








Classifying Melanoma in ISIC Dermoscopic Images Using Efficient Convolutional Neural Networks and Deep Transfer Learning

Habeba Mahmoud¹, Osama A. Omer¹, Shimaa Ragab¹, Hamada Esmail^{1,2}, Mohamed Abdel-Nasser^{1*}

¹ Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University, Aswan 81542, Egypt

² Department of Electronics and Communication Engineering, College of Engineering, A'Sharqiyah University, P.O. Box 42, Ibra 400, Oman

Corresponding Author Email: mohamed.abdelnasser@aswu.edu.eg

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410211>

ABSTRACT

Received: 16 April 2023

Revised: 29 September 2023

Accepted: 16 December 2023

Available online: 30 April 2024

Keywords:

skin cancer, image classification, precise Computer-Aided Diagnosis (CAD), deep learning, Convolutional Neural Network (CNN), dermoscopic images

Melanoma, recognized as the most life-threatening form of skin cancer, poses a significant threat to life expectancy. The timely identification of melanoma plays a crucial role in mitigating the morbidity and mortality associated with skin cancer. Dermoscopic images, acquired through advanced dermoscopic tools, serve as vital resources for the early detection of skin cancer. Hence, there is an urgent need to develop a reliable and accurate Computer-Aided Diagnosis (CAD) system capable of autonomously discerning skin cancer. This study focuses on the meticulous construction of diverse skin cancer classification models, specifically employing various Convolutional Neural Network (CNN) architectures configured across four distinct layer arrangements. Additionally, a transfer learning approach is explored, leveraging robust pre-trained deep CNN models extensively trained on the comprehensive ISIC dermoscopic image dataset, known for its diversity in skin lesions. Utilizing the ISIC dataset as the foundation of our analysis, the CNN model's performance is systematically evaluated with varying numbers of layers—ranging from 15 to 27. Results indicate that the CNN model comprising 15 layers achieves an accuracy of 89.55%, while the model with 27 layers exhibits the highest performance, attaining an accuracy of 90.85%. In the realm of transfer learning, ten baseline CNN models pre-trained on ImageNet are employed. All baseline models demonstrate accuracies surpassing 80%, with SqueezeNet recording the lowest accuracy at 80.89%. In contrast, the ResNet-50 model consistently outperforms other models in transfer learning, achieving an accuracy of 92.98%. These findings underscore the efficacy of the proposed models in melanoma classification and highlight the superior performance of the ResNet-50 model in the context of transfer learning.

1. INTRODUCTION

Skin cancer results from modifications in normal skin cells, initiating their conversion into malignant cells that undergo uncontrolled proliferation, assuming distorted shapes due to DNA damage. Histologically, skin cancer presents as an irregular structure characterized by chromatin, nuclei, and various stages of cell differentiation within the cytoplasm [1]. Recent studies emphasize an annual increase in skin cancer diagnoses, surpassing incidences of other cancer types [2].

There are two primary categories of skin cancer—Non-Melanoma Skin Cancer (NMSC) and Melanoma Skin Cancer (MSC), as illustrated in Figure 1. Within the realm of Melanoma Skin Cancer (MSC), multiple subtypes exist, including lentigo maligna, acral lentiginous, and nodular melanoma [3]. NMSC is primarily associated with heightened exposure to UV radiation and ozone depletion, influenced by factors such as sun-seeking behaviors, cumulative UV exposure, and extended lifespans [4]. NMSC further divides into three main categories: Basal Cell Carcinoma (BCC), the most prevalent type of NMSC (constituting 75%), followed by

Squamous Cell Carcinoma (SCC) (24%), with rarer variants making up a minor fraction (1%), such as Sebaceous Gland Carcinoma (SGC) [5]. BCC, SGC, and SCC originate in the intermediate and upper layers of the epidermis, displaying a lower tendency for metastasis compared to melanoma. Generally, non-melanoma cancers carry a more favorable prognosis than melanoma cancers [6].

Melanoma skin cancer, though constituting only 1% of total cases, is associated with a higher mortality rate, as reported by the American Cancer Society [6]. Primarily affecting melanocytes, the cells comprising the skin's surface [7], melanoma can manifest in various hues, ranging from colorless to shades like rose pink, royal purple, or azure [8]. Its heightened fatality and aggressiveness are attributed to rapid metastasis [9]. Melanoma originates when normal melanocytes undergo uncontrolled proliferation, leading to the formation of malignant tumors. While it can affect any part of the body, it predominantly appears in sun-exposed areas such as the hands, face, neck, lips, and other exposed regions.

Timely detection of skin cancer is crucial in mitigating associated risk factors; otherwise, the condition may

metastasize to other body sites, leading to considerable distress and potentially fatal outcomes [10]. Physicians typically resort to a biopsy procedure to diagnose skin cancer, involving the extraction of a tissue sample from a suspicious skin lesion for subsequent examination. While indispensable, this procedure is often complex, time-consuming, and protracted. Thus, early identification and classification of skin cancer are vital for improving survival rates [8, 9].

In the contemporary medical landscape, the use of CAD systems is imperative for diagnosing and evaluating medical images. These systems offer a convenient, cost-effective, and expedient means of diagnosing skin cancer symptoms. Several non-invasive methodologies have been proposed for investigating skin cancer symptoms and differentiating between melanoma and non-melanoma. These methods encompass image acquisition, preprocessing, post-acquisition image segmentation, feature extraction, and classification as fundamental steps in skin cancer detection [11].

The literature presents various hand-crafted feature-based methods for the classification of malignant melanoma and benign skin lesions, including the ABCD rule-based method [8]. More recently, diverse deep learning-based approaches have emerged for skin cancer detection, as evidenced by references [5, 12].

In this study, we introduce a suite of highly efficient skin cancer classification models specifically crafted for detecting melanoma skin cancer in dermoscopic images of skin lesions. At the core of our methodology lies the utilization of CNNs and leveraging transfer learning techniques, consistently proven in their efficacy in image analysis tasks. Our approach

encompasses a range of CNN models with different network depths—15, 19, 23, and 27 layers. Furthermore, we employ transfer learning, utilizing ten robust pre-trained CNN architectures: SqueezeNet, GoogLeNet, AlexNet, ResNet-18, ResNet-50, MobileNet-v2, ShuffleNet, NASNet-Mobile, EfficientNetB0, and VGG-19. The selection of these models is based on their adeptness in extracting intricate image features, aligning with our goal of enhancing the precision and reliability of melanoma detection.

The primary aim of this study is to create efficient models for detecting melanoma skin cancer. These classification models are designed to achieve high accuracy while maintaining lower complexity and reduced computational cost. These characteristics make them suitable for resource-constrained devices and environments with limited computational resources. Moreover, these attributes significantly contribute to early diagnosis, thereby enhancing patient outcomes in the healthcare domain.

The subsequent sections of this paper are organized as follows: Section 2 offers a brief discussion of related works in the existing literature. Section 3 comprehensively explains the proposed methods for skin cancer classification. Section 4 presents the performance evaluation metrics used throughout this study. In Section 5, we not only present our research results but also conduct an in-depth analysis of skin cancer detection techniques, comparing them with relevant prior studies. Finally, in Section 6, we draw conclusions, emphasizing the significant implications derived from our study's findings, and outline potential directions for future research.

