Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Melanoma-Sensitive Deep Learning Model for Multi-Class Skin Cancer Classification Using Dermoscopic Images

Dr.Reshma Pise

reshma.pise@vupune.ac.in

Vishwakarma university

Prof.Sumaira Shaikh

Vishwakarma university

Research Article

Keywords: Skin Cancer Classification, Deep Learning, Convolutional Neural Networks, Transfer Learning, Data Augmentation, Dermoscopic Images, Melanoma Detection

Posted Date: April 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9449618/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations:
Competing interests: None declared.

Abstract

Skin cancer is a serious health problem that affects the world and melanoma is the most dangerous type of the disease because it is fast growing and causes high mortality in case of late diagnosis. Deep learning (DL) methods have become the focus of attention in recent years because they can be used to conduct automated dermoscopic image analysis to provide support in making early and objective clinical decisions. The article introduces the creation of a DL-based skin can-cer classification system based on Convolutional Neural Networks (CNNs) to distinguish between various types of skin lesions. The common image process-ing pipeline was adopted then the feature extraction was done through transfer learning (TL) on the pretrained models. In order to counter the issues brought about by the class imbalance and limited labelled information, data augmentation (DA) methods were included during training. A number of backbones architec-tures were experimentally evaluated to support the classification performance, and especially optimize melanoma recognition. The results of the experimental assessment suggest the high accuracy of the overall outcome and the high sensitivity of melanoma reflected by the effectiveness of the adopted training strategy. The findings indicate that the use of DL-based methods is promising as a reliable method to diagnose skin cancer, but also contribute to the fact that larger datasets and better model generalization need to be considered in further research.

1 Introduction

One of the most common malignancies is skin cancer that continues to be a major concern to the general population in many countries. It has been on a gradual but steady increase in the past decades due to the exposure to ultraviolet radiations, environmental factors, and altering trends of lifestyle. The disease presents a wide range of the types of lesions, including benign pigmented nevi, as well as very malignant lesions like melanoma. Despite a comparatively low percentage of skin cancer incidences, melanoma causes most of the mortality related to skin cancer as it is invasive and highly likely to spread through metastasis. Early and accurate diagnosis is therefore very critical in enhancing patient survivability and minimizing healthcare cost.

The common non-invasive imaging method applied in the clinical examination of skin lesions is dermoscopic. It improves the skin structures of the subsurface, thus allow-ing dermatologists to differentiate visual patterns that relate to malignancy. Although dermoscopic diagnosis has clinical utility, it is still highly dependent on clinician experience, and thus it is highly subjective. There is significant interpractitioner variability in diagnostic performance, and experienced dermatologists are not available in most areas, especially in resource marginalized settings.

The recent advancements in the field of deep learning (DL) have impacted the medical image analysis domain profoundly and especially in the image classification and the pattern recognition tasks. Convolutional Neural Networks (CNNs) have shown the capability to discover discriminative features directly on raw image data, and eliminate manually engineered features. CNN-based models have shown good results in the field of skin cancer testing, with the best performance on large scale benchmark datasets, suggesting that deep learning as a diagnostic tool can be extensively studied. However, such obstacles as a high degree of class imbalance, extensive visual similarity of lesion types, and the lack of high-quality annotated medical data persist.

A key challenge that is especially critical to overcome is the reliability of performance of classes of clinical interest but underrepresented like melanoma. Most of the available literature is mainly aimed at maximizing the overall classification accuracy and not sensitivity of rare but high-risk lesions. Models that are trained using lop-sided datasets are likely to be biased in majority classes, which increases the chances of false-negative predictions of melanoma. Such errors are particularly alarming in a clinical environment, where a misdiagnosis may result in a treatment delay and poor outcomes. To solve this problem, model design choices, loss functions and training strategies must be explicitly chosen and consider the class distribution.

This paper describes a deep learning-based framework of skin cancer classification that focuses on the strong model construction as opposed to deployment factors. The approach proposed is a transfer learning of contemporary architecture to extract high-level representations of dermoscopic images. It has a multi-phase training plan, in which classifier-head optimization is done initially and then fine-tuning of the backbone network gradually. To further increase the sensitivity of melanoma, class-weighted optimization is added to the overall performance. It is based on this systematic methodology that the study will establish a successful and repeatable procedure in the development of high-performance skin lesion classification models that can be used in future research and validation in clinical practice.

2 Literature Review

The intensive growth of deep learning (DL) methods has radically transformed the process of automated skin cancer diagnosis, especially using dermoscopic images. The latest research by Akinrinade and Du (2025) [1] introduces an in-depth study of the deep machine learning methods to detect skin cancer, and in particular, address the issue of class imbalance and insufficient datasets. Their experiment compared various convolutional neural network (CNN) backbones with transfer learning and data augmentation and Generative Adversarial Networks (GANs). The performance increases were significant and particularly in melanoma classification that is generally under-represented

in clinical datasets. However, the authors highlighted the ongoing issues with overfitting, data heterogeneity, and population generalization, which places their article as a good modern standard of work in the implementation-oriented research.

