JOURNAL
COMPUTER SCIENCES INSTITUTE

JCSI 37 (2025) 492–499
Received: 8 July 2025
Accepted: 7 September 2025

Influence of activation function in deep learning for cutaneous melanoma identification

Wpływ funkcji aktywacji w uczeniu głębokim na identyfikowanie czerniaka skóry

Adrian Szymczyk*, Maria Skublewska-Paszkowska

Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract
Malignant melanoma is an aggressive skin cancer requiring early detection for effective treatment. In this study, it is hypothesized that the choice of activation function affects the classification performance of pre-trained models in melanoma detection, and that the optimal activation function varies across deep CNN architectures. The impact of various activation functions (ReLU, LeakyReLU, ELU, GELU, Swish, Mish, PReLU) on the diagnostic accuracy of ResNet152, DenseNet201, and EfficientNet-B4 models was investigated. The study was conducted using a combined ISIC dataset, comprising dermoscopic images collected between 2018 and 2020. Findings indicate EfficientNet-B4 with LeakyReLU achieved the highest accuracy of 90.5%, while DenseNet201 benefited most from ReLU (90.3%). Results confirm the influence of activation function selection, demonstrating architecture-specific optimal choices for enhanced classification.

Keywords: melanoma; skin cancer; convolutional neural networks; activation function

Streszczenie
Czerniak złośliwy jest agresywnym nowotworem skóry, którego skuteczne leczenie wymaga wczesnego wykrycia. W niniejszym badaniu postawiono hipotezę, że wybór funkcji aktywacji wpływa na skuteczność klasyfikacji wstępnie wytrenowanych modeli w wykrywaniu czerniaka oraz że optymalna funkcja aktywacji różni się w zależności od architektury głębokich sieci CNN. Zbadano wpływ różnych funkcji (ReLU, LeakyReLU, ELU, GELU, Swish, Mish, PReLU) aktywacji na dokładność diagnostyczną modeli ResNet152, DenseNet201 i EfficientNet-B4. Badanie przeprowadzono przy użyciu połączonego zbioru danych ISIC, zawierającego obrazy dermatoskopowe zebrane w latach 2018–2020. Wyniki wskazują, że model Effi-cientNet-B4 z funkcją LeakyReLU osiągnął najwyższą dokładność wynoszącą 90,5%, natomiast model DenseNet201 uzyskał najlepsze wyniki przy użyciu funkcji ReLU (90,3%). Wyniki potwierdzają wpływ wyboru funkcji aktywacji, wykazując optymalne wybory specyficzne dla architektury w celu ulepszenia klasyfikacji.

Słowa kluczowe: czerniak; rak skóry; konwolucyjne sieci neuronowe; funkcja aktywacji

*Corresponding author
Email address: **adrian.szymczyk@pollub.edu.pl** (A. Szymczyk)

1. Introduction

Cutaneous melanoma remains among the most dangerous and rapidly progressing skin cancers globally, with its occurrence steadily increasing, especially in populations with lighter skin tones [1]. Though it is less common than non-melanoma skin cancers, melanoma is responsible for a disproportionately high number of skin cancer-related fatalities. Globally, over 325,000 new cases are identified each year, and nearly 57,000 people lose their lives to the disease annually [2]. Projections from the International Agency for Research on Cancer (IARC) suggest that the number of new melanoma cases could grow by more than 50% from 2020 to 2040, eventually exceeding 500,000 cases per year. Deaths attributed to melanoma are also expected to rise beyond 100,000 annually [2]. Contributing factors to this upward trend include excessive exposure to ultraviolet (UV) radiation, often due to tanning behaviours and increased outdoor activity during summer months, as well as an aging global population. Early-stage diagnosis is critical for improving survival rates, yet this remains difficult due to the disease's often silent progression and the subtlety of its early signs, which may be overlooked by both individuals and medical professionals [3].

While conventional diagnostic practices such as dermoscopic assessment and clinical visual examination continue to be essential tools, they are inherently limited by subjective interpretation and inter-observer variability. Dermoscopy improves the visualization of skin lesion structures, enhancing diagnostic reliability, but still achieves only about 80% accuracy in real-world clinical environments [4]. These limitations underscore the urgent need for automated and more consistent diagnostic approaches.

In recent years, deep learning - particularly through Convolutional Neural Networks (CNNs) has shown exceptional promise in the field of medical imaging. CNNs can extract complex features from dermoscopic images, often delivering diagnostic results that rival or exceed those of dermatologists. Previously, models like EfficientNet, DenseNet201, and ResNet152 have been successfully applied to the classification of benign versus malignant skin lesions. However, performance in such tasks is not solely dictated by architecture or dataset; the

choice of activation function within a neural network plays a critical role in its ability to learn non-linear representations and generalize to new data [5]. The motivation behind this research lies in leveraging advanced deep learning technologies to enhance the early and accurate diagnosis of malignant melanoma, a potentially deadly skin cancer. By exploring the role of activation functions across high-performing CNN architectures, this study focuses on evaluating the effects of different activation functions - including ReLU, LeakyReLU, ELU, GELU, Swish, Mish, and PReLU when applied as the final activation layer in pre-trained deep learning models. By systematically analyzing these functions in ResNet152, DenseNet201, and EfficientNet-B4 architectures, the research aim to determine how the activation mechanism influences classification outcomes on dermoscopic datasets. This represents a nuanced approach to optimizing melanoma detection pipelines, targeting the model's internal transformation process rather than altering the architecture or input data alone.

