CNN-BASED ENSEMBLE ARCHITECTURES WITH EXPLANAINABLE AI FOR CUTANEOUS MELANOMA IDENTIFICATION

Adrian SZYMCZYK*, Maria SKUBLEWSKA-PASZKOWSKA*, Pawel POWROZNIK*

*Department of Computer Science, Lublin University of Technology, Nadbystrzycka 38A, 20-618 Lublin Poland

adrian.szymczyk@pollub.edu.pl, maria.paszkowska@pollub.pl, p.powroznik@pollub.pl

received 23 December 2025, revised 29 April 2026, accepted 02 May 2026

Abstract: Malignant melanoma, a highly aggressive form of skin cancer, poses a significant global health challenge due to its rapid progression and high mortality rate if not detected on time. Early diagnosis is crucial for improving patient outcomes. The effectiveness of skin cancer detection still faces serious challenges, like visual inspection that is less accurate and time-consuming. However, deep learning-based models provide early and accurate diagnosis, serving as a supporting tool for dermatologists. Thus, this study focuses on indicating the most suitable model for skin diseases identification. Three prominent, pre-trained deep learning models, ResNet152, DenseNet201 and EfficientNet-B4, were involved in order to detect benign and malignant melanoma skin lesions. The study was performed utilizing a combined ISIC datasets gathered between 2018 and 2020 that consist of dermoscopic images. The above-mentioned deep learning algorithms were verified using accuracy, precision, recall, and F1-score metrics. Moreover, in this study the performance of skin cancer detection was enhanced utilizing soft, hard voting, and XGBoost ensemble learning methods. Combining two and three models were verified. The single models obtained accuracy at the level of 89.20%, 88.20%, and 90.40% for ResNet152, DenseNet201 and EfficientNet-B4, respectively. The soft voting ensemble, merging ResNet-152 with EfficientNet-B4 or all three models, achieved the highest absolute accuracy of 91.30%, demonstrating superior performance in melanoma diagnosis compared to individual models. Hard voting and XGBoost stated to be less effective in melanoma diagnosis. To confirm that the models were making decisions based on the significant image regions representing skin lesions, a visual explainable technique was applied. Gradient-weighted Class Activation Mapping proved the models to focus their attention to the relevant disease features. These findings underscore the potential of combining individual model strengths through ensemble learning to achieve superior diagnostic performance in melanoma detection, supporting clinicians in making more accurate and timely diagnosis.

Key words: melanoma, skin cancer identification, deep learning, convolutional neural networks, ensemble learning, Grad-CAM

1. INTRODUCTION

Malignant melanoma is one of the most aggressive and deadly forms of skin cancer, with its incidence rising rapidly worldwide, particularly among fair-skinned populations [1]. According to recent data, melanoma accounts for a significant portion of skin cancer-related deaths despite being less common than non-melanoma skin cancers. More than 325,000 cases have been diagnosed worldwide, with approximately 57,000 deaths. On the basis of International Agency for Research on Cancer (IARC) scientists, the annual incidence of new cutaneous melanoma cases is expected to rise by over 50% between 2020 and 2040, surpassing 500,000 cases per year [2]. Additionally, melanoma-related deaths are projected to increase to over 100,000 annually [3]. The increasing prevalence of melanoma can be attributed to various factors, including greater sun exposure particularly in the summer months, tanning practices, and a growing aging population. Early detection is crucial, as the prognosis for patients significantly improves when the tumor is identified at the initial stage [1,4]. Despite advancements in diagnostic technologies, melanoma remains challenging to detect early due to its asymptomatic nature and subtle visual cues that often go unnoticed by patients and even healthcare professionals [4].

Traditional diagnostic methods, such as visual examination and dermoscopy, although valuable, have limitations in accuracy and are subject to human error and variability. Dermoscopy, a noninvasive technique that enhances the visualization of skin lesions, improves diagnostic accuracy but still falls short, with around 80% accuracy in routine clinical settings [5]. This highlights the need for more reliable and efficient diagnostic tools.

In recent years, the application of deep learning (DL), particularly Convolutional Neural Networks (CNNs), have revolutionized medical image analysis by significantly enhancing diagnostic accuracy and efficiency [6–8]. CNNs have demonstrated superior performance in detecting and classifying melanoma from dermoscopic images, often surpassing human expertise [9–11]. EfficientNet, a state-of-the-art CNN architecture, has shown promise in melanoma classification tasks by effectively learning complex and fine-grained patterns from dermoscopic images [12].

Recent high-impact studies have further advanced these capabilities, applying CNN architectures to diverse high-precision medical tasks ranging from skeletal action recognition to retinal disease detection [6-8, 11]. Despite these successes, significant challenges persist. Beyond the inherent difficulty of extracting subtle, deeper features in complex lesions [9, 13], models remain highly sensitive to noise [14] and artifacts characteristic of dermoscopic data. Furthermore, the lack of interpretability in standard CNNs remains a major barrier to their full clinical adoption. This underscores the urgent need for diagnostic systems that combine high accuracy with visual explanations, ensuring they are both robust and transparent

for dermatological practice.

However, the CNNs approaches often suffer from extracting deeper features [9], Modern techniques such as ensemble learning allow one to combine single models by leveraging their strengths of feature extraction, and as a result, obtaining more reliable architecture This may lead to achieve higher efficiency, reducing overfitting, underfitting, and susceptibility to noise. Thus, the main motivation of this study is to verify how effective the single, pre-trained CNN models with transfer learning are for melanoma identification tasks. Moreover, the analysis concerning what type of model aggregations enables skin cancer detection gets reinforced.

The aim of this study is to develop and evaluate the effectiveness of deep neural networks, specifically state-of-the-art deep learning models - ResNet152, DenseNet201, and EfficientNet-B4, for the automatic diagnosis and classifying benign and malignant melanoma. Additionally, an ensemble learning approach using different techniques is examined to leverage the strengths of individual models for superior diagnostic performance. This study seeks to improve the accuracy and support melanoma diagnosis, contributing to earlier detection and better patient outcomes.

The main contributions of this study are as follows:

- Comparing the effectiveness of pre-trained ResNet152, DenseNet201, and EfficientNet-B4 architectures for automatic benign and malignant melanoma identification. Accuracy, precision, recall and F1-score are chosen to verify the performance. The last model stated to be the most suitable, achieving 90.40% accuracy.
- Preparing the combining dataset consisting of multiple International Skin Imaging Collaboration (ISIC) datasets from 2018, 2019, and 2020 to create a balanced and reliable resource.
- Combining various variants the above-mentioned models utilizing soft voting, hard voting, and stacking with XGBoost ensemble learning techniques. The soft voting improved the skin cancer detection up to 0.90% while merging ResNet152 together with EfficientNet-B4 or all models altogether.
- Applying Grad-CAM to highlight the image areas that are the most significant to contribute to the way a model performs. The EfficientNet-B4 model concentrates on the lesion's irregular edges and darker regions that reflect features characteristic of melanoma, which often presents with asymmetry, irregular borders, and variable pigmentation.

To systematically evaluate these contributions, this study aims to answer the following research questions:

- How do different state-of-the-art CNN architectures (ResNet152, DenseNet201, and EfficientNet-B4) compare in their ability to extract diagnostic features from dermoscopic images?
- To what extent can ensemble learning techniques improve the identification of malignant melanoma compared to single-model approaches?
- Which ensemble strategy (soft voting, hard voting, or stacking) is most effective for handling the complexities of skin lesion classification?

This study explores the capabilities of state-of-the-art models and their enhanced performance using ensemble learning to improve melanoma detection, with the goal of establishing a robust, automated diagnostic system that assists clinicians in making more accurate and timely diagnoses. The findings of this study have the potential to significantly impact the field of dermatology by providing a reliable tool for early melanoma detection, ultimately improving patient survival rates.

2. RELATED WORKS

In recent years, significant advancements have been made in the application of deep learning techniques, particularly convolutional neural networks, for the classification and detection of melanoma and other skin lesions. Various research studies have proposed and validated different models and methodologies to enhance the accuracy and efficiency of these systems.

One noteworthy approach is the EfficientNet architecture, presented by Runyuan Zhang, who demonstrated that employing EfficientNet-B6, optimized through neural architecture search, significantly improved the accuracy of melanoma detection, achieving an AUC-ROC of 0.917 [12]. Similarly, S M Jaisakthi utilized EfficientNet with transfer learning and ranger optimizer, achieving a superior AUC-ROC score of 0.9681 [15]. These studies highlight the effectiveness of EfficientNet models in learning complex patterns and improving classification performance. The comprehensive study with pre-trained VGG19, ResNet18, and MobileNet_V2 utilizing the ISIC 2018 dataset was applied for benign and malignant type cancer detection in [4]. The above-mentioned architectures were used for feature extraction, while the detection was performed utilizing Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes, and k-Nearest Neighbors (kNN). MobileNet with SVM achieved the highest effectiveness, reaching 92% accuracy. The ensemble learning combining ResNet-18 and MobileNet models enhances the cancer classification up to 92.87%.

