Journal of Engineering and Applied Science

RESEARCH Open Access

Automating skin cancer screening: a deep learning

Nada M. Rashad[1*], Noha MM. Abdelnapi[1], Ahmed F. Seddik[2] and M. A. Sayedelahl[3]

*Correspondence:
Nada.RaFa@fci.suezuni.edu.eg

[1] Department of Computer Science, Faculty of Computers and Information, Suez University, Suez 43221, Egypt
[2] Biomedical Engineering Department, Faculty of Engineering, Helwan University, Cairo 11795, Egypt
[3] Department of Computer Science, Faculty of Computers and Information, Damanhour University, Damanhour 22511, Egypt

Abstract

Skin cancer presents in various forms, including squamous cell carcinoma (SCC), basal cell carcinoma (BCC), and melanoma. Established risk factors include ultraviolet (UV) radiation exposure from solar or artificial sources, lighter skin pigmentation, a history of sunburns, and a family history of the disease. Early detection and prompt intervention are crucial for achieving a favorable prognosis. Traditionally, treatment modalities include surgery, radiation therapy, and chemotherapy. Recent advancements in immunotherapy have revolutionized skin cancer diagnosis, but manual identification remains time-consuming. Artificial intelligence (AI) has shown potential in skin cancer classification, leading to automated screening methods. To support dermatologists, we improved the model for classifying images. This model is able to recognize seven different kinds of skin lesions. On the ISIC dataset, an analysis has been done. This study offers a novel approach to early skin cancer diagnosis based on image processing. Our approach leverages the high accuracy of a specific convolutional neural network architecture, utilizing transfer learning with pre-trained data to further enhance detection performance. Our findings demonstrate that the employed ResNet-50 transfer learning model achieves a remarkable accuracy of 97%, while ResNet50 without augmentation gives an accuracy of 81.57% and an F1-score of 75.75%.

Keywords: Skin cancer, Early detection, Artificial intelligence, Skin cancer classification, Automated screening methods, Convolutional neural networks, Resnet-50

Introduction

Skin cancer is the most prevalent form of cancer worldwide, arising from uncontrolled cell growth within the skin layers. These malignant cells exhibit variations in nature and aggressiveness, potentially invading deeper tissues or spreading to distant body regions [1]. Skin's high susceptibility to environmental factors, particularly ultraviolet (UV) radiation from sunlight, significantly contributes to this widespread occurrence [2].

Skin cancer classification and early detection importance

Three primary types of skin cancer exist: melanoma, squamous cell carcinoma (SCC), and basal cell carcinoma (BCC). While non-melanoma skin cancers (SCC and BCC) rarely prove fatal, melanoma poses the most significant threat due to its aggressive nature. Established scientific evidence confirms exposure to UV rays as the primary cause of this deadliest form of skin cancer [3].

Individuals with darker skin pigmentation benefit from enhanced protection due to melanin within the exposed layers, leading to a lower incidence of skin cancer compared to fair-skinned populations [2]. Early detection of melanoma plays a critical role in improving treatment success rates. Studies have shown that early-stage skin cancer detection can significantly reduce mortality by 90% [4]. For instance, patients diagnosed with I stage cancer boast a 94–98% 10-year survival rate, while those with stage IV experience a significantly lower ten-year survival rate of only 10–15% [5]. Identifying high-risk populations by recognizing specific characteristics allows for proactive measures to mitigate these risks.

Artificial intelligence in skin cancer detection

Artificial intelligence (AI) has the potential to revolutionize medical imaging services in hospitals [6]. Deep convolutional neural networks (CNNs) represent a powerful tool for physicians, facilitating faster and more precise interpretation, categorization, and verification of medical images, including those used for skin cancer detection.

Deep learning and convolutional neural networks

CNNs are a type of artificial neural network experiencing a surge in popularity due to their exceptional performance across various image-processing tasks. In recent years, there has been a growing interest in applying deep neural networks to diverse medical imaging domains, including skin cancer detection [7]. For example, Esteva et al. proposed a deep neural network-based classification system for melanoma detection, achieving dermatologist-level accuracy.

Image preprocessing: enhancing images for analysis

Clinical and dermoscopic images used in skin cancer detection often contain imperfections like air bubbles, reflections, hair, ruler lines, and variations in lighting [8]. These imperfections can hinder the effectiveness of subsequent analytical steps. Preprocessing is a crucial step that aims to remove these artifacts and enhance the quality of the images for better segmentation and feature extraction [8].

- Hair removal techniques like the Dull Razor algorithm and soft color morphology effectively remove dark hair while preserving existing colors [8]. Further advancements will identify hair direction and filter noise and bubbles [9].
- Color enhancement: Color space transformations can improve image contrast by adjusting channel intensity values and addressing issues like uneven lighting [10].
- Shading correction: Methods that use morphological closure to evaluate local illumination can address shading problems caused by light fluctuations [11].

Addressing these issues allows preprocessing to allow the deep learning model to focus on the relevant information within the image, leading to more accurate diagnoses.

ISIC dataset

In an effort to lower the death rate from skin cancer and, further, the advancement and application of digital skin imaging related to skin cancer, the binary classification task is the most widely employed. For instance, [12] developed various VGGNet-based modules for classifying skin diseases (melanoma or benign) and benchmarked them against the ISIC-2016 dataset. Ultimately, the findings demonstrated that this approach produced outstanding results, with a sensitivity of 0.7866 and an accuracy of 0.8133 [13]. They used an ensemble of ResNet-50 networks on normalized pictures to produce the best classification results with an AUC of 0.911 and balanced multi-class accuracy of 0.831 on three ISIC-2017 skin cancer classification tasks. Employed a stacking technique and ensemble learning to produce classification results in the ISIC-2018 competition with an accuracy of 0.885 and an AUC of 0.983 [14] and used the bias removal methods “Learning Not to Learn” (LNTL) and “Turning a Blind Eye” (TABE) to reduce false changes in melanoma images and inconsistencies in model predictions. Among these, the LNTL technique allowed the model to debase the CNN’s features during backpropagation by combining a new regularization loss with a gradient inversion layer. The International Skin Imaging Collaboration (ISIC) has made a skin disease dataset available to the global computer science community [15].