The hypothesis is that certain advanced activation functions, particularly those with adaptive or smooth non-linearity like Mish or Swish, will provide superior performance in terms of classification metrics (such as precision and F1-score) compared to traditional activations like ReLU. Evaluations will be conducted using merged ISIC datasets to ensure a robust and diverse image set. The outcomes of this comparative study may point toward optimal activation function choices that enhance generalization, especially when paired with state-of-the-art CNNs in medical image classification tasks.

The ultimate goal of this research is to identify activation functions that maximize diagnostic accuracy for melanoma detection in pretrained networks, thereby contributing to the development of a more reliable and fully automated screening system. The anticipated benefits include reducing human error in clinical diagnosis, expediting early detection, and ultimately improving patient prognoses through timely treatment interventions.

2. Literature review

Recent progress in the use of deep learning, especially convolutional neural networks (CNNs), has greatly advanced the automatic classification and detection of melanoma and other skin-related conditions. Numerous studies have demonstrated the effectiveness of deep models by applying various architectural innovations and optimization techniques to improve classification performance.

EfficientNet, known for its scalable and well-optimized design, has emerged as a particularly effective architecture in this domain. For instance, Runyuan Zhang (2019) showed that the EfficientNet-B6 variant, enhanced through neural architecture search, achieved a strong AUC-ROC score of 0.917 in melanoma detection tasks [6]. Similarly, S. M. Jaisakthi (2023) applied transfer learning with the EfficientNet architecture and employed the Ranger optimizer, which further boosted performance, yielding an AUC-ROC of 0.9681 [7]. These results emphasize EfficientNet's ability to capture intricate dermoscopic image features. However, beyond the architecture and optimization methods, one of the often-overlooked yet impactful components in such networks is the activation function, which fundamentally affects learning dynamics and classification outcomes.

Activation functions are crucial for deep learning models, enabling them to learn complex non-linear representations and significantly impacting their training and classification performance. Functions like ReLU, while popular for their computational efficiency and ability to introduce sparsity [8], can suffer from the "dying ReLU" problem where neurons become inactive due to zero gradients for negative inputs or the derivative being zero for negative inputs, hindering model training and overall classification accuracy. Advanced activation functions such as LeakyReLU, ELU, GELU, Swish, Mish, and PReLU were developed to mitigate these issues by allowing small gradients for negative inputs or by introducing smoother, non-monotonic properties, which can lead to faster convergence [9], improved generalization [8], and potentially higher classification results in tasks like image-based diagnosis of cutaneous melanoma.

The introduction of activation functions into neural networks is fundamental for enabling nonlinear expression capabilities, which in turn enhances accuracy. Different activation functions, including ReLU, LeakyReLU, ELU, Swish, Mish, and PReLU, exhibit varying performance across diverse neural network architectures and datasets, directly influencing the model's ability to fit results and improve classification accuracy. Wang Hao (2020) demonstrated that selection of an appropriate activation function is critical as it can help balance accuracy and speed, and even improve model performance without requiring an increase in the dataset size [10]. Deep residual learning has likewise proven useful. Lequan Yu (2016) utilized fully convolutional residual networks and multi-scale contextual cues to achieve highly accurate segmentation and classification of skin lesions, which are important for distinguishing malignant from benign findings [11].

To deal with the often limited availability of annotated medical images, researchers have employed strategies like data augmentation and transfer learning. Hosny Khalid M. (2019) used AlexNet in conjunction with image augmentation techniques to build a robust lesion classification model, improving its reliability across diverse datasets [12]. Hybrid approaches have also shown potential. Mahbod Amirreza (2019) extracted deep features from pretrained CNNs such as VGG16, ResNet-18, and AlexNet, and used them in tandem with support vector machines (SVMs) for final classification. This combination leveraged the strengths of both deep and traditional machine learning techniques, achieving high predictive accuracy [13].

Activation functions are vital components in Convolutional Neural Networks (CNNs) for tasks like cutaneous melanoma diagnosis, as they introduce non-linearity and significantly influence the model's accuracy by determining information flow between layers. This study specifically demonstrates that the choice of activation

function directly impacts a CNN's performance in classifying skin lesions, with a parameterized Leaky ReLU function outperforming other nonlinear activation functions in a proposed CNN model for melanoma recognition, even with limited datasets. Therefore, selecting the appropriate activation function is crucial for effective model training and achieving higher prediction accuracy in machine learning-based melanoma diagnosis [14].