Ensemble methods have also proven beneficial in melanoma detection. Dominika Kwiatkowska et al. employed an ensemble of CNN models, including ResNet-101, ResNeXt, SE-ResNet, and SE-ResNeXt, to classify malignant melanoma in dermoscopic images [16]. By leveraging the strengths of multiple architectures, the ensemble approach demonstrated higher prediction accuracy, with ResNeXt achieving the highest performance metrics. Deep residual networks have been another area of focus. Lequan Yu proposed using very deep fully convolutional neural networks with residual networks (FCRN) for accurate skin lesion segmentation and classification [17]. This method achieved high accuracy by integrating multi-scale contextual features, which are crucial for distinguishing between melanoma and non-melanoma lesions. Innovative segmentation techniques have also been explored. Ali Rizwan introduced a fully convolutional encoder-decoder network (FCEDN) optimized by the Sparrow Search Algorithm (SpaSA) for precise segmentation, coupled with an adaptive CNN for classification [18]. This approach achieved impressive segmentation and classification accuracies, showcasing the potential of combining FCEDN with optimization algorithms for enhanced performance. Combining traditional methods with advanced deep learning techniques has also yielded promising results. Nida Nudrat integrated a deep region-based convolutional neural network (RCNN) with fuzzy C-means (FCM) clustering for melanoma lesion detection and segmentation [19]. This method outperformed traditional and existing deep learning approaches, particularly in handling artifacts and achieving precise segmentation.

Data augmentation and transfer learning have been extensively utilized to address the challenge of limited labeled medical images. Hosny Khalid M. leveraged AlexNet with transfer learning and data augmentation to develop a robust skin lesion classification system [20]. The use of image augmentation significantly improved classification accuracy across multiple datasets. Additionally, Qin Zhiwei employed a style-based GAN for generating high-quality synthetic images, which were used to augment the training dataset of a

ResNet50 model, leading to improved performance metrics [21]. Hybrid models combining deep learning and traditional machine learning techniques have also shown efficacy. Mahbod Amirreza utilized pre-trained CNNs (AlexNet, VGG16, ResNet-18) as feature extractors, combined with support vector machines (SVMs) for classification [22]. This fusion of deep features from multiple networks achieved high performance, demonstrating the advantage of hybrid approaches in melanoma detection. Other notable contributions include the development of novel network architectures and optimization techniques. Veeramani Nirmala introduced a Double Decker Convolutional Neural Network (DDCNN) feature fusion framework, which combined advanced preprocessing and feature extraction techniques, resulting in improved specificity and accuracy [23]. Wu Huisi proposed FAT-Net, a dual encoder architecture integrating CNNs and transformers, achieving state-of-the-art segmentation performance by capturing both local and global features effectively [24]. Muhammad Amir Khan et al. proposed the intelligent computer-aided system for melanoma and non-melanoma skin cancer identification [25]. The fully connected layer was modified by utilizing principal component analysis. This approach enables the discriminative feature extraction and reduces overfitting. Rakin Sad Aftab introduced a new SkinScanNet, deep learning-based model, for high effective melanoma identification [26]. The study utilized Melanoma Skin Cancer dataset. The DenseNet121 and a proprietary Convolutional Neural Network were applied for skin cancer identification [27]. The following variety of disease was included in the study: actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, and vascular lesions. The DenseNet121 obtained 89% accuracy, while the custom model that combines CNN and DenseNet121 utilizing augmenting data and synthetic minority oversampling technique for class weighing outperforms the pre-trained architecture, reaching 95% accuracy. Explainable AI techniques in deep learning play a pivotal role in understanding the key image regions based on which the decision-making is performed. They give visual interpretation into the model's prediction. Gradient-weighted Class Activation Mapping (Grad-CAM) has been widely applied in skin cancer detection due to the need to develop a reliable and trustworthy supportive system for dermatologists [26–28]. Grad-CAM can generate high-resolution heatmaps that highlight key regions as pixel-level or segmented explanations.

Overall, these studies collectively illustrate the advancements in melanoma detection using CNNs and other deep learning methods. By exploring various innovative techniques, such as EfficientNet architectures, ensemble learning methods, deep residual networks, segmentation techniques, data augmentation, transfer learning, and hybrid models, researchers have significantly improved the accuracy, efficiency, and robustness of melanoma detection from dermoscopic images.

3. MATERIALS AND METHODS

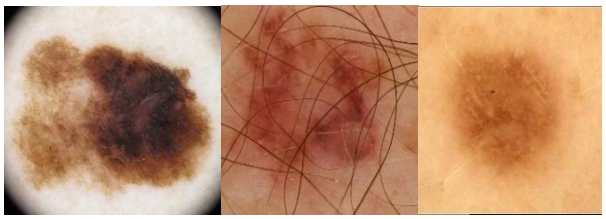
The primary objective of this study is to assess the efficacy of transfer learning techniques across various deep learning architectures for the classification of skin lesions using dermoscopic images. This section begins by introducing the dataset of dermoscopic images, followed by a detailed explanation of the experimental setup employed in this study. Subsequently, the selected Convolutional Neural Networks architectures applied to tackle the skin lesion classification task, along with the training configuration are described. The applied architectures are followed by the use of ensemble learning techniques to improve the results. Finally, the results of a basic feature-based classification approach used for comparison are discussed.

3.1. Dataset

The dataset used for the study consists of 17,114 carefully curated images, specifically designed to facilitate the advancement of the field of dermatology and computational diagnostics in melanoma detection [29]. It combines multiple International Skin Imaging Collaboration (ISIC) datasets from 2018, 2019, and 2020 to create a balanced and reliable resource for training and evaluation [30]. A total of 1,249 images were collected from the 2018 dataset (174 malignant, 1,075 benign), 5,660 from the 2019 dataset (4,439 malignant, 1,221 benign), and 10,205 from the 2020 dataset (582 malignant, 9,623 benign). These images were taken directly from the official ISIC website and were selected to minimize data imbalance and enhance the representativeness of both malignant and benign cases. Although the images vary in size, they were resized during training to suit the requirements of each architecture. Although additional examples from the ISIC datasets could have been included, this would have increased the data imbalance.

The images are available in both DICOM [29,31] and JPEG formats and are accompanied by metadata (Fig. 1). This metadata includes patient ID, gender, age, and the general anatomical location, along with the target value. The training dataset consists of two categories of images: benign and malignant, each labelled with its ground truth.

a) benign skin lesions



b) malignant skin lesions

Fig. 1. Example of skin lesions in combined ISIC datasets: a) benign and b) malignant cases

For the melanoma classification study conducted using the ISIC datasets, the metadata and image distribution across the training, test, and validation sets are outlined in a Tab. 1. The dataset maintains an approximate 70%-30% split between benign and malignant cases. Of the 17,114 images, 14,204 were allocated to training, while 1,455 images were allocated for testing and an additional 1,455 for validation.

DOI: 10.65731/ama/2026-0033

Tab. 1. Number of the images from ISIC-Combined Dataset for skin cancer identification

Class	Total	Training	Testing	Validation
Malignant	5192	4312	441	439
Benign	11922	9892	1014	1016
Total	17114	14204	1455	1455

3.2. Data preprocessing

To ensure the effectiveness of the melanoma detection model, a comprehensive data pre-processing pipeline was implemented. The primary goal of the preprocessing was to standardize and enhance the input images to improve the model's ability to generalize to unseen data while mitigating overfitting. The preprocessing was tailored to each model paying particular attention to the size of the input image sizes.

During the training phase, a sequence of dynamic, on-the-fly transformations was employed to virtually increase data diversity without altering the static size of the dataset or the class distribution. This approach ensures that the model encounters slightly different versions of the training images in each epoch [32].

The input images were first resized to fit the requirements of each architecture. To simulate real-world clinical conditions, such as varying handheld dermatoscope orientations, random horizontal and vertical flips were applied, followed by random rotations of up to 30 degrees. This specific rotation range was chosen to introduce sufficient variability while preventing excessive information loss or artifacts caused by padding and cropping, which could otherwise obscure critical diagnostic features like irregular lesion borders. Additionally, color jittering was applied to adjust brightness and contrast, simulating lighting variations during image acquisition. Finally, images were normalized using ImageNet's mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225] to facilitate stable and efficient training.

These preprocessing steps not only standardized the input data, but also introduced controlled variability into the training data set. This ultimately improved the model's robustness and performance of the model in melanoma detection tasks.

3.3. Deep learning architectures

The fundamental reason for selecting these three specific models was their complementarity. Each architecture interprets skin lesions differently based on its unique design. By combining them into an ensemble using the soft voting method, we successfully mitigated individual model weaknesses and leveraged their collective strengths.

3.3.1. ResNet

The ResNet (Residual Network) architecture is a deep neural network designed to solve the vanishing gradient problem in very deep networks, allowing hundreds or even thousands of layers to be effectively trained [33]. It introduces the concept of residual learning through the use of residual blocks, where each block adds a "shortcut connection", shown in Fig. 2, which skips one or more layers. This connection directly propagates the input of a block directly to its output, bypassing intermediate transformations [32].