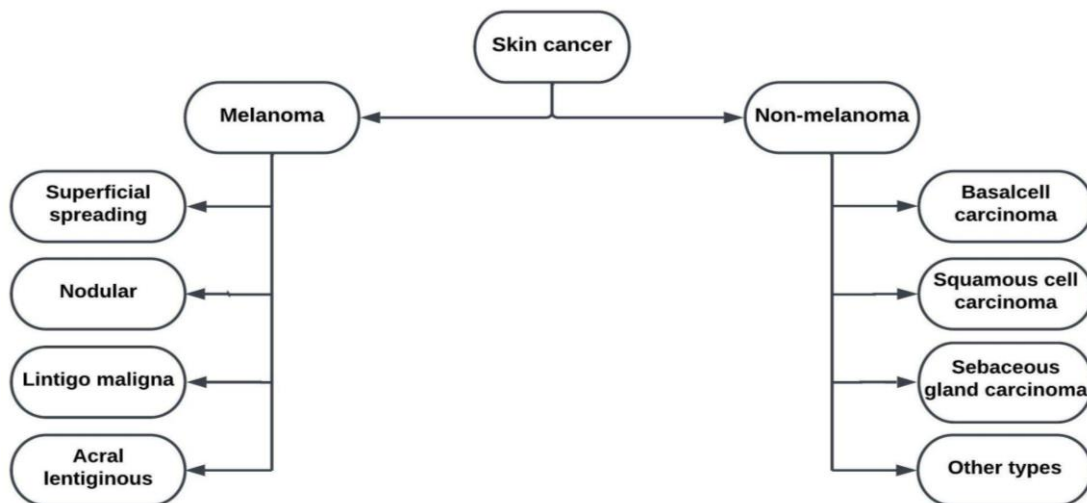


Figure 1. Skin cancer types

2. RELATED WORKS

This section endeavors to provide a comprehensive overview and analysis of prior research pertaining to the classification of skin cancer, with a specific focus on melanoma within images. Our examination encompasses a spectrum of studies, incorporating approaches utilizing handcrafted features as well as those leveraging advanced deep learning methodologies.

2.1 Classification of melanoma using handcrafted features

The process of diagnosing malignant skin cancer often

commences with extracting various handcrafted features from lesion images, including shape, size, color, and texture. These extracted features are subsequently refined through feature ranking and optimization algorithms before undergoing classification.

In the realm of detecting melanoma skin cancer in dermoscopic images, literature has extensively employed a range of handcrafted features and diverse machine learning techniques. Below, we present some related studies, categorized by the employed machine learning technique, features, or dataset:

- **Genetic Algorithm and Support Vector Machine (SVM) Approach:** Tan et al. [13] enhanced classification

accuracy to 88% on the Dermofit dataset by utilizing a genetic algorithm coupled with an SVM classifier.

- **Feature Ranking Methods:** Mukherjee et al. [14] utilized feature ranking methods to isolate the top-performing 163 features.
- **Color and Texture Features with K-Nearest Neighbor (KNN) Algorithm:** Ballerini et al. [15] achieved 93% accuracy using color and texture features in conjunction with a KNN classifier on the Dermofit dataset.
- **Multilayer Perceptron (MLP) Classifier and Dermofit:** Mukherjee et al. [16] attained an 83.33% classification accuracy by employing the top 10 conventional features and MLP on the Dermofit dataset.
- **Combining Dermofit and MEDNODE:** Mukherjee et al. [17] reported a 91.02% accuracy using 1886 features from both Dermofit and MEDNODE datasets.

Although handcrafted features have demonstrated their utility in preliminary endeavors to classify melanoma, they are constrained in their capacity to encapsulate the nuanced and diverse characteristics inherent in melanoma skin lesions. As a result, deep learning methodologies, particularly CNNs, have arisen as a promising alternative. These methodologies showcase the capacity to autonomously learn discriminative features from data and dynamically adapt to the intricate complexities associated with melanoma classification.

2.2 Classification of melanoma using deep CNNs

Advancements in deep CNNs have demonstrated notable efficacy in object recognition tasks, prompting their exploration in medical image processing, specifically in the categorization of melanoma. Table 1 provides a comprehensive summary of various deep learning-based methods, delineating the strengths and limitations of deep CNN networks for classifying melanoma in dermoscopic images [18]. Researchers commonly rely on well-established

datasets such as ISIC and PH2, underscoring the pivotal role of benchmark datasets in this specialized domain. To address challenges related to dataset size and diversity, the adoption of data augmentation techniques has become a prevalent trend.

In addressing the unique intricacies of melanoma classification, researchers often employ customized CNN architectures or modifications of established models. This tailored approach is crucial for adapting solutions to the specific nuances inherent in melanoma detection. While there is variability in achieved classification accuracy across studies, collective findings consistently underscore the promising potential of CNN-based methodologies in melanoma detection, with several studies reporting accuracy rates surpassing 80%, indicative of the robust capabilities of CNNs in this medical imaging domain.

The prevailing trajectory in the field leans towards exploring increasingly complex CNN architectures and incorporating advanced techniques, all aimed at further enhancing the accuracy of melanoma classification. This inclination reflects the ongoing commitment of researchers to push the boundaries of innovation, ultimately contributing to the refinement of diagnostic tools in the critical domain of melanoma detection.

The studies outlined in Table 1 highlight the potential of deep CNNs in melanoma classification. These investigations employ diverse methodologies to effectively address challenges associated with dataset limitations. Researchers experiment with various data augmentation techniques and leverage a spectrum of CNN architectures to attain competitive results. The selection of a particular method often depends on factors such as dataset size, available computational resources, and the specific research objectives. This nuanced decision-making process reflects the dynamic landscape of melanoma classification research, where tailored approaches are crafted to align with the unique demands of each study's context and goals.

Table 1. A summary of some deep learning based methods for classifying melanoma in dermoscopic images

Ref.	Methods and Datasets	Strengths	Limitations
[19]	A CNN on the MEDNODE dataset with data augmentation.	Data augmentation to increase dataset size, use of color images.	Limited dataset (170 images), relatively simple CNN architecture.
[20]	Deep CNNs and augmented images from DermIS and DermQuest datasets.	Extensive data augmentation, high classification accuracy.	Data augmentation can be computationally expensive and may not always generalize well.
[21]	Data augmentation with a CNN on the ISIC dataset.	Exploration of data augmentation impact, use of the ISIC dataset.	Moderate classification accuracy without augmentation.
[22]	Different CNN architectures and achieved accuracies of 81.2%, 75.5%, and 80.7% for VGG-19, ResNet-50, and VGG-19-SVM, respectively.	Comparative analysis of multiple CNN architectures, various augmentation techniques.	Imbalanced dataset, moderate to good accuracy.
[23]	VGG-19 CNN on the PH2 dataset.	High classification accuracy (92.5%) on the PH2 dataset.	Limited dataset size (200 images).
[24]	Deep CNN based on the VGG-16 architecture with the ISBI 2016 challenge dataset.	Use of a well-established VGG-16 architecture, ISBI 2016 challenge dataset.	Moderate classification accuracy (81.33%).
[25]	A modified LightNet architecture on the ISBI 2016 challenge dataset	A modified architecture, achieving a competitive classification accuracy.	Limited dataset, moderate accuracy.

3. METHODS

3.1 Convolutional Neural Network (CNN)

The CNN stands out as an innovative paradigm within the realm of multilayer perceptrons (MLP), meticulously designed for the processing of two-dimensional data. CNNs have

garnered substantial attention owing to their widespread application and inherent capability to accommodate high-depth networks, making them particularly adept at analyzing image data [26]. The architectural framework of CNNs shares similarities with general neural networks, featuring neurons equipped with weights, biases, and activation functions. A typical CNN structure encompasses a convolution layer

employing Rectified Linear Unit (ReLU) activation, succeeded by a pooling layer that serves as the feature extraction stage, culminating in a fully connected layer employing softmax activation for classification purposes.

3.1.1 Convolution Layer

At the core of CNNs lies the fundamental operation of convolution, primarily executed within the convolution layer. Operating as the initial processing layer for input images, this layer utilizes a convolutional filter, typically with dimensions such as 3×3 , 5×5 , or 7×7 , to convolve the image. This convolution process generates multiple feature maps, often referred to as the feature map set [27], as depicted in Eq. (1).

$$F_j^{l+1} = \sum_i w_{ij}^l F_i^l + b_j^l \quad (1)$$

Here:

- F represents the j -th feature map of the l -th layer.
- w signifies the convolutional kernel filter.
- b represents the bias.

In the training phase of CNNs employing the Backpropagation technique, convolutional kernel filters (w), responsible for extracting features from input images, undergo initialization with small random values before the commencement of training. Throughout the forward pass of the backpropagation process, these filters are applied to input dermoscopic images, resulting in the generation of feature maps. Each feature map captures unique image patterns. The optimization technique of gradient descent is then deployed to adjust the filter weights, with gradients supplying crucial information on how the filters should be updated to minimize the loss function. The learning rate, recognized as a hyperparameter, dictates the step size during this process. Filters undergo iterative updates across epochs by subtracting a fraction of the gradient from their current values. This iterative process enables the neural network to progressively learn and refine features tailored specifically for the task at hand, which, in this study, is melanoma classification.