In addition to model architecture, the size and structure of the datasets are also very important in defining the performance of DL. The SLICE-3D dataset was intro-duced by Kurtansky et al. (2024) in [2], which is a high-volume dataset, consisting of about 400,000 lesion image crops, taken as a result of three-dimensional total body photography. The work is a break in the traditional single-lesion dermoscopic datasets and represents spatially regular and longitudinal skin representations. They discov-ered that huge, massive datasets were beneficial to detecting in large scales and more thorough screening. Nevertheless, it was also pointed out by the authors that three-dimensional data involves more computationa.l load, which requires effective training methods and streamlined pipelines to be used in real life.

Recent studies have strongly criticized the use of AI-based diagnostic systems in clinical settings. The article by Wei et al. (2024) [3] assessed the incorporation of artificial intelligence in diagnosing skin cancer and made the findings that despite the frequent high-performance of DL models in a controlled experimental environment, the models are hard to implement in a clinical context. Notable obstacles found are dataset bias, lack of external validation and the lack of interpretability of model decisions. The paper highlighted that implementation-based research would focus more on strong and clinically relevant measures, especially the sensitivity of melanoma, instead of focusing on aggregate accuracy.

Systematic reviews have also expounded existing patterns and methodological shortcomings in this sphere. Choy et al. (2023) [4] used the extensive review of the DL-based methods of analyzing images to diagnose and monitor skin diseases. Their comparison affirmed the superiority of CNN-based networks in the classification of lesion and transfer learning is a common practice in this case because of the limited number of annotated medical images. Notably, the review described the differences in assessment schemes used in studies, which makes it challenging to primarily com-pare the level of performance. The authors have promoted standardized experimental pipelines and open reporting practices in order to enhance reproducible research an issue that is especially important in the case of implementation-oriented research.

Explainability has been one of the essential needs of medical AI systems partic-ularly in high stakes diagnostic situations. The systematically reviewed explainable artificial intelligence (XAI) techniques of skin cancer recognition described in Hauser et al. (2022) [5] Gradient-weighted Class Activation Mapping (Grad-CAM) as the

most commonly used and identified techniques to visualize model attention (Hauser et al., 2022). It was, however, noted in the review that most of the studies are based on a qualitative inspection of explanation maps, with little quantitative or clinical validation. This loophole explains why interpretability and performance optimization should be incorporated to come up with reliable diagnostic systems.

Simultaneously, the value of analysis of clinical context in DL models has been studied using dataset-centric research. Rotemberg et al. (2021) [6] described such a dataset in a patient-focused dataset as the combination of dermoscopic images and metadata including age of the patient, sex, and location of the lesion. Their research showed that melanoma can be detected more accurately on the basis of contextual information, and the false-positives will be minimized. However, to the extent of the caution by the authors that such enriched datasets should be carefully curated to avoid the risk of demographic or sampling bias, the authors validated the significance of a balanced data design in the DL-based deployment.

In addition to the issues of performance and datasets, the importance of explainable decision support systems has received growing popularity in the field of dermatological AI studies. The article by Lucieri et al. (2021) [7]offered an in-depth overview of the currently existing decision support systems based on DL to diagnose skin cancer and claim that clinically viable systems should go beyond autonomous prediction to provide explanations compatible with the cognition aspects of dermatologists. Their categorized work of explanability methods includes localization of visual relevance, prediction and localization of dermoscopic features, similarity-based retrieval, and intervention-based approaches, observing that most of the current systems are too dependent on saliency-based visualization techniques. Elaborating this point of view, Lucieri et al. (2021) [8] insisted on the fact that the successful implementation of clinical deployment must be based on user-friendly explanations, the combination of local and global interpretability, and the mechanisms that will enable intervention to remove the learned biases. These studies together point out that explainability ought to be used as part and parcel of decision support and no longer as a post hoc or purely visualization tool.

Altogether, the literature review indicates that the sphere of DL-based skin cancer classification has made significant strides in terms of the novel network structures, big datasets, and transfer learning methodology. Nevertheless, some serious gaps can be seen. There are numerous studies that focus on the overall classification accu-racy but pay inadequate attention to the melanoma related sensitivity and the risk of false-negatives. Moreover, explainability and implementation robustness are some-times discussed as non-priority issues. Inspired by these constraints, the given work is aimed at the systematic creation and optimization of a deep learning model to explicitly capture melanoma

detection performance without losing interpretability and methodological quality, and thus responding to major concerns that have been raised in recent studies.

3 Dataset Preparation

The data used in this task was the HAM10000-style skin lesion dataset, which was downloaded on Kaggle. The data set consists of dermoscopic pictures of various clinically-diverse pigmented skin lesion types such as melanoma, melanocytic nevi, basal cell carcinoma and other diseases. The augmented dataset in this paper is more balanced in the distribution of samples than the original HAM10000 dataset, which largely nullifies the strong class bias characteristic of skin lesion datasets. The initial collection of HAM10000 comprises about 10,015 dermoscope images of seven different types of skin lesions:

- Actinic Keratoses / Intraepithelial Carcinoma (AKIEC)
- Basal Cell Carcinoma (BCC)
- Benign Keratosis-Like Lesions (BKL)
- Dermatofibroma (DF)
- Melanoma (MEL)
- Melanocytic Nevi (NV)
- Vascular Lesions (VASC)

These classes represent a wide range of benign and malignant conditions and hence, the dataset can be used to create a multi-class diagnostic-modeling.