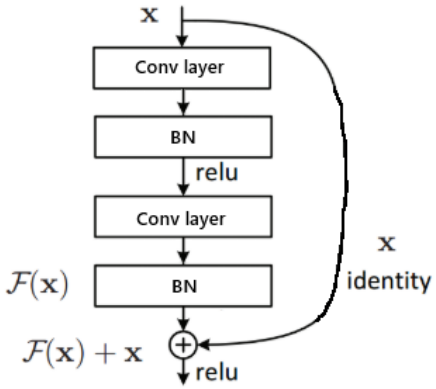


Fig. 2. Residual blocks - building block of ResNet [34]. $\mathcal{F}(x)$ indicates the input signal after processing signal through all layers

Residual blocks typically consist of convolutional layers with batch normalization and ReLU activations, and the shortcut connections can either be identity mappings or involve a linear transformation to match dimensions. ResNet variants (e.g., ResNet-50, ResNet-101) differ in the number of layers and block structures, often utilizing bottleneck designs to reduce computation. This architecture has demonstrated state-of-the-art performance on numerous image recognition benchmarks, showcasing significant improvements in training deeper CNNs, facilitating better performance and accuracy in various computer vision tasks.

In this study, the ResNet-152 architecture, a deep convolutional neural network with 152 layers was used. It employs bottleneck blocks, each consisting of three convolutional layers, along with identity or shortcut connections. This design achieves high performance on complex tasks like image recognition while maintaining manageable computational complexity.

3.3.2. DenseNet

The DenseNet architecture is designed to address the vanishing gradient problem and improve feature propagation by connecting each layer to every other layer in a feed-forward manner [35]. Each layer receives feature maps from all preceding layers and passes its own feature maps to subsequent layers, facilitating feature reuse and reducing the number of parameters (Fig. 3).

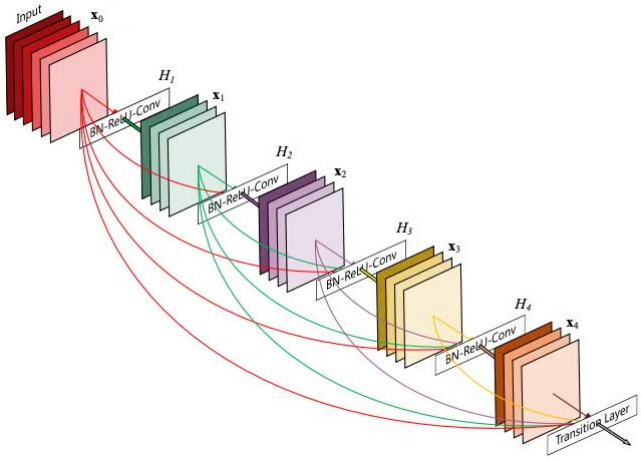


Fig. 3. Dense Convolutional Network with 5-layer dense block [35]

DenseNet divides the network into multiple dense blocks, maintaining the same feature map size within blocks and using downsampling operations outside these blocks. This architecture supports hundreds of layers, enabling efficient training and high performance in deep networks. DenseNet201, a variant selected in this study containing 201 layers, utilizes this dense connectivity to achieve excellent accuracy on image recognition tasks while being computationally efficient compared to similarly deep architectures.

3.3.3. Efficient-Net

EfficientNet is a family of convolutional neural networks designed to achieve state-of-the-art accuracy while being significantly smaller and faster than previous models [36]. These models use a novel scaling method that uniformly scales all dimensions of depth, width, and resolution using a compound coefficient, enabling a principled way to scale up ConvNets for better performance. This approach balances computational efficiency and accuracy, addressing the challenges of scaling deep networks. The schematic representation of EfficientNet shown in Fig. 4.

EfficientNet-B4 is a mid-sized model in the EfficientNet family, optimized for a balance between accuracy and computational cost. It maintains the core characteristics of EfficientNet, including the compound scaling method, inverted bottleneck residual blocks, and Squeeze-and-Excitation (SE) blocks. EfficientNet-B4 employs Mobile Inverted Bottleneck layers (MBConv), which are based on depth wise separable convolutions combined with inverted residuals. These blocks also include SE modules that improve performance by emphasizing important features while de-emphasizing irrelevant ones.

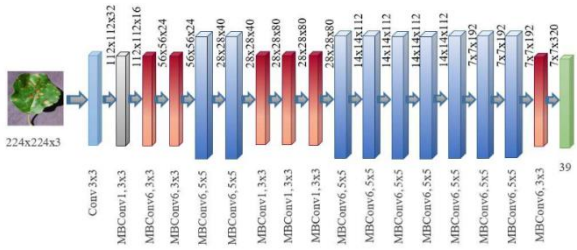


Fig. 4. EfficientNet – schematic representation [37]

3.3.4. Ensemble Learning

Ensemble learning is a powerful machine learning paradigm that combines the predictions of multiple models to improve overall performance and robustness. Ensembles, especially when individual models have complementary strengths, often achieve higher accuracy than individual models. They are also more robust to data outliers and noise, as errors in individual models are often corrected by others. There are several types of ensemble techniques, including voting, stacking, bagging and boosting [38,39]. Weighted majority voting (also known as soft voting) and majority (hard) voting methods to aggregate predictions from the ensemble were applied in this study (Fig. 5).

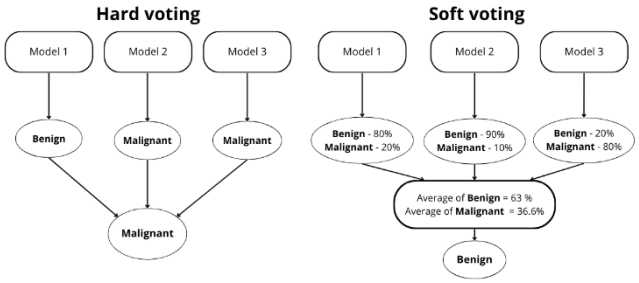


Fig. 5. Example of an ensemble learning with hard and soft voting

In hard voting, the final prediction is determined by a majority vote, with each model contributing equally to the decision. In contrast, soft voting averages the predicted probabilities from all models, using confidence scores to produce a more nuanced and potentially more accurate result, especially for unbalanced datasets [39]. These techniques capitalize on the complementary strengths of individual models, ensuring a well-rounded and robust performance in dermoscopic image classification.

The mathematical formula for voting [40]**:**

$$\hat{y} = arg \max_{k\in\{1,\ldots,K\}} \sum_{i=1}^{N} \omega_i m_i^j(x) \quad (1)$$

where: K stands for the number of classes, ω_i represents the weight assigned to the *i*-th model m_i, $m_i^j(x)$ is the output of the *i*-th model for the *j*-th category label. Hard voting can be readily modeled by assigning the same positive value to each individual weight ω_i such as setting ω_i= 1 for all models.

Two main types of voting approaches exist: hard voting and soft voting. In hard voting, each model's prediction counts as a single vote, and the final output is determined by the majority or plurality of votes. While straightforward, hard voting can disregard useful information about classifier confidence. Soft voting, on the other hand, integrates class probabilities or confidence scores, weighting each classifier's vote according to how certain it is about its prediction. This approach can lead to more nuanced ensemble decisions, particularly for datasets with overlaps or class imbalances.

In addition to evaluating the performance of individual models, ensemble methods such as soft voting, hard voting, and stacking were applied to further enhance classification performance. The ensemble was composed of diverse deep learning architectures trained independently on the same dataset. In case of soft voting, the predicted class probabilities from each base model were averaged to produce the final prediction, allowing the ensemble to reflect the relative confidence of each model. For hard voting, the final class was selected based on the majority class label predicted by each model.

Stacking was also explored as a higher-level ensemble strategy. In this approach, the outputs of multiple base learners are combined using a meta-learner trained to optimize the final prediction. The XGBoost algorithm, an efficient and scalable implementation of gradient boosting, was selected as the meta-classifier due to its high predictive performance and ability to model complex feature interactions. XGBoost builds an ensemble of decision trees in a sequential manner, where each new tree corrects the errors made by the previous ones, and is particularly effective in handling structured data and reducing overfitting through regularization [41].

In this study, XGBoost was employed as a meta-classifier in a stacking ensemble framework. Specifically, the output probabilities obtained by trained ResNet152, DenseNet201, and EfficientNet-B4

models were used as input features for the XGBoost model, which learned to optimize the final prediction by leveraging the individual strengths and compensating for the weaknesses of each base model. This hierarchical ensemble strategy aimed to improve generalization and overall classification accuracy by combining diverse model architectures. The predictive performance of the stacked model was compared against that of the individual CNNs to evaluate the added value of this ensemble learning approach.

3.4. Evaluation metrics

The trained model's performance was evaluated using balanced accuracy (2), precision (3), recall (4), and F1 score (5) on the test dataset. Due to the imbalanced nature of the dataset, accuracy was used. The other metrics were calculated using the parameter average=weighted to account for class imbalance. True Positives (TP) and True Negatives (TN) denote the number of correctly predicted subjects, while False Positives (FP) and False Negatives (FN) correspond to the incorrectly predicted subjects.

$$Balanced\ Accuracy = \frac{1}{2} * (\frac{TP}{TP+FN} + \frac{TN}{TN+FP}) \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall\ (sensitivity) = \frac{TP}{TP+FN} \quad (4)$$

$$F1\ score = \frac{2TP}{2TP+FP+FN} \quad (5)$$

3.5. Experimental Setup

All experiments were conducted on the same hardware setup to ensure consistency and reproducibility. Model training was carried out entirely on local equipment, allowing for greater control over the process.

