applied system innovation

MDPI

Article

Exploring Pre-Trained Models for Skin Cancer Classification

Abdelkader Alrabai [1,*], Amira Echtioui [1] and Fathi Kallel [1,2]

[1] Advanced Technologies for Medicine and Signals Laboratory 'ATMS', National Engineering School of Sfax (ENIS), Sfax University, Sfax 3038, Tunisia
[2] National School of Electronics and Communications, Sfax University, Sfax 3018, Tunisia
* Correspondence: abdelkader.alrabai@enis.tn

Abstract: Accurate skin cancer classification is essential for early diagnosis and effective treatment planning, enabling timely interventions and improved patient outcomes. In this paper, the performance of four pre-trained models—two convolutional neural networks (ResNet50 and VGG19) and two vision transformers (ViT-b16 and ViT-b32)—is evaluated in distinguishing malignant from benign skin cancers using a publicly available dermoscopic dataset. Among these models, ResNet50 achieved the highest performance across all the evaluation metrics, with accuracy, precision, and recall of 89.09% and an F1 score of 89.08%, demonstrating its ability to effectively capture complex patterns in skin lesion images. While the other models produced competitive results, ResNet50 exhibited superior robustness and consistency. To enhance model interpretability, two eXplainable Artificial Intelligence (XAI) techniques, Local Interpretable Model-Agnostic Explanations (LIME) and integrated gradients, were employed to provide insights into the decision-making process, fostering trust in automated diagnostic systems. These findings underscore the potential of deep learning for automated skin cancer classification and highlight the importance of model transparency for clinical adoption. As AI technology continues to evolve, its integration into clinical workflows could improve diagnostic accuracy, reduce the workload of healthcare professionals, and enhance patient outcomes.

Keywords: classification; skin cancer; CNN; ViT; XAI

Academic Editor: Friedhelm Schwenker

Received: 18 December 2024
Revised: 23 February 2025
Accepted: 10 March 2025
Published: 13 March 2025

Citation: Alrabai, A.; Echtioui, A.; Kallel, F. Exploring Pre-Trained Models for Skin Cancer Classification. *Appl. Syst. Innov.* **2025**, *8*, 35. https://doi.org/10.3390/asi8020035

1. Introduction

Skin cancer is a prevalent condition that can manifest with symptoms such as burning, bleeding, and itching. While survival rates remain high, the increasing number of diagnoses has become a global concern [1]. It is estimated that approximately one in five individuals will develop skin cancer during their lifetime. Early detection and effective management have significantly reduced mortality rates. Skin cancers are primarily categorized into two types: non-melanoma and melanoma [2]. Early diagnosis substantially increases the likelihood of successful treatment for all skin cancer types. However, distinguishing between benign and malignant skin lesions remains a challenge, even for experienced dermatologists [3].

Recent advancements in computer vision and deep learning have demonstrated promising results in localizing and classifying skin lesions, enabling faster and more accurate analysis of dermatological images. These technologies aim to enhance skin cancer detection and facilitate timely, personalized treatment [4]. The conventional process for diagnosing skin cancer involves image acquisition, preprocessing, segmentation, feature extraction, and classification. Artificial intelligence (AI) has made early skin cancer detection increasingly feasible, reducing cancer-related morbidity and mortality [5].

Over the past decades, various Computer-Aided Diagnosis (CAD) systems have been developed for skin cancer detection. Traditional computer vision techniques primarily relied on feature extraction based on texture, color, and shape. More recently, deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful tool, achieving remarkable performance in skin cancer classification [6]. Moreover, research and development efforts have increasingly focused on integrating attention mechanisms into CNN-based architectures. These attention-driven transformer models have gained traction due to their ability to capture long-range dependencies and learn complex feature representations [7].

Transformers, originally developed for sequence modeling and machine translation, have become the leading deep learning models for numerous Natural Language Processing (NLP) tasks. The latest advancements in this domain include self-supervised transformers, which are pre-trained on large datasets and later fine-tuned on smaller datasets for specific applications [8]. Following their success in NLP, transformers have been increasingly adopted in computer vision. Vision transformer (ViT) models were among the first to achieve performance comparable to CNNs, processing images as sequences of patches. Inspired by ViT's success, several modified architectures and training strategies have been proposed, leading to significant improvements in transformer-based vision models [9].

With the growing interest in machine learning, deep learning, and neural networks, automated skin cancer classification has become a crucial area of research [10]. However, a major challenge in AI-based medical applications is the lack of transparency in algorithmic decision making. To address this issue, eXplainable Artificial Intelligence (XAI) techniques have been introduced to enhance model interpretability. XAI provides insights into the reasoning behind model predictions, improving trust in AI-driven diagnostic systems and fostering collaboration between AI experts and medical professionals [11]. In precision medicine, explainability is essential, as clinicians require more than just binary classifications—they need detailed insights to support their diagnostic decisions [12].

Accurately distinguishing malignant from benign skin lesions remains a significant challenge in dermatology. Deep learning and eXplainable Artificial Intelligence (XAI) have shown considerable potential in improving diagnostic accuracy and transparency. This study presents a comprehensive evaluation of four pre-trained deep learning models—two convolutional neural networks (CNNs) (ResNet50 and VGG19) and two vision transformers (ViTs) (ViT-b16 and ViT-b32)—for skin lesion classification using a publicly available dermoscopic dataset. While CNNs have been widely applied in dermatological diagnostics, the potential of ViTs remains underexplored. This study provides a comparative analysis of CNNs and ViTs, assessing their respective effectiveness in skin cancer detection.

A key novelty of this research is the systematic integration of XAI techniques, specifically Local Interpretable Model-Agnostic Explanations (LIME) and integrated gradients, to enhance model transparency and interpretability. Unlike previous studies that primarily focus on CNN-based architectures or apply XAI techniques to a single model, this research evaluates multiple deep learning models across different architectures, offering a broader and more in-depth analysis of model explainability. By incorporating XAI, this study ensures that model decisions are interpretable, thereby improving clinician trust in AI-driven diagnostics. The primary contributions of this study are as follows:

- Comparative evaluation of CNNs (ResNet50 and VGG19) and vision transformers (ViT-b16 and ViT-b32) for skin lesion classification, addressing a gap in prior research that predominantly focused on CNNs.
- Improving the performance of the models employed and benchmarking them against models from previous research that utilized the same dataset in this analysis.

- This study emphasizes the crucial role of interpretation methods, highlighting their importance in enhancing model transparency and building confidence in the results. By applying both (LIME and integrated gradients) techniques to the best-performing model, the study strengthens the interpretation and validation of the model's outcomes.
- Exploration of transformer-based models in dermatology, assessing their viability against traditional CNNs.

2. Related Works

Deep learning has significantly advanced machine learning, with deep neural networks playing a crucial role in skin cancer detection through various computer-based approaches [13]. Gouda et al. [14] applied convolutional neural networks (CNNs), utilizing transfer learning models such as ResNet50, InceptionV3, and Inception-ResNet, to distinguish between malignant and benign skin lesions. Their approach demonstrated high classification performance in identifying skin lesions. Similarly, Kousis et al. [15] evaluated multiple CNN architectures and employed data augmentation, transfer learning, and fine-tuning to address challenges such as class imbalance and image similarity. Their best-performing model was integrated into a mobile application for real-time skin lesion classification, also offering personalized sun exposure recommendations. Additionally, Salma and Eltrass [16] developed an automated Computer-Aided Diagnosis (CAD) system for skin lesion classification, achieving high accuracy while maintaining low computational complexity. Several pre-trained CNNs were assessed, with ResNet50 combined with a Support Vector Machine (SVM) yielding the best results.

Deep learning has also been widely applied in oncology to improve diagnosis and treatment planning. However, human involvement remains essential for clinical adoption. To be effectively integrated into medical workflows, deep learning models must not only achieve high predictive accuracy but also ensure interpretability and provide an estimate of prediction uncertainty [17]. A common challenge is that deep learning models are often perceived as black-box systems, raising concerns regarding potential biases, particularly in medical applications. This has led to a growing demand for eXplainable Artificial Intelligence (XAI) techniques to enhance the interpretability of these models [18]. Several explanation methods have been developed to generate heatmaps, highlighting important image regions by backpropagating gradients and analyzing how changes in individual pixels influence the model's decision-making process [19].

Regarding XAI techniques in skin cancer detection, Hauser et al. [20] conducted a systematic review on the integration of XAI in deep neural networks (DNNs) for skin lesion classification, focusing on common visualization techniques. However, there remains a lack of comprehensive evaluations on the effectiveness of XAI in deep learning-based cancer detection, particularly concerning its impact on human decision making and confidence in CAD systems.