Similarly, biases (b) also commence with small random values and are incorporated into convolutional outputs and layer activations during the forward pass. Loss gradients associated with biases are computed, and gradient descent is utilized to update biases, aiming to minimize the loss.

3.1.2 Batch normalization layer

The primary goal of batch normalization is to alleviate issues related to data saturation, enabling neural networks to achieve faster convergence rates and greater resilience against problems associated with parameter initialization [27]. By integrating batch normalization layers between convolutional layers and nonlinearities, such as ReLU layers, the network training process is accelerated, and overall stability is enhanced.

3.1.3 Rectified Linear Units (ReLUs)

Rectified Linear Units (ReLUs) function as activation layers within CNNs, expediting the neural network training phase while minimizing errors. In instances where a pixel's value (x) within an image falls below zero, the ReLU activation promptly sets all pixel values to zero [28], as illustrated in Eq. (2).

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (2)$$

The pivotal characteristic of Rectified Linear Units (ReLUs) lies in their ability to introduce non-linearity into deep neural networks, a crucial aspect enabling the network to recognize and comprehend complex patterns and relationships within the input data. This non-linearity is vital for CNNs, as it enables the modeling and learning of intricate features within input images. Additionally, ReLUs play a role in mitigating the vanishing gradient problem, a significant issue often encountered in deep neural networks [28].

3.1.4 Pooling layer

The pooling layer provides several advantages, including controlling the size of the output volume on the feature map to prevent overfitting. Commonly positioned after multiple convolution layers within the CNN architecture, this layer performs data reduction through mean- or max-pooling. Mean-pooling calculates the average value, while max-pooling selects the highest value among elements within a small neighborhood.

3.1.5 Fully Connected (FC) layer

In the + data through a flattening process [29]. Following this process, the logistic regression technique can be utilized with softmax activation for the classification of more than two groups.

3.1.6 Optimizer

Hyperparameters play a crucial role in model performance as they require adjustment during model training. In this study, we utilize the Stochastic Gradient Descent (SGD) optimizer to fine-tune the CNN networks. This optimization technique leverages each training sample as a parameter to improve the model's performance by utilizing advanced mathematical functions such as derivatives or subderivatives [30]. The SGD update equation can be expressed as follows:

$$\theta_{(i+1)} = \theta_{i-\alpha} \nabla L(\theta_i) + \gamma(\theta_i - \theta_{(i-1)}) \quad (3)$$

Here:

- $\theta_{(i+1)}$ represents the updated parameter vector.
- θ_i denotes the current parameter vector.
- α signifies the learning rate.
- $\nabla L(\theta_i)$ is the gradient of the loss function with respect to the current parameters θ_i .
- γ controls the momentum and influences how the previous gradient step contributes to the current iteration.
- $\theta_{(i-1)}$ represents the parameter vector from the previous iteration.

3.1.7 The proposed CNN network for classifying melanoma in dermoscopic images

Figure 2 illustrates the structure and configuration of the proposed network, consisting of 27 layers. Operating on skin images with a resolution of 128×128 pixels, the network incorporates three hidden layers. The number of output channels in each hidden layer is derived by convolving the image using 3×3 filters, resulting in respective values of 8, 16, 32, 64, 128, and 256. At each layer, the activation process involves the application of ReLU activation and Max pooling, which serves to downsize the image.

The output of the fully connected layer undergoes normalization through the softmax activation function. This function generates positive values that sum up to one,

providing classification probabilities used by the classification layer for categorizing inputs into Melanoma and nevi, and for calculating loss.

No	Layers	Shape
1:	Input Image	(128, 128, 3)
2:	Convolution	(3, 8)
3:	BN	BN
4:	ReLU	(3, 8)
5:	Max-Pooling	(2x2)
6:	Convolution	(3, 16)
7:	BN	BN
8:	ReLU	(3, 16)
9:	Max-Pooling	(2x2)
10:	Convolution	(3, 32)
11:	BN	BN
12:	ReLU	(3, 32)
13:	Max-Pooling	(2x2)
14:	Convolution	(3, 64)
15:	BN	BN
16:	ReLU	(3, 64)
17:	Max-Pooling	(2x2)
18:	Convolution	(3, 128)
19:	BN	BN
20:	ReLU	(3, 128)
21:	Max-Pooling	(2x2)
22:	Convolution	(3, 256)
23:	BN	BN
24:	ReLU	(3, 256)
25:	FC	2
26:	Softmax	Softmax
27:	Classification	Classification

Figure 2. The architecture of the proposed CNN network for classifying melanoma in dermoscopic images. It should be noted that this CNN network includes 27 layers

The CNN network proposed in this study is trained using melanoma images obtained from the SIC 2019 and ISIC 2020 challenge datasets. Detailed information about the training and testing images will be expounded upon in Section 5.1.

The training process of the network is facilitated by utilizing the SGD optimizer with momentum, where a predefined number of epochs is set at 50. To improve training effectiveness, the learning rate undergoes reduction by a factor of 0.2 after every 5 epochs. It's important to note that all experiments conducted for this research have been performed using MATLAB.

3.2 Deep transfer learning-based approach

In this study, we utilize transfer learning based on robust pre-trained CNNs to classify melanoma in dermoscopic images. Our approach involves employing ten pre-trained CNN architectures, each renowned for its specific characteristics and advantages: SqueezeNet, GoogLeNet, AlexNet, ResNet-18, ResNet-50, MobileNet-v2, ShuffleNet, NASNet-Mobile, EfficientNetB0, and VGG-19.

The selection of these ten pretrained CNN models in this study was made after careful consideration of their individual strengths and weaknesses, aligning them with the goals of classifying skin cancer in dermoscopic images. Each of these pretrained models has been widely employed in various

applications. Models such as MobileNet and EfficientNet are recognized for their highly efficient use of computational resources. On the other hand, models like ResNet and NASNet-Mobile excel in capturing deep features, potentially improving melanoma classification accuracy.

However, some of the pretrained models mentioned above come with limitations. For instance, VGGNet, GoogleNet, and ResNet demand substantial computational resources, limiting their applicability in resource-constrained environments. Additionally, pretrained models like GoogleNet and ResNet possess complex architectures that can lead to longer training times and convergence challenges.

In the subsequent section, we delve into the rationale behind the selection of these models and provide a more detailed analysis of their strengths and weaknesses for the specific task of melanoma classification in dermoscopic images.

3.2.1 SqueezeNet

SqueezeNet, a formidable architecture, comprises fifteen layers, which include two convolution layers, three max-pooling layers, eight fire layers, one global average pooling layer, and a final output layer with softmax activation [31]. Operating with an input size of 227×227 RGB channels, SqueezeNet incorporates max-pooling after convolution to generalize input images. Its convolution layers employ 3×3 kernels, employing an element-wise activation function that sets all values less than zero to zero [28]. SqueezeNet further leverages fire layers, each encompassing squeeze and expansion stages, maintaining consistent input and output tensor scales.

The inclusion of SqueezeNet in this study stems from its lightweight architecture, rendering it suitable for resource-constrained environments. Its efficient parameter usage and ability to sustain high accuracy are significant strengths. Nonetheless, one limitation is its potential to be less effective in capturing complex features compared to deeper networks.

3.2.2 AlexNet

AlexNet, a widely recognized model for classification and pattern recognition, consists of eight layers, including five convolutional and three fully connected layers. In our adaptation, we align input images with the specifications of AlexNet to ensure compatibility with our data, where the number of classes matches the number of inputs in the original model. The model integrates max-pooling layers positioned between the first two convolutional layers, aiding in the reduction of feature map sizes.

Throughout the training process, a low learning rate has been applied, contributing to smaller weight updates [32]. Unlike some of its counterparts, AlexNet is less computationally intensive, making it an appealing choice. As a pioneering model in the field of deep learning, it was selected for its robustness and established performance. Its relatively simpler architecture and fewer layers contribute to its computational efficiency. However, its architecture might be considered shallow compared to more recent and deeper models, potentially constraining its capacity for feature extraction.