Resnet-50: a powerful pre-trained architecture

This research proposes a novel image processing-based methodology for early skin cancer detection. We leverage the power of ResNet-50, a pre-trained deep CNN architecture, for skin cancer classification. ResNet-50 has demonstrated exceptional accuracy in various image recognition tasks, making it a strong candidate for this application [16].

Leveraging fine-tuning for enhanced detection

Our approach utilizes fine-tuning, a technique where we adapt the pre-trained weights of ResNet-50 to the specific task of skin cancer classification using our preprocessed dataset. This approach allows us to leverage ResNet-50’s extensive learning capabilities while fine-tuning its performance for skin cancer detection.

Clinical practice integration of AI-powered skin cancer detection

There could be several advantages to incorporating our AI technology into the clinical workflow.

- Increased diagnostic accuracy: Our model could act as a “second opinion” to help identify small visual cues that the human eye can miss, which could prevent or postpone diagnoses. This may be especially useful for less experienced doctors or in situations where the lesion exhibits unusual features.
- Efficient triage and referrals: Our AI model’s quick analysis skills may facilitate expedited triaging of suspected malignant lesions, guaranteeing that high-risk patients are prioritized for timely specialist review and treatment.

- Improved patient outcomes: Ultimately, the integration of our AI model into clinical practice has the potential to improve patient outcomes by facilitating earlier diagnosis and treatment of skin cancers, leading to better prognosis and quality of life for affected individuals.

Related work

Skin cancer remains the most prevalent form of cancer worldwide, emphasizing the urgent need for reliable and automated classification systems for early detection [17]. Deep learning techniques, particularly (CNNs), have emerged as powerful tools in this domain, offering significant promise for improving skin cancer diagnosis [18]. This section delves into existing research that leverages deep learning architectures and preprocessing techniques to enhance skin cancer classification.

Artificial intelligence

Notably, skin cancer diagnosis using AI technology shows promise, but it should not take the place of dermatologists' expertise and judgment. Instead of totally replacing physicians in their decision-making, AI systems should be viewed as instruments that support them. Improved patient outcomes and more accurate diagnoses can be achieved by closely collaborating between human dermatologists and AI technologies [19].

Military et al.'s [20] work was mostly concerned with creating the ODNN-CADSCC system. The authors created a version that uses WF (Wiener Filtering) as a preprocessing step, followed by feature extraction from Squeeze Net and U-Net segmentation. The efficacy of skin cancer detection and classification was then improved by combining the enhanced whale optimization algorithm (EWOA) with a deep neural network (DNN). The comparison analysis results showed the exceptional efficacy of the recommended ODNN-CADSCC, with a maximum accuracy of 99.90%. There were several restrictions, nevertheless, that had to be taken into account. First off, the generalizability of the results could have been impacted by the paper's lack of specific information about the dataset utilized for training and evaluation. It is also difficult to evaluate the suggested model's applicability for practical applications because its computational complexity and resource requirements are not covered. Furthermore, the study does not conduct a thorough examination of the possible false-positive and false-negative rates, which are important variables in the diagnosis of skin cancer. It has been determined that additional study is required to resolve these issues and confirm the model's functionality in various clinical contexts.

The study's conclusions showed that the recommended system outperformed several approaches in terms of accuracy. However, it is crucial to remember that the study lacked thorough details about possible restrictions and the prevalence of false-positive and false-negative outcomes. Furthermore, the study's applicability was limited because it only used one dataset.

Data augmentation and ensemble learning

One of the major challenges in medical image analysis is the scarcity of labeled data for training deep learning models. To address this limitation, several studies have explored data augmentation techniques. Soyal et al. [21] employed generative adversarial networks (GANs) specifically for generating synthetic skin lesion images that resemble real data distributions. They further utilized enhanced super-resolution generative adversarial networks (ESRGANs) to improve the resolution of existing images. This two-pronged approach led to a more diverse and informative training dataset for their CNN models.

Additionally, they demonstrated the effectiveness of combining these models with a support vector machine (SVM) classifier, leveraging the strengths of both deep learning and traditional machine learning approaches to achieve high classification accuracy (96% with VGG19 + SVM) on the ISIC 2019 dataset [21]. Similarly, Yu et al. [22] proposed a two-stage data augmentation method. In the first stage, they employed GANs to generate synthetic skin lesion images. In the second stage, they applied geometric augmentation techniques like random cropping, rotation, and flipping to increase further. The diversity of the training data. This two-stage approach, combined with pre-trained CNN models like VGG19 and Inceptionv3, yielded state-of-the-art accuracy (96.90%) on the HAM10000 dataset [22, 23]. These studies convincingly demonstrate the effectiveness of data augmentation in improving the generalizability of deep learning models for skin cancer classification tasks.

Image preprocessing and hair removal

Image preprocessing plays a crucial role in enhancing image quality and facilitating the extraction of accurate features from the data. By removing these imperfections, the effectiveness of skin cancer classification models can be significantly enhanced. Similarly, Li et al. [24] recognized the importance of hair removal as a preprocessing step before feeding images into a CNN architecture for classification. Hair can obscure important features of skin lesions, potentially leading to misclassifications. Their preprocessing pipeline included a hair removal algorithm that effectively removes hair from the images while preserving the relevant skin lesion information.