In order to guarantee dependable model training and testing it was decided first to consolidate the augmented datasets folders and to test the integrity and consistency of all images. Typical preprocessing steps were then used such as resizing all images to a shared resolution of 224 × 224 pixels and pixel intensity normalization using ImageNet statistics. These were the actions which guaranteed compatibility with the pretrained neural network backbone and encouraged stable training behavior.

A stratified partitioning technique was used to avoid the leakage of data as well as ensure that the split distributions of classes were consistent. The data was broken down into three mutually exclusive subsets in the following proportions:

- **Training set**: 70%
- **Validation set**: 10%
- **Test set**: 20%

This division approach gave an equal representation of each lesion type in all subsets that is crucial to stable learning on the model and the objective evaluation of gener-alization

performance. Patient-level constraints were honored during the partitioning where possible to prevent data leakage, so that one may not find images of the same lesion in more than one subset.

The ensuing distributions of the classes were indicative of balanced representation of the seven types of lesions in the training sample, and the same was the case in the validation as well as the test sets. The training subset was fed only data augmentation strategies such as random horizontal flipping, random rotation, and color jittering to make the network more robust and minimise overfitting due to the lack of visual diversity.

Such a well-crafted dataset constituted a valid and balanced basis of training the deep learning classification model, which was able to produce an accurate assessment of the overall predictive accuracy and clinically significant class-specific outcomes, specifically melanoma sensitivity.

4 Methodology

4.1 System Overview

The suggested method presents a multiclass dermoscopic skin cancer classification framework based on the deep learning. The general pipeline is designed in such a way that it provides high classification capabilities and directly focuses on melanoma sensitivity, which is very important to clinical screening and early diagnosis. The framework will follow a chronological workflow as shown in Fig. 1 and consists of dataset preprocessing, controlled data augmentation, deep feature learning, staged training, and extensive performance assessment.

Figure 1 Fig. 1 Proposed deep learning architecture of multiclass skin cancer classification with the ConvNeXt-Tiny.

In the center of the system, there is a modern convolutional neural network architecture, ConvNeXt-Tiny, which is the backbone of the feature extraction. ConvNeXt-Tiny is an inductive convolutional model that is based on Vision Trans-formers and preserves the efficiency and inductive biases of convolutional networks. It has four sequential architectures that have feature dimensions that are larger and larger in that order, and large-kernel depthwise convolutions, layer normalization, and residual connections that are used to improve the representation of features and training stability. Such design options allow successful modeling of fine grained visual features like lesion texture, color change and border abnormalities.

Transfer learning is used in the case of initializing the ConvNeXt-Tiny backbone with ImageNet-pretrained weights, which enables the network to borrow the high-level visual

features and rich low-level and mid-level visual features trained on a large-scale natural image. The head of the original classification is substituted with a task specific head that consists of global average pooling, then a fully connected layer with a softmax activation function to give class probability estimates of the seven skin lesion categories. This architecture has a trade-off between representational capability and computing efficiency, and thus it is suitable to medical image analysis scenarios where the amount of data is constrained, yet fine feature selection is needed.

4.2 Image Preprocessing

Dermoscopic images were all standardized before being trained so that the input representation of the dataset remains constant. Images were resized then to 256 x 256 pixels then to 224 x 224 pixels which is the input resolution of the ConvNeXt-Tiny backbone. The ImageNet mean and standard deviation were then used to normalize the pixel values in order to be compatible with the pretrained weight initialization.

To ensure that important visual features such as lesion texture, heterogeneity of colors and edges, which are clinically relevant, were preserved, the decision was to use a purposefully lightweight preprocessing strategy. The model was inhibited to learn discriminative patterns using realistic dermoscopic images, which inhibited excessive artifact removal, thus favoring strong features learning in settings that are highly in line with real world clinical data.

4.3 Data Augmentation Strategy

All the dermoscopic images were standardized before the training so as to provide similar input representation throughout the dataset. The images were then resized to 256 × 256 and then center-cropped to 224 × 224, which is the input resolution of the ConvNeXt-Tiny backbone. Normalization of pixel values was then done based on ImageNet mean and standard deviation in order to match the pretrained weight initialisation.

A preprocessing strategy use was carefully made light in order to retain clinically significant visual data, such as lesion texture, color heterogeneity, and boundary fea-tures. The model was trained not to overfit by removing artifacts, thus helping it to learn discriminative patterns of realistic dermoscopic images, which promotes strong feature learning under realistic conditions that are close to actual clinical data.

4.4 Model Architecture Design

All the dermoscopic images were standardized before the training so as to provide similar input representation throughout the dataset see Fig. 2. The images were then upsampled to 256 × 256 pixels and then center-cropped to 224 × 224 pixels, which is the input resolution of the ConvNeXt-Tiny backbone. Normalization of pixel values was then done based on ImageNet mean and standard deviation in order to match the pretrained weight

initialization A preprocessing approach specifically designed to be lightweight was implemented, allowing clinically relevant visual text to be retained, such as texture of the lesion, heterogeneity of colors and boundary. The model was trained not to overfit by removeing artifacts, thus helping it to learn discriminative patterns of realistic dermoscopic images, which promotes strong feature learning under realistic conditions that are close to actual clinical data.

4.5 Multi-Phase Training Strategy

An organized multi-phase education program was embraced to facilitate consistent convergence and achieve optimization of clinically significant performance. This plan was aimed at gradually transforming the model of generic visual feature learning to domain-specific discrimination with a specific focus on the detection of melanoma.

