медична візуалізація.

Маріам Умулілт – аспірант у галузі обробки зображень, Національна школа прикладних наук Аль-Хосейма, Лабораторія прикладних наук, команда: SDIC, Університет Абдельмалека Есааді, Тетуан, Марокко.

Алі Омарі Аллауї – аспірант у галузі обробки зображень у L-STI, T-IDMS, FST Еррачідія, Університет Мулая Ісмаїла Мекнес, Марокко.

Юсеф Фархауї – д-р наук з комп'ютерної безпеки, проф. факультету науки і технологій FSTE, Еррачідія, Марокко.

Ахмад Ель Аллауї – д-р наук з інформаційних технологій, проф. факультету науки і технологій FSTE, Еррачідія, Марокко.

Абдельхалак Бахрі – проф., Національна школа прикладних наук Аль-Хосейма, Лабораторія прикладних наук, команда: SDIC, Університет Абдельмалека Есааді, Тетуан, Марокко.

Mariame Oumoulylte – PhD student in the field of image processing at the National School of Applied Sciences Al-Hoceima, Laboratory of Applied Sciences, Team: SDIC, Abdelmalek Esaadi University, Tétouan, Morocco,

e-mail: mariameoumoulylte@gmail.com, ORCID: 0009-0005-8542-317X.

Ali Omari Alaoui – PhD student in the field of image processing at L-STI, T-IDMS, FST Errachidia, Moulay Ismail University of Meknes, Morocco,

e-mail: alialaouii9@gmail.com, ORCID: 0009-0003-9210-7875.

Yousef Farhaoui – Doctor in Computer Security, Professor at the Faculty of Science and Technology FSTE, Errachidia, Morocco,

e-mail: y.farhaoui@fste.umi.ac.ma, ORCID: 0000-0003-0870-6262, Scopus Author ID: 57193499468.

Ahmad El Allaoui – Doctor in Information Technology, Professor at the Faculty of Science and Technology FSTE, Errachidia, Morocco,

e-mail: a.elallaoui@umi.ac.ma, ORCID: 0000-0002-8897-3565, Scopus Author ID: 57195110231.

Abdelkhalak Bahri – Professor, National School of Applied Sciences Al-Hoceima, Laboratory of Applied Sciences, Team: SDIC, Abdelmalek Esaadi University, Tétouan, Morocco,

e-mail: abahri@uae.ac.ma, ORCID: 0000-0002-8527-7281.