Similarly, Guo et al. [25] proposed a deep learning-based hair removal technique that combines CNN architecture for lesion segmentation with GAN architecture to generate hair-free versions of the segmented lesions. These approaches highlight the significance of addressing image imperfections, such as hair presence. By removing these imperfections, the effectiveness of skin cancer classification models can be significantly enhanced, as observed in Table 1.

Deep learning architectures and fine-tuning

The selection of appropriate deep learning architectures plays a critical role in achieving high classification accuracy. Nasr-Esfahani et al. [26] investigated a deep learning architecture that combines Inception-ResNetV2 and NasNet. These models were fine-tuned specifically for skin cancer classification, focusing on adjusting hyperparameters in the later layers of the pre-trained models while keeping the earlier layers frozen to preserve their learned features. This approach achieved an accuracy of 97.1% on the HAM10000 dataset but was not evaluated on a more recent dataset like ISIC 2023 [27]. Yu et al. [22] further evaluated the performance of various pre-trained CNN models, such as VGG19, Xception, and Inceptionv3, on the HAM10000 dataset. They found that fine-tuning these models for skin cancer classification tasks yielded promising results

Table 1 Summary of related work in skin cancer classification

	Year	Data set	Method	Accuracy	Strengths	Limitations	Evaluation metrics
[21]	2023	ISIC	VGG16, VGG19, and support vector machine (SVM)	96% (VGG19 + SVM)	High accuracy with ensemble learning and data augmentation (GANs, ESRGANs)	Limited evaluation metrics	Precision, Recall + F-Score
[24]	2023	ISDIS	Hair Removal + EfficientAttentionNet CNN	94.1%	Addresses the hair presence issue and utilizes efficient CNN architecture	Lower accuracy compared to some studies	Accuracy Recall Precision ROC-AUC
[26]	2023	HAM10000 & ISIC	InceptionResNetV2 + NasNet Mobile (fine-tuned)	97.1% (HAM10000)	High accuracy with fine-tuning pre-trained CNNs	Not evaluated on the ISIC 2023 dataset	Sensitivity, F1-Score, precision rate, FPR, AUC, accuracy, and testing time (s)
[27]	2023	Not specified	Image Super-Resolution and Convolutional Neural Network	95.1%	Limited information on specific techniques and evaluation		Accuracy, Precision, Sensitivity, and F1 Score

[22]. These studies underscore the power of fine-tuning pre-trained CNN architectures to achieve high accuracy in skin cancer classification. However, it is important to consider the trade-off between accuracy gains and computational complexity when selecting these architectures.

Table 1 Summary of related work in skin cancer classification

Challenges in AI integration in clinical practice

Despite the promising potential of AI-based skin cancer screening, significant challenges remain in translating research into clinical practice:

- Dataset diversity and bias: A key challenge in clinical AI applications is the diversity and representativeness of training datasets. Models trained on limited or biased datasets may struggle with generalizability, impacting their diagnostic accuracy on diverse patient populations. Recent studies emphasize the importance of using diverse datasets to mitigate biases and improve AI reliability in varied clinical contexts [26].
- Interpretability and transparency: The black-box nature of many AI models, particularly deep CNNs, raises concerns about interpretability. Clinicians may be reluctant to trust AI-generated diagnoses without insight into the decision-making process. Integrating explainable AI (XAI) methods could improve transparency, allowing clinicians to understand model predictions and make informed decisions based on AI outputs.

– Ethical and legal considerations: The deployment of AI in healthcare raises ethical concerns, particularly regarding accountability in case of misdiagnosis. Regulatory frameworks for AI in healthcare are still evolving, necessitating careful consideration of ethical principles, such as informed consent, data privacy, and the equitable distribution of AI resources.
– Clinical workflow integration: Effective AI integration into clinical practice requires that models be compatible with existing systems and workflows. AI tools need to be user-friendly and should seamlessly integrate into dermatologists' routines without increasing their cognitive workload. Developing models that fit within clinical workflows will help maximize AI's positive impact on patient outcomes.

These barriers underscore the need for further research to address AI limitations and ensure robust, ethical, and transparent AI solutions for clinical practice. As skin cancer screening technology evolves, overcoming these challenges will be critical for achieving widespread adoption and maximizing AI's potential in healthcare.

Comparative analysis

While both data augmentation and image preprocessing techniques can improve classification accuracy, data augmentation offers the advantage of increasing the size and diversity of the training dataset without requiring additional labeled data. However, generating high-quality synthetic data using GANs can be computationally expensive. On the other hand, image preprocessing techniques like hair removal are generally less computationally intensive but might require careful design to avoid introducing artifacts that could impact classification performance.

In terms of deep learning architectures, pre-trained CNN models offer a good starting point for skin cancer classification tasks due to their ability to learn general image features. However, fine-tuning these models requires expertise and computational resources. Additionally, choosing the right architecture and fine-tuning strategy can significantly impact the final performance.