Collectively, these works reflect the expanding toolkit available for improving melanoma diagnosis using CNNs. While architectures like EfficientNet and ResNet, along with data augmentation, transfer learning, and hybrid models, contribute significantly to model performance, the research centres on a critical yet often underanalysed component - the activation function.

3. Material and Methods

This study aims to assess the impact of different activation functions on the performance of pretrained deep learning models for classifying skin lesions from dermoscopic images. It begins with a description of the dataset, followed by an overview of the experimental setup. The analysis focuses on pretrained architectures—ResNet152, DenseNet201, and EfficientNet-B4—with various activation functions (ReLU, LeakyReLU, ELU, GELU, Swish, Mish, PReLU) applied as the final non-linear layer. Details of the training pipeline, including preprocessing, augmentation, and evaluation metrics, are provided. Finally, the results are compared across activation functions to evaluate their influence on classification accuracy and convergence.

3.1. Dataset

This study utilizes a comprehensive dataset composed of 17,114 high-quality dermoscopic images, curated to support research in dermatological image analysis and improve computational diagnostic methods for melanoma detection [15]. The dataset is formed by aggregating multiple subsets from the International Skin Imaging Collaboration (ISIC) challenges conducted in 2018, 2019, and 2020, resulting in a balanced and diverse image repository suitable for training and evaluating deep learning models [16]. All images were sourced directly from the official ISIC platform, with careful selection criteria applied to reduce the risk of class imbalance presented in Figure 1.

Available in both JPEG and DICOM formats [17], the images are accompanied by structured metadata, including patient-specific information such as ID, gender, age, and lesion location, along with the binary ground truth labels indicating whether the lesion is benign or malignant. This structured labelling supports supervised learning and allows for a consistent evaluation across experimental setups.

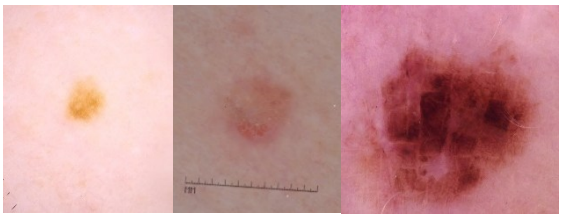

Figure 1: Representative examples of skin lesions from the merged ISIC datasets.

To investigate the effect of activation functions on melanoma classification performance, the dataset was split into training, validation, and testing sets. Of the total 17,114 images, 14,204 were used for training, and the remaining 2,910 were equally split between the validation and testing phases, as detailed in Table 1. The class distribution maintained a 70–30 ratio in favour of benign lesions, reflecting the prevalence seen in real-world clinical datasets. The data partitioning strategy was based on the approach of Serra Aksoy, who used a similar 83–17% training-to-evaluation split and reported encouraging classification outcomes [18].

Table 1: Number of the image data from ISIC-Combined dataset

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 support the evaluation of activation functions in deep learning models for melanoma classification, a robust data preprocessing pipeline was developed to enhance model generalization and reduce the risk of overfitting. The preprocessing steps were designed to standardize the dermoscopic images and introduce realistic variability, thereby simulating the diversity of real-world clinical scenarios. Specific attention was paid to resizing procedures to match the input requirements of each pretrained architecture under investigation.

The data augmentation pipeline included several transformations aimed at increasing the representational diversity of the dataset. All images were resized to a fixed dimension according to the input size expected by each CNN model. To simulate natural variability in skin lesion imaging, random horizontal and vertical flips were applied, along with random rotations to account for shifts in camera angle and patient positioning. Color jittering techniques were used to alter image brightness and contrast, emulating differences in lighting during image capture. Following augmentation, the images were converted to tensors and normalized based on the mean and standard deviation values of the ImageNet dataset: [0.485, 0.456, 0.406] for mean and [0.229, 0.224, 0.225] for standard deviation [19]. This normalization ensured compatibility with the pretrained networks and contributed to more stable and efficient training dynamics across experiments involving different activation functions.

3.3. Activation Functions

3.3.1. Rectified Linear Unit, Leaky ReLU, and Parametric ReLU

These activation functions are foundational in deep learning, primarily addressing the vanishing gradient problem. ReLU is defined as Formulae 1, outputting the input directly if positive, and zero otherwise. While simple and computationally efficient, it can suffer from the "dying ReLU" problem, where neurons become inactive for negative inputs [20].

$$f(x) = \max(0, x) \tag{1}$$

LeakyReLU mitigates this by introducing a small, non-zero slope for negative inputs, defined as Formulae 2 where α is a small constant (e.g., 0.01). This allows gradients to flow even when the input is negative. PReLU extends LeakyReLU by making α a learnable parameter, allowing the network to adaptively determine the optimal slope for negative inputs, potentially improving model performance [8].

$$f(x) = \max(\alpha x, x) \tag{2}$$

3.3.2. Exponential Linear Unit and Gaussian Error Linear Unit

These activation functions aim to produce more robust and stable learning. ELU computes by Formulae 3 and 4. It combines the benefits of ReLU (for positive inputs) with a small, negative output for negative inputs, which can push mean activations closer to zero, leading to faster learning by reducing the "bias shift" problem [8].