The hardware specifications used for training are as follows:

- Graphics Card: NVIDIA GeForce RTX 3060 Ti, equipped with 8GB VRAM;
- Processor: Intel Core i5-10600K, 4.1 GHz base clock speed;
- RAM: 16GB DDR4, operating at 3200 MHz.

This setup provided sufficient computational power for training moderately sized deep learning models while addressing the challenges posed by memory limitations.

The preparation of the study environment involved installing Python version 3.11, along with essential libraries such as PyTorch (version 2.4.1), NumPy (version 1.23.5), and Pandas (version 2.2.2). Additionally, libraries for image processing and result visualization, such as Matplotlib (version 3.9.2), were imported. The environment configuration also included setting the GPU as the primary device for model training, enabling faster computation and parallel processing to speed up the learning process.

3.6. Training models

The model training process began with data pre-processing, which included image resizing and normalization. The choice of image size was determined by the selected architecture, 224x224 for ResNet152 and DenseNet201, and 380x380 for EfficientNet-B4. The pre-trained architectures, adapted for the skin image classification task by tuning the final network layers, were trained on 14,204 images from the ISIC database. To address the dataset imbalance, a cost-sensitive learning approach was adopted by applying class weights to the loss function. The weights were calculated based on the inverse frequency of each class (1.44 for benign and 3.29 for malignant), ensuring that the minority class significantly influenced the model's optimization process. The training was conducted using the AdamW optimizer with weight decay to further enhance regularization and prevent overfitting. The training process was monitored using callbacks such as EarlyStopping, which halts training when the model's performance stops improving, and weight decay regularization, which helps to prevent overfitting by penalizing large weights during training. These parameters, summarized in Table 2, were selected based on preliminary tuning to maximize the accuracy and stability of each model. Additionally, the technique of freezing and unfreezing layers in pre-trained models was applied; however, it led to performance improvement only in the case of EfficientNet-B4.

Following the training process, the models were tested on a separate test set of 1,455 images. Testing the model involves assessing its ability to correctly diagnose melanoma based on previously unknown data. Each image was classified as 'malignant' or 'benign' and the results were saved for later analysis.

The final step of the study involved comparing the performance of the models created using ensemble learning techniques with other popular CNN architectures, such as ResNet152, DenseNet201, and EfficientNet-B4. This analysis aimed to determine which architecture performs best in the automatic diagnosis of cutaneous melanoma. Additionally, the performance of the ensemble learning model was compared to that of the individual models to assess the potential advantages of combining multiple architectures.

Tab. 2. Optimized training hyperparameters for each deep learning architecture

Hyperparameter	ResNet152	DenseNet201	Efficient-Net-B4
Input Size	224x224	224x224	380x380
Optimizer	AdamW	AdamW	AdamW
Learning Rate	0.05	0.05	0.05
Batch size	32	32	16
Weight Decay	0.001	0.001	0.0001
Early Stopping Patience	5	8	5

3.7. Visualization technique: Grad-CAM

To enhance the transparency and trustworthiness of the proposed diagnostic system, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. This technique produces a localized heatmap by utilizing the gradients of a specific target class (e.g., 'Malignant') flowing into the final convolutional layer of the network [42]. By calculating the importance of each neuron through global average pooling of the gradients, Grad-CAM highlights the specific pixels and regions that most significantly influenced the model's prediction.

In this study, Grad-CAM was applied to each individual base model (ResNet152, DenseNet201, and EfficientNet-B4) rather than the ensemble as a whole. This approach allows for a granular

analysis of how different architectures perceive dermatological features. By verifying that each network in the ensemble focuses on clinically relevant areas, such as irregular borders or pigment distribution, we ensure that the final combined decision (via soft voting) is based on meaningful medical patterns rather than image artifacts.

4. RESULTS AND DISCUSSION

4.1. Single deep learning architectures studies

Plots of loss and accuracy curves for each model provide visual insight into the training process, illustrating stability and rate of convergence (Fig. 6-8). The graphs provide a better understanding of the performance of each architecture in the melanoma identification task. In addition, these plots allow one to see the differences that occur during training due to the use of the layer freezing technique of the pre-trained model.

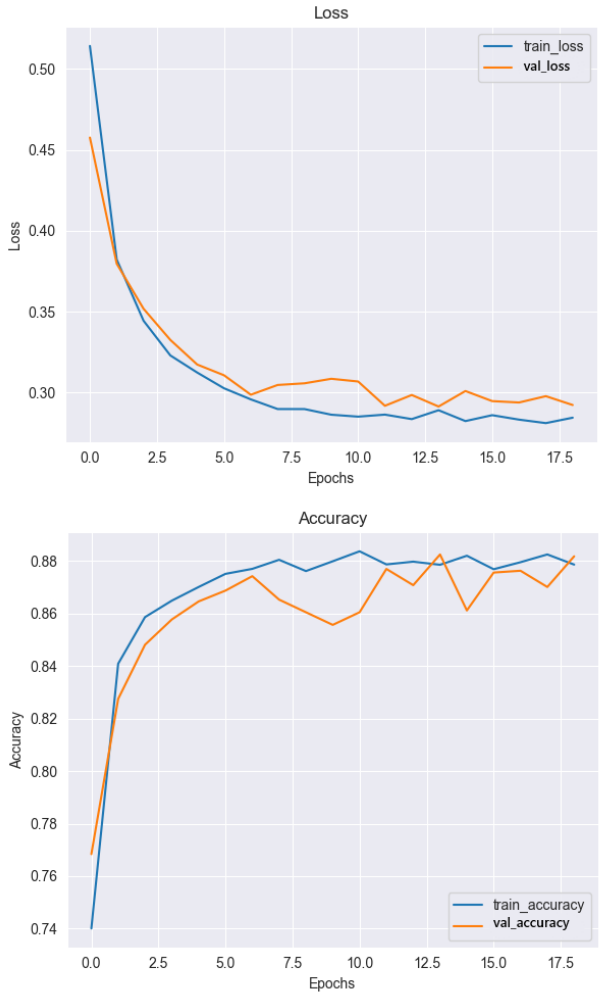


Fig. 6. Loss function and accuracy curves for DenseNet201 model

Analysis of the plots indicates that all models learn efficiently, and that the differences between loss and accuracy on the training and validation sets are within acceptable limits. In the early epochs, we observe a significant reduction in loss for both the training and validation sets, indicating that all models are learning effectively. As training progresses, the differences between the loss for the validation and training sets become small, indicating good generalization of the model. The small differences in accuracy between the test set and the training set indicate an appropriate balance between learning and generalization. In case of the EfficientNet-B4 model, the loss and accuracy curves have the most dynamic start due to the layer freezing technique and its unfreezing after the 10th epoch of training. However, in the further stages of training, the differences decrease and the accuracy on the test set reaches a high level. The higher test accuracy compared to training in some epochs may be indicative of the effectiveness of the regularization used in this particular model of the model.

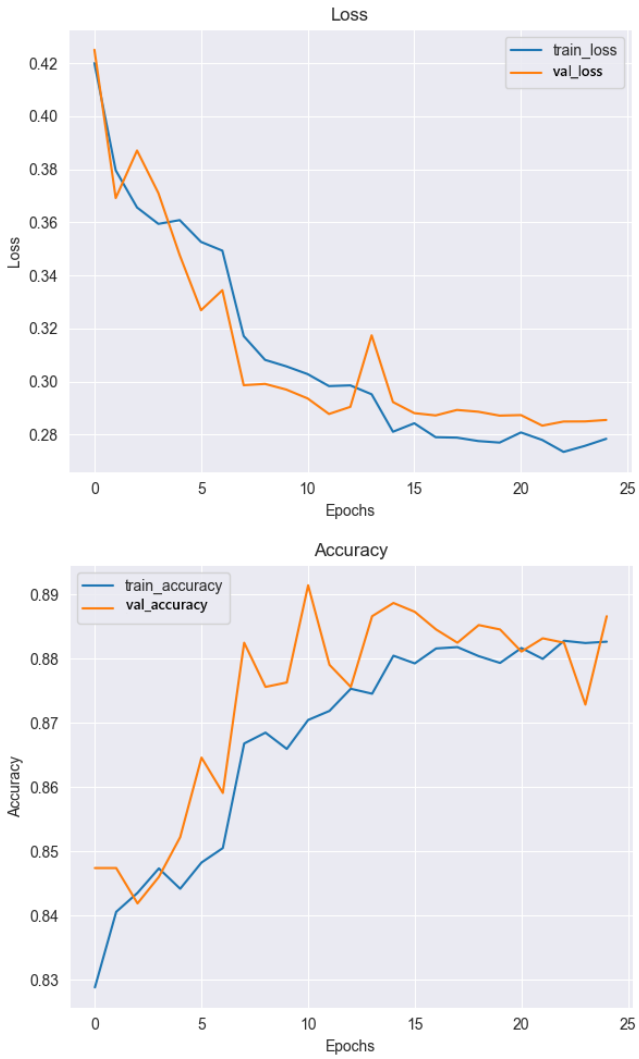


Fig. 7. Loss function and accuracy curves for ResNet152 model

After the learning process, each model was evaluated to assess its performance and reliability. Model performance was determined by making predictions on test data and comparing the results to actual labels (Tab. 3). To ensure the robustness and consistency of the evaluation, each model was tested across 10 independent runs, with performance metrics averaged to account for variability and improve result reliability. The comprehensive evaluation performed on the test dataset yielded insightful results, with the performance metrics for each model summarized in Tab. 3. In addition, Fig. 9, 10, and 11 present the corresponding confusion matrix, providing a detailed breakdown of each model's classification performance. Tables 4, 5, and 6 further present the inference results

DOI: 10.65731/ama/2026-0033

for the trained models, with a breakdown for each class.