3.2.3 VGGNet

The VGGNet, developed by researchers at Oxford University, is recognized for its pyramidal structure [33]. It comprises a series of convolutional layers followed by pooling layers, contributing to its distinctive architecture. VGGNet is

known for its adaptability and suitability for benchmarking across various tasks. Its pre-trained models are commonly utilized in diverse applications. However, it's important to note that VGGNet can be computationally demanding, particularly when initiated from scratch.

The pyramidal structure and simplicity of VGGNet render it an excellent choice for benchmarking purposes. It excels in feature extraction and finds widespread application. Nonetheless, its computational intensity becomes apparent, especially when dealing with large datasets and in resource-constrained environments.

3.2.4 GoogleNet

GoogleNet, also recognized as the inception architecture, is focused on efficiently estimating and distributing dense components within the sparse structure of a convolutional network. It specifically addresses the redundancy issue within deep network activations, emphasizing that not all connections between input and output channels need to be present in the network's design. GoogleNet utilizes convolutions of different sizes (5×5 , 3×3 , 1×1) to capture data and features at varying scales. It also introduces bottleneck convolutional layers (1×1), which play a pivotal role in its design [34].

The inception architecture of GoogleNet is structured to capture features at different scales, offering a comprehensive understanding of images. It adeptly employs various kernel sizes for convolution. However, its complexity can pose challenges to training convergence and demand significant computational resources.

3.2.5 ResNet

ResNet, which stands for residual network, is distinguished by its use of residual modules as the fundamental building blocks. These residual modules are organized in a stacked manner to constitute a complete end-to-end network. One of the unique aspects of the ResNet architecture is its incorporation of tens of thousands of residual layers, enhancing its performance and efficacy in network training. In comparison to AlexNet and VGG, ResNet is notably deeper, being 20 and 8 times deeper, respectively [35].

The deep architecture of ResNet allows it to effectively capture intricate patterns and features in images. Its residual modules play a crucial role in mitigating the vanishing gradient problem, enabling the training of exceptionally deep networks. However, a notable drawback is the increased computational cost associated with its depth.

3.2.6 MobileNet

MobileNet is distinctive for its efficiency, making it well-suited for deployment on devices with limited computational resources, such as mobile devices and low-powered computers. Its architecture is characterized by a minimalistic design, which has proven effective for various tasks, including Palmprint Recognition [36]. It employs various convolutional layers and abstraction layers that utilize depthwise convolutions. MobileNet integrates ReLU activation components and residual layers with specific stride values, contributing to its unique design [37].

This model is ideal for resource-constrained devices due to its low computational requirements. It maintains good performance while minimizing memory and power consumption. However, due to its compact architecture, it may not capture fine-grained details as effectively as larger and more complex models.

3.2.7 ShuffleNet

ShuffleNet introduces a novel ShuffleNet unit aimed at optimizing small networks by integrating the channel shuffle function. It commences with a bottleneck unit, utilizing a 3×3 depthwise convolution in the residual branch, followed by pointwise group convolution. The goal of this approach is to align the channel dimension between the residual and shortcut paths. The architecture integrates mixed-grouped convolution, point-grouped convolution, and profoundly separable convolution, maintaining accuracy while significantly reducing computational expenses [38].

The distinctive channel shuffle function of ShuffleNet makes it well-suited for small networks. It effectively balances accuracy and computational efficiency. However, it may not perform as strongly as larger models on more complex tasks.

3.2.8 NASNet-Mobile

The Neural Architecture Search (NASNet) is a contemporary CNN architecture developed by GoogleBrain. It utilizes a reinforcement learning search strategy to discover an efficient building block on a small dataset, which is subsequently transferred to a larger dataset, resulting in state-of-the-art performance with a smaller model size and complexity. In contrast to traditional CNN designs, NASNet explores cells that can form a high-performance block, with the internal structure of these cells determined by a Recurrent Neural Network. This approach involves constructing Normal and Reduction cells and iteratively repeating the building block, with the value of 'n' being automatically calculated.

NASNet's automated architecture search generates efficient models tailored to specific datasets, effectively balancing performance and model size, rendering it suitable for various applications. However, it may require significant computational resources during the search process.

3.2.9 EfficientNet

EfficientNet diverges from traditional CNN design approaches by focusing on augmenting network depth, width, and input resolution in the baseline network. It employs a multi-objective neural architecture search to maximize both accuracy and constrained computational resources. The baseline network, EfficientNetB0, is crafted through this process by utilizing a slightly larger version of the mobile inverted bottleneck convolution (MBConv) and scaling it to create the EfficientNet family of models. This innovative methodology facilitates improved accuracy without significantly inflating computational demands [39].

EfficientNet's strategy for scaling network depth, width, and resolution strikes an optimal balance between accuracy and computational efficiency. It excels in both feature extraction and model size. However, fine-tuning may be necessary to tailor it to specific tasks or datasets.

3.3 Deep transfer learning-based approach for melanoma classification

Training a deep CNN model demands access to a vast image repository [40]. Regrettably, datasets comprising a substantial collection of labeled skin lesion images remain limited. Leveraging the capabilities of deep CNNs often necessitates training on expansive datasets like ImageNet, a process fraught with significant challenges. However, we overcame this challenge by employing transfer learning with robust pre-trained CNN models, including SqueezeNet, AlexNet, VGG-

16, GoogleNet, ResNet, MobileNet, ShuffleNet, NASNet-Mobile, and EfficientNet, for the task of melanoma classification.

Notably, our approach introduces a two-classification model specifically tailored to differentiate between malignant and benign skin lesions, contrary to CNN models trained on ImageNet, which are designed for distinguishing among 1000 classes. Thus, three critical stages are involved in achieving effective transfer learning for melanoma classification in dermoscopic images:

1. The initial stage involves selecting a pre-trained model, such as SqueezeNet, AlexNet, VGG-16, GoogleNet, ResNet, MobileNet, ShuffleNet, NASNet-Mobile, EfficientNet. A summary of the strengths and limitations of each model is provided in Section 3.2.
2. In the second stage, we replace the model's classification layer with a SoftMax layer, which involves adjusting weights to substitute class-specific attributes with new layers. This step also encompasses the relocation and replacement of feature maps in convolutional layers with class-specific counterparts.
3. The final stage involves deploying fine-tuned layers that are reused and adjusted to suit the requirements of the melanoma classification task.

4. PERFORMANCE EVALUATION

To evaluate the performance of the proposed methods in classifying melanoma and benign skin lesions, we calculate the confusion matrix presented in Figure 3. We then compute various metrics, including accuracy, recall, precision, F1 score, and the area under the curve (AUC). Eqs. (4)-(7) detail the formulation of these metrics [1].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$F - score = 2 \times \frac{Recall+Precision}{Recall \times Precision} \tag{7}$$

In this context, True Positive (TP) signifies accurate positive predictions, while True Negative (TN) represents cases where negatives were correctly predicted. False Positive (FP) indicates instances where negatives were erroneously identified as positives, and False Negative (FN) pertains to positive data mistakenly considered negative. A favorable outcome is characterized by a high true negative rate and low false-positive rates, positioning most points in the left section of the receiver operating characteristic (ROC) curve [41].

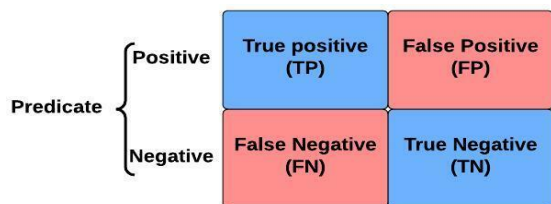


Figure 3. An illustration for the confusion matrix calculations

5. RESULT AND DISCUSSION

5.1 Dataset

The dataset was procured from the Kaggle platform, primarily sourced from the International Skin Imaging Collaboration (ISIC), a global initiative dedicated to advancing melanoma diagnosis and hosting the most extensive publicly accessible repository of high-quality dermoscopic images of skin lesions within the ISIC Archive. The acquired dataset comprises images of melanoma (malignant) and nevus (benign) from the ISIC 2019 and ISIC 2020 Challenge Datasets. Notably, the ISIC 2020 dataset associates each image with a unique patient identity. In total, our dataset comprises 11,449 dermoscopic images encompassing 5,106 melanoma cases and 6,343 nevus cases.