Phase 1 - Training on Classifier Head

At this stage, the child presents images and words to the classifier head which in turn relates them to the words. During the first step, all the layers of pretrained backbone were frozen, and training was limited to the newly added classification head. A relatively large learning rate was used to allow the classifier to quickly learn the target dataset with the features learned in ImageNet being used without change.

Phase 2A - Full Model Fine-Tuning

Thereafter, each network layer was unfrozen followed by fine-tuning of all network layers with a smaller learning rate. This step made it possible to perform a gradual refinement of deep feature representations in order to more effectively represent dermoscopic patterns that are unique to the skin lesion classification. Optimization of the whole network resulted in a significant enhancement in the validation error and the general network resilience.

Phase 2B – Melanoma-Specific Fine-Tuning

To further deal with the prob-lem of melanoma false negative which is a critical issue, another targeted fine-tuning step was included. A class-weighted loss function was used to impose more penalty to misclassification of melanoma, therefore, making it more sensitive to this high-risk category. This step was done with very low learning rate in order to improve the boundaries of classes and still have balanced performance across the lesion categories.

4.6 Loss Function and Optimization

The main objective function used in the skin lesion multiclass classification task was categorical cross-entropy loss. To reduce the residual imbalance in the training data in terms of class, the weights associated with the classes were calculated depending on the incidence of each category of lesion in the training data. These weights were added to the loss functional and later tuned during the melanoma-specific fine-tuning phase to up-weight the misclass cation of clinically critical melanoma cases.

The AdamW optimizer was used to perform the optimization of the model, and it is a combination of both adaptive learning rate updates and decoupled weight decay to enhance convergence stability and performance in generalization. Mixed-precision training was used to achieve high training efficiency and decrease the amount of used gpu memory. Such a solution made computation faster, retained the numerical stability, and model accuracy.

4.7 Evaluation Methodology

Independent test set was used in evaluating model performance which was not exposed either in training or validation. Measures of standard classification such as overall accuracy, precision, recall, and F1-score were calculated in order to obtain a detailed assessment of prediction performance. To study the behavior of per-class classification, a confusion matrix was created and melanoma sensitivity was considered since it is a highly clinically important disease.

In addition to quantitative measurements, the model interpretability was measured by Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAMheatmaps were obtained of correctly classified samples, false positive and false negative predi-cations. These graphical descriptions made it possible to qualitatively verify that the attention of the network was focused on the areas of diagnostic interest in lesions, and thus prove clinical plausibility of the learned representations.

5 Experimental Setup

5.1 Dataset Description and Splitting

All the experiments were performed on the basis of the Skin Cancer MNIST: HAM10000 Augmented Dataset, a popular dermoscopic skin lesion classification benchmark. This dataset has seven categories of lesions of clinical interest, which include Actinic Keratoses and Intraepithelial Carcinoma, Basal Cell Carcinoma, Benign Keratosis-like Lesions, Dermatofibroma, Melanocytic Nevi, Melanoma, and Vascular Lesions.

In order to have a clean and biased-free experimental protocol, it was reorganized before training. The original training and validation directories were merged in terms of classes

and then repartitioned into new training, validation and test subsets. This strategy ensured that in this manner all data were strictly separated among data consumed in model learning, hyperparameter search and testing data which was then conducted on the final evaluation making sure that there was no data leakage.

The last dataset was divided into two parts with the approximate percentage of 70% training, 10% validation, and 20% testing. This division gave adequate samples to optimize the model appropriately and maintain an independent test set to evaluate the performances without any bias. It was also of special concern that constant dis-tributions of classes were made in all splits, especially minority and high-risk groups like melanoma.

5.2 Training Configuration

All the experiments were carried out with PyTorch code. A workstation with an NVIDIA RTX 3050 (4 GB VRAM) and 8 GB of system memory was used to model train on, which is a realistic and available hardware setup in most academic and clinical research environments.

The mini-batch optimization was used to train and the size of the batch was chosen depending on the memory capabilities of the GPUs to keep the training stable. Mixed-precision training was also allowed during the experiments in order to increase the computational efficiency and minimize the use of memory. The approach ensured a quicker training and better use of resources without affecting numerical stability and model performance.

5.3 Hyperparameter Settings

To arrive at stable convergence and optimum performance, different hyperparameter settings were used on training phases:

- **Optimizer**: AdamW
- **Loss Function**: Categorical Cross-Entropy (with class weighting during Phase 2B)
- **Learning Rate**:
- – Phase 1: 1×10^{-3}
- – Phase 2A: 1×10^{-5}
- – Phase 2B: 2×10^{-6}
- **Weight Decay**: 1×10^{-5}

6 Results

6.1 Training Performance

The proposed skin cancer classification model was optimized with the help of struc-tured multi-phase training strategy, which was aimed at guaranteeing the stable convergence and the improvement of performance. In Phase 1, the only part of the training involved a fresh classification head as the pretrained ConvNeXt-Tiny back-bone was kept frozen. In this model, the validation accuracy of the model was about 76.7% and the results of the model showed that the model successfully adapted the classifier to the dermoscopic image characteristics without changing the pretrained representations.