$$f(x) = \text{x for x} > 0 \tag{3}$$

$$f(x) = \alpha(e^x - 1) \text{ for x} \leq 0 \tag{4}$$

GELU is a more recent and highly effective activation function, particularly in transformer-based models [8]. It is defined as Formulae 5, where Φ(x) is the cumulative distribution function for the standard Gaussian distribution. GELU incorporates stochasticity, scaling the input by its probability of being greater than zero, which can lead to smoother and more accurate gradient estimates.

$$f(x) = \text{x}\Phi(\text{x}) \tag{5}$$

3.3.3.Swish and Mish

These are self-gated activation functions known for their smoothness and strong performance. **Swish** is defined as Formulae 6, where β is often a learnable parameter or set to 1. Its smooth, non-monotonic shape allows for better information propagation and can help in avoiding dead neurons [8].

$$f(x) = x \cdot sigmoid(\beta x) \tag{6}$$

Mish is another highly effective and smooth activation function, defined as Formulae 7, where $softplus(x) = \ln(1 + e^x)$ [8]. Mish's non-monotonicity and smoothness are believed to contribute to its strong empirical performance across various tasks, often outperforming ReLU and Swish in deeper networks due to its ability to retain small negative inputs.

$$f(x) = x \cdot tanh(softplus(x)) \tag{7}$$

3.4. Evaluation metrics

To assess the impact of different activation functions on the classification of dermoscopic images, each pretrained model's performance was evaluated using a set of standard metrics: accuracy, precision, recall, and F1-score. These metrics were computed based on Formulas 8-11. Considering the class imbalance present in the dataset, all metrics were calculated using the *weighted* parameter to ensure that underrepresented classes were appropriately factored into the overall evaluation. In this context, True Positives (TP) and True Negatives (TN) refer to instances where the model correctly identified lesions, whereas False Positives (FP) and False Negatives (FN) represent misclassified cases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall\ (sensitivity) = \frac{TP}{TP + FN} \tag{10}$$

$$F1\ score = \frac{2TP}{2TP + FP + FN} \tag{11}$$

3.5. Experimental setup

Every experiment was performed using an identical local hardware configuration. Given the limitations imposed by the 8GB VRAM available on the NVIDIA GeForce RTX 3060 Ti, the study focused on using pretrained models with moderate memory demands. The model fine-tuning, including the integration and testing of various activation functions as the final non-linear layer, was entirely carried out on a local machine to retain full control over training parameters.

The detailed hardware environment was as follows:

- **Graphics Card**: NVIDIA GeForce RTX 3060 Ti, equipped with 8GB VRAM,
- **Processor**: Intel Core i5-10600K, 4.1 GHz
- **Memory**: 16GB DDR4, clocked at 3200 MHz

This configuration offered sufficient performance for evaluating the impact of various activation functions without exceeding memory constraints, enabling consistent testing across all models.

3.6. Hyper-parameter Configuration

For all evaluated architectures—DenseNet-121, ResNet-50, and EfficientNet-B0—model fine-tuning was conducted using the AdamW optimizer with a learning rate set to 0.0001. The loss function employed was CrossEntropyLoss, adjusted with class weights [1.44, 3.29] to mitigate class imbalance. A weight decay of 0.001 was applied to prevent overfitting. Due to differing memory

requirements, batch size was set to 32 for both DenseNet and ResNet models, while for EfficientNet-B0, a reduced batch size of 16 was necessary to accommodate GPU memory limitations. All other training parameters and data preprocessing procedures remained consistent across architectures to ensure comparability of results.

3.7. Research scenario

1. Environment Setup

To establish a consistent framework for experimentation, the research environment was configured with Python version 3.11 and a set of essential libraries tailored for deep learning and image processing. Key dependencies included PyTorch (2.4.1) for model implementation, NumPy (1.23.5) for numerical operations, and Pandas (2.2.2) for data management. Visualization of training dynamics and evaluation outcomes was supported by Matplotlib (3.9.2). The system was explicitly set to leverage GPU acceleration, designating the CUDA-enabled device as the default compute backend to enhance training efficiency, particularly during the evaluation of multiple activation functions.

2. Model Fine-Tuning with Activation Functions

The model adaptation phase began with preprocessing steps such as resizing dermoscopic images to architecture-specific dimensions (ranging from 224×224 to 380×380), normalization, and augmentations to increase training set variability. Pretrained models were modified only at their final activation layer to incorporate and compare different activation functions: ReLU, LeakyReLU, ELU, GELU, Swish, Mish, and PReLU, respectively. The models were trained on a curated subset of 14,204 labeled images from the ISIC dataset, with training governed by techniques like weight decay to optimize learning and mitigate overfitting across function variants.