These images are in the ".jpg" color format and exhibit variations in pixel sizes, as depicted in Figure 4. Figure 4(a) provides an example of melanoma images, while Figure 4(b) displays a nevus skin lesion.

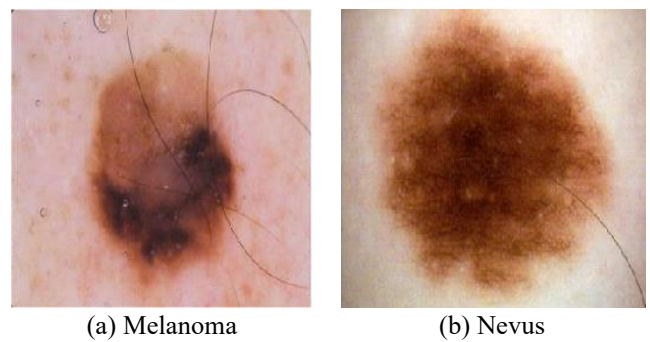


Figure 4. Examples of melanoma and nevus skin lesions from the ISIC dataset

In our experiments, we divided the dataset into 75% for training and 25% for testing. The dataset exhibits a semi-balanced distribution, with 3,830 melanoma images and 4,757 nevus images allocated for training, and 1,276 melanoma images and 1,586 nevus images designated for testing.

The decision to utilize a semi-balanced dataset in our experiments is motivated by the need to strike a harmonious balance between mitigating class imbalance issues and optimizing the overall performance of the model. In the context of melanoma classification, there typically exists a substantial class imbalance between malignant (melanoma) and benign (nevus) skin lesion cases. In real-world scenarios, the occurrence of melanoma cases is significantly less frequent compared to benign cases. Consequently, training a model on a severely imbalanced dataset may result in suboptimal performance, as the model might develop a bias towards the majority class (nevus) and encounter challenges in effectively learning patterns associated with the minority class (melanoma).

By adopting a semi-balanced dataset, our goal is to address this class imbalance challenge while still maintaining a representation of real-world conditions. This approach involves stratified sampling to ensure that both melanoma and nevus cases are adequately represented in both the training and validation sets. By doing so, we provide the model with a more balanced exposure to both classes during training, fostering better generalization and improved performance in identifying melanoma cases.

5.2 Results of the proposed CNN model

In this investigation, we systematically trained four distinct configurations of the proposed CNN model, each distinguished by varying numbers of layers: 15, 19, 23, and 27. The primary objective was to ascertain the optimal configuration that would maximize classification accuracy. As detailed in Section 3.1.7, we introduced the comprehensive CNN model for melanoma classification (Configuration 1), comprising 27 layers, as illustrated in Figure 2. For the remaining three configurations (Configuration 2, Configuration 3, and Configuration 4), we progressively removed Convolution, ReLU, Batch Normalization (BN), and Max-Pooling layers from the complete 27-layer CNN network. Specifically, for Configuration 2 (23 layers), we eliminated layers 18-21 (four layers) from the complete network, as depicted in Figure 2. In the case of Configuration 3 (19 layers), we pruned eight layers from the complete network (layers 14-21). Finally, for Configuration 4, we pruned 12 layers from the complete network (layers 10-21).

The performance of these four CNN configurations was meticulously evaluated, and the results are presented in Table 2. Each configuration underwent assessment using fundamental metrics, including accuracy, AUC, recall, precision, and F1-score, to gauge their suitability for the task of skin cancer classification.

Significantly, Configuration 1 emerged as the top performer, showcasing an impressive accuracy rate of 90.85%. Following closely, Configuration 2 achieved an accuracy of 90.74%. While Configurations 3 and 4 exhibited slightly lower accuracies of 90.32% and 89.55%, respectively, it is noteworthy that even the least performing Configuration 4 demonstrated commendable levels of accuracy. Regarding the Area Under the Curve (AUC), both Configuration 1 and Configuration 2 took the lead, registering an AUC of 0.907. This implies that these configurations adeptly discriminated between malignant and benign skin lesions.

Interestingly, Configuration 4, boasting the fewest layers, secured the highest recall rate at 94.83%, indicating its

effectiveness in capturing malignant lesions. However, this heightened sensitivity came at the cost of lower precision. Conversely, Configuration 1 achieved a balanced recall rate of 91.80%, while Configurations 2 and 3 achieved recall rates of 91.17% and 89.72%, respectively. Precision, a metric reflecting the ability to correctly classify instances as positive, revealed that Configuration 2 exhibited the highest precision at 92.58%, closely followed by Configuration 3 at 92.04%.

Table 2. The performance of the four configurations of the proposed CNN model, in terms of accuracy (Acc.), AUC, Recall (Re.), Precision (Pre.) and F1-score (F1)

Configuration	Acc.	AUC	Re.	Pre.	F1
C1: 27 layers	90.85	0.907	91.80	91.69	91.75
C2: 23 layers	90.74	0.907	91.17	92.04	91.61
C3: 19 layers	90.32	0.904	89.72	92.58	91.13
C4: 15 layers	89.55	0.889	94.83	87.39	90.96

Figure 5 further illustrates the confusion matrix of these four CNN configurations. Notably, Configuration 4 produced the highest false-positive rate, incorrectly classifying 217 nevus images as melanoma. Conversely, Configuration 3 demonstrated the highest false-negative rate, with 136 melanoma cases misclassified as nevus.

These results underscore Configuration 1, comprising 27 layers, as the most balanced performer, excelling in accuracy and maintaining a harmonious balance between precision and recall. Configuration 2, featuring 23 layers, closely follows Configuration 1 in performance and offers the advantage of reduced complexity. Configuration 3, with 19 layers, emphasizes precision, while Configuration 4, with 15 layers, prioritizes recall. The choice of configuration should be guided by specific application requirements, such as the importance of minimizing false positives or maximizing sensitivity. Nevertheless, all configurations underscore the remarkable potential of the proposed CNN model in skin lesion classification, demonstrating its effectiveness for this critical medical task.

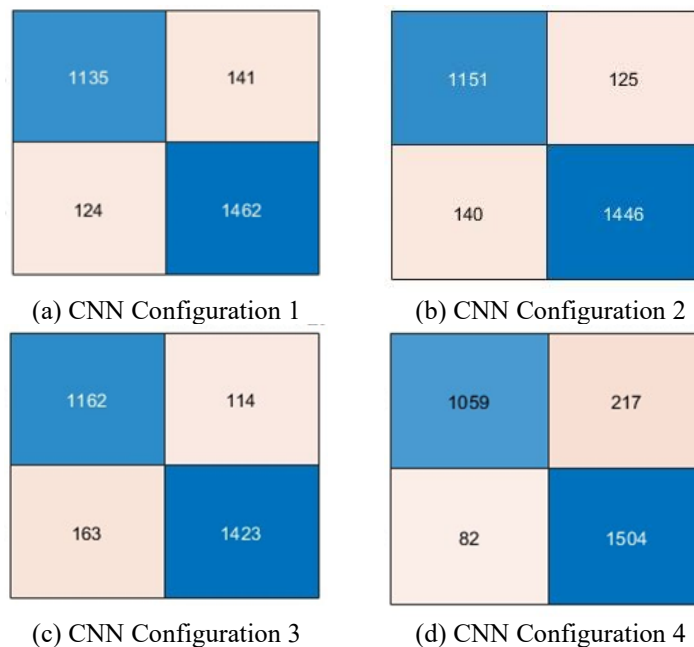


Figure 5. The confusion matrix of the four CNN configurations: (a) Configuration 1 (C1) with 27 layers, (b) Configuration 2 (C2) with 23 layers, (c) Configuration 3 (C3) with 19 layers, and (d) Configuration 4 (C4) with 15 layers