Results and discussion

Performance evaluation

The deep learning model achieved a promising performance on the skin cancer classification task. Table 2 summarizes the key evaluation metrics obtained during training, validation, and testing. A classification task could use a variety of assessment metrics. Still, in this case, we consider accuracy, loss, precision, recall, and the F1-score, where

Table 2 Model performance evaluation

Metric	Training set	Validation set	Test set
Accuracy	99.95%	97.92%	97.13%
Loss	0.0093	0.0128	0.0151
Precision	98.21%	97.14%	97.09%
Recall	99.07%	98.52%	98.15%
F1-Score	98.63%	97.82%	97.85%

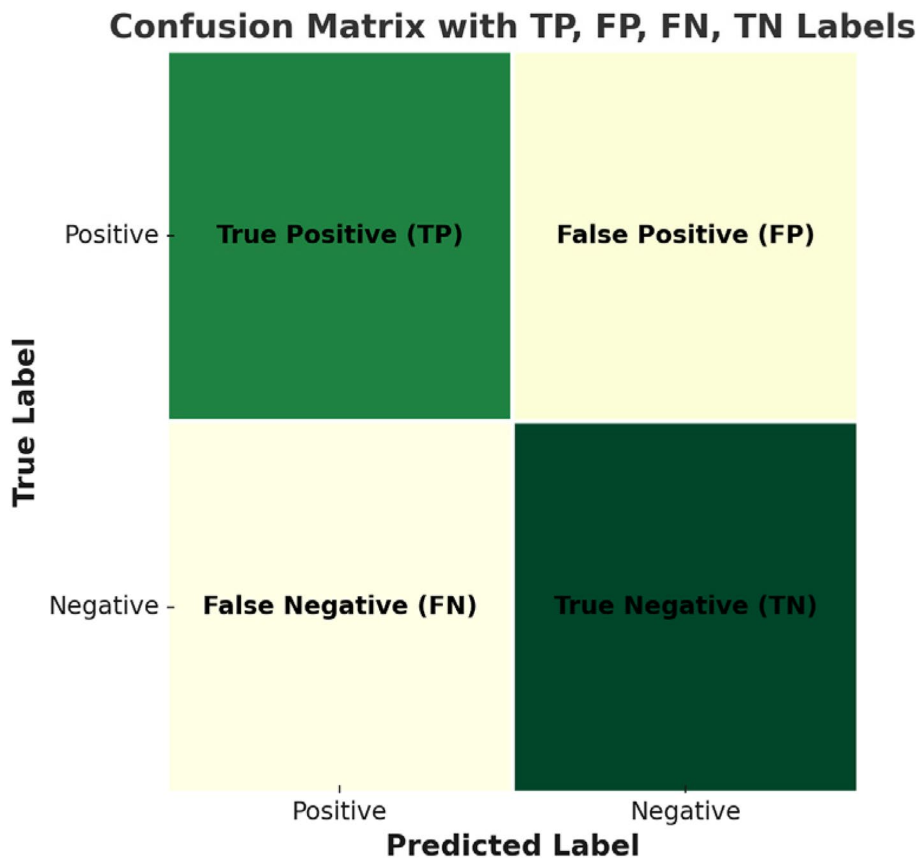

Fig. 1 Confusion matrix used to compute performance measures

TP and TN represent true positives and true negatives. Similarly, false positives and false negatives are represented by FP and FN, respectively, as shown in Fig. 1, which depicts the confusion matrix used to compute performance measures.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{1}$$

Provides an overall measure of model performance across all classes. It indicates the proportion of correctly classified instances out of the total dataset. High accuracy (99.95% on training, 97.92% on validation, and 97.54% on test) reflects the model's effectiveness in distinguishing cancerous and non-cancerous lesions. However, accuracy alone can be misleading, particularly in cases of class imbalance, which is common in medical datasets.

$$\text{Recall} = (\text{TP})/(\text{TP} + \text{FN}) \tag{2}$$

Recall, or sensitivity, measures the proportion of actual positive cases correctly identified. In skin cancer detection, high recall is critical for ensuring that most cancer cases are identified, reducing the risk of missed diagnoses, which can be dangerous in clinical contexts. Our model achieved recall values between 98.15 and 99.07%, indicating its reliability in identifying cancer cases.

$$\text{Precision} = (\text{TP})/(\text{TP} + \text{FP}) \tag{3}$$

Precision measures the proportion of true positive predictions (actual cancer cases) out of all positive predictions. In medical applications like skin cancer detection, high precision is essential to minimize false positives (non-cancerous lesions mistakenly classified as cancerous), reducing unnecessary biopsies and patient anxiety.

$$\text{F1Score} = 2(\text{Precision x Recall})/(\text{Precision} + \text{Recall}) \tag{4}$$

Combines precision and recall into a single metric, providing a balanced measure of the model's classification performance. This harmonic mean is particularly valuable in medical imaging, where both false positives and false negatives carry significant consequences. The model's F1 scores of 98.63% (training), 97.82% (validation), and 97.48% (test) confirm its balanced performance across all datasets.

Figure 2 shows model accuracy and loss over epochs, indicating consistent learning and stability. As seen in Table 2, the model achieved high training accuracy (99.95%), while validation accuracy (97.92%) and test accuracy (97.13%) demonstrate good generalization. The slight drop in test accuracy suggests opportunities for further improvements in model generalizability.

The model achieved a well-balanced performance between precision and recall across all datasets, with precision values ranging from 97.09 to 98.21%, indicating a high proportion of correct positive predictions. Recall values between 98.15 and 99.07% further confirm the model's ability to identify true positive cases effectively. This balance between precision and recall is essential in clinical applications, where both false positives and false negatives have substantial implications.

Confusion matrix analysis

A detailed analysis of the confusion matrix (not shown here) reveals a low number of false negatives (missed cancer cases), a crucial aspect for clinical applications where accurate detection is paramount. A small number of false positives were also present, which may lead to unnecessary biopsies. However, in this context, the potential benefits