Several researchers have explored various XAI methodologies to enhance model interpretability. Nigar et al. [21] proposed an XAI-based skin lesion classification system aimed at improving the early detection of skin cancer. By incorporating LIME, the model provided interpretable visual explanations, enhancing its applicability in clinical environments. Saarela and Georgieva [22] compared two local explanation techniques—integrated gradients and LIME—for skin lesion classification. Their findings indicated that while both methods demonstrated strong local fidelity, integrated gradients exhibited greater stability and robustness, whereas LIME provided more intuitive explanations. Hurtado et al. [23] analyzed the interpretability of LIME and SHAP (Shapley Additive Explanations) in melanoma classification, assessing their reproducibility and execution time. Their results

indicated that LIME outperformed SHAP in both aspects. Bokadia et al. [24] examined six saliency methods (integrated gradients, Kernel SHAP, Occlusion, RISE, Grad-CAM, and Attention) on a melanoma dataset, introducing two novel metrics: perceptual interpretability (visual coherence) and semantic interpretability (overlap with expert-identified features). Similarly, Shivhare et al. [25] evaluated three XAI techniques—integrated gradients, LIME, and Grad-CAM—on CNN models for lung cancer classification. Their study found that Grad-CAM achieved the lowest computational cost, while LIME required the highest execution time. These works provided diverse perspectives on the use of XAI for skin cancer classification, emphasizing the balance between accuracy, stability, interpretability, and clinical applicability. To make XAI techniques truly reliable in clinical practice, there is still a need for further research on ensuring the reproducibility and stability of explanations across different datasets and imaging conditions.

Recently, transformers have emerged in medical image analysis, being increasingly utilized for disease diagnosis and clinical applications [26]. Several studies have investigated transformer-based architectures in comparison to CNNs. Nikitin and Shapoval [27] evaluated vision transformers (ViTs) against convolutional models for skin cancer classification, suggesting that ViTs may outperform CNNs, particularly on large datasets. Dagnaw et al. [28] explored skin lesion classification using various pre-trained models, including ViTs, Swin Transformers, and CNNs. Aslan [29] compared ResNet18 and ResNet50 with ViTs for skin lesion classification, demonstrating that ViTs achieved superior performance metrics; however, a major limitation was their longer training time. Zhao [30] investigated deep learning methods for automatic skin lesion classification, comparing CNN architectures (VGGNet and ResNet) with transformer-based models (ViT and DeepViT). Their findings indicated that CNNs still outperformed ViTs in skin cancer classification, despite ViTs' promising capabilities.

The reviewed studies underscore the critical role of pre-trained deep learning models in skin cancer classification and their potential for early detection. The continuous advancement of these models holds significant promise for improving diagnostic accuracy, ultimately enhancing patient survival rates and expediting recovery. Furthermore, most studies highlight the importance of eXplainable AI (XAI) techniques in ensuring transparent and interpretable model predictions, thereby increasing confidence in automated diagnostic systems. However, further research and continuous innovation are essential to fully harness the potential of deep learning and XAI in skin cancer diagnosis.

Previous works have primarily concentrated on CNNs, with only a few studies exploring ViTs and comparing them to CNNs. These studies have highlighted discrepancies in the results of CNN and ViT comparisons, influenced by factors such as data size and type, preprocessing methods, and variations in model training parameters. In addition, there is an ongoing need to identify strategies to improve these models for better performance, particularly in the detection and diagnosis of skin cancers. This area remains underexplored and warrants further investigation. The proposed method aims to address this gap by evaluating both CNN-based models (ResNet50 and VGG19) and transformer-based architectures (ViT-b16 and ViT-b32) for skin lesion classification. This approach contributes to the enhancement of pre-trained models, the advancement of ViT techniques, and their application in skin cancer diagnosis, ultimately leading to a more comprehensive understanding of their clinical relevance.

3. Materials and Methods

The proposed method follows a structured pipeline for skin lesion classification, integrating deep learning models and eXplainable AI (XAI) techniques, as illustrated in Figure 1. Initially, raw skin lesion images undergo preprocessing, which includes resizing to

ensure uniform dimensions, normalization to standardize pixel values for improved model convergence, and data augmentation to enhance model generalization. The study evaluates both convolutional neural network (CNN)-based architectures (ResNet50 and VGG19) and vision transformer (ViT) models (ViT-b16 and ViT-b32) to assess their performance in skin lesion classification.

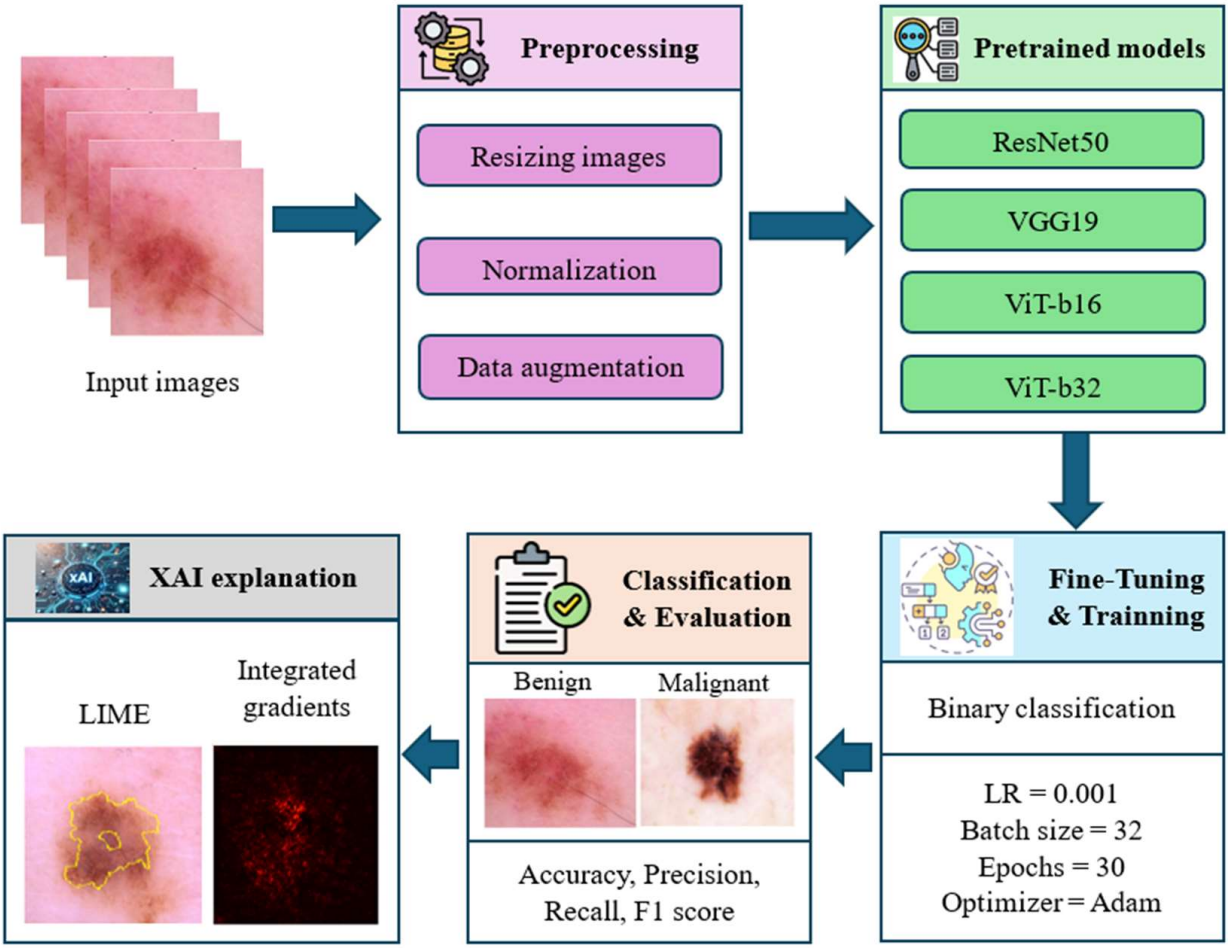

Figure 1. Overview of the proposed method for skin lesion classification.

To enhance interpretability and provide transparency in decision making, the study employs XAI techniques, specifically LIME and integrated gradients, to highlight the critical regions influencing model predictions. The integration of XAI techniques aims to enhance both the classification accuracy and the reliability of automated skin lesion diagnosis. These techniques are applied to the top-performing model among those utilized in the study.

3.1. Dataset

The study utilized a publicly accessible skin cancer dataset available on Kaggle [31], which is a part of the ISIC Archive [32]. This dataset consists of two distinctive classes, namely benign and malignant. To provide a visual understanding, Figure 2 showcases a sample of images from both classes. Furthermore, Figure 3 presents detailed information concerning the distribution, statistics, and characteristics of the images contained by this dataset, highlighting compositions of the employed dataset for this study.

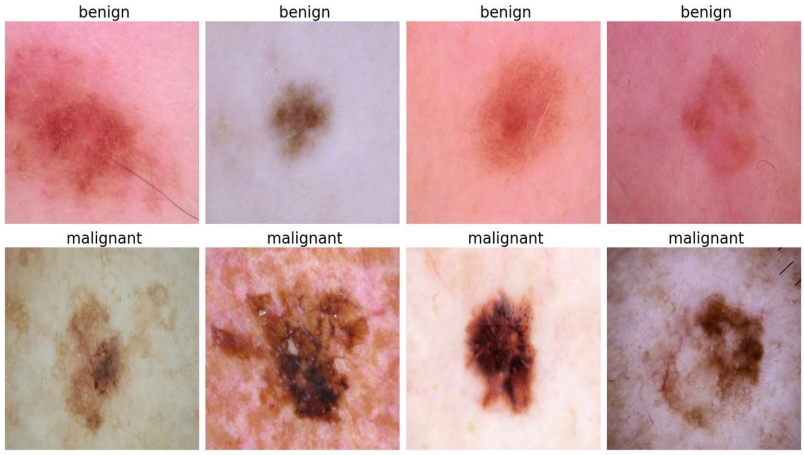

Figure 2. Selected sample.