In Phase 2A, all network layers were unfrozen and optimized with a smaller learning rate. This step allowed greater feature descriptions to be more flexible to domain-specific lesions. This led to significant improvement in the accuracy of validation to about 97.2%, which is a good sign of successful transfer and refinement of visual features that can be used in classifying skin lesions.

To further discuss the clinical significance of the detection of melanoma, Phase 2B proposed the melanoma-oriented fine-tuning by class-weighted optimization. This training stage was conservative and focused on minimizing false negatives in melanoma and balanced performance in all the classes. The method resulted in better general-ization and a maximum validation accuracy of 98.25% which was a measure of the success of the focused fine-tuning.

6.2 Test Set Evaluation

The last model assessment was performed on an external test set of 7,902 non-training/non-validation dermoscopic images. The trained model recorded the overall test accuracy of 98.01% and was found to exhibit high levels of generalization and strength in case of different lesion types.

- **Macro-average F1-score**:98.19%
- **Weighted-average F1-score**:98.01%

It is important to note that the model detected melanoma with a recall of 96.14% and this way significantly lowered the risk of false-negative prediction rates in this clinically imperative category.

6.3 Class-wise Performance Analysis

All the seven categories of skin lesions showed a high precision and recall in the pro-posed model. Other lesions like dermatofibroma and vascular lesions obtained almost the perfect classification of the lesion, which shows a high level of discriminative abil-ity of visually different classes. Benign lesions of keratosis type and melanocytic nevi demonstrated a bit lower but still strong performance, which can be explained by the visual similarity of these lesions types with malignant ones.

Remarkably, the Phase 2B melanoma-centered fine-tuning strategy proved to be of a great help in the detection of melanoma. This specific optimization enhanced sensitivity of melanoma and maintained equal performance in the rest of the lesion categories, which validates the efficacy of class-sensitive training in considering clinical classes of importance.

6.4 Confusion Matrix Analysis

An independent test set that was not visible in the process of training and validation was used to assess model performance. Several related measurements were used to get a confident and clinically significant evaluation of classification performance. The percentage of samples that were accurately classified under all lesion categories was used to determine overall accuracy. Precision was used to measure accuracy of positive forecasts by showing the number of forecasted samples of a certain type used correctly. Recall (sensitivity) was used to quantify the model in terms of its capacity to recognize the samples of each category correctly, but more specifically the melanoma because it is a very important clinical issue. F1-score which is the Harmonic of precision and recall gave a balanced analysis as both false positives and false negatives were considered.

To further examine the behavior by classes a confusion matrix was created and it is provided in Fig. 3 with the raw counts of prediction, and also with the normalized values of recall. This visualization allows to inspect in detail the misclassification patterns and to note the efficiency of the model in minimizing false-negative prediction of melanoma and to keep the performance constant when dealing with other types of lesions.

Besides quantitative evaluation, the interpretability of the model was evaluated with the help of Gradient-weighted Class Activation Mapping (Grad-CAM). Heatmaps were created of correctly classified cases and false positive and false negative predic-tion. These visual explanations confirmed that the model was always concerned with the clinically relevant lesion areas including irregular borders, asymmetry, and color change thus proving the clinical plausibility and transparency of the representations learnt.

6.5 Explainability Results

Grad-CAM representations gave clear indications that the model was always directed to clinically significant areas of lesion instead of the normal skin. The activation maps in the appropriately classified cases highlighted the features that are irregular bor-ders, color heterogeneity, and asymmetric structures, and they are well-established predictors of malignancy.

Additional comparison of false-positive and false-negative predictions provided use-ful information about model properties. Misclassifications were generally linked with delicate

or vague visual patterns and not with misdirected attention and this implied that any remaining errors would be as a result of inherent visual overlap between lesion types and not necessarily shortcoming on the model attention mechanism.

6.6 Qualitative Results

To supplement the quantitative assessment, the qualitative outcome was assayed by visualizing the model predictions on the model on typical dermoscopic and real-world skin lesion images, as demonstrated in Fig. 4. These examples consist of the external images and sample of the test dataset, which illustrates the ability of the proposed model to generalize to the test distribution.

The model was able to detect benign keratosis-like lesions and dermatofibroma with a high level of confidence and a probability of more than 96% in all the depicted cases. The visual inspection shows that the predictions are consistent with the clinically identifiable characteristics like lesion texture, pigmentation patterns, and boundary features. The ability of the learned feature representations to predict with consistency when changing skin tones, lighting conditions or the appearance of lesions underscores the strength of the learned feature representations.

These qualitative findings also confirm the usefulness of the ConvNeXt-Tiny archi-tecture and the multi-stage training plan. The high confidence predictions on the curated test samples and real-life images indicate that the model has acquired both discriminative and transferable lesion features, which supports the idea that the model can be relied on with regard to skin cancer classification tasks and validates the quantitative results reported above.

7 Conclusion

This research paper introduced a deep learning-based system of multiclass skin cancer classification using dermoscopic images with the main emphasis on the melanoma detection improvement. The proposed method can generalize well and attain high-quality diagnostic results on various types of lesions with a modern ConvNeXt-Tiny backbone and a multi-phase training scheme.