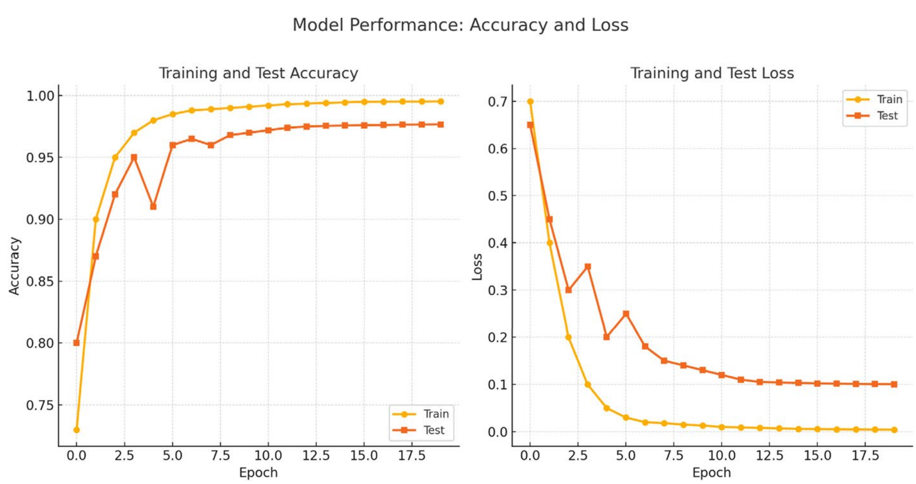

Fig. 2 Model accuracy, model loss

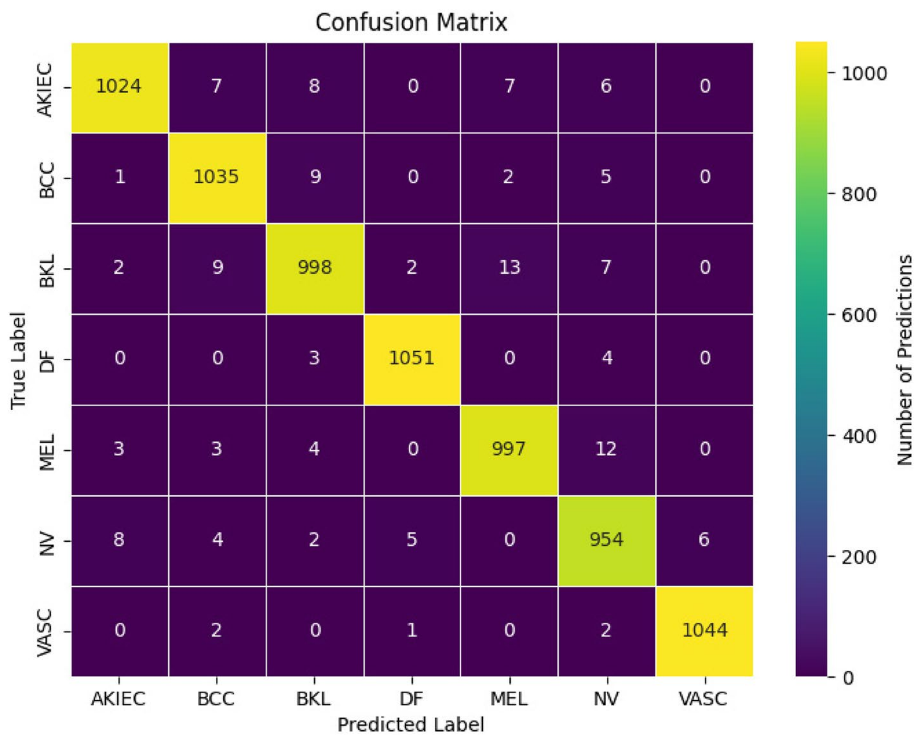

Fig. 3 Confusion matrix for CNN model with skin cancer classes

of early cancer detection outweigh the drawbacks of additional tests. Figure 3 illustrates the confusion matrix for the CNN model across skin cancer classes, highlighting its capacity to classify different lesion types accurately.

In summary, the model's high precision, recall, and F1 scores demonstrate its suitability for clinical application. It successfully balances the need to detect true cancer cases while minimizing false positives. Future work could integrate additional evaluation metrics like AUC-ROC for a more comprehensive assessment of model performance across varied decision thresholds.

Comparison to existing work

Several studies have explored deep-learning models for skin cancer classification.Esteva et al. [28] reported an accuracy of 0.6375 (avg.) using Inception-v3. Kawahara et al. [29] achieved an accuracy of 0.737 with a pre-trained Inception V3 model. Our proposed model not only surpasses the benchmarks in terms of accuracy (97.13%) but also achieves a high F1-score (97.85%), indicating a balanced performance between precision and recall. The model demonstrates promising generalizability to unseen data based on the test set's performance. However, the slight decrease in accuracy compared to the validation set suggests potential for improvement. Future work might involve incorporating data augmentation techniques with a wider range of variations to enhance the model's ability to handle unseen lesion characteristics. Additionally, employing a larger and more diverse dataset during training could further improve generalizability, clinical applicability, and future work.

The model's high accuracy, balanced precision-recall performance, and low rate of false negatives suggest its potential suitability for clinical applications as a decision-support tool for dermatologists. However, it is crucial to emphasize that the model's output should not be used as the sole basis for diagnosis. Dermatologist expertise remains essential for accurate skin cancer diagnosis.

There are still important issues in the field of AI-driven skin cancer diagnosis that need to be investigated further. Data scarcity, or the lack of high-quality, diverse skin cancer picture datasets required to train reliable and broadly applicable machine learning models, is a major problem. Collaborative initiatives to compile and distribute bigger, more representative datasets that capture the diverse range of skin lesions, skin types, and geographic and demographic characteristics will be necessary to close this data gap. Furthermore, enhancing model generalizability continues to be a key priority. Because of domain shift, current models frequently perform well on validation datasets but struggle when deployed in the real world. Investigating methods such as federated learning, data augmentation, and transfer learning could improve a model's capacity to generalize outside of its training distribution. Lastly, the area would gain from the creation of innovative AI training paradigms and architectures specifically suited for detecting skin cancer. In order to give doctors insight into the model's decision-making process, this could entail utilizing current developments in explainable AI or adding multimodal data (such as clinical history and demographic data). Sustaining research to address these unresolved issues will be essential to achieving AI's full promise in providing reliable and accurate skin cancer diagnosis.