5.3 Results of deep transfer learning-based approach

Table 3. The performance of the transfer learning approach for melanoma classification utilizing the SqueezeNet, GoogLeNet, AlexNet, ResNet-18, ResNet-50, MobileNet-v2, ShuffleNet, NASNet-Mobile, EfficientNetB0, and VGG-19 pretrained CNN models, in terms of accuracy (Acc.), AUC, Recall (Re.), Precision (Pre.) and F1-score (F1)

Method	Acc.	AUC	Re.	Pre.	F1
SqueezeNet	80.89	0.787	98.36	74.96	85.08
GoogLeNet	80.96	0.789	97.16	75.50	84.97
AlexNet	92.56	0.923	94.77	92.04	93.38
ResNet-18	91.89	0.916	93.82	91.74	92.77
ResNet-50	92.98	0.927	95.08	92.46	93.75
MobileNet-v2	90.43	0.904	90.61	92.00	91.30
ShuffleNet	89.73	0.900	87.39	93.65	90.41
NASNet	91.26	0.908	94.83	89.95	92.33
EfficientNet	91.20	0.908	94.20	90.33	92.22
VGG-19	91.79	0.911	97.54	88.76	92.94

Table 3 provides a comprehensive overview of the outcomes derived from the deep transfer learning approach employed for melanoma classification in dermoscopic images. The experimentation involved leveraging ten pretrained CNN models, all of which had undergone training on ImageNet. These models encompassed SqueezeNet, GoogLeNet, AlexNet, ResNet-18, ResNet-50, MobileNet-v2, ShuffleNet, NASNet-Mobile, EfficientNetB0, and VGG-19. The foundation for transfer learning was laid using the dermoscopic image dataset outlined in Section 5.1, thoughtfully partitioned into a 75% training set and a 25%

testing set. All images underwent resizing to dimensions of $227 \times 227 \times 3$.

As evidenced by Table 3, ResNet-50 emerges as the leading performer among the models, showcasing a remarkable classification accuracy of 92.98%, an AUC of 0.927, and an F1-score of 93.75%. Following closely, AlexNet secures the second-best results with a classification accuracy of 92.56%. Concurrently, NASNet and EfficientNetB0 achieve F1-scores that are approximately one point lower than ResNet-50.

It is noteworthy that models such as SqueezeNet and GoogLeNet exhibit relatively lower classification performance, with accuracy rates falling below 81% and AUC values dropping below 0.79. The architectural designs and depth of ResNet-50 and AlexNet play a significant role in their superior performance in melanoma classification. These models demonstrate proficiency in capturing both low-level and high-level features crucial for this intricate task. Conversely, the intricacy of models like GoogleNet and SqueezeNet, coupled with their emphasis on parameter efficiency, may not ideally align with the characteristics of the dataset, resulting in comparatively lower performance.

Moreover, Figure 6 provides a graphical representation of the confusion matrix for each pretrained CNN model. Notably, SqueezeNet demonstrates the lowest false negative (FN) rate, misclassifying only 26 melanoma images as nevus cases. In contrast, ShuffleNet produces the highest FN rate, misclassifying 200 melanoma images as nevus cases. Both SqueezeNet and GoogLeNet exhibit the highest false positive (FP) rates, inaccurately categorizing over 500 nevus images as melanoma. Among the top-performing models, ResNet-50 and AlexNet display FP rates of 123 and 130, respectively.

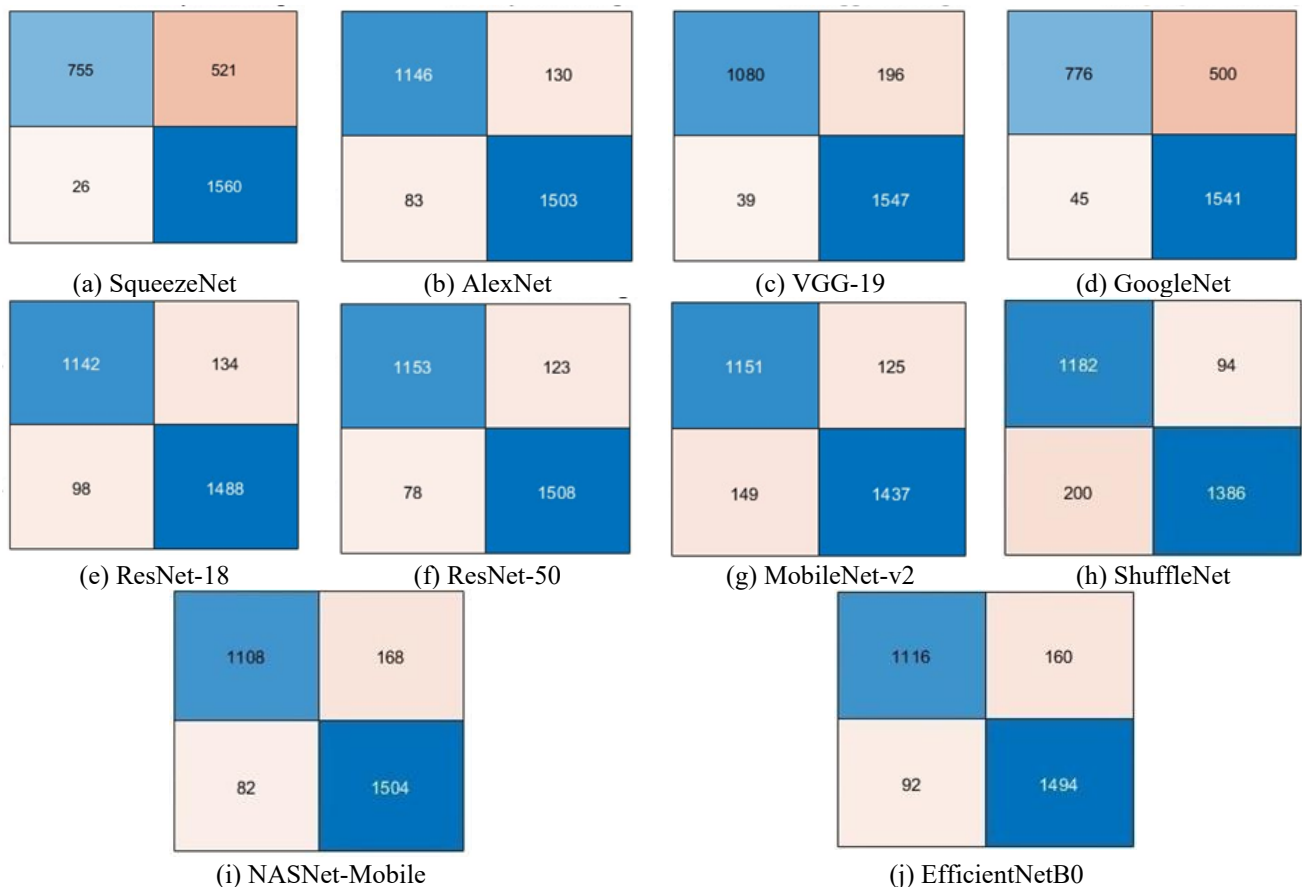


Figure 6. The confusion matrix of each pretrained CNN model: (a) SqueezeNet, (b) AlexNet, (c) VGG-19, (d) GoogleNet, (e) ResNet-18, (f) ResNet-50, (g) MobileNet-v2, (h) ShuffleNet, (i) NASNet-Mobile, and (j) EfficientNetB0

Significantly, ShuffleNet achieves the lowest FP rate, erroneously classifying 94 nevus images as melanoma. It's crucial to highlight that the relatively high FN rate of ShuffleNet impacts its overall performance. ShuffleNet's architectural choices may not optimally align with the characteristics of the dataset used. Dermoscopic images exhibit significant variation in texture, color, and lesion size. A conservative model like ShuffleNet might struggle to generalize effectively across this diversity, leading to misclassifications.

In summary, ResNet-50 and AlexNet emerge as the top-performing CNN models for melanoma classification. Both models excel in accuracy, AUC, recall, precision, and F1-score, demonstrating their effectiveness. The architectural designs and depth of these models significantly contribute to their superior performance, enabling them to capture both low-level and high-level features crucial for melanoma detection. On the other hand, SqueezeNet and GoogLeNet, while achieving high recall rates, exhibit lower precision and, consequently, lower F1-scores. This implies that SqueezeNet and GoogLeNet produce a higher rate of false positives, a critical consideration in diagnosing melanoma, where minimizing false positives is essential. The choice of CNN architecture proves pivotal in obtaining precise and reliable melanoma classification results, with ResNet-50 and AlexNet exhibiting strong suitability for the melanoma classification task.