Experimental analysis showed that the progressive fine-tuning is an important factor to enhance the outcome of classification. The last model had a total test accuracy of 98.01% and a melanoma recall of 96.14% which shows that high-risk lesions would be discriminated. The Phase 2B training phase that utilized the melanoma-centric false-negative prediction significantly minimized the false-negative prediction alongside

the other classes highlighting the importance of class-sensitive optimization in the clinically sensitive setting.

Moreover, explainability by means of Grad-CAM offered a visual confirmation of how the model made decisions. The predictions were clinically relevant and inter-pretable with the activation maps successively showing lesion-related areas that were consistently related to malignancy. In general, the framework suggested provides a strong and clear method of automated skin cancer classification and leads to the development of reliable AI-based diagnostic systems.

Declarations

Research Funding Statement:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics Statement:

This study presents a deep learning model for multi-class skin cancer classification using dermoscopic images. The research was conducted using publicly available dermoscopic image datasets and did not involve direct interaction with human participants or animals. All images used in the study were obtained from open-access medical image repositories that follow ethical data collection practices and patient consent protocols. The study strictly adheres to ethical guidelines for research and data usage. No personal or identifiable patient information was used during the development and evaluation of the proposed model.

Consent to Participate:

Not applicable.

This study does not involve human participants directly, as it is based entirely on publicly available datasets with pre-existing ethical approvals and consent.

Acknowledgment:

The authors would like to thank Vishwakarma University for supporting this research work.

References

1. Akinrinade O, Du C (2025) Skin cancer detection using deep machine learning techniques, *Intelligence-Based Medicine*, vol. 11, Art. no. 100191
2. Kurtansky NR et al (2024) The SLICE-3D dataset: 400,000 skin lesion image crops extracted from 3D total body photography for skin cancer detection, *Scientific Data*,

vol. 11, no. 1, Art. no. 884

3. Wei ML, Tada M, So A, Torres R (2024) Artificial intelligence and skin cancer. Front Med, 11
4. Choy SP et al (2023) Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease, *npj Digital Medicine*, vol. 6, Art. no. 180
5. Hauser K et al (2022) Explainable artificial intelligence in skin cancer recognition: A systematic review. Eur J Cancer 167:54–69
6. Rotemberg V et al (2021) A patient-centric dataset of images and metadata for identifying melanomas using clinical context, *Scientific Data*, vol. 8, no. 1, Art. no. 34
7. Furriel BCRS, Oliveira BD, Prˆoa R, Paiva JQ, Loureiro RM, Calixto WP, Reis MRC, Giavina-Bianchi M (2024) Artificial intelligence for skin cancer detection and classification for clinical environment: A system-atic review, *Frontiers in Medicine*, vol. 10, Art. no. 1305954, Jan. 10.3389/fmed.2023.1305954
8. Lucieri A, Dengel A, Ahmed S (2021) Deep learning based decision support for medicine: A case study on skin cancer diagnosis, *Artificial Intelligence in Medicine*, preprint, Mar
9. Roky AH et al (2025) Overview of skin cancer types and prevalence rates across continents. Cancer Pathogenesis Therapy 3(1):89–100. 10.1016/j.cpt.2024.08.002
10. Yousef H, Alhajj M, Fakoya AO, Sharma S Anatomy, Skin (Integu-ment), Epidermis, in *StatPearls*. Treasure Island, FL, USA: StatPearls Publish-ing, Jun. 8, 2024. Available: https://www.ncbi.nlm.nih.gov/books/NBK470464/
11. Zhou L, Zhong Y, Han L, Xie Y, Wan M (2025) Global, regional, and national trends in the burden of melanoma and non-melanoma skin cancer: insights from the global burden of disease study 1990–2021. Sci Rep 15(5996):1–14. 10.1038/s41598-025-90485-3
12. Arnold M et al (May 2022) Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. JAMA Dermatology 158(5):495–503. 10.1001/jamadermatol.2022.0160
13. Kavitha C, Priyanka S, Kumar MP, Kusuma V (2024) Skin Cancer Detection and Classification using Deep Learning Techniques. Procedia Comput Sci 235:2793–2802. 10.1016/j.procs.2024.04.264
14. Pan Y, Tang B, Guo Y, Cai Y, Li Y (2025) Global burden of non-melanoma skin cancers among older adults: a comprehensive analysis using machine learn-ing approaches. Sci Rep 15(15266):1–21. 10.1038/s41598-025-99605-5
15. Hasan N et al (Oct. 2023) Skin cancer: understanding the journey of transformation from conventional to advanced treatment approaches. Mol Cancer 22(168):1–70. 10.1186/s12943-023-01854-3
16. Latoni DI, Semenov YR, Tsao H (2024) Differential time series analysis shows deceleration in melanoma mortality prior to the widespread use of highly effective

therapies, *Journal of the American Academy of Dermatology*, vol. 91, no. 5, pp. 872–879, Nov. 10.1016/j.jaad.2024.07.1457