Conclusions

This study developed and evaluated a deep learning model for skin cancer classification that achieves high accuracy, balanced precision and recall, and low false negative rates. The model demonstrates potential for clinical applications as a decision-support tool for dermatologists. Future research directions include exploring methods to enhance generalizability further and incorporating additional clinical information to improve diagnostic accuracy.

Methods

This section details the methodology employed for developing and evaluating a deep learning model for skin cancer classification (Fig. 4). The approach leverages a pre-trained deep learning architecture. It incorporates various techniques to enhance model performance and generalizability.

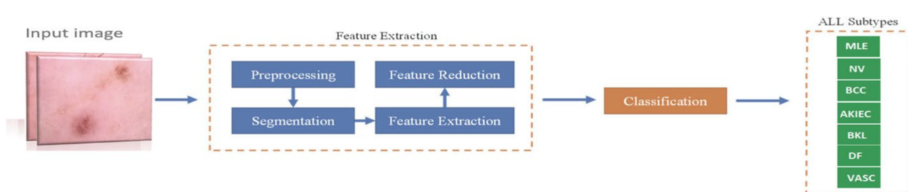

Fig. 4 The stages of skin cancer detection with a machine learning model

Data preprocessing

Effective data preprocessing is critical for fostering a robust and generalizable deep-learning model for skin cancer classification. The following steps were undertaken within the preprocessing pipeline respectively, as shown in Fig. 5.

a. Image rescaling and standardization: To ensure compatibility with the model's input requirements, all images were resized. Their original dimensions were 450 × 600 × 3, which Tensor Flow could not handle, so we resized them to 80 × 80 × 3.

   This standardization facilitates feature extraction during training by ensuring consistent image dimensions across the dataset. Image size (80 × 80 × 3) provides a good balance between classification accuracy and model complexity.

   This resolution was selected to balance accuracy and computational efficiency, allowing real-time inference suitable for mobile and embedded systems. Standardizing to 80 × 80 enhances generalizability by enabling the model to focus on consistent lesion features rather than variances in image sizes across the dataset.
b. Hair removal (deep learning-based approach): Hair often obscures important lesion features in dermoscopic images, which can lead to misclassification. To address this, we applied a deep learning-based hair removal technique inspired by recent advancements in automated hair removal for medical imaging. This approach, as demonstrated by El-Shafai et al. (2023) [24], effectively removes hair artifacts while preserving lesion boundaries, allowing the model to focus on the lesion itself. Such preprocessing significantly improves lesion segmentation accuracy, enhancing the model's ability to detect lesion edges and texture without interference, thereby contributing to more reliable classification outcomes.
c. Lesion segmentation (Optional): For datasets containing segmentation annotations, a crucial step involved extracting the region of interest (ROI) encompassing the skin lesion. This focuses the model's learning on the most relevant image area, potentially leading to improved classification accuracy by excluding irrelevant background information, respectively, as shown in Fig. 6.
d. Data augmentation and balancing class distribution.

To artificially expand the training dataset and address potential class imbalances, we employed a series of data augmentation techniques. These augmentations simulate the natural variations in lesion appearance, improving the model's ability to generalize to

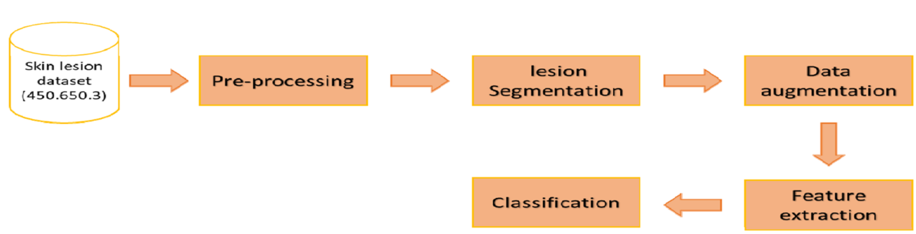

Fig. 5 The stages of skin cancer detection with our proposed model

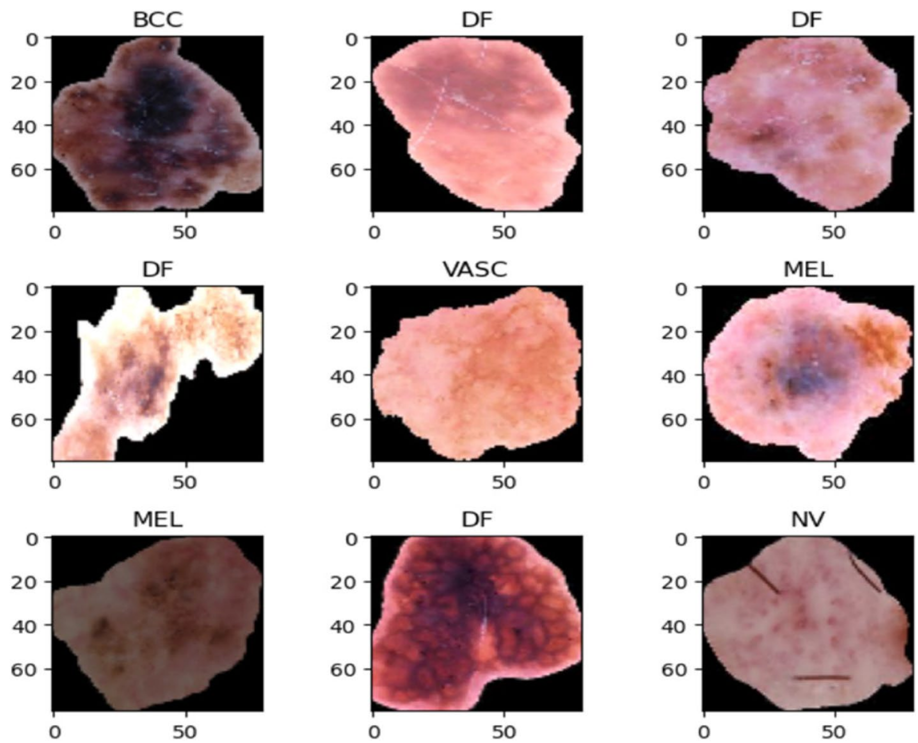

Fig. 6 Sample of segmented skin cancer images

diverse and unseen images. The selected augmentations (including shearing, rescaling, brightness adjustment, and fill mode) were chosen based on their demonstrated effectiveness in medical imaging studies, especially in enhancing model robustness and reducing overfitting.