3. Model Evaluation

Each trained model configuration was tested using a hold-out test set of 1,455 dermoscopic images, none of which were seen during training. The objective was to determine how each activation function influenced the model's ability to correctly classify lesions as either benign or malignant.

4. Performance Evaluation and Comparative Analysis

Model performance under each activation function was evaluated using standard classification metrics, including accuracy, precision, recall (sensitivity), and F1-score. To address class imbalance within the dataset, all metrics were calculated using weighted averaging, ensuring a fair assessment across different lesion types. A comprehensive comparative analysis was then conducted across all selected pre-trained architectures: ResNet152, DenseNet201, and EfficientNet-B4 to examine the influence of activation function choice on diagnostic accuracy. This analysis aimed to identify activation functions that consistently enhance predictive performance in the context of automated melanoma detection.

4. Results

Following the training phase, each model configuration was evaluated to determine its classification performance and consistency. Predictions were generated on the test dataset and compared against the corresponding ground truth labels. The analysis focused on assessing the influence of different activation functions on model behavior when integrated into the final layer of pretrained architectures.

To ensure a consistent evaluation protocol, various activation functions were integrated into the final layer of pretrained CNN models, followed by additional training for 5 epochs. This brief fine-tuning was conducted to allow the models to adjust to the new nonlinearities and to assess whether further training would improve performance. The goal was to evaluate how different activation functions influence inference performance on dermoscopic test data. Each model was trained and evaluated once per activation function.

To enhance the reliability and consistency of the evaluation, each model configuration was evaluated over 10 independent runs, and the resulting performance metrics were averaged to mitigate random variation. A thorough assessment conducted on the test dataset revealed meaningful insights, with the results for the best-performing model–activation function pairs clearly outlined in Table 2. Among these, EfficientNet-B4 paired with LeakyReLU achieved the highest overall results across all evaluated metrics, including accuracy (90.5%) and F1-score (90.4%). DenseNet201 using the standard ReLU function followed closely, also delivering strong performance. Interestingly, ResNet152 combined with LeakyReLU despite not being the leading model, still produced competitive scores across the board. These findings underscore the significant influence that activation functions can have, even at the inference stage, on the predictive capabilities of pretrained convolutional networks in skin lesion classification tasks.

Table 2: Performance models with best activation functions across 10 test repetitions (mean ± standard deviation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet152 with Leaky-ReLU	89.0 ±0.08	88.9 ±0.10	89.0 ±0.11	89.0 ±0.10
DenseNet201 with ReLU	90.3 ±0.04	90.2 ±0.06	90.3 ±0.06	90.3 ±0.04
EfficientNet-B4 with Leaky-ReLU	90.5 ±0.15	90.5 ±0,12	90.5 ±0.13	90.4 ±0.15

These results highlight how different activation functions can influence not just overall performance metrics but also specific error types, which is particularly important in medical diagnostics.

Although Table 3 shows that the original ResNet152 model (with its default configuration) slightly outperformed the LeakyReLU-modified version in overall accuracy (89.2% vs. 89.0%) and F1-score (89.4% vs. 89.0%). These results suggest that while default ReLU remains a strong baseline, LeakyReLU provides a comparable and in some broader comparisons, slightly better alternative, particularly when evaluated in isolation across diverse activation setups.

Table 3: Performance comparison of ResNet152 before and after applying the LeakyReLU activation function

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet152	89.2 ±0.014	89.8 ±0.012	90.2 ±0.014	89.4 ±0.013
ResNet152 with Leaky-ReLU	89.0 ±0.08	88.9 ±0.10	89.0 ±0.11	89.0 ±0.10

The comparison of activation functions applied to the ResNet152 architecture reveals that LeakyReLU delivers the most consistent and high-performing results among all tested functions. As shown in the bar charts (Figure 2 and 3), it achieved the highest F1-score (0.890) and accuracy (0.890), slightly outperforming the standard ReLU function (F1 = 0.887, Acc = 0.888). Other activation functions such as GELU and Swish+SiLU demonstrated moderate results, while PReLU, SELU, and Mish performed significantly worse across both metrics.

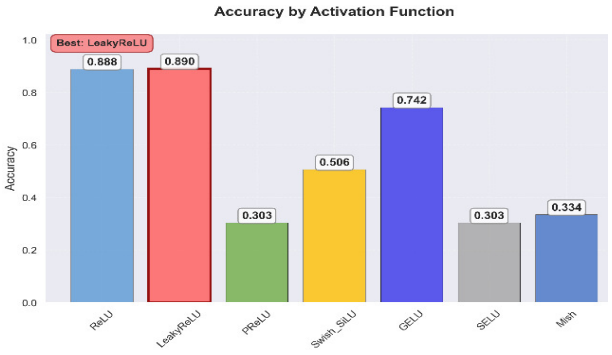

Figure 2: ResNet152 – accuracy by activation function.