5.4 Comparisons

In Table 4, we present a thorough comparison between our optimized CNN model's best configuration and the deep transfer learning approach utilizing ResNet-50 pretrained on ImageNet. This comparative analysis illuminates the strengths and trade-offs inherent in each approach.

Table 4. Comparing the proposed CNN model with the deep transfer learning approach based on ResNet-50, in terms of accuracy (Acc.), AUC, Recall (Re.), Precision (Pre.) and F1-score (F1)

No. of Layers	Acc.	AUC	Re.	Pre.	F1
Proposed CNN	90.85	0.907	91.80	91.69	91.75
ResNet-50	92.98	0.927	95.08	92.46	93.75

ResNet-50 achieves an impressive accuracy of 92.98%, underscoring its robust classification capabilities. This accuracy is approximately 2 percentage points higher than that of our proposed CNN model, which attains 90.85%. It is noteworthy that although there exists a discrepancy in accuracy, the margin is not substantially large, especially considering the significant contrast in complexity between the two models.

Importantly, in terms of the recall rate, ResNet-50 exhibits superior performance compared to our CNN model, indicating its enhanced ability to correctly identify melanoma cases. However, both methods exhibit similar precision rates, emphasizing their capability to accurately classify melanoma cases while minimizing false positives.

This comparative analysis establishes that our proposed CNN model yields result on par with the transfer learning approach employing ResNet-50. What distinguishes our CNN model is its notably lower complexity and lighter computational cost. These attributes render it particularly

well-suited for resource-limited devices or environments where computational resources are constrained. Consequently, the choice between the two approaches should hinge on the specific requirements of the application at hand.

Additionally, we conducted a thorough comparison to assess the accuracy of our proposed deep transfer learning-based method against the findings from a previous study. In Table 5, we juxtapose the accuracies achieved by our deep transfer learning approach, utilizing ten pretrained CNN models, with the results obtained by Fraiwan and Faouri [12]. This comparative analysis provides valuable insights into the effectiveness of our approach in relation to existing research.

Notably, Fraiwan and Faouri [12] reported their best accuracy scores with ResNet-18 and ShuffleNet, both achieving an accuracy of 79%. In contrast, our study attains the highest classification accuracy with ResNet-50, reaching an impressive accuracy rate of 92.98%. It's important to highlight that out of the ten pretrained CNN models utilized in our research, three models—AlexNet, NASNet, and VGG-19—were not employed in study of Fraiwan and Faouri [12]. The accuracy rates of all melanoma classification models developed in our study using these ten pretrained CNN models are notably higher, ranging from 5 to 17 percentage points above the results reported by Fraiwan and Faouri [12]. This substantial performance gap underscores the efficacy of the proposed method in enhancing melanoma classification.

Table 5. A comparative analysis of the accuracies achieved by our deep transfer learning approach utilizing ten pretrained CNN models and the accuracies reported by Fraiwan and Faouri [12]

Model	Our Accuracy	Accuracy by Fraiwan and Faouri [12]
SqueezeNet	80.89	75.00
GoogLeNet	80.96	73.40
AlexNet	92.56	Not Used
ResNet-18	91.89	79.00
ResNet-50	92.98	77.80
MobileNet-v2	90.43	74.90
ShuffleNet	89.73	79.00
NASNet	91.26	Not Used
EfficientNet	91.20	76.70
VGG-19	91.79	Not Used

6. CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In this comprehensive investigation of melanoma classification in dermoscopic images, several noteworthy findings have surfaced. The study employed both custom CNN models and pretrained CNN models with transfer learning, enabling a thorough performance evaluation. The proposed CNN model demonstrated a noteworthy accuracy of 90.85%, effectively distinguishing between malignant and benign skin lesions. Conversely, ResNet-50, a deep transfer learning-based model pretrained on ImageNet, achieved a slightly higher accuracy of 92.98%, surpassing the proposed CNN model by approximately 2 percentage points. This discrepancy, although significant, remains modest when considering the substantial difference in model complexity. Furthermore, ResNet-50 exhibited a superior recall rate compared to the

proposed CNN model, while both models displayed comparable precision rates.

The study's outcomes underscore the potential of deep transfer learning methods, particularly with models such as ResNet-50 and AlexNet, to enhance melanoma classification accuracy in dermoscopic images. Noteworthy is the proposed CNN model, which, while delivering competitive results, distinguishes itself through its lighter and less complex architecture, rendering it suitable for resource-limited devices. In a comparative analysis, our results were juxtaposed with a study by Fraiwan et al. [12], where ResNet-18 and ShuffleNet achieved the highest accuracy at 79%. In contrast, our study outperformed ResNet-50, achieving an exceptional accuracy rate of 92.98%. These considerable performance improvements underscore the ongoing significance of research efforts in leveraging deep learning for melanoma diagnosis and skin cancer in general. This research has the potential to enhance early detection, thereby improving patient outcomes.

The implications of this study in the healthcare and dermatology fields are substantial. These implications encompass: 1) Advancements in skin cancer diagnosis, 2) The adaptability of the proposed CNN model's lighter architecture for deployment on resource-limited devices, including smartphones, 3) A potential reduction in biopsy rates, and 4) Progress in telemedicine applications. These findings hold the promise to revolutionize dermatology by enhancing skin cancer diagnosis, particularly in cases of potentially life-threatening melanoma. Furthermore, the deployment of accurate and resource-efficient deep learning models, such as the proposed CNN model, could lead to improved patient care, reduced healthcare costs, and increased accessibility to skin lesion assessments. Ultimately, this would benefit both patients and healthcare providers.

6.2 Future work

The future endeavors stemming from this study will concentrate on several key aspects:

1. Extending Classification to Various Skin Diseases: The deep learning models elucidated in this study will be harnessed for the classification of diverse skin conditions, encompassing basal cell carcinoma, squamous cell carcinoma, and a spectrum of dermatological ailments such as psoriasis and eczema. Recognizing the unique challenges presented by each skin disease in dermoscopic images, our focus will be on acquiring more extensive and varied datasets. This expansion aims to ensure accurate classification across a spectrum of skin diseases.
2. Architectural Refinement and Hyperparameter Optimization: The architecture of the CNN models, alongside critical hyperparameters such as learning rates, will be subject to meticulous tuning and optimization. This refinement process seeks to enhance the efficacy of melanoma classification results, pushing the boundaries of model performance.
3. Enhancing Model Interpretability: Strategies for enhancing the interpretability of the developed models in melanoma classification will be explored. This involves employing diverse methods to make the decision-making process of CNN models more transparent. One such approach includes generating heatmaps to highlight regions of interest within skin lesion images. Additionally, textual explanations for the decisions made by each CNN model will be explored, contributing to a more

comprehensive understanding of the model's classification rationale.