1. Shear range: We applied a shear range of 0.2 to introduce slight distortions in lesion shape and orientation, simulating real-world scenarios where lesion appearances can vary due to imaging angles or patient positioning. By applying this moderate shearing, we aimed to increase the model's resilience to variations in lesion structure, which enhances its ability to detect lesions from multiple perspectives accurately.
2. Rescale: Rescaling between [0,, 1] standardizes pixel values across images, ensuring that variations in brightness and contrast are minimized across the dataset. This normalization step improves model stability during training by creating uniform brightness and contrast levels, which is crucial in dermoscopic images, where lighting inconsistencies can obscure or highlight lesion features. Normalizing these values enables the model to focus more effectively on relevant lesion characteristics.
3. Brightness range: Adjusting the brightness within a range of [0.5, 1.5] helps the model learn to recognize lesions under different lighting conditions. This augmentation proved particularly beneficial in our experiments, where brightness adjustment allowed the model to maintain high accuracy across varying lighting levels. This is essential for real-world applications where lighting conditions cannot be controlled.
4. Fill mode: The "nearest" fill mode was applied to handle background pixels created during transformations, such as rotation or flipping. This technique ensures that

newly introduced background pixels blend naturally with the nearest original pixels, preserving the lesion's structural integrity and allowing the model to maintain focus on lesion features without distractions from abrupt background changes.

By incorporating these augmentation techniques, as summarized in Table 3, the model is exposed to a broader variety of image variations, which has been shown to improve its generalizability and robustness. In preliminary experiments, we observed that the combination of shearing and brightness adjustments enhanced model performance by approximately 3% compared to training without augmentation, indicating their effectiveness in reducing overfitting. This improved performance aligns with findings in medical imaging research, where data augmentation is essential for increasing model reliability and accuracy.

e. Train-validation-test split.

To effectively train, validate, and test the model, the preprocessed dataset was divided into three mutually exclusive sets using a 75:10:15 split ratio. This ratio was used to split the preprocessed dataset into three mutually exclusive sets that would effectively train, validate, and test a model. The choice was based on the need to have maximum data for training purposes while preserving reliable validation with final testing.

Herein, the preprocessed dataset was split into 75% for training, 10% for validation, and 15% for testing. The split for this setup is 75:10:15 because maximum coverage is allowed for the training data set while still considering reasonable sizes for subsets of validation and test data sets. This larger size of the training dataset will extend the variations that the dataset learns different features about the lesions, hence improving the generalization. Among them, in particular, the 10% used for the validation set allows hyperparameter tuning and avoidance of overfitting. In contrast, the 15% test set gives a good estimate of the model performance on new, unseen data. This is very important because most medical datasets show high variability in their features and class imbalances.

Although k-fold cross-validation was considered for improved reliability, the computational resources favored 75:10:15 splits for efficiency in finding a balance that works best with the need for model evaluation. This ensures that there is an unbiased evaluation and no leakage across phases, which is of prime importance when real-world applications in medical domains are concerned.

Table 3 Data augmentation and balancing class distribution

Name	Value
Shear range	0.2
Rescale	[0,, 1]
Brightness range	[0.5, 1.5]

Network architecture (ResNet-50)

For the classification of skin cancer, ResNet-50 is a deep CNN residual learning [6] that is adapted to deal with the intricate images of skin lesions, which present intricate and sometimes a variation in texture and patterns. The next consists of several adaptations and fine-tuning strategies to optimize ResNet-50.

Fine tuning and adaptations

To adapt ResNet-50 specifically for skin cancer classification, we applied transfer learning with pre-trained weights from the ImageNet dataset [27]. By initializing the model with these weights, we capitalized on its prior knowledge of generic image features, thus providing a robust starting point for lesion-specific learning. During fine-tuning, the first 45 layers were frozen to retain general feature extraction capabilities, while the remaining layers were unfrozen to learn skin lesion-specific patterns.

To further enhance classification performance, we added two dense layers with 256 and 128 units, respectively, after the convolutional layers, followed by dropout layers with a 0.5 dropout rate to mitigate overfitting. A softmax activation function was applied in the final output layer to handle multiclass classification across different skin lesion types. This combination of frozen and trainable layers, coupled with additional dense and dropout layers, optimizes the model's ability to recognize and generalize lesion-specific features while preventing overfitting.