This study identified EfficientNet-B4 as the optimal pre-trained model for melanoma classification, providing the highest accuracy and stability on the test dataset. These results highlight the potential of deep learning models to improve diagnostic accuracy and assist dermatologists in detecting malignant melanoma lesions.

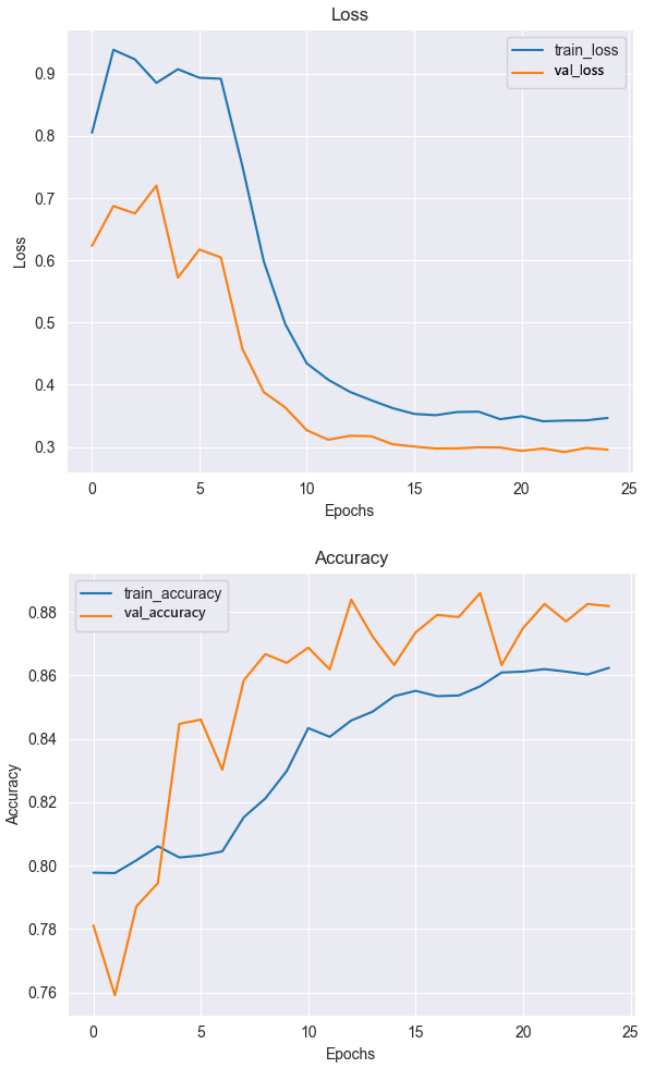


Fig. 8. Loss function and accuracy curves for EfficientNet-B4 model

Tab. 3. Average performance on the 10-iteration test set for single DL models (mean ± standard deviation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Res-Net152	89.2±0.014	89.8±0.012	90.2±0.014	89.4±0.013
Dense-Net201	88.3±0.004	89.3±0.003	88.2±0.006	88.5±0.005
Efficient-Net-B4	90.4±0.003	90.8±0.002	90.4±0.002	90.6±0.003

Tab. 4. Performance for each class for ResNet152 model (mean ± standard deviation)

Class	Precision (%)	Recall (%)	F1-Score (%)
Bening	95.8±0.007	88.4±0.020	91.8±0.012
Malignant	77.3±0.021	91.2±0.008	83.3±0.020

Tab. 5. Performance for each class for DenseNet-201 model (mean ± standard deviation)

Class	Precision (%)	Recall (%)	F1-Score (%)
Bening	95.2±0,003	87.6±0,004	91.3±0,002
Malignant	75.9±0,002	89.8±0,002	82.5±0,003

Tab. 6. Performance for each class for EfficientNet-B4 model (mean ± standard deviation)

Model	Precision (%)	Recall (%)	F1-Score (%)
Bening	94.8±0,004	91.3±0,005	93.0±0,005
Malignant	81.6±0,006	88.4±0,005	84.7±0,005

The analysis performed on the test set (Tab. 4-6) provided detailed results, comparing the performance metrics of different models to assess their effectiveness in melanoma classification. EfficientNet-B4 outperformed all other models in key metrics, achieving an F1-score of 90.6%, indicating an excellent balance between precision (90.8%) and sensitivity (90.4%). This confirms its effectiveness in classifying melanoma as malignant or benign, as it identifies positive cases exceptionally well as true positives, effectively detecting malignant melanoma lesions. ResNet152 performed similarly but slightly lower, while DenseNet201 had the weakest performance in each category, although it also achieved high scores. The small differences in results between the models highlight the superiority of EfficientNet-B4 in analyzing complex dermoscopic patterns.

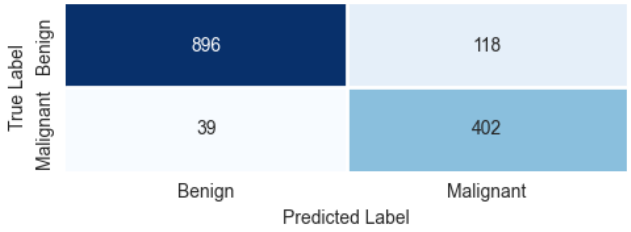


Fig. 9. Confusion matrix for ResNet152 model

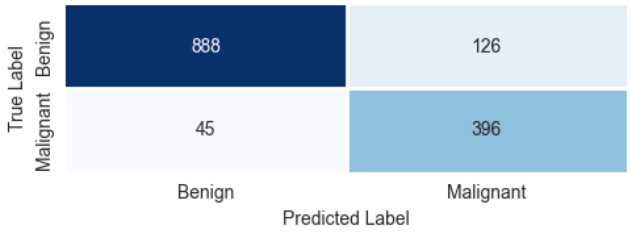


Fig. 10. Confusion matrix for DenseNet201 model

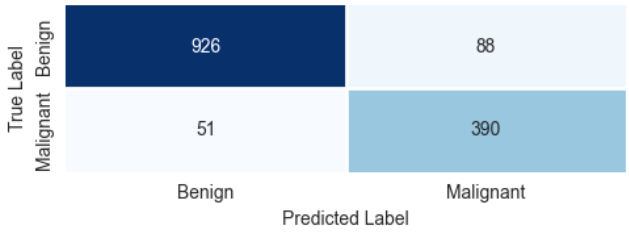


Fig. 11. Confusion matrix for EfficientNet-B4 model

Analysis of confusion matrix (Fig. 9-11) reveals that all models have a high number of correct predictions for the 'Benign' class. The main difference between the models is the number of misclassified malignant lesions (false negatives). EfficientNet-B4 recorded 51 false negatives (FN), indicating a need for further optimization, particularly for melanoma diagnosis. Based on the results for each class (Tab. 4-6), the average precision for the 'Benign' class was 95 ± 1%, demonstrating all models' ability to accurately diagnose benign lesions. Conversely, for the 'Malignant' class, the precision was lower, 78 ± 4%, highlighting the challenge of diagnosing malignant lesions, which often have subtle and variable characteristics. The F1-score for the 'Malignant' class averaged 83 ± 2%, indicating a good balance between precision and sensitivity, although there is still potential for improvement.

The results, gathered in Tab. 4-6, suggest that the EfficientNet-B4 architecture, with frozen layers during training, is the most suitable for skin cancer identification, combining high precision with exceptional stability. This model is particularly notable for its ability to balance precision and recall for malignant cases (the minority class), making it the most reliable model for this task.

4.2. Ensemble learning studies

This study focuses on evaluating how various ensemble learning techniques influence on enhancing the performance of melanoma skin cancer detection. The soft voting, hard voting and XGBoost methods were compared. The use of a soft-voting approach, which combines the probabilities predicted by different models, produced better results than individual models (Tab. 7). This approach allows the strengths of each model to be exploited.

Tab. 7. Performance of soft voting ensemble model on the 10-iteration test set (mean ± standard deviation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
All models	91.3± 0.08	91.2± 0.07	91.3± 0.07	91.2± 0.07
DenseNet-201 + EfficientNet-B4	91.0± 0.06	91.2± 0.06	91.0± 0.05	91.1± 0.05
ResNet152 + EfficientNet-B4	91.3± 0.06	91.2± 0.05	91.3± 0.05	91.2± 0.05
ResNet152 + DenseNet201	89.5± 0.06	89.4± 0.08	89.5± 0.07	89.2± 0.08

The peak accuracy of 91.3% was shared by two ensemble variants: the combination of all three models and the pairing of ResNet152 with EfficientNet-B4. Notably, the latter combination proved to be slightly more robust, yielding a lower standard deviation of ±0.06 compared to ±0.08 for the three-model ensemble, suggesting that a two-model synergy is sufficient to achieve optimal results in this task. The combination of ResNet152 and EfficientNet-B4 also achieved an accuracy of 91.3%, suggesting that these models classify exceptionally well together. The ensemble of DenseNet201 and EfficientNet-B4 achieved an accuracy of 91.0%, which is slightly lower, but still indicates a very good generalization. On the other hand, the combination of ResNet152 and DenseNet201 resulted in an accuracy of 89.5%, which is the lowest of the ensembles tested. However, it still shows a slight improvement over the individual models, by 0.3% and 1.2% for ResNet152 and DenseNet121, respectively. All variants of the ensembles outperformed the accuracy of the individual models, highlighting the advantages of ensemble learning in the classification of dermoscopic images.