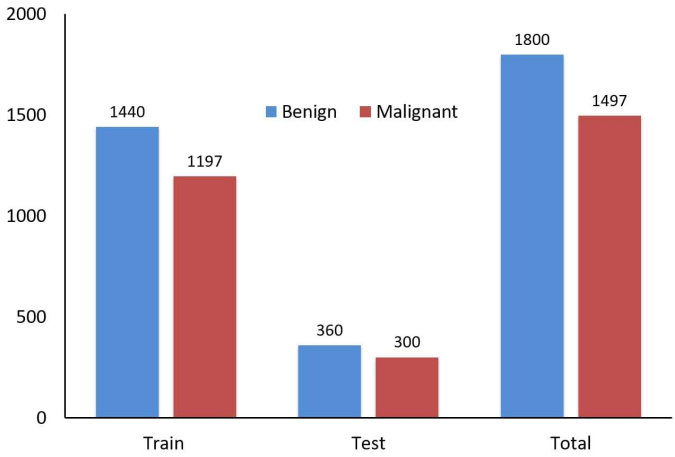

Figure 3. Dataset statistics.

3.2. Preprocessing

Data preprocessing is essential for optimal classification outcomes in machine learning. It includes tasks such as discretizing data, removing outliers, integrating data from a variety of sources, handling missing values, and normalizing features. Normalization is particularly crucial as it standardizes features to a common range, ensuring that larger values do not dominate smaller ones [33]. Data augmentation is a technique utilized to expand and diversify datasets while maintaining their labels. It generates more data from a small dataset and reduces the risk of overfitting [34]. The majority of effective data augmentation aspects are applied to image datasets, where the original images can be altered through geometric transformations (contrast adjustments, rotation, scaling, and translation) while preserving their class labels [35].

Before model training began, we applied preprocessing actions to standardize the dataset and ensure its compatibility with model architectures. This involved resizing all the images to a uniform 224 $\times$ 224 pixels and normalizing their pixel values. To enhance the diversity of the training data without requiring further data collection and mitigate overfitting, different augmentation aspects were implemented. The employed augmentation pipeline consists of a sequence of transformations including horizontal and vertical flips, brightness and contrast adjustments, and 90-degree rotations. These preprocessing steps ensured uniformity in image size and intensity, preparing the dataset for successful training.

3.3. Considered Pre-Trained Models

This study employs four pre-trained deep learning models for skin lesion classification: two convolutional neural networks (CNNs), ResNet50 and VGG19, and two transformer-based architectures, ViT-b16 and ViT-b32. These models were selected based on their strong performance in image classification tasks and their potential to enhance both accuracy and interpretability in automated skin cancer detection.

3.3.1. ResNet50

ResNet50 (Residual Network with 50 layers) is a deep CNN that addresses the vanishing gradient problem through residual connections, which allow gradients to bypass certain layers, facilitating more efficient training of deep networks. The architecture consists of bottleneck residual blocks, each comprising convolutional layers, batch normalization, and ReLU activations. Due to its ability to extract hierarchical features, ResNet50 has been widely used in medical imaging, particularly for distinguishing subtle differences between benign and malignant lesions.

3.3.2. VGG19

VGG19 is a deep CNN with 19 layers, characterized by its uniform architecture, where small 3×3 convolutional filters are stacked sequentially. The model employs max pooling layers for downsampling, enabling it to capture complex patterns in medical images. While VGG19 is computationally intensive due to its depth, its structured architecture makes it effective for fine-grained image classification tasks, such as distinguishing different types of skin lesions.

3.3.3. ViT-b16 (Vision Transformer—Base, 16×16 Patch Size)

ViT-b16 represents a shift from CNN-based architectures by treating images as sequences of non-overlapping 16×16 patches. Instead of relying on local receptive fields, ViT-b16 employs self-attention mechanisms to capture global dependencies across an image. The architecture consists of a patch embedding layer, multiple transformer encoder layers, and a classification head. By leveraging self-attention, ViT-b16 can effectively capture long-range dependencies, making it particularly suitable for analyzing complex and irregular skin lesion structures.

3.3.4. ViT-b32 (Vision Transformer—Base, 32×32 Patch Size)

ViT-b32 follows the same principles as ViT-b16 but processes larger 32×32-pixel patches, reducing the number of tokens and thereby lowering computational cost. While this approach may result in a loss of fine-grained details compared to ViT-b16, it improves efficiency and generalization, making it suitable for large-scale medical image analysis. Like ViT-b16, ViT-b32 benefits from contextual feature extraction through self-attention mechanisms, enhancing its applicability to skin lesion classification.

3.4. Implementation and Evaluation

Four pre-trained models, namely ResNet50, VGG19, ViT-b16, and ViT-b32, were utilized for classifying skin cancer. The utilized models were customized for specific tasks, specifically binary classification. Various hyperparameters were fine-tuned, and the model's performance was closely monitored to ensure effective discrimination between benign and malignant skin cancers. The training process used Adam optimizer, with a learning rate of 0.001 and a batch size of 32. The models were trained over 30 epochs, providing sufficient iterations for weight adjustment and improving classification accuracy.

Performance evaluation was performed using metrics such as accuracy, precision, recall, and F1 score, as following equations [36]:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{FN + TP} \tag{3}$$

$$F1\ \ score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

where TN true negatives, TP true positives, FP false positives, and FN false negatives.

3.5. XAI Techniques

To enhance the interpretability of model predictions and provide insights into classification decisions, XAI techniques were employed. Specifically, integrated gradients (IG) and LIME were utilized to identify the key features influencing the model's predictions. These techniques contribute to greater transparency and trust in automated skin lesion classification systems by revealing the underlying decision-making process.

Notably, the study utilized the ResNet50, VGG19, ViT-b16, and ViT-b32 models. These architectures, incorporating convolutional and transformer-based mechanisms, enable robust feature extraction and the classification of skin lesions.

3.5.1. Integrated Gradients (IG)

Integrated gradients is an attribution method designed to quantify the contribution of each input feature to the model's prediction. It calculates the average gradient of the model's output as the input transitions along a linear path from a baseline reference. The baseline, often set to zero or a neutral image, serves as a starting point for measuring feature importance. The attribution score for each feature is computed using the following formulation:

$$IG_i(x) = (x_i - \acute{x_i}) \times \int_{\alpha=0}^{1} \frac{\vartheta F(\acute{x} + \alpha(x - \acute{x}))}{\vartheta x_i} d\alpha \tag{5}$$

where

- $IG_i(x)$ represents the attribution score for feature;
- x is the input instance;
- $\acute{x}$ is the baseline input;
- F(x) is the model's output function;
- α is a scaling factor interpolating between the baseline and the input.

Integrated gradients satisfies two key properties:

- Sensitivity: If a feature significantly affects the model's prediction, it should receive a non-zero attribution score.
- Implementation Invariance: The attributions remain consistent across functionally equivalent neural networks.

3.5.2. Local Interpretable Model-Agnostic Explanations (LIME)

LIME provides local explanations by perturbing input features and analyzing their impact on the model's output. In image classification, LIME segments an input image into superpixels and selectively alters them to observe the effect on classification probability.

A surrogate linear model is then trained to approximate the model's decision boundary within a local neighborhood of the input. The importance of each superpixel is determined based on its contribution to the locally approximated model.

Mathematically, LIME constructs a local surrogate model g(x) that approximates the complex model F(x) within a neighborhood π_x around the input x. The optimal explanation is obtained by solving the following:

$$\hat{g} = \arg\, min_{g \in G}\, L(F,\, g,\, \pi_x) + \Omega(\mathrm{g}) \tag{6}$$

where

- G is the family of interpretable models;
- $L(F,\, g,\, \pi_x)$ is a loss function quantifying how well g(x) approximates F(x);
- Ω(g) imposes a complexity constraint to maintain interpretability.

By selectively modifying image regions and analyzing their influence on predictions, LIME highlights the superpixels that contribute most significantly to classification outcomes. This technique enhances understanding of model decisions and aids in distinguishing between malignant and benign lesions.

4. Results and Discussions

Table 1 presents the performance metrics of the proposed models, including ResNet50, VGG19, ViT-b16, and ViT-b32, in terms of accuracy, precision, recall, and F1 score. The results indicate that ResNet50 achieves the highest performance across all the metrics, with an accuracy of 0.8909, precision of 0.8909, recall of 0.8909, and an F1 score of 0.8908. This suggests that ResNet50 effectively captures discriminative features, leading to superior classification performance.

Table 1. Results obtained from the proposed models.

Metrics	ResNet50	VGG19	ViT-b16	ViT-b32
Accuracy	0.8909	0.8621	0.8697	0.8333
Precision	0.8909	0.8665	0.8711	0.8499
Recall	0.8909	0.8621	0.8697	0.8333
F1 score	0.8908	0.8624	0.8699	0.8332

VGG19 follows with slightly lower performance, achieving an accuracy of 0.8621, making it a competitive alternative among CNN-based models. In contrast, ViT-b16 outperforms ViT-b32 among transformer-based models, reaching an accuracy of 0.8697 compared to 0.8333. However, both ViT-b16 and ViT-b32 exhibit lower performance compared to ResNet50, which may be attributed to the higher data requirements of vision transformers for effective generalization. The weakest performance is observed with ViT-b32, which achieves the lowest scores across all the metrics, indicating its limitations for the given classification task.