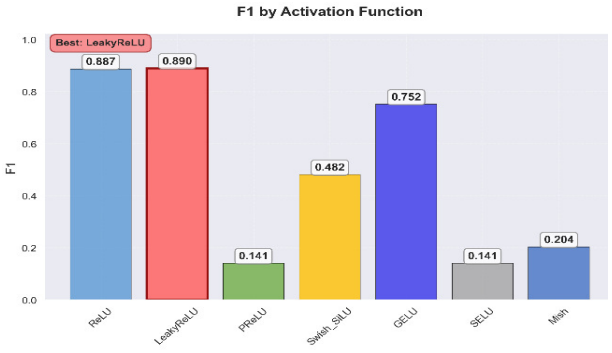

Figure 3: ResNet152 – F1-score by activation function.

In conclusion, LeakyReLU emerges as the most robust activation function for this classification task when applied to the pretrained ResNet152 model, maintaining high predictive performance while slightly reducing the risk of neuron inactivity compared to traditional ReLU.

When comparing to the baseline performance shown in Table 4, the application of the ReLU function clearly improves the model's predictive ability. Accuracy increased from 88.3% to 90.3%, and F1-score rose from 88.5% to 90.3%, indicating a consistent gain in both precision and recall.

Table 4: Performance comparison of DenseNet201 before and after applying the ReLU activation function

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet201	88.3 ±0.004	89.3 ±0.003	88.2 ±0.006	88.5 ±0.005
DenseNet201 with ReLU	90.3 ±0.04	90.2 ±0.06	90.3 ±0.06	90.3 ±0.04

The comparison of activation functions for DenseNet201 shows consistently high performance across all tested options, with ReLU slightly outperforming others. As illustrated in the Figure 4, ReLU achieved the highest F1-score (0.903) and accuracy (0.903), closely followed by LeakyReLU (F1 = 0.901, Acc = 0.901) and PReLU (F1 = 0.898, Acc = 0.899). The differences between the functions are minimal, with all metrics staying within a narrow performance range, suggesting that DenseNet201 is relatively robust to the choice of activation.

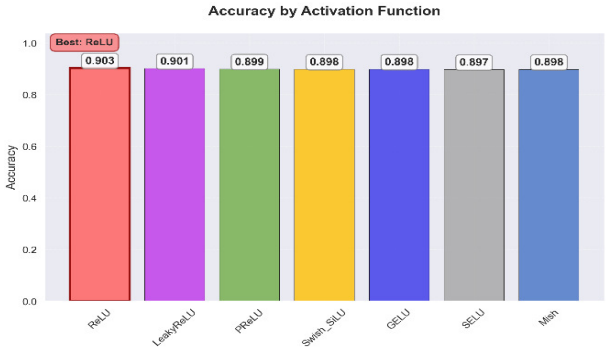

Figure 4: DenseNet201– accuracy by activation function.

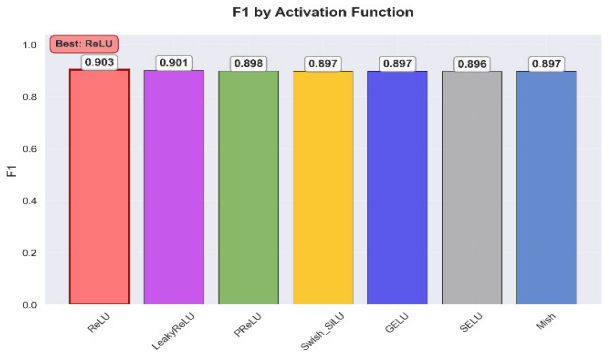

Figure 5: DenseNet201– F1-score by activation function.

These results suggest that while DenseNet201 performs reliably across various nonlinearities, ReLU remains the most effective activation function for this architecture in the context of melanoma classification, offering slight yet consistent improvements in key metrics.

Interestingly, as shown in Table 5, the use of LeakyReLU led to a marginal performance shift compared to the default EfficientNet-B4 setup. While the baseline model achieved a slightly higher F1-score (90.6% vs. 90.4%) and precision, the differences across all metrics remain negligible ($\leq 0.2\%$). This indicates that the switch to LeakyReLU does not introduce meaningful performance degradation, and may even provide slight advantages in specific use cases.

Table 5: Performance comparison of EfficientNet-B4 before and after applying the LeakyReLU activation function

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNet-B4	90.4 ±0.003	90.8 ±0.002	90.4 ±0.002	90.6 ±0.003
EfficientNet-B4 with Leaky-ReLU	90.5 ±0.15	90.5 ±0.12	90.5 ±0.13	90.4 ±0.15

The evaluation of activation functions for EfficientNet-B4 highlights LeakyReLU as the top-performing option, achieving the highest accuracy (0.905) and F1-score (0.905) among all tested nonlinearities. ReLU and GELU followed closely with nearly identical scores (F1 ≈ 0.902–0.904), suggesting a high level of stability in this architecture regardless of activation function.