In contrast to the results obtained with soft voting, the performance of the ensemble models with hard voting and the XGBoost stacking ensemble is noticeably lower across all metrics (Tab. 8-9). Only combination of all deep learning models, or DenseNet201 together with EfficientNet-B4 with hard voting provided improvements according to the single architecture performance. In the case of XGBoost this technique is stated to slightly improve skin cancer detection according to DenseNet201.

Tab. 8. Performance of hard voting ensemble model on the 10-iteration test set (mean ± standard deviation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
All models	90.7±0.13	90.6±0.12	90.7±0.13	90.6±0.13
DenseNet201 + EfficientNet-B4	90.5±0.08	90.4±0.08	90.5±0.11	90.4±0.09
ResNet152 + EfficientNet-B4	88.1±0.09	88.1±0.1	88.0±0.09	87.5±0.11
ResNet152 + DenseNet201	87.4±0.07	87.7±0.09	87.4±0.08	86.9±0.1

Tab. 9. Performance of XGBoost ensemble model on the 10-iteration test set (mean ± standard deviation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
XGBoost	88.9±0.08	89.2±0.07	88.9±0.06	89.0±0.07

These results clearly demonstrate the superiority of soft voting for skin cancer classification tasks. The improved performance of soft voting is likely due to its ability to leverage the probabilistic outputs of individual models. This allows for a more nuanced and effective combination of predictions [43]. In contrast, both hard voting and the XGBoost stacking ensemble, which rely on discrete or weighted decisions, appear to be less effective at capturing the complementary strengths of individual models, resulting in lower accuracies and balanced accuracies. This further highlights the importance of using soft voting in ensemble learning to achieve optimal results in dermoscopic image classification.

DOI: 10.65731/ama/2026-0033

4.3. Explainable of deep learning models

Machine learning models used for melanoma detection often leverage visualization techniques like Grad-CAM. These techniques help understand which areas of an image influence the model's decisions.

The ResNet152 model primarily focuses on the central part of the skin lesion, particularly areas with the highest color intensity (Fig. 12). It emphasizes the dark center while largely ignoring the surrounding skin. This suggests that the model effectively isolates key diagnostic regions, making it particularly useful for detecting lesions with well-defined borders and intense pigmentation.

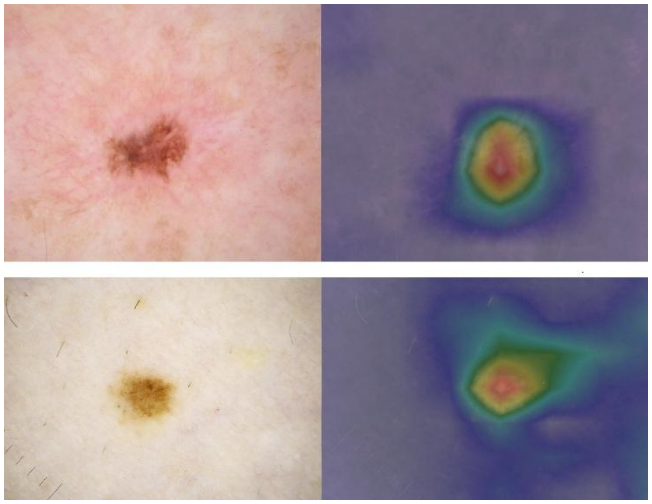


Fig. 12. Analysis of cutaneous melanoma with Grad-CAM mapping using the ResNet152 model

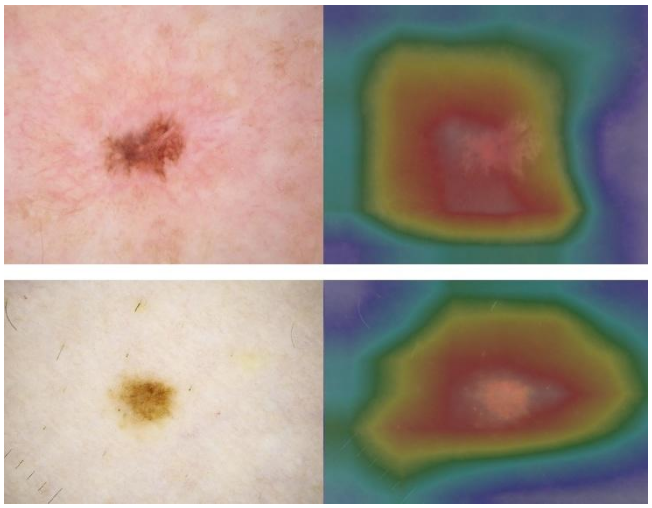


Fig. 13. Analysis of cutaneous melanoma with Grad-CAM mapping using the DenseNet201 model

The DenseNet201 model distributes its attention more evenly across the lesion, a pattern commonly observed in benign lesions with regular shapes and uniform color (Fig. 13). Its broader focus suggests an analysis of the lesion's overall context, including surrounding tissue. The absence of sharply highlighted areas indicates lesion consistency, making this model suitable for identifying lesions with unclear boundaries that require a more comprehensive assessment.

The EfficientNet-B4 model concentrates on the lesion's irregular edges and darker regions, features characteristic of melanoma, which often presents with asymmetry, irregular borders, and variable pigmentation (Fig. 14). This model exhibits a more complex pattern of interest, analyzing multiple features simultaneously. Such an approach is valuable for detecting intricate and irregular lesions that require multifaceted evaluation.

Each of these models effectively identifies key diagnostic regions in skin lesions, though their interpretations vary based on architectural differences. Grad-CAM visualizations not only provide insights into model performance but also enhance interpretability.

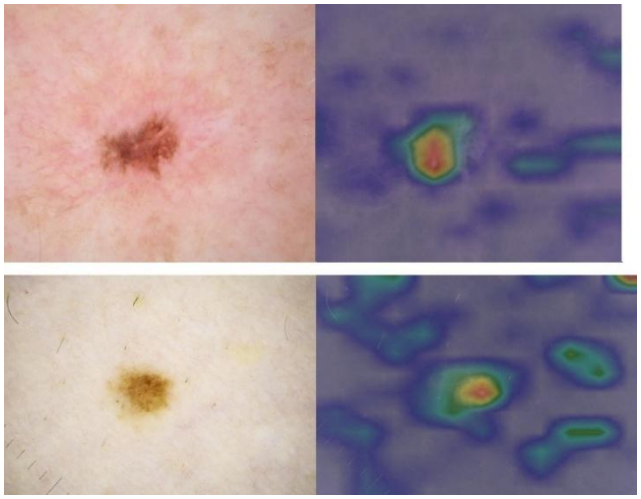


Fig. 14. Analysis of cutaneous melanoma with Grad-CAM mapping using the EfficientNet-B4 model

4.4. Training stability and generalization

A critical challenge in deep learning for medical imaging is maintaining an optimal balance between learning complex patterns and ensuring generalization. In this study, the risk of overfitting was mitigated through a combination of weight decay regularization and EarlyStopping callbacks, which prevented the networks from overoptimizing on the training data. Furthermore, the use of dynamic data augmentation simulated real-world variability, ensuring that the models remained robust to noise and minor image artifacts. As shown in the training curves (Fig. 6-8), the convergence of training and validation metrics indicates that the models successfully avoided significant overfitting. Conversely, the risk of underfitting was addressed by employing powerful pre-trained architectures and fine-tuning their deep layers to capture the subtle, irregular features characteristic of malignant melanoma. The ensemble approach further stabilized these results, as the soft voting mechanism helped compensate for individual architectural biases.

4.5. Comparison with the state-of-the-art

The state-of-the-art studies (Tab. 10) provide an in-depth insight into skin cancer identification utilizing various types of CNN and deep learning architectures. The studies performed using the ISIC datasets were only considered. The majority of papers involve the single models. The pre-trained architectures assure accuracy between 71.19% for VGG-19 and 90.40% for the EfficientNet-B4. Among the common deep learning architectures, ResNet-based

models are the most applied for melanoma detection. However, various CNN-based solutions provide quite good performance that vary between 71% and 97.41% accuracy, depending on their configurations and hyper parameters. The minority of studies concern ensemble learning for melanoma identification. Three types of ensemble learning methods, hard voting, soft voting, and stacking, were analyzed. The highest performance, 93% accuracy, was obtained for the stacking model. It should be emphasized that ensemble learning is a promising direction of studies in enhancing skin cancer identification. That is why the studies presented in this paper are of great values.