Figure 4 presents the accuracy and loss curves for the proposed models (ResNet50, VGG19, ViT-b16, and ViT-b32) over 30 training epochs. The accuracy graphs (left column) indicate that ResNet50 exhibits stable learning, with training and validation accuracy closely following each other, suggesting effective generalization. VGG19 initially experiences fluctuations, particularly around epoch 3, but later stabilizes and achieves competitive performance. ViT-b16 demonstrates a smooth and consistent increase in accuracy, with minimal deviation between training and validation curves, reflecting strong model stabil-

ity. In contrast, ViT-b32 shows noticeable fluctuations in validation accuracy, suggesting potential overfitting or difficulty in capturing complex patterns.

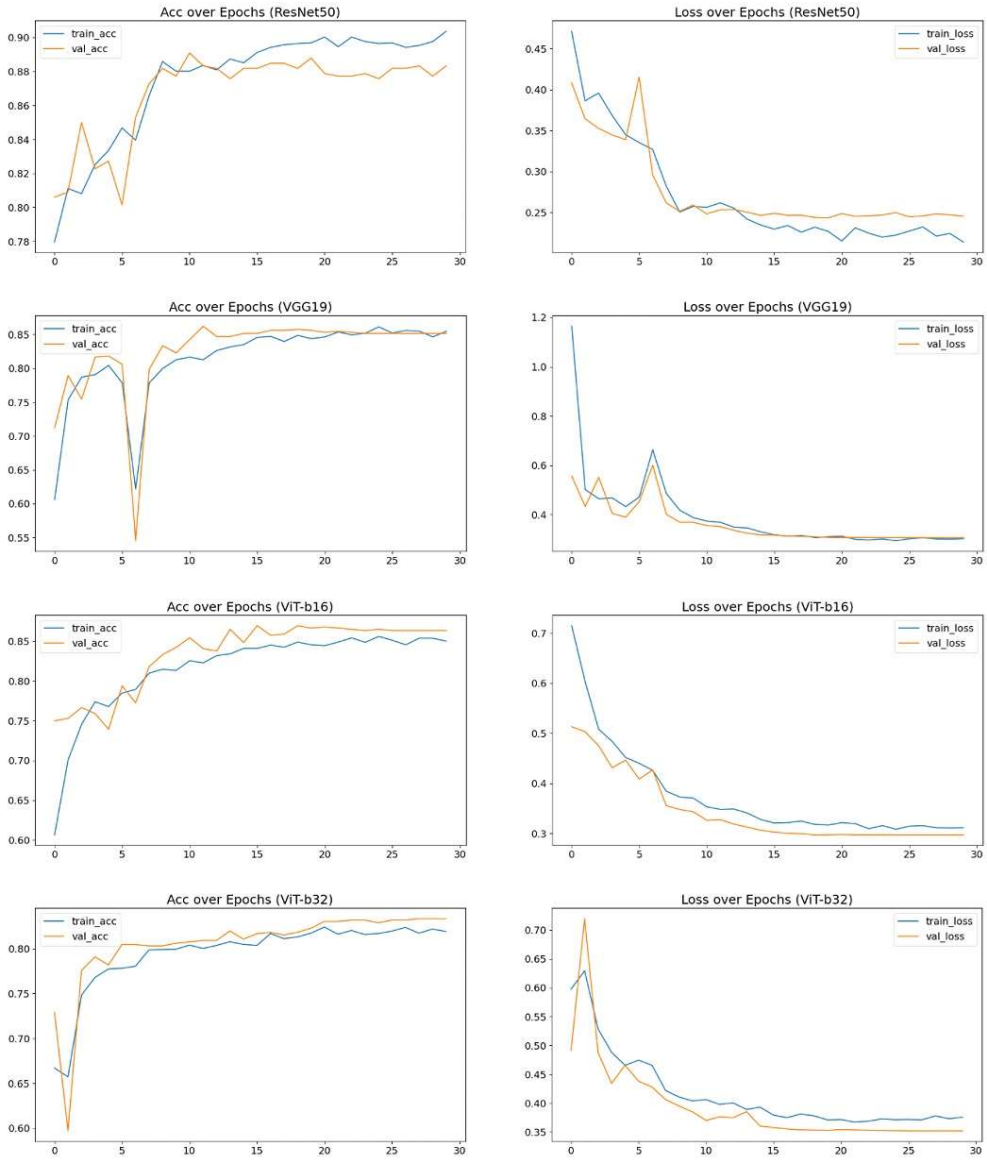

Figure 4. Accuracy and loss graphs of the proposed models.

The loss graphs (right column) further highlight the learning behavior of the models. ResNet50 and ViT-b16 exhibit a steady decline in both training and validation loss, indicating stable optimization and effective learning. VGG19, despite its early instability, eventually aligns training and validation loss, suggesting improved convergence. However, ViT-b32 experiences significant fluctuations in validation loss, implying inconsistencies in learning and possible overfitting. Overall, ResNet50 emerges as the most stable and effective model, followed closely by ViT-b16. VGG19 recovers from its initial instability, while ViT-b32 demonstrates the highest variability, indicating challenges in feature extraction. These findings align with the performance metrics in Table 1, further supporting ResNet50's superiority among the evaluated models.

Figure 5 illustrates the confusion matrices for the proposed models (ResNet50, VGG19, ViT-b16, and ViT-b32) in classifying benign and malignant cases. These matrices provide a detailed breakdown of each model's classification performance, highlighting the number of true positives, true negatives, false positives, and false negatives. ResNet50 demonstrates the most balanced performance, correctly identifying 328 benign and 260 malignant

cases, with 32 benign cases misclassified as malignant and 40 malignant cases misclassified as benign. This indicates that the model effectively differentiates between both classes with minimal misclassification. VGG19 also performs well, achieving high accuracy in detecting malignant cases (270 correctly classified). However, it exhibits a higher false positive rate, misclassifying 61 benign cases as malignant, while only 30 malignant cases are misclassified as benign. ViT-b16 achieves a competitive performance, correctly classifying 309 benign and 265 malignant cases. However, it misclassifies 51 benign cases as malignant and 35 malignant cases as benign, indicating a slightly reduced specificity compared to ResNet50. ViT-b32, while achieving a strong recall for malignant cases (278 correctly classified), exhibits the highest misclassification rate for benign samples. The model incorrectly classifies 88 benign cases as malignant, suggesting a tendency to overestimate malignancy.

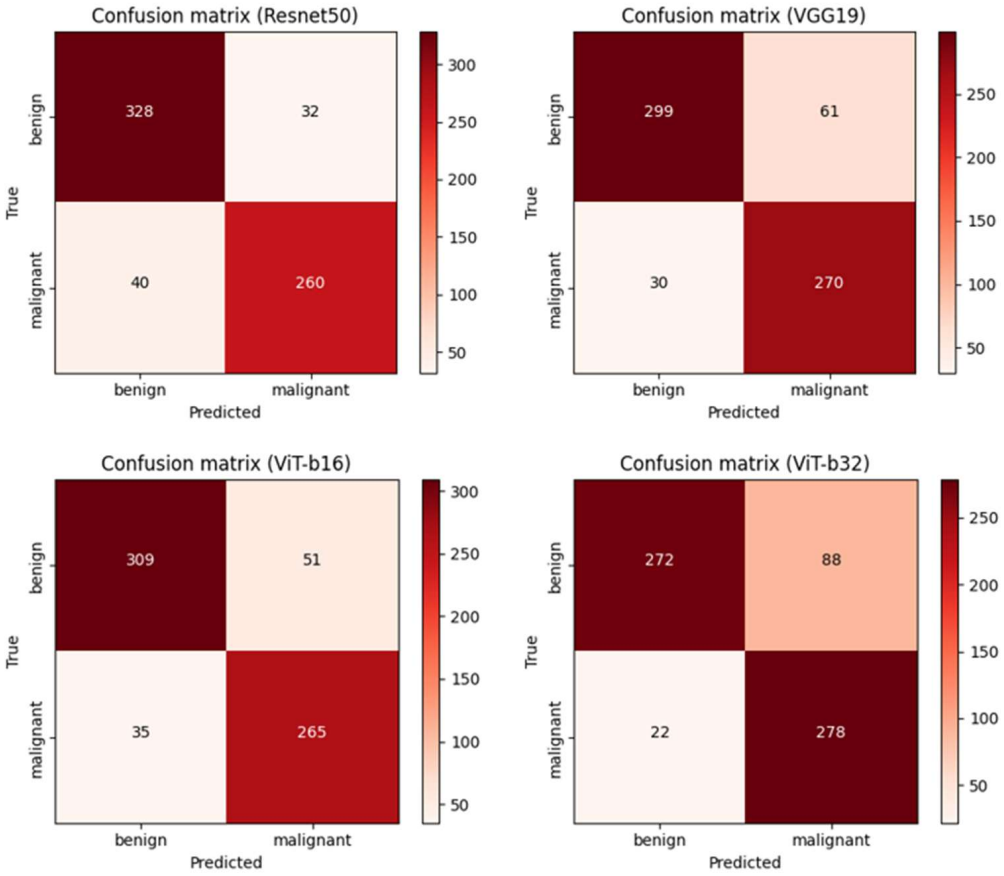

Figure 5. Confusion matrices of the proposed models.