17. Jones OT, Matin RN, Walter FM (Jan. 2025) Using artificial intelligence tech-nologies to improve skin cancer detection in primary care. Lancet Digit Health 7(1):e8–e10. 10.1016/S2589-7500(24)00216-4
18. Davar D et al (Nov. 2024) Neoadjuvant vidutolimod and nivolumab in high-risk resectable melanoma: A prospective phase II trial. Cancer Cell 42(11):1898–1918. 10.1016/j.ccell.2024.10.007
19. Lee JH et al (Apr. 2020) Transcriptional downregulation of MHC class I and melanoma de-differentiation in resistance to PD-1 inhibition. Nat Commun 11(1897):1–12. 10.1038/s41467-020-15726-7
20. Eggermont AMM et al (2020) Longer follow-up confirms recurrence-free survival benefit of adjuvant pembrolizumab in high-risk stage III melanoma: updated results from the EORTC 1325-MG/KEYNOTE-054 trial, *Journal of Clinical Oncology*, vol. 38, no. 33, pp. 3925–3936, Nov. 10.1200/JCO.20.02110
21. Janda M et al (2020) Early detection of melanoma: a consensus report from the Australian Skin and Skin Cancer Research Centre Melanoma Screening Summit, *Australian and New Zealand Journal of Public Health*, vol. 44, no. 2, pp. 111–115, Apr. 10.1111/1753-6405.12972
22. Boleti APDA et al (2024) Cutaneous melanoma: An overview of physiological and therapeutic aspects and biotechnological use of serine protease inhibitors, *Molecules*, vol. 29, no. 16, p. 3891, Aug. 10.3390/molecules29163891
23. Hasan N et al (Oct. 2023) Skin cancer: understanding the journey of transformation from conventional to advanced treatment approaches. Mol Cancer 22(168):1–70. 10.1186/s12943-023-01854-3
24. Walsh NM, Cerroni L (2021) Merkel cell carcinoma: A review, *Journal of Cuta-neous Pathology*, vol. 48, no. 3, pp. 411–421, Mar. 10.1111/cup.13910
25. Naik PP, Desai MB (2022) Basal cell carcinoma: A narrative review on con-temporary diagnosis and management, *Oncology and Therapy*, vol. 10, no. 2, pp. 317–335, Dec. 10.1007/s40487-022-00201-8

Figures

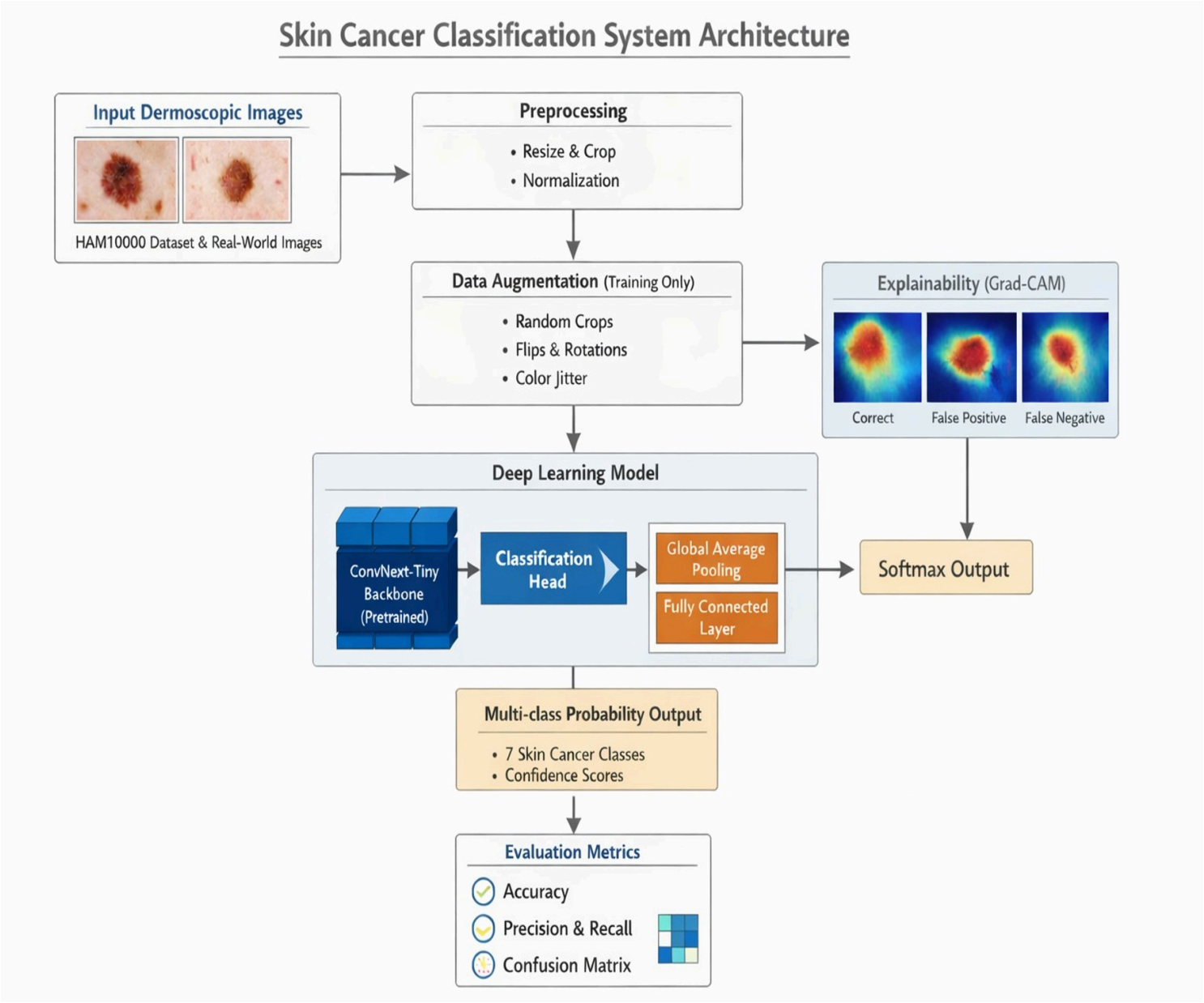


Figure 1

Proposed deep learning architecture of multiclass skin cancer classification with the ConvNeXt-Tiny.