Tab. 10. Comparison with the-state-of-the-art of DL models utilizing Skin Imaging Collaboration datasets

Dataset	Model	Ensemble	Acc (%)	Ref.
ISIC 2017	ResNet50	-	81.57	[41]
ISIC 2017	Inception-v3	-	81.29	[41]
ISIC 2017	Dense-Net-201	-	73.44	[44]
ISIC Archive	CNN	-	85.50	[45]
ISIC Archive	KNN, SVM, CNN	Hard voting	88.40	[42]
ISIC 2018	ResNet50	-	87.10	[46]
ISIC 2018	Inception V3	-	89.70	[46]
ISIC 2018	Res-Net50, Inception V3	Soft voting (average)	89.90	[46]
ISIC 2018	AlexNet	-	84.00	[47]
ISIC Archive	AlexNet	-	84.70	[48]
ISIC Archive	VGG16	-	84.43	[48]
ISIC Archive	ResNet50	-	86.54	[48]
ISIC Archive	Res-Net101	-	86.81	[48]
ISIC Archive	MLP	-	83.00	[49]
ISIC Archive	LeNet	-	84.00	[49]
ISIC Archive	VGG-11	-	89.00	[49]
ISIC Archive	MLP, LeNet, VGG-11, kNN	Stacking	93.00	[49]
ISIC 2019	MobileNet	-	85.00	[50]
ISIC 2019	VGG16	-	87.00	[50]
ISIC 2019	Inception V3	-	90.00	[50]
ISIC 2019	Inception Resnet	-	91.00	[50]
ISIC Archive	Efficient-Net-B7	-	85.62	[50]
ISIC Archive	Efficient-Net-B5	-	84.27	[50]
ISIC Archive	Res-Net50V2	-	83.13	[50]
ISIC Archive	VGG16	-	84.07	[50]
ISIC	CNN	-	79.45	[51]
ISIC	VGG-16	-	69.57	[51]
ISIC	VGG-19	-	71.19	[51]
ISIC	CNN	-	71.00	[52]
ISIC 2018-2020	Res-Net152	-	89.20	Own
ISIC 2018-2020	Dense-Net201	-	88.30	Own
ISIC 2018-2020	Efficient-Net-B4	-	90.40	Own
ISIC 2018-2020	Res-Net152, Dense-Net201, Efficient-Net-B4	Soft voting	91.30	Own
ISIC 2018-2020	ResNet-152, Effi-cientNet-B4	Soft voting	91.30	Own
ISIC 2018-2020	Res-Net152, Dense-Net201, Efficient-Net-B4	Hard voting	90.70	Own

This study highlights the great potential of advanced deep learning models in classifying skin lesions for melanoma diagnosis, improving accuracy of detection. The results show that deep learning models such as ResNet152, DenseNet201 and EfficientNet-B4 achieve high accuracy in melanoma identification. In particular, the application of an ensemble learning technique, using soft-voting combining all three architectures, improved the precision rates and F1-score up to 91.3%.

In addition, soft-voting proved to be much more effective on an unbalanced dataset than the hard-voting technique. In contrast, the stacking method using XGBoost did not produce the expected improvement in performance, suggesting that this technique may be less effective for dermoscopic images.

The results align with existing literature (Tab. 10), highlighting the high performance of EfficientNet models in melanoma diagnosis, with precision and F1-scores exceeding 85%. Similar to this study, other research has demonstrated improved classification rates using ensemble learning techniques, particularly voting, which confirms the accuracy of this approach. Conversely, the results for stacking suggest that this method may not be optimal for classifying highly complex images. This study confirms the effectiveness of deep learning models and ensemble learning techniques in skin melanoma classification. In particular, EfficientNet-B4 models and the use of an ensemble model with soft-voting proved to be particularly effective, achieving higher precision and F1-scores than individual models.

The results of this study successfully address the formulated research questions. While all models showed high competence, EfficientNet-B4 emerged as the most reliable single architecture, achieving 90.40% accuracy. Its success is attributed to its compound scaling and SE-blocks, which effectively capture asymmetric borders. For research question 2 and 3, our findings demonstrate that ensemble learning, specifically soft voting, significantly enhances performance, reaching an absolute accuracy of 91.30%. This confirms that leveraging probabilistic confidence scores is more effective than the discrete logic of hard voting or the meta-classification approach of XGBoost stacking in this domain.

5. CONCLUSION

In this study, three robust convolutional neural network models based on popular architectures ResNet152, DenseNet201, and EfficientNet-B4 for dermoscopic image identification were introduced. Extensive experiments were conducted to evaluate the performance of the models with different architectures. Given the complexity of dermoscopic images, deeper and more powerful versions of these architectures were employed to achieve optimal results, demonstrating their effectiveness in improving diagnostic accuracy. Among them, the EfficientNet-B4 emerged as the most reliable model, striking a balance between precision and recall, both of which are critical for reliable medical diagnosis.

Moreover, the application of ensemble techniques, such as voting proved beneficial, as they combine the strengths of individual models to deliver superior performance over single model approaches. In particular, soft voting emerged as the most effective ensemble method, outperforming both hard voting and the XGBoost stacking ensemble. The XGBoost model, while demonstrating potential in other contexts, performed comparatively poorly in this study, further emphasizing the advantage of soft voting in achieving higher accuracy and balanced performance for dermoscopic image identification. This may be attributed to limitations such as the relatively small dataset size, high dimensionality of the input features, or the model's sensitivity to parameter tuning. Additionally, tree-based models like XGBoost may not capture spatial patterns in image-derived data [41] as effectively as CNN-based models.

Expanding the training data to include a more balanced and diverse representation of lesion types represents a potential area for future research aimed at improving model robustness and generalization across diverse datasets. As machine learning continues to rapidly evolve, the potential of these and emerging models to transform melanoma diagnosis remains promising.

While the proposed ensemble model demonstrates high diagnostic accuracy, its transition into actual clinical practice requires addressing several implementation challenges. A primary barrier to AI adoption in medicine is the transparency of the decision-making process; however, our integration of Grad-CAM facilitates a more interpretable diagnostic workflow by providing visual justifications that are essential for medical accountability. To bridge the gap between research and practice, future work will focus on implementing an automated pre-segmentation module using architectures like U-Net. This stage would allow the system to isolate the lesion as a specific Region of Interest (ROI), effectively filtering out non-diagnostic elements such as hair, clinical markers, or air bubbles, thereby enabling the ensemble to focus exclusively on fine-grained diagnostic features.

Furthermore, we intend to evolve the system into a multimodal framework by incorporating clinical metadata, including patient age, gender, and anatomical location, which are already provided within the ISIC datasets. We also envision deploying the model as a mobile or web-based decision support tool designed to assist general practitioners during early screening phases. These advancements, coupled with rigorous clinical validation and continued collaboration between AI researchers and healthcare specialists, hold the potential to significantly improve early detection and treatment outcomes for malignant melanoma.

The ultimate validation of this approach lies in its alignment with expert medical judgment. Future efforts will involve qualitative clinical studies where dermatologists evaluate the correlation between Grad-CAM visualizations and established diagnostic criteria, such as the ABCDE rule. By comparing the ensemble's 91.30% accuracy against traditional dermoscopic accuracy, which often falls short at 75% in routine settings [53], we aim to demonstrate the system's utility as a robust decision-support tool. This collaborative validation is essential to mitigate human error and variability, particularly in identifying asymptomatic lesions with subtle visual cues that might otherwise go unnoticed.

REFERENCES

1. Meedeniya D, De Silva S, Gamage L, Isuranga U. (2024). Skin cancer identification utilizing deep learning: A survey. /ET Image Processing. 2024; 18(13): 3731-3749.
2. Arnold M, Singh D, Laversanne M, Vignat J, Vaccarella S, Meheus F, Cust AE, De Vries E, Whiteman DC, Bray F. Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. JAMA Dermatol. 2022;158(5):495.
3. Melanoma Awareness Month 2022 [Internet]. Available from: https://www.iarc.who.int/news-events/melanoma-awareness-month-2022/
4. Shakya M, Patel R, Joshi S. A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification. Sci Rep [Internet]. 2025;15(1). Available from: https://www.nature.com/articles/s41598-024-82241-w
5. Morton, Mackie. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. Br J Dermatol. 1998;138(2):283–7.
6. Powroznik P, Skublewska-Paszkowska M, Dziedzic K, Barszcz M. Feature Fusion Graph Consecutive-Attention Network for Skeleton-Based Tennis Action Recognition. Appl Sci. 2025;15(10):5320.
7. Powroźnik P, Skublewska-Paszkowska M, Nowomiejska K, Aristidou A, Panayides A, Rejdak R. Deep convolutional generative adversarial networks in retinitis pigmentosa disease images augmentation and detection. Adv Sci Technol Res J. 2024;19(2):321–40.
8. Skublewska-Paszkowska M, Powroznik P, Rejdak R, Nowomiejska K. Application of Convolutional Gated Recurrent Units U-Net for Distinguishing between Retinitis Pigmentosa and Cone–Rod Dystrophy. Acta Mech Autom. 2024;18(3):505–13.
9. Nawaz K, Zanib A, Shabir I, Li J, Wang Y, Mahmood T, Rehman A. Skin cancer detection using dermoscopic images with convolutional neural network. Sci Rep [Internet]. 2025;15(1). Available from: https://www.nature.com/articles/s41598-025-91446-6
10. Ahmad M, Ahmed I, Chehri A, Jeon G. Fusion of metadata and dermoscopic images for melanoma detection: Deep learning and feature importance analysis. Inf Fusion. 2025;124:103304.
11. Esmaeili V, Mohassel Feghhi M, Seyedarabi H. Automatic melanoma detection using an optimized five-stream convolutional neural network. Sci Rep [Internet]. 2025;15(1). Available from: https://www.nature.com/articles/s41598-025-05675-w
12. Zhang R. Melanoma Detection Using Convolutional Neural Network. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE) [Internet]. Guangzhou, China: IEEE; 2021;75–8. Available from: https://ieeexplore.ieee.org/document/9342142/
13. Ahmed A, Sun G, Bilal A, Li Y, Ebad SA. (2025). Precision and efficiency in skin cancer segmentation through a dual encoder deep learning model. Scientific Reports. 2025; 15(1): 4815.
14. Ozdemir B, Pacal I. A robust deep learning framework for multiclass skin cancer classification. Scientific Reports. 2025; 15(1): 4938.
15. SMJ, PM, Aravindan C, Appavu R. Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimed Tools Appl. 2023;82(10):15763–78.
16. Kwiatkowska D, Kluska P, Reich A. Convolutional neural networks for the detection of malignant melanoma in dermoscopy images. Adv Dermatol Allergol. 2021;38(3):412–20.
17. Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated Melanoma

Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Trans Med Imaging. 2017;36(4):994–1004.
18. Ali R, Manikandan A, Lei R, Xu J. A novel SpaSA based hyper-parameter optimized FCEDN with adaptive CNN classification for skin cancer detection. Sci Rep [Internet]. 2024;14(1). Available from: https://www.nature.com/articles/s41598-024-57393-4
19. Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. Int J Med Inf. 2019;124:37–48.
20. Hosny KM, Kassem MA, Foaud MM. Classification of skin lesions using transfer learning and augmentation with Alex-net. Zhang J. PLOS ONE. 2019;14(5):e0217293.
21. Qin Z, Liu Z, Zhu P, Xue Y. A GAN-based image synthesis method for skin lesion classification. Comput Methods Programs Biomed. 2020;195:105568.
22. Mahbod A, Schaefer G, Wang C, Ecker R, Ellinge I. Skin Lesion Classification Using Hybrid Deep Neural Networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [Internet]. Brighton. United Kingdom: IEEE. 2019; 1229–33. Available from: https://ieeexplore.ieee.org/document/8683352/
23. Veeramani N, Jayaraman P, Krishankumar R, Ravichandran KS, Gandomi AH. DDCNN-F: double decker convolutional neural network “F” feature fusion as a medical image classification framework. Sci Rep [Internet]. 2024;14(1). Available from: https://www.nature.com/articles/s41598-023-49721-x
24. Wu H, Chen S, Chen G, Wang W, Lei B, Wen Z. FAT-Net: Feature adaptive transformers for automated skin lesion segmentation. Med Image Anal. 2022;76:102327.
25. Khan MA, Mazhar T, Ali MD, Khattak UF, Shahzad T, Saeed MM, Hamam H. Automatic melanoma and non-melanoma skin cancer diagnosis using advanced adaptive fine-tuned convolution neural networks. Discov Oncol [Internet]. 2025;16(1). Available from: https://link.springer.com/10.1007/s12672-025-02279-8
26. Aftab RS, Hamim SA, Ahmed MM, Shafi SMA, Mazid-Ul-Haque Md. SkinScanNet: A CNN-Based Model with Explainable AI for Reliable and Transparent Skin Cancer Detection. In: 2024 27th International Conference on Computer and Information Technology (ICCIT) [Internet]. Cox’s Bazar. Bangladesh: IEEE. 2024; 2846–51. Available from: https://ieeexplore.ieee.org/document/11021833/
27. Kumar MK, Vaishnavi J, Anjibabu T, Jayasri V. Skin Cancer Detection: A Hybrid Approach Combining CNNs and Explainable AI. In: 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS) [Internet]. Prawet, Thailand: IEEE; 2025;1102–7. Available from: https://ieeexplore.ieee.org/document/10968907/
28. Gamage L, Isuranga U, Meedeniya D, De Silva S, Yogarajah P. Melanoma Skin Cancer Identification with Explainability Utilizing Mask Guided Technique. Electronics. 2024;13(4):680.
29. Que SKT. Research Techniques Made Simple: Noninvasive Imaging Technologies for the Delineation of Basal Cell Carcinomas. J Invest Dermatol. 2016;136(4):e33–8.
30. The International Skin Imaging Collaboration [Internet]. Available from: https://www.isic-archive.com/
31. Caffery LJ, Rotemberg V, Weber J, Soyer HP, Malvehy J, Clunie D. The Role of DICOM in Artificial Intelligence for Skin Disease. Front Med [Internet]. 2021;7. Available from: https://www.frontiersin.org/articles/10.3389/fmed.2020.619787/full
32. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data [Internet]. 2019;6(1). Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0
33. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge [Internet]. arXiv; 2014. Available from: https://arxiv.org/abs/1409.0575
34. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition [Internet]. arXiv. 2015. Available from: https://arxiv.org/abs/1512.03385
35. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks [Internet]. arXiv; 2016. Available from: https://arxiv.org/abs/1608.06993
36. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019. Available from: https://arxiv.org/abs/1905.11946
37. Atila Ü, Uçar M, Akyol K, Uçar E. Plant leaf disease classification using EfficientNet deep learning model. Ecol Inform. 2021;61:101182.
38. Zhang C, Ma Y, editors. Ensemble Machine Learning: Methods and Applications [Internet]. New York, NY: Springer New York. 2012. Available from: https://link.springer.com/10.1007/978-1-4419-9326-7
39. Mienye ID, Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications and Prospects. IEEE Access. 2022;10:99129–49.
40. Rokach L. Ensemble-based classifiers. Artif Intell Rev. 2010 Feb;33(1–2):1–39.
41. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. San Francisco California USA: ACM; 2016; 785–94. Available from: https://dl.acm.org/doi/10.1145/2939672.2939785
42. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D.. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision. 2017; 618-626.
43. Dogan A, Birant D. A Weighted Majority Voting Ensemble Approach for Classification. In: 2019 4th International Conference on Computer Science and Engineering (UBMK) [Internet]. Samsun. Turkey: IEEE; 2019; 1–6. Available from: https://ieeexplore.ieee.org/document/8907028/
44. Al-masni MA, Al-antari MA, Park HM, Park NH, Kim TS. A Deep Learning Model Integrating FrCN and Residual Convolutional Networks for Skin Lesion Segmentation and Classification. In: 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) [Internet]. Okinawa, Japan: IEEE. 2019; 95–8. Available from: https://ieeexplore.ieee.org/document/8807441/
45. Daghrir J, Tlig L, Bouchouicha M, Sayadi M. Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In: 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) [Internet]. Sousse. Tunisia: IEEE; 2020;1–5. Available from: https://ieeexplore.ieee.org/document/9231544/
46. Shahin AH, Kamal A, Elattar MA. Deep Ensemble Learning for Skin Lesion Classification from Dermoscopic Images. In: 2018 9th Cairo International Biomedical Engineering Conference (CIBEC) [Internet]. Cairo. Egypt: IEEE. 2018; 150–3. Available from: https://ieeexplore.ieee.org/document/8641815/
47. Kaymak S, Esmaili P, Serener A. Deep Learning for Two-Step Classification of Malignant Pigmented Skin Lesions. In: 2018 14th Symposium on Neural Networks and Applications (NEUREL) [Internet]. Belgrade: IEEE. 2018; 1–6. Available from: https://ieeexplore.ieee.org/document/8587019/
48. Yu Z, Jiang X, Zhou F, Qin J, Ni D, Chen S, Lei B, Wang T. Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features. IEEE Trans Biomed Eng. 2019;66(4):1006–16.
49. Kumar, V, Choudhury, T. Real-Time Recognition of Malignant Skin Lesions using Ensemble Modeling. J Sci Ind Res. 2029;78:148–53.
50. Deep Learning Solutions for Skin Cancer Detection and Diagnosis. In: Learning and Analytics in Intelligent Systems [Internet]. Cham: Springer International Publishing; 2020; 159–82. Available from: http://link.springer.com/10.1007/978-3-030-40850-3_8
51. Rezaoana N, Hossain MS, Andersson K. Detection and Classification of Skin Cancer by Using a Parallel CNN Model. In: 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) [Internet]. Bhubaneswar, India: IEEE. 2020; 380–6. Available from: https://ieeexplore.ieee.org/document/9397987/

DOI: 10.65731/ama/2026-0033

52. Sedigh P, Sadeghian R, Masouleh MT. Generating Synthetic Medical Images by Using GAN to Improve CNN Performance in Skin Cancer Classification. In: 2019 7th International Conference on Robotics and Mechatronics (ICRoM) [Internet]. Tehran. Iran: IEEE. 2019; 497–502. Available from: https://ieeexplore.ieee.org/document/9071823/
53. Winkler JK, Blum A, Kommoss K, Enk A, Toberer F, Rosenberger A, Haenssle HA. Assessment of diagnostic performance of dermatologists cooperating with a convolutional neural network in a prospective clinical study: human with machine. JAMA dermatology. 2023; 159(6): 621-627.

Andrian Szymczyk: https://orcid.org/0009-0009-7576-0243

Maria Skublewska-Paszkowska: https://orcid.org/0000-0002-0760-7126

Pawel Powroznik: https://orcid.org/0000-0002-5705-4785