Overall, ResNet50 emerges as the most reliable model, offering a well-balanced trade-off between sensitivity and specificity. VGG19 and ViT-b16 provide strong performance, with VGG19 favoring malignant case detection but at the cost of increased false positives. ViT-b32, while effective in detecting malignant cases, demonstrates a higher misclassification rate for benign samples. These findings are consistent with the performance metrics presented in Table 1, further validating the robustness of ResNet50 for this classification task.

Figure 6 presents the Receiver Operating Characteristic (ROC) curves for the proposed models, including ResNet50, VGG19, ViT-b16, and ViT-b32. The ROC curve illustrates the relationship between the true positive rate (sensitivity) and the false positive rate, providing insight into each model's ability to distinguish between benign and malignant cases. The Area Under the Curve (AUC) serves as a key performance metric, where a higher AUC indicates better classification performance.

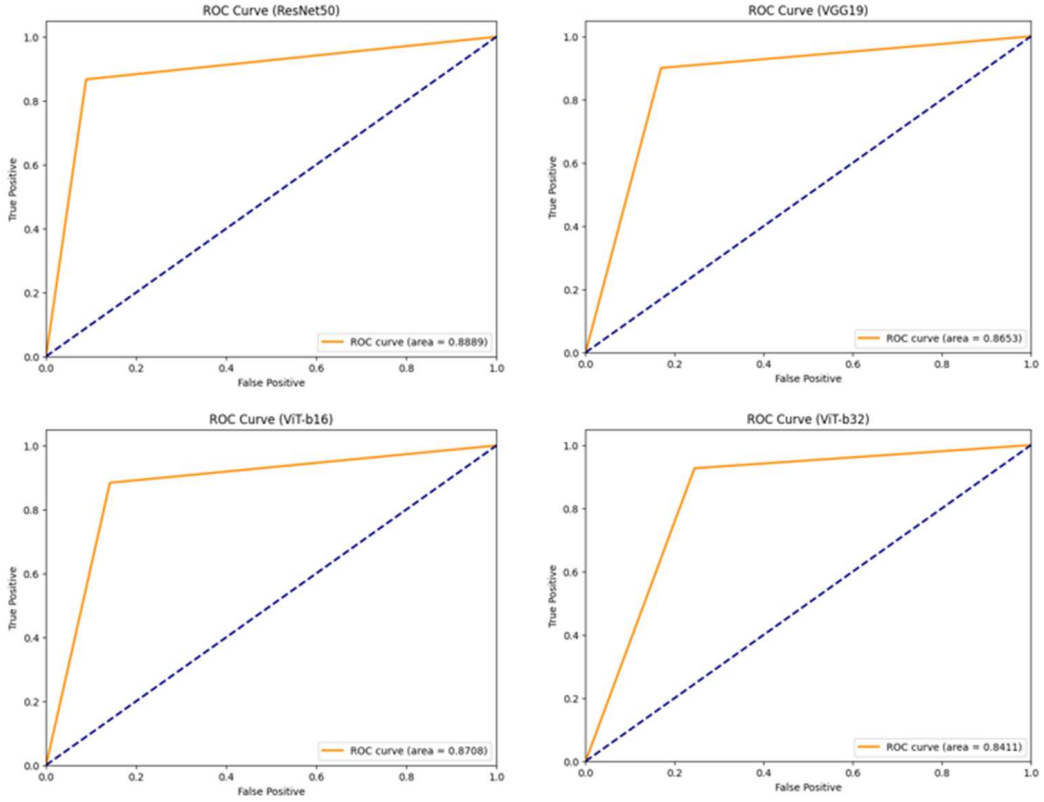

Figure 6. ROC curves of the proposed models.

Among the evaluated models, ResNet50 achieves the highest AUC of 0.8889, indicating superior discrimination between the two classes. Its ROC curve remains close to the upper-left corner, signifying a strong balance between sensitivity and specificity. VGG19 follows with an AUC of 0.8653, demonstrating competitive performance, albeit with a slightly higher false positive rate compared to ResNet50. ViT-b16 attains an AUC of 0.8708, performing comparably to VGG19, suggesting a robust classification capability. However, its ROC curve indicates minor trade-offs in specificity. ViT-b32, with an AUC of 0.8411, exhibits the lowest classification performance among the evaluated models. Although it remains above the diagonal baseline, its ability to distinguish between benign and malignant cases is relatively weaker than the other models. Overall, ResNet50 emerges as the most effective model, followed by ViT-b16 and VGG19, while ViT-b32 demonstrates the least favorable performance. These findings align with the confusion matrix results in Figure 5, further validating ResNet50's effectiveness in achieving a strong balance between sensitivity and specificity.

Figure 7 illustrates the performance of various models in classifying cases as either benign or malignant. All the models, including ResN50, VGG19, ViT-b16, and ViT-b32, accurately predicted the benign cases. The ResN50 and VGG19 models exhibited high and consistent confidence levels, with probabilities of being benign ranging from 0.94 to 1.00 for ResN50 and 0.98 to 0.99 for VGG19. In contrast, the ViT models displayed greater variability in confidence. The ViT-b16 model's probabilities ranged from 0.70 to 1.00, while the ViT-b32 model showed even more significant variation, with probabilities between 0.43 and 0.99. These results indicate that while all the models achieved accurate predictions, the ResN50 and VGG19 models provided more reliable confidence estimates. The ViT models, particularly ViT-b32, may benefit from further refinement to improve the consistency of their confidence levels.

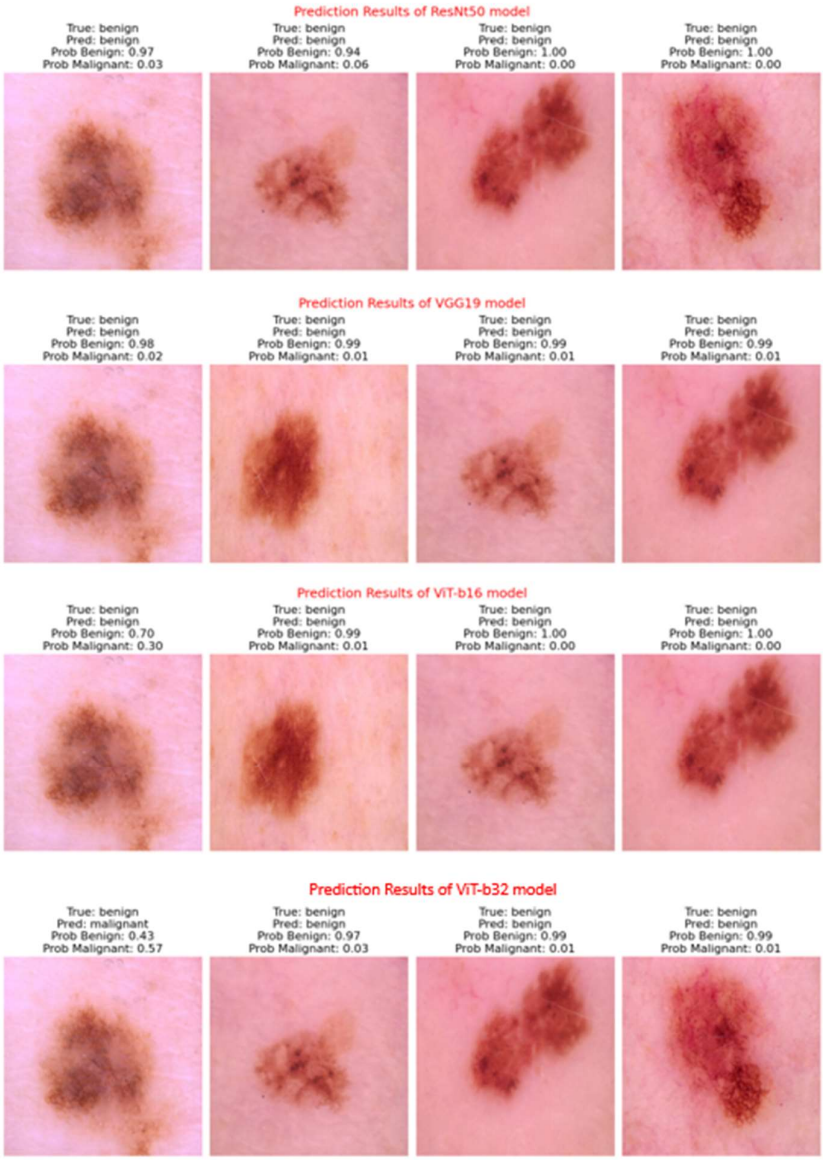

Figure 7. Prediction results of the proposed models.

The results of applying the XAI techniques are shown in Figures 8–11, which highlight the use of the LIME and IG techniques on the best-performing model (ResNet50) among all the models utilized.

Figures 8 and 9 demonstrate the application of the integrated gradients technique to provide interpretability in the model's predictions for benign and malignant skin lesions. In Figure 8, which presents benign samples, the original images show lesions with relatively uniform color distribution and well-defined borders. The integrated gradients heatmaps reveal sparse and less intense red regions, indicating that the model primarily focuses on specific, localized features without complex patterns. The overlay images further confirm that these highlighted areas are concentrated and less irregular, consistent with the characteristic appearance of benign lesions.