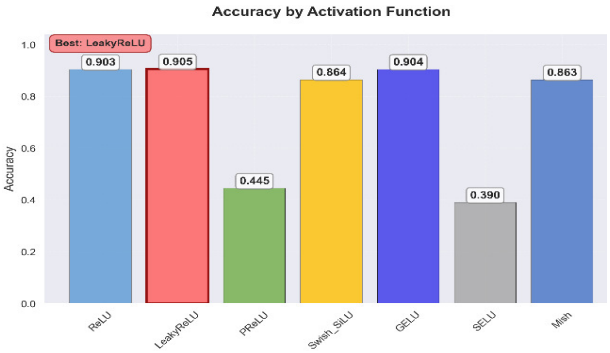

Figure 6: EfficientNet-B4 – accuracy by activation function.

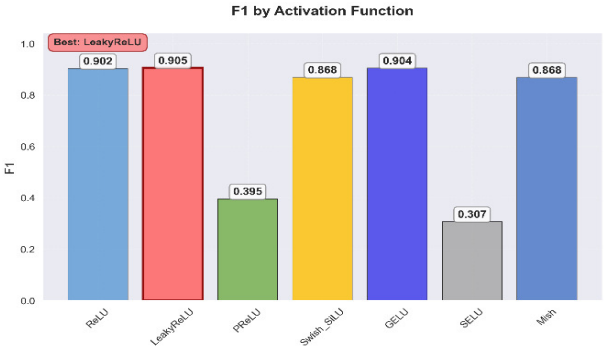

Figure 7: EfficientNet-B4 – F1-score by activation function.

Activation functions such as PReLU and SELU performed significantly worse in this context, with accuracy scores falling below 50% (Figure 6-7), reinforcing that not all advanced activations are well-suited for every pretrained architecture. In conclusion, EfficientNet-B4 paired with LeakyReLU maintains strong, balanced performance, making it a reliable choice for skin lesion classification tasks, especially in settings where small performance gains are meaningful.

5. Discussion

This study investigated the influence of various activation functions on the classification performance of pretrained convolutional neural networks (CNNs) in the context of melanoma detection. By substituting the original activation functions with alternatives such as LeakyReLU, PReLU, GELU, Swish, SELU, and Mish. The experiments provided valuable insight into how different nonlinear transformations affect model behavior during inference, without the need for additional training or fine-tuning.

The results offer strong empirical support for the initial hypothesis, demonstrating that activation function selection has a measurable effect on key performance metrics, even when applied to fixed, pretrained architectures.

Marked differences were observed across activation functions. In particular, PReLU and SELU consistently yielded lower performance across two tested models, underperforming in both accuracy and F1-score when compared to ReLU, LeakyReLU, and GELU. This consistent underperformance suggests that certain activation functions may be suboptimal for facilitating feature propagation or maintaining gradient stability in pretrained networks used for this task.

The findings also reinforce the second hypothesis: the optimal activation function varies across CNN architectures. For instance, ResNet152 achieved its highest accuracy (89.0%) with LeakyReLU, closely followed by ReLU. DenseNet201, on the other hand, performed best with ReLU, improving upon its baseline by approximately 2 percentage points. EfficientNet-B4 is the highest-performing model overall, with 90.5% accuracy showed optimal results with LeakyReLU, although both GELU and ReLU remained highly competitive. The relatively small performance gap among these top activations for EfficientNet-B4 may point to its robustness and adaptability with respect to activation function choice.

These results underscore the architecture-dependent nature of activation function effectiveness. Traditional activations such as ReLU continue to demonstrate strong performance in medical image classification, while the introduction of more complex or less conventional nonlinearities does not necessarily lead to improvements and may, in fact, hinder performance in certain settings.

In line with current literature, this study reaffirms the competitiveness of EfficientNet and DenseNet architectures in dermatological diagnostics. However, it adds a novel contribution by systematically comparing activation functions at the inference level, without model retraining.

6. Conclusion

Although medical technologies have advanced significantly, there remains a pressing need for automated systems to assist in disease diagnosis. Modern algorithms and models continue to enhance the precision of medical data classification [21]. This study investigated the influence of various activation functions on the performance of pretrained convolutional neural networks in the task of binary skin lesion classification. The empirical results demonstrate that the activation function plays a significant role in determining the effectiveness of a model, even when the architecture and weights remain unchanged.

ReLU and LeakyReLU consistently achieved the highest classification metrics across most architectures, indicating their robustness and reliability in this medical imaging context. In contrast, functions such as PReLU

and SELU were associated with noticeably lower performance, suggesting limited compatibility with the tested CNNs in this specific task.

Importantly, the optimal activation function was not consistent across all models, but instead varied depending on the architecture. For instance, LeakyReLU yielded the best results in ResNet152 and EfficientNet-B4, while DenseNet201 performed best with ReLU. This architecture-specific sensitivity underscores the importance of aligning nonlinear transformations with model structure to maximize performance.

These findings highlight the need for deliberate and context-aware selection of activation functions when deploying pretrained models in sensitive applications such as medical diagnostics. Even when architectures are fixed and no retraining is performed, internal components like activation functions can be optimized to enhance accuracy and reliability. Future research may explore adaptive or learnable activations, test broader datasets, or assess how these functions perform in full end-to-end training pipelines.