REFERENCES

- [1] Fu'adah, Y.N., Pratiwi, N.C., Pramudito, M.A., Ibrahim, N. (2020). Convolutional Neural Network (CNN) for automatic skin cancer classification system. In IOP Conference Series: Materials Science and Engineering 982(1): 012005. <https://doi.org/10.1088/1757-899X/982/1/012005>
- [2] Kong, B., Sun, S., Wang, X., Song, Q., Zhang, S. (2018). Invasive cancer detection utilizing compressed Convolutional Neural Network and transfer learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 156-164. https://doi.org/10.1007/978-3-030-00934-2_18
- [3] Elgamel, M. (2013). Automatic skin cancer images classification. International Journal of Advanced Computer Science and Applications, 4(3): 287-294.
- [4] Murphy, M., Mabruk, M.J.E.M.F., Lenane, P., Liew, A., McCann, P., Buckley, A., Murphy, G.M. (2002). Comparison of the expression of p53, p21, Bax and the induction of apoptosis between patients with basal cell carcinoma and normal controls in response to ultraviolet irradiation. Journal of Clinical Pathology, 55(11): 829-833. <https://doi.org/10.1136/jcp.55.11.829>
- [5] Salah, B., Alshraideh, M., Beidas, R., Hayajneh, F. (2011). Skin cancer recognition by using a neuro-fuzzy system. Cancer Informatics, 10: CIN-S5950. <https://doi.org/10.4137/CIN.S5950>
- [6] Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., Mahnashi, M.H. (2021). Skin cancer detection: A review using deep learning techniques. International Journal of Environmental Research and Public Health, 18(10): 5479. <https://doi.org/10.3390/ijerph18105479>
- [7] Zhang, Y., Wang, S., Phillips, P., Dong, Z., Ji, G., Yang, J. (2015). Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC. Biomedical Signal Processing and Control, 21: 58-73. <https://doi.org/10.1016/j.bspc.2015.05.014>
- [8] Kasmir, R., Mokrani, K. (2016). Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Processing, 10(6): 448-455. <https://doi.org/10.1049/iet-ipr.2015.0385>
- [9] Friedman, R.J., Rigel, D.S., Kopf, A.W. (1985). Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA: A Cancer Journal for Clinicians, 35(3): 130-151. <https://doi.org/10.3322/canjclin.35.3.130>
- [10] Khan, M.Q., Hussain, A., Rehman, S.U., Khan, U., Maqsood, M., Mehmood, K., Khan, M.A. (2019). Classification of melanoma and nevus in digital images for diagnosis of skin cancer. IEEE Access, 7: 90132-90144. <https://doi.org/10.1109/ACCESS.2019.2926837>
- [11] Masood, A., Ali Al-Jumaily, A. (2013). Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms. International Journal of Biomedical Imaging, 2013: 323268. <https://doi.org/10.1155/2013/323268>

- [12] Fraiwan, M., Faouri, E. (2022). On the automatic detection and classification of skin cancer using deep transfer learning. *Sensors*, 22(13): 4963. <https://doi.org/10.3390/s22134963>
- [13] Tan, T.Y., Zhang, L., Jiang, M. (2016). An intelligent decision support system for skin cancer detection from dermoscopic images. In 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, pp. 2194-2199. <https://doi.org/10.1109/FSKD.2016.7603521>
- [14] Mukherjee, S., Adhikari, A., Roy, M. (2019). Malignant melanoma identification using best visually imperceptible features from Dermofit dataset. In *Advances in Computer, Communication and Control: Proceedings of ETES 2018*, pp. 263-274. https://doi.org/10.1007/978-981-13-3122-0_25
- [15] Ballerini, L., Fisher, R.B., Aldridge, B., Rees, J. (2013). A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. *Color Medical Image Analysis*, 63-86. https://doi.org/10.1007/978-94-007-5389-1_4
- [16] Mukherjee, S., Adhikari, A., Roy, M. (2020). Malignant melanoma detection using multi layer perceptron with visually imperceptible features and PCA components from MED-NODE dataset. *International Journal of Medical Engineering and Informatics*, 12(2): 151-168. <https://doi.org/10.1504/IJMEI.2020.106899>
- [17] Mukherjee, S., Adhikari, A., Roy, M. (2019). Melanoma identification using MLP with parameter selected by metaheuristic algorithms. In *Intelligent Innovations in Multimedia Data Engineering and Management*, pp. 241-268. <https://doi.org/10.4018/978-1-5225-7107-0.ch010>
- [18] Ali, R., Hardie, R.C., De Silva, M.S., Kebede, T.M. (2019). Skin lesion segmentation and classification for ISIC 2018 by combining deep CNN and handcrafted features. *arXiv preprint arXiv:1908.05730*. <https://doi.org/10.48550/arXiv.1908.05730>
- [19] Nasr-Esfahani, E., Samavi, S., Karimi, N., Soroushmehr, S.M.R., Jafari, M.H., Ward, K., Najarian, K. (2016). Melanoma detection by analysis of clinical images using Convolutional Neural Network. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, pp. 1373-1376. <https://doi.org/10.1109/EMBC.2016.7590963>
- [20] Pomponiu, V., Nejati, H., Cheung, N.M. (2016). Deepmole: Deep neural networks for skin mole lesion classification. In 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp. 2623-2627. <https://doi.org/10.1109/ICIP.2016.7532834>
- [21] Ayan, E., Ünver, H.M. (2018). Data augmentation importance for classification of skin lesions via deep learning. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, pp. 1-4. <https://doi.org/10.1109/EBBT.2018.8391469>
- [22] Kwasigroch, A., Mikołajczyk, A., Grochowski, M. (2017). Deep neural networks approach to skin lesions classification—A comparative analysis. In 2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, Poland, pp. 1069-1074. <https://doi.org/10.1109/MMAR.2017.8046978>
- [23] Maia, L.B., Lima, A., Pereira, R.M.P., Junior, G.B., de Almeida, J.D.S., de Paiva, A.C. (2018). Evaluation of melanoma diagnosis using deep features. In 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), Miedzyzdroje, Poland, pp. 1-4. <https://doi.org/10.1109/IWSSIP.2018.8439373>
- [24] Lopez, A.R., Giro-i-Nieto, X., Burdick, J., Marques, O. (2017). Skin lesion classification from dermoscopic images using deep learning techniques. In 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), Maribor, Slovenia, pp. 49-54. <https://doi.org/10.2316/P.2017.852-053>
- [25] Ali, A.A., Al-Marzouqi, H. (2017). Melanoma detection using regular convolutional neural networks. In 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Innsbruck, Austria, pp. 1-5. <https://doi.org/10.1109/ICECTA.2017.8252041>
- [26] Kim, P. (2017). *Convolutional Neural Network*. In: *MATLAB Deep Learning*. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-2845-6_6
- [27] Wu, S., Zhong, S., Liu, Y. (2018). Deep residual learning for image steganalysis. *Multimedia Tools and Applications*, 77: 10437-10453. <https://doi.org/10.1007/s11042-017-4440-4>
- [28] Agarap, A.F. (2018). Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*. <https://doi.org/10.48550/arXiv.1803.08375>
- [29] Khan, S., Rahmani, H., Shah, S.A.A., Bennamoun, M., Medioni, G., Dickinson, S. (2018). A guide to Convolutional Neural Networks for computer vision. *Synthesis Lectures on Computer Vision*, 8(1): 1-207. <https://doi.org/10.1007/978-3-031-01821-3>
- [30] Robbins, H., Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400-407. <https://www.jstor.org/stable/2236626>
- [31] Ucar, F., Korkmaz, D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical Hypotheses*, 140: 109761. <https://doi.org/10.1016/j.mehy.2020.109761>
- [32] Ashraf, R., Afzal, S., Rehman, A.U., Gul, S., Baber, J., Bakhtyar, M., Maqsood, M. (2020). Region-of-interest based transfer learning assisted framework for skin cancer detection. *IEEE Access*, 8: 147858-147871. <https://doi.org/10.1109/ACCESS.2020.3014701>
- [33] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- [34] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [35] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [36] Michele, A., Colin, V., Santika, D.D. (2019). MobileNet Convolutional Neural Networks and support vector

- machines for palmprint recognition. *Procedia Computer Science*, 157: 110-117. <https://doi.org/10.1016/j.procs.2019.08.147>
- [37] Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, 21(8): 2852. <https://doi.org/10.3390/s21082852>
- [38] Zhang, X., Zhou, X., Lin, M., Sun, J. (2018). Shufflenet: An extremely efficient Convolutional Neural Network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, pp. 6848-6856. <https://doi.org/10.1109/CVPR.2018.00716>
- [39] Miglani, V., Bhatia, M.P.S. (2020). Skin lesion classification: A transfer learning approach using efficientnets. In *International Conference on Advanced Machine Learning Technologies and Applications*, pp. 315-324. https://doi.org/10.1007/978-981-15-3383-9_29
- [40] Hosny, K.M., Kassem, M.A., Foad, M.M. (2018). Skin cancer classification using deep learning and transfer learning. In *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, Cairo, Egypt, pp. 90-93. <https://doi.org/10.1109/CIBEC.2018.8641762>
- [41] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8): 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>