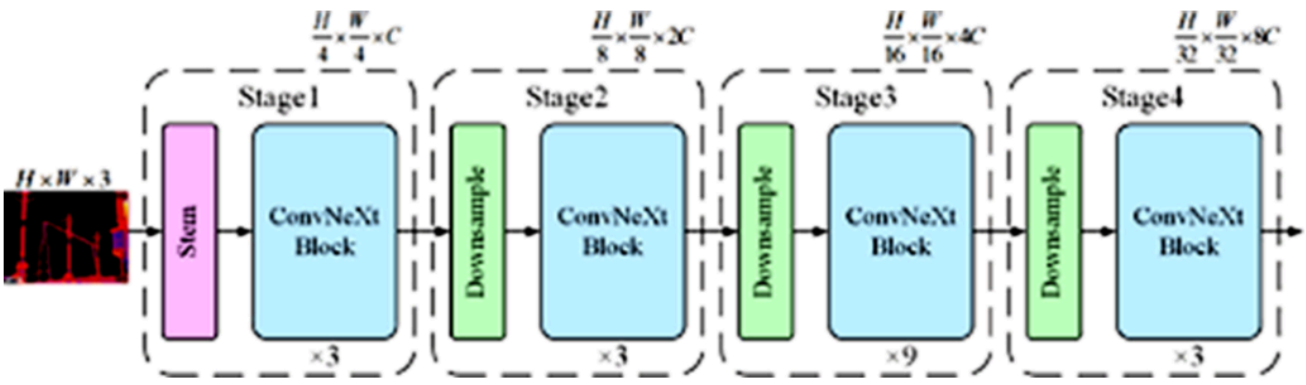


Figure 2

Architecture of the ConvNeXt-Tiny backbone illustrating multi-stage hierarchical feature extraction.

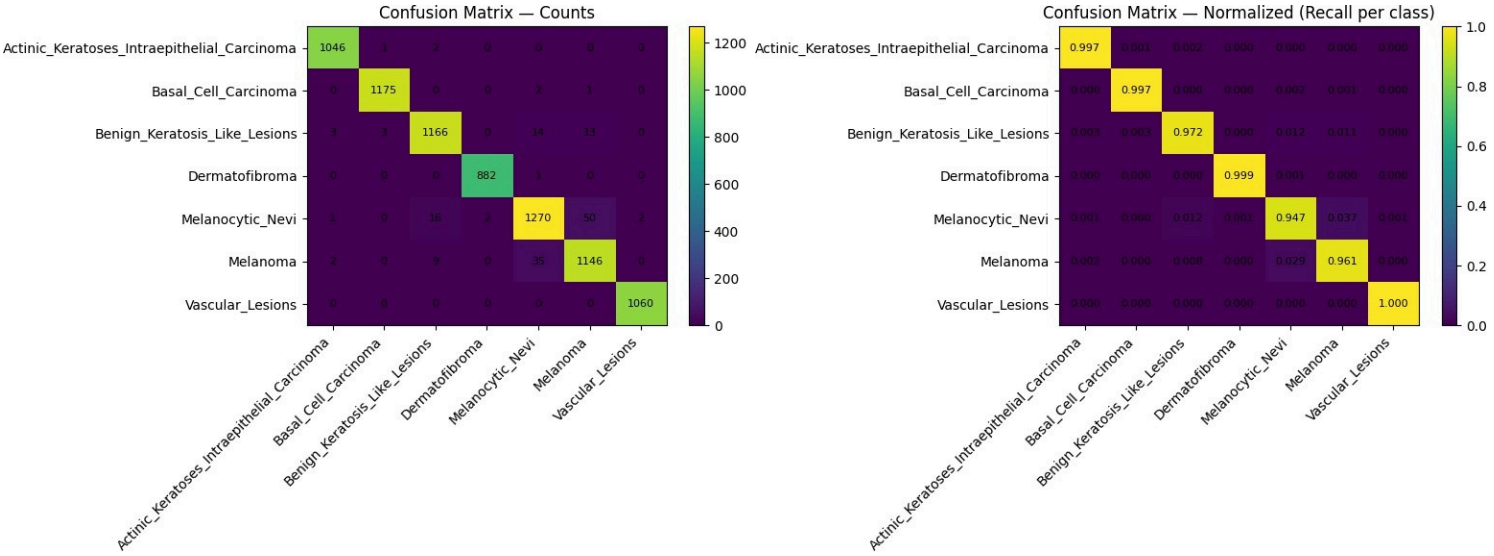


Figure 3

Confusion matrices of the proposed ConvNeXt-Tiny model showing (left) raw prediction counts and (right) normalized class-wise recall on the test set.



Figure 4

Qualitative classification outcomes of representative test and external skin lesion images. The model can be confident in predicting the classes of lesions and it is robust and can be generalized.