References

[1] D. C. Whiteman, A. C. Green, C. M. Olsen, The growing burden of invasive melanoma: projections of incidence rates and numbers of new cases in six susceptible populations through 2031, Journal of Investigative Dermatology 136(6) (2016) 1161-1171, https://doi.org/10.1016/j.jid.2016.01.035.

[2] Global impact of skin cancer, https://www.iarc.who.int/news-events/melanoma-awareness-month-2022/, [03.02.2025].

[3] P. P. Naik, Cutaneous malignant melanoma: a review of early diagnosis and management, World journal of oncology 12(1) (2021) 7-19, https://doi.org/10.14740/wjon1349.

[4] C. Morton, R. M. Mackie, Clinical accuracy of the diagnosis of cutaneous malignant melanoma, British Journal of Dermatology 138(2) (1998) 283–287, https://doi.org/10.1046/j.1365-2133.1998.02075.x.

[5] P. Carcagnì, M. Leo, G. Celeste, C. Distante, A. Cuna, A systematic investigation on deep architectures for automatic skin lesions classification, In 2020 25th International Conference on Pattern Recognition (ICPR) (2021) 8639-8646.

[6] R. Zhang, Melanoma detection using convolutional neural network, In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), IEEE (2021) 75-78, https://doi.org/10.1109/ICCECE51280.2021.9342142.

[7] J. SM, C. Aravindan, R. Appavu, Classification of skin cancer from dermoscopic images using deep neural network architectures, Multimedia Tools and Applications 82 (2023) 15763–15778, https://doi.org/10.1007/s11042-022-13847-3.

[8] A. D. Rasamoelina, F. Adjailia, P. Sinčák, A review of activation function for artificial neural network, In 2020 IEEE 18th world symposium on applied machine intelligence and informatics IEEE (2020) 281-286, https://doi.org/10.1109/SAMI48414.2020.9108717.

[9] Y. Wang, Y. Li, Y. Song, X. Rong, The influence of the activation function in a convolution neural network model of facial expression recognition, Applied Sciences 10(5) (2020) 1897.

[10] W. Hao , W. Yizhou , L. Yaqin, S. Zhili, The role of activation function in CNN, In 2020 2nd international conference on information technology and computer application IEEE (2020) 429-432.

[11] L. Yu, H. Chen, Q. Dou, J. Qin, P. A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE transactions on medical imaging 36 (2017) 994–1004.

[12] K. M. Hosny, M. A. Kassem, M. M. Foaud, Classification of skin lesions using transfer learning and augmentation with Alex-net, PloS one 14(5) (2019) e0217293, https://doi.org/10.1371/journal.pone.0217293.

[13] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, I. Ellinge, Skin lesion classification using hybrid deep neural networks, In ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE (2019) 1229–1233, http://dx.doi.org/10.1109/ICASSP.2019.8683352.

[14] M. A. Rasel, U. H. Obaidellah, S. A. Kareem, Convolutional neural network-based skin lesion classification with Variable Nonlinear Activation Functions, IEEE Access 10 (2022) 83398-83414, https://doi.org/10.1109/ACCESS.2022.3196911.

[15] S. K. T. Que, Research techniques made simple: noninvasive imaging technologies for the delineation of basal cell carcinomas, Journal of Investigative Dermatology 136(4) (2016) 33–38, https://doi.org/10.1016/j.jid.2016.02.012.

[16] The International Skin Imaging Collaboration, https://www.isic-archive.com/, [03.02.2025].

[17] L. J. Caffery, V. Rotemberg, J. Weber, H. P. Soyer, J. Malvehy, D. Clunie, The role of DICOM in artificial intelligence for skin disease, Frontiers in medicine 7 (2020) 619787, https://doi.org/10.3389/fmed.2020.619787.

[18] S. Aksoy, P. Demircioglu, I. Bogrekci, Enhancing melanoma diagnosis with advanced deep learning models focusing on vision transformer, swin transformer, and convnext, Dermatopathology 11(3) (2024) 239–252, https://doi.org/10.3390/dermatopathology11030026.

[19] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, Communications of the ACM 60(6) (2017) 84-90, https://doi.org/10.1145/3065386.

[20] P. Powroznik, M. Skublewska-Paszkowska, R. Rejdak, K. Nowomiejska, Automatic method of macular diseases detection using deep CNN-GRU network in oct images, acta mechanica et automatica 18(4) (2024) 697-706, https://doi.org/10.2478/ama-2024-0074.

[21] P. Powroznik, M. Skublewska-Paszkowska, K. Nowomiejska, B. Gajda-Deryło, M. Brinkmann, M. Concilio, R. Rejdak, Residual self-attention vision transformer for detecting acquired vitelliform lesions and age-related macular drusen, Scientific Reports 15(1) (2025) 1-22, https://doi.org/10.1038/s41598-025-02299-y.