In contrast, Figure 9 presents malignant samples, which exhibit irregular borders, heterogeneous pigmentation, and more complex structures. The corresponding integrated gradients heatmaps display more extensive and intense red regions, suggesting that the model identifies intricate patterns and variations in pigmentation commonly associated with malignancy. The overlay images show broader and more dispersed areas of interest, particularly around irregular and asymmetric patterns.

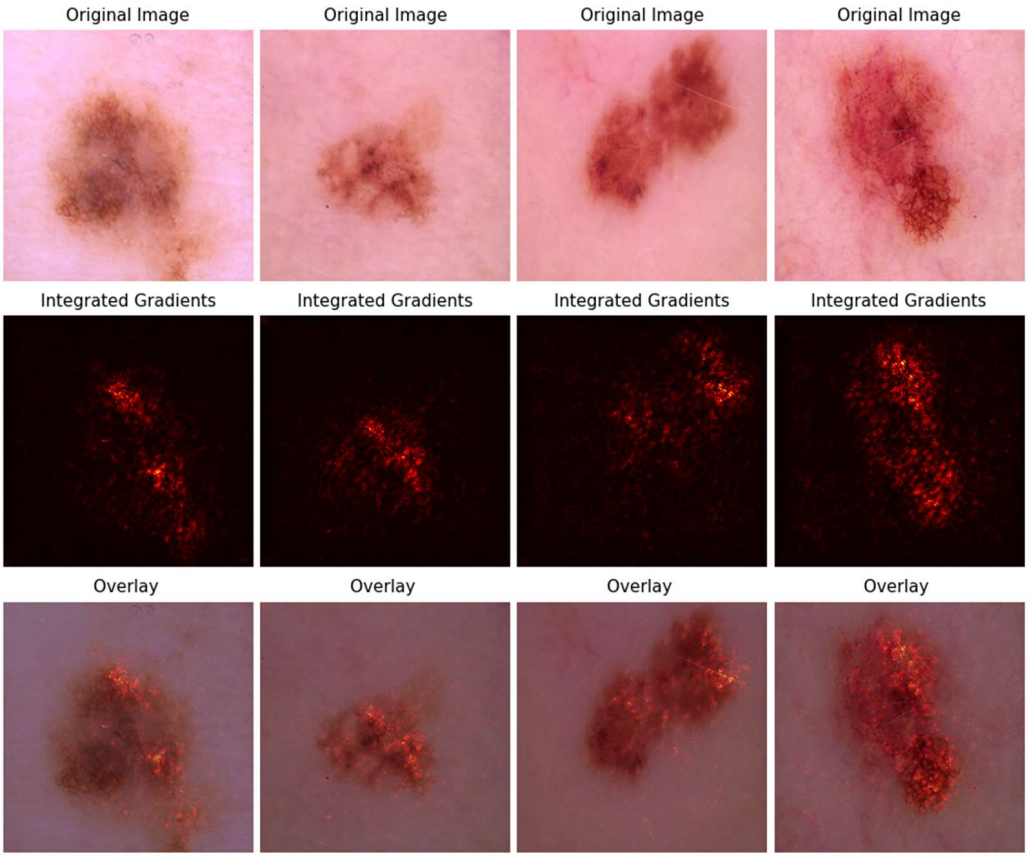

Figure 8. IG technique applied to ResNet50 model for benign sample.

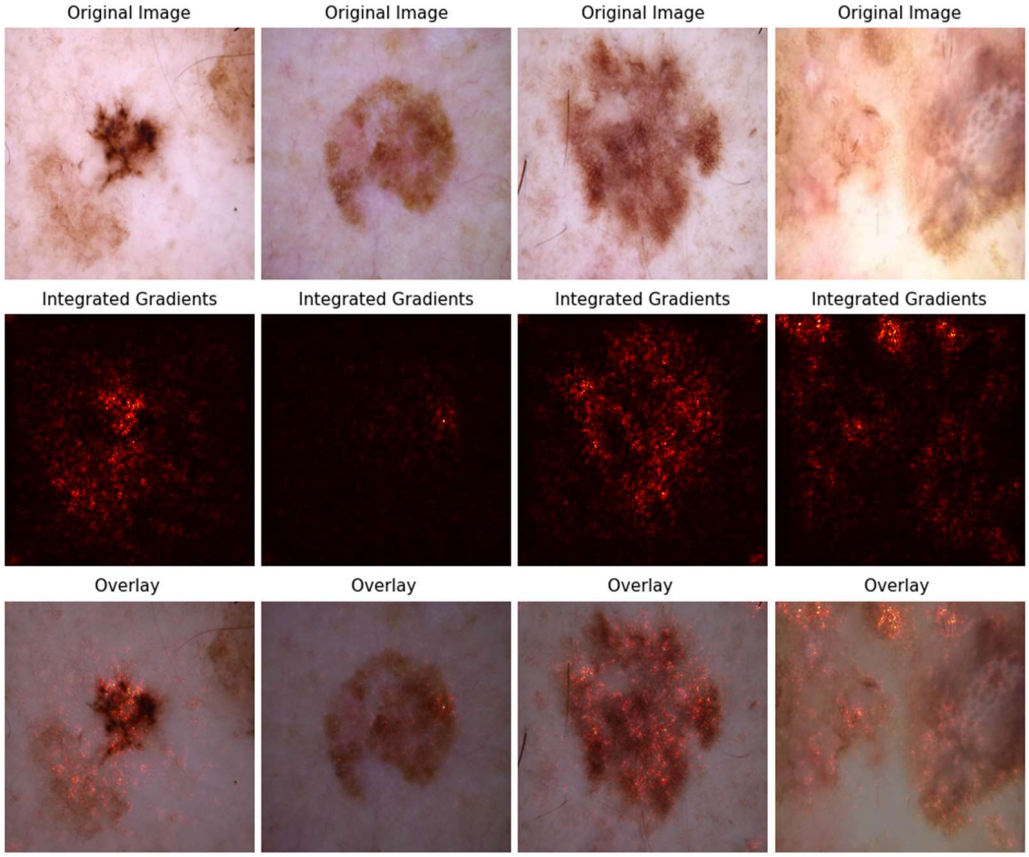

Figure 9. IG technique applied to ResNet50 model for malignant sample.

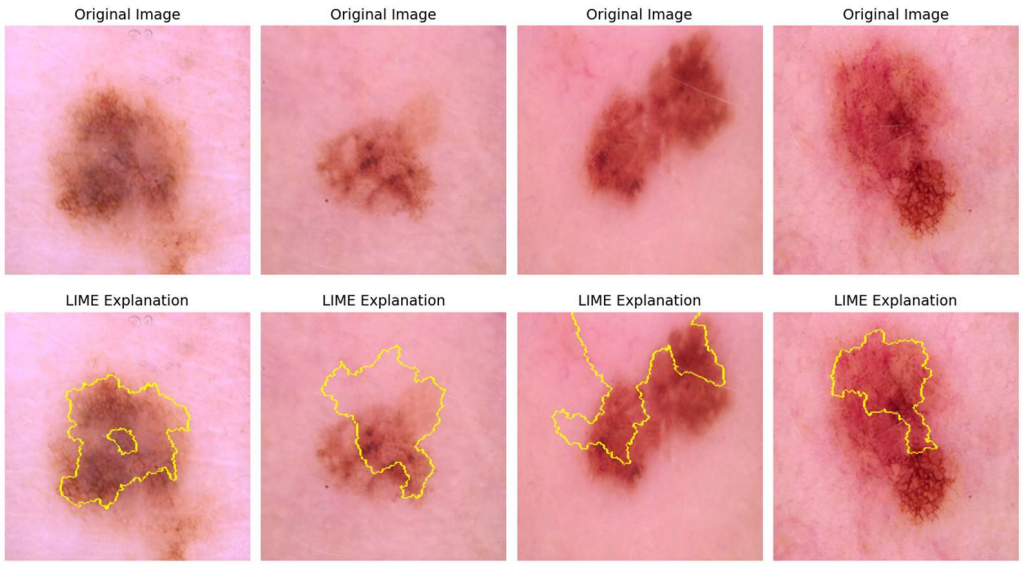

Figure 10. LIME technique applied to ResNet50 model for benign sample.

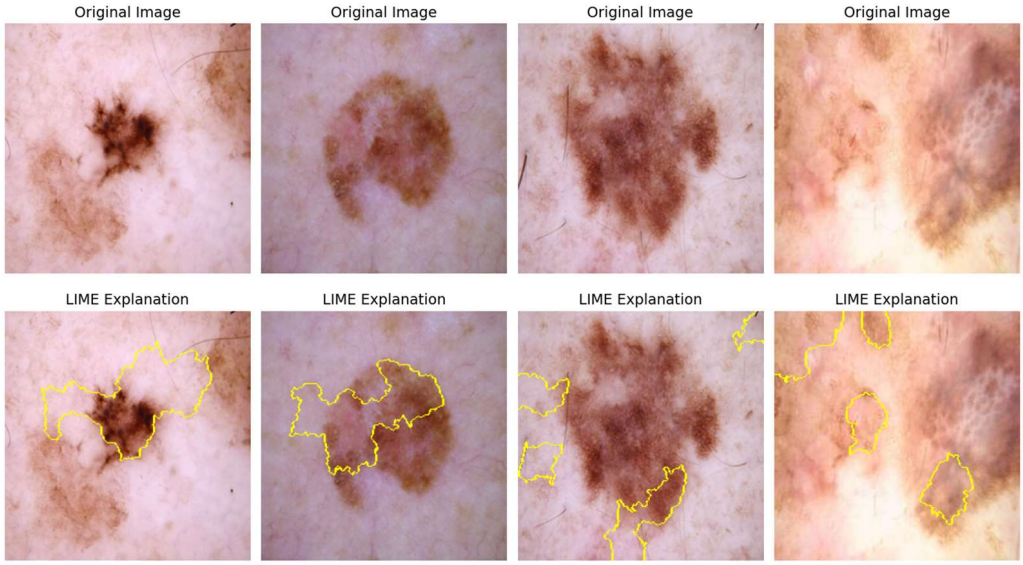

Figure 11. LIME technique applied to ResNet50 model for malignant sample.

A comparison of Figures 8 and 9 indicates that the model activation is more pronounced and widespread in malignant lesions compared to benign lesions, reflecting the distinct morphological differences between these two categories. This analysis highlights the effectiveness of the integrated gradients technique in enhancing the transparency and interpretability of the model's decision-making process for skin lesion classification.

Figures 10 and 11 demonstrate the application of the LIME (Local Interpretable Model-agnostic Explanations) technique to provide interpretability for the model's predictions of benign and malignant skin lesions. In Figure 10, which presents benign samples, the original images show lesions with relatively uniform pigmentation and well-defined borders. The LIME, highlighted with yellow outlines, indicates the regions that contributed most significantly to the model's prediction. These highlighted areas are primarily localized and concentrated around distinct lesion boundaries, suggesting that the model focuses on consistent and uniform features typically associated with benign lesions.

In contrast, Figure 11 illustrates malignant samples, which are characterized by irregular borders, heterogeneous pigmentation, and structural asymmetry. The LIME for these samples shows more dispersed and irregularly shaped highlighted regions. The yellow

outlines often encompass multiple areas within the lesion, indicating that the model identifies complex patterns such as uneven pigmentation, irregular borders, and asymmetric structures as significant for malignancy.

A comparison between Figures 10 and 11 reveals a distinct difference in the model's behavior when predicting benign versus malignant lesions. The benign samples exhibit more compact and localized areas of interest, whereas the malignant samples display more extensive and fragmented highlighted regions. This pattern aligns with the clinical characteristics of these lesion types, demonstrating the effectiveness of the LIME technique in providing insights into the model's decision-making process for skin lesion classification.

When comparing the integrated gradients and LIME techniques, notable differences emerge in the model's interpretability patterns. Integrated gradients highlight pixel-level attributions with a focus on continuous color and texture variations within the lesion, which is useful for identifying subtle differences in structure. In contrast, LIME provides a more segmented explanation by identifying larger, discrete areas that influence the model's prediction. Integrated gradients tend to produce more granular and detailed heatmaps, whereas LIME focuses on broader, contiguous regions. This distinction makes integrated gradients more suitable for analyzing fine-grained patterns, while LIME offers more intuitive visual explanations by outlining specific lesion areas. Together, these techniques provide complementary insights, enhancing the overall interpretability of the model's predictions.

Table 2 presents a comparative analysis of previous studies that utilized the same dataset for skin lesion classification, highlighting the models and their corresponding accuracy rates. The results demonstrate a noticeable variation in performance across different architectures and timeframes.

Table 2. Comparative analysis of related previous studies utilized the same dataset.

Study and (Year)	Model and (Accuracy)
Demir et al. [37] (2019)	ResNet-101 (84.09%) and Inceptionv3 (87.42%)
Lembhe et al. [38] (2023)	VGG16 (70.17%), ResNet50 (86.57%), and InceptionV3 (91.26%)
Hiswati [39] (2021)	CNN (54%)
Farooq et al. [40] (2019)	Inception-v3 (86%)
Aydin [41] (2023)	CNN (80%) and Xception (80%)
Hussein et al. [42] (2023)	AlexNet (88.33%), ResNet18 (89.39%), SqueezeNet (85.91%), and ShuffleNet (87.42%)
Dagnaw et al. [28] (2024)	Vgg19 (85.6%), ResNet18 (82.4%), ResNet50 (8.8%), DensNet201 (83.2%), MobileNetv2 (85.3%), ViT_b (88.6%), ViT_L (88.6%), Swin_s (87.7%), Swin_b (87.8%), and Swin_T (87.7%)
This study	ResNet50 (89.09%), VGG19 (86.21%), ViT_b16 (86.97%), and ViT_b32 (83.33%)

Early studies, such as Demir et al. [37] (2019), achieved relatively high performance with InceptionV3 (87.42%), surpassing ResNet-101 (84.09%). Farooq et al. [40] (2019) also employed InceptionV3, reporting an accuracy of 86%, which aligns closely with Demir et al.'s findings [37]. However, Hiswati [39] (2021) obtained a significantly lower accuracy of 54% using a CNN model, possibly indicating limitations in model architecture or data preprocessing techniques.

More recent studies exhibit improved performance, suggesting advancements in model architectures and training strategies. Lembhe et al. [38] (2023) achieved 91.26% with InceptionV3, outperforming older models and demonstrating the ongoing relevance of this architecture. Hussein et al. [42] (2023) explored multiple architectures, with ResNet18 achieving the highest accuracy of 89.39%, followed by AlexNet (88.33%), SqueezeNet (85.91%), and ShuffleNet (87.42%), indicating the effectiveness of lightweight models. Dagnaw et al. [28] (2024) presented a comprehensive evaluation across various models, with ResNet50 (88.8%) and vision transformers (ViT_b and ViT_L) both achieving 88.6%, showcasing the potential of transformer-based architectures in medical imaging tasks.

This study is consistent with [28] in comparing the CNN and ViT techniques. However, we have noted improvements in the performance of certain models in our work. These enhancements are due to differences in the initial data processing and the approach to training the models. This suggests that performance can be further enhanced by incorporating different methods and modifications. Moreover, we have applied alternative interpretation techniques, providing a wider view and a more thorough comparison for understanding the same models.

In this study, the best performance was achieved with ResNet50 (89.09%), exceeding the accuracies reported by previous works, including Lembhe et al. [38] (2023) and Hussein et al. [42] (2023). The ViT_b16 model achieved 86.97%, demonstrating competitive performance compared to traditional CNNs.

Although pre-trained CNNs and ViTs are highly effective at classifying skin cancer, some limitations remain. These models may still have problems classifying images with subtle or complex variations, such as small lesions or rare skin cancers, which may lead to misclassification. In addition, while techniques such as LIME and integrated gradients can help explain model predictions, they may not always provide fully accurate or human-understandable explanations, especially for highly complex models. The computational cost of using pre-trained models and generating explanations with these methods can also be high, limiting their real-time applicability in clinical settings. In addition, the reliance on large, high-quality datasets to fine-tune these models may reduce their adaptability in resource-limited settings.

From a broader perspective, these findings suggest that CNN-based models like ResNet50 continue to perform effectively in this domain. However, the competitive performance of vision transformers indicates a promising direction for future research. Future work could explore hybrid architectures that combine CNN- and transformer-based models to leverage their respective strengths. Additionally, investigating the impact of advanced augmentation techniques, transfer learning strategies, and model interpretability methods may further enhance model performance and reliability in clinical applications.

Furthermore, the analysis underscores the importance of selecting the appropriate model architecture based on the task requirements. While CNN models like ResNet50 and InceptionV3 have consistently delivered strong performance, vision transformers have shown competitive results and may offer advantages in capturing complex spatial relationships. Future studies could focus on optimizing hyperparameters, incorporating ensemble learning, and leveraging larger, more diverse datasets to improve model generalization and robustness in real-world dermatological applications.

5. Conclusions

Skin cancer is a significant public health issue, making early detection vital for improving patient outcomes. This study assessed four pre-trained models for classifying skin lesions as benign or malignant, with the ResNet50 model outperforming the utilized models across all the evaluation metrics. All the utilized models showcased their classi-

fication capabilities, reinforcing their role in assisting healthcare professionals with early detection. Moreover, the use of the integrated gradients and LIME techniques as XAI tools provided insights into the models' decision-making processes, which is essential for clinical acceptance and fosters trust between technology and healthcare providers. The study emphasized the need for continued research into automating skin cancer diagnosis using advanced machine learning techniques to enhance diagnostic accuracy and reduce the workload on medical staff. Ultimately, developing reliable and interpretable AI models is crucial for bridging the gap between technology and clinical practice, leading to improved health outcomes for patients with skin cancer.

Future work should focus on exploring a wider array of models and datasets to enhance the classification of various skin lesions. By incorporating different architectures, we can assess their performance and robustness across diverse lesion types. In addition, utilizing various datasets that encompass a broader range of skin lesions will help improve generalizability and accuracy. It is also essential to investigate other XAI techniques to provide deeper insights into model predictions and decision-making processes. This multifaceted approach will not only advance our understanding of lesion classification but also enhance the interpretability of AI tools in clinical settings, ultimately leading to better diagnostic support for healthcare professionals.

Author Contributions: Conceptualization, A.A., A.E. and F.K.; methodology design, A.A., A.E. and F.K.; writing—original draft preparation, A.A.; writing—review and editing, A.E. and F.K.; supervision and project administration, F.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are publicly accessible and available on Kaggle at https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign (accessed on 7 March 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

CNN	convolutional neural network
IG	integrated gradients
LIME	Local Interpretable Model-Agnostic Explanations
ViT	vision transformer
XAI	eXplainable Artificial Intelligence

References

1. Kwon, J.; Kethar, J.; Appavu, R. Skin Cancer: The Ozone Layer and UV Radiation. *J. Stud. Res.* **2022**, *11*. [CrossRef]
2. Hasan, N.; Nadaf, A.; Imran, M.; Jiba, U.; Sheikh, A.; Almalki, W.H.; Ahmad, F.J. Skin cancer: Understanding the journey of transformation from conventional to advanced treatment approaches. *Mol. Cancer* **2023**, *22*, 168. [CrossRef] [PubMed]
3. Hohn, J.; Hekler, A.; Krieghoff-Henning, E.; Kather, J.N.; Utikal, J.S.; Meier, F.; Brinker, T.J. Integrating patient data into skin cancer classification using convolutional neural networks: Systematic review. *J. Med. Internet Res.* **2021**, *23*, e20708. [CrossRef] [PubMed]
4. Magdy, A.; Hussein, H.; Abdel-Kader, R.F.; Abd El Salam, K. Performance Enhancement of Skin Cancer Classification Using Computer Vision. *IEEE Access* **2023**, *11*, 72120–72133. [CrossRef]
5. Shah, A.; Shah, M.; Pandya, A.; Sushra, R.; Sushra, R.; Mehta, M.; Patel, K. A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN). *Clin. Ehealth* **2023**, *6*, 76–84. [CrossRef]
6. Pacheco, A.G.; Krohling, R.A. Recent advances in deep learning applied to skin cancer detection. *arXiv* **2019**, arXiv:1912.032802019.
7. Shamshad, F.; Khan, S.; Zamir, S.W.; Khan, M.H.; Hayat, M.; Khan, F.S.; Fu, H. Transformers in medical imaging: A survey. *Med. Image Anal.* **2023**, *88*, 102802. [CrossRef] [PubMed]
8. Liu, Y.; Zhang, Y.; Wang, Y.; Hou, F.; Yuan, J.; Tian, J.; He, Z. A survey of visual transformers. *IEEE Trans. Neura Netw. Learn. Syst.* **2023**, *35*, 7478–7498. [CrossRef]

9. de Lima, L.M.; Krohling, R.A. Exploring advances in transformers and CNN for skin lesion diagnosis on small datasets. In *Brazilian Conference on Intelligent Systems*; Springer International Publishing: Cham, Switzerland, 2022; pp. 282–296.
10. Ali, K.; Shaikh, Z.A.; Khan, A.A.; Laghari, A.A. Multiclass skin cancer classification using EfficientNets—A first step towards preventing skin cancer. *Neurosci. Inform.* **2022**, *2*, 100034. [CrossRef]
11. Saranya, A.; Subhashini, R. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decis. Anal. J.* **2023**, *7*, 100230.
12. Arrieta, A.B.; Diaz-Rodriguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Herrera, F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
13. Dildar, M.; Akram, S.; Irfan, M.; Khan, H.U.; Ramzan, M.; Mahmood, A.R.; Mahnashi, M.H. Skin cancer detection: A review using deep learning techniques. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5479. [CrossRef]
14. Gouda, W.; Sama, N.U.; Al-Waakid, G.; Humayun, M.; Jhanjhi, N.Z. Detection of skin cancer based on skin lesion images using deep learning. *Healthcare* **2022**, *10*, 1183. [CrossRef] [PubMed]
15. Kousis, I.; Perikos, I.; Hatzilygeroudis, I.; Virvou, M. Deep learning methods for accurate skin cancer recognition and mobile 348 application. *Electronics* **2022**, *11*, 1294. [CrossRef]
16. Salma, W.; Eltrass, A.S. Automated deep learning approach for classification of malignant melanoma and benign skin lesions. *Multimed. Tools Appl.* **2022**, *81*, 32643–32660. [CrossRef]
17. Tran, K.A.; Kondrashova, O.; Bradley, A.; Williams, E.D.; Pearson, J.V.; Waddell, N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* **2021**, *13*, 152. [CrossRef]
18. Van der Velden, B.H.; Kuijf, H.J.; Gilhuijs, K.G.; Viergever, M.A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* **2022**, *79*, 102470. [CrossRef]
19. Graziani, M.; Andrearczyk, V.; Marchand-Maillet, S.; Muller, H. Concept attribution: Explaining CNN decisions to physicians. *Comput. Biol. Med.* **2020**, *123*, 103865.
20. Hauser, K.; Kurz, A.; Haggenmuller, S.; Maron, R.C.; von Kalle, C.; Utikal, J.S.; Brinker, T.J. Explainable artificial intelligence in skin cancer recognition: A systematic review. *Eur. J. Cancer* **2022**, *167*, 54–69. [CrossRef]
21. Nigar, N.; Umar, M.; Shahzad, M.K.; Islam, S.; Abalo, D. A deep learning approach based on explainable artificial intelligence for skin lesion classification. *IEEE Access* **2022**, *10*, 113715–113725. [CrossRef]
22. Saarela, M.; Georgieva, L. Robustness, stability, and fidelity of explanations for a deep skin cancer classification model. *Appl. Sci.* **2022**, *12*, 9545. [CrossRef]
23. Hurtado, S.; Nematzadeh, H.; Garcia-Nieto, J.; Berciano-Guerrero, M.A.; Navas-Delgado, I. On the use of explainable artificial intelligence for the differential diagnosis of pigmented skin lesions. In *International Work-Conference on Bioinformatics and Biomedical Engineering*; Springer International Publishing: Cham, Switzerland, 2022; pp. 319–329.
24. Bokadia, H.; Yang, S.C.H.; Li, Z.; Folke, T.; Shafto, P. Evaluating perceptual and semantic interpretability of saliency methods: A case study of melanoma. *Appl. AI Lett.* **2022**, *3*, e77. [CrossRef]
25. Shivhare, I.; Jogani, V.; Purohit, J.; Shrawne, S.C. Analysis of explainable artificial intelligence methods on medical image classification. In Proceedings of the 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 5–6 January 2023; pp. 1–5.
26. He, K.; Gan, C.; Li, Z.; Rekik, I.; Yin, Z.; Ji, W.; Gao, Y.; Wang, Q.; Zhang, J.; Shen, D. Transformers in medical image analysis. *Intell. Med.* **2023**, *3*, 59–78. [CrossRef]
27. Nikitin, V.; Shapoval, N. Vision transformer for skin cancer classification. *Sci. Collect. ≪InterConf+≫* **2023**, *33*, 449–460. [CrossRef]
28. Dagnaw, G.H.; El Mouhtadi, M.; Mustapha, M. Skin cancer classification using vision transformers and explainable artificial intelligence. *J. Med. Artif. Intell.* **2024**, *7*, 14. [CrossRef]
29. Aslan, M.F. Comparison of vision transformers and convolutional neural networks for skindisease classification. *Proc. Int. Conf. New Trends Appl. Sci.* **2023**, *1*, 31–39.
30. Zhao, Z. Skin cancer classification based on convolutional neural networks and vision transformers. *J. Phys. Conf. Ser.* **2022**, *2405*, 012037. [CrossRef]
31. Fanconi, C. Skin Cancer: Malignant vs. Benign. Available online: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign (accessed on 7 March 2024).
32. International Skin Imaging Collaboration ISIC Archive. Available online: https://www.isic-archive.com/ (accessed on 7 March 2024).
33. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [CrossRef]
34. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [CrossRef]
35. Moreno-Barea, F.J.; Jerez, J.M.; Franco, L. Improving classification accuracy using data augmentation on small data sets. *Expert Syst. Appl.* **2020**, *161*, 113696. [CrossRef]

36. Alzahrani, S.; Al-Bander, B.; Al-Nuaimy, W. A comprehensive evaluation and benchmarking of convolutional neural networks for melanoma diagnosis. *Cancers* **2021**, *13*, 4494. [CrossRef] [PubMed]
37. Demir, A.; Yilmaz, F.; Kose, O. Early detection of skin cancer using deep learning architectures: ResNet-101 and Inception-V3. In Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 3–5 October 2019; pp. 1–4.
38. Lembhe, A.; Motarwar, P.; Patil, R.; Elias, S. Enhancement in skin cancer detection using image super resolution and convolutional neural network. *Procedia Comput. Sci.* **2023**, *218*, 164–173. [CrossRef]
39. Hiswati, M.E. DeepSkin: Robust skin cancer classification using convolutional neural network algorithm. *Int. J. Inform. Comput. (IJICOM)* **2021**, *3*. [CrossRef]
40. Farooq, M.A.; Khatoon, A.; Varkarakis, V.; Corcoran, P. Advanced deep learning methodologies for skin cancer classification in prodromal stages. *arXiv* **2020**, arXiv:2003.06356.
41. Aydin, Y. A comparative analysis of skin cancer detection applications using histogram-based local descriptors. *Diagnostics* **2023**, *13*, 3142. [CrossRef]
42. Hussein, H.; Magdy, A.; Abdel-Kader, R.F.; Ali, K.A.E. Binary classification of skin cancer using pretrained deep neuralnetworks. *Suez Canal Eng. Energy Environ. Sci.* **2023**, *1*, 10–14. [CrossRef]