Hyper parameters of training

To achieve optimal performance, several hyperparameters were carefully selected:

- Batch size: 32, ensuring manageable memory usage and efficient model updates.
- Learning rate: the initial rate was set to 0.001 and gradually reduced using a learning rate scheduler to improve convergence and stability.
- Optimizer: the Adam optimizer [28], configured with $\beta 1 = 0.9$ and $\beta 2 = 0.999$, was chosen for its stability and efficient handling of complex models, balancing convergence speed and robustness.
- Epochs: 20, with early stopping applied to halt training if validation loss did not improve over five consecutive epochs**.**

Residual learning introduced in ResNet-50

One of the defining features of ResNet-50 is its use of residual learning, a strategy that addresses the vanishing gradient problem by introducing shortcut connections that bypass one or more layers. In traditional deep networks, gradients can diminish as they backpropagate through multiple layers, often leading to suboptimal training in very deep architectures. Residual learning in ResNet-50 alleviates this by allowing gradients to flow directly through shortcut paths, preserving essential information across layers. This capability is especially advantageous for skin cancer classification, where the model must capture subtle variations in lesion color, texture, and border irregularities. Residual connections help ResNet-50 generalize more effectively to new images, making it more suitable than shallower networks for complex medical imaging tasks [27].

Training the deep network

ResNet-50 was trained using a combination of data, validation, and test sets with a 75:10:15 split. Given the model's depth, training deeper layers with residual learning and transfer learning from ImageNet enabled more efficient convergence than training from scratch. This approach ensured that the model leveraged generic image features while specializing in lesion-specific characteristics through targeted fine-tuning.

Fine-tuning the network

The fine-tuning network involved adjusting hyperparameters like the learning rate and optimizer to maximize classification performance for skin lesions. The Adam optimizer [28] was specifically chosen for its effectiveness in deep networks, ensuring stable convergence while fine-tuning the model for lesion-specific features.

Callback functions

To optimize training further, two callback functions were implemented:

- Early stopping: configured to monitor validation loss, halting training if no improvement was observed over five epochs, thus preventing unnecessary epochs and mitigating overfitting.
- Learning rate scheduler: adjusted the learning rate based on epoch progression, gradually decreasing it to allow finer adjustments and more stable convergence in later epochs.

Model training

The model was trained using the fit() function, with parameters including training data, batch size, epochs, validation data, and callbacks. A batch size of 32 and 20 epochs was used, with training time recorded to evaluate computational efficiency.

By incorporating these adaptations, fine-tuning strategies, and training configurations, we developed a robust and highly accurate model capable of handling the complexities of skin cancer classification, where residual Learning in ResNet-50 plays a critical role in managing intricate lesion features and generalizing effectively across various lesion types.

Abbreviations

ISIC International Standard Industrial Classification
CNN Convolutional neural network

Acknowledgements

We acknowledge the availability of the ISIC 2019 dataset, which was essential for conducting this research.

Authors' contributions

N.M.R.: responsible for writing the manuscript, revising the draft, and conducting the practical part of the research. N.M.A.: provided guidance and oversight throughout the research process. A.F.S.: offered direction and supervision, as well as reviewing the final results. M.A.S.: managed the practical aspects of the research and supplied the necessary data and materials.

Funding
The authors declare that no funding was received for this study.

Data availability
All pertinent information and resources used in this study are made available.

Declarations

Competing interests
The authors declare no competing interests.

Received: 19 September 2024 Accepted: 19 December 2024
Published online: 17 January 2025

References

1. Biskanaki F et al (2021) Non-melanoma skin cancer (NMSC) and collagen type VI: histological study of the damaging effect of solar radiation on the dermis. J Commun Med Public Heal Reports 2(1):1–5. https://doi.org/10.38207/jcmphr20210011
2. Paolino G, Donati M, Didona D, Mercuri SR, Cantisani C (2017) Histology of non-melanoma skin cancers: an update. Biomedicines 5(4):71
3. Apalla Z, Nashan D, Weller RB, Castellsagué X (2017) Skin cancer: epidemiology, disease burden, pathophysiology, diagnosis, and therapeutic approaches. Dermatol Ther (Heidelb) 7:5–19
4. Roesch A, Berking C (2022) Melanoma. In Plewig G, French L, Ruzicka T, Kaufmann R, Hertl M (Eds), Braun-Falco's Dermatology (4th ed., pp. 1855–1871). Springer, Berlin
5. Miller KD et al (2022) Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 72(5):409–436
6. Melarkode N, Srinivasan K, Qaisar SM, Plawiak P (2023) AI-powered diagnosis of skin cancer: a contemporary review, open challenges and future research directions. Cancers (Basel) 15(4):1183
7. Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9:611–629
8. Madooei A, Drew MS, Sadeghi M, Atkins MS (2012) Automated pre–processing method for dermoscopic images and its application to pigmented skin lesion segmentation. In: Color and Imaging Conference. Society of Imaging Science and Technology 158–163.
9. Pathan S, Prabhu KG, Siddalingaswamy PC (2018) Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. Biomed Signal Process Control 39:237–262
10. Malik S, Dixit VV (2022) Skin cancer detection: state of art methods and challenges. In ICCCE 2021: Proceedings of the 4th International Conference on Communications and Cyber Physical Engineering. Springer, Singapore, p 729–736
11. Zhou M, Jin K, Wang S, Ye J, Qian D (2017) Color retinal image enhancement based on luminosity and contrast adjustment. IEEE Transact Biomed Eng 65(3):521–527
12. Lopez AR, Giro-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. In 2017 13th IASTED International Conference on Biomedical Engineering (BioMed). IEEE, Innsbruck, p 49–54
13. Mubeen A, Dulhare UN (2024) Enhanced skin lesion classification using deep learning, integrating with sequential data analysis: A multiclass approach. Eng Proc 78(1):6
14. John-Otumu AM, Ekemonye RO, Ewunonu TC, Aniugo VO, Okonkwo OT (2024) Optimizing CNN Kernel sizes for enhanced melanoma lesion classification in dermoscopy images. Mach Learn Res 9(2):26–38
15. Wu Y, Chen B, Zeng A, Pan D, Wang R, Zhao S (2022) Skin cancer classification with deep learning: a systematic review. Front Oncol. 12. https://doi.org/10.3389/fonc.2022.893972.
16. Diab AG, Fayez N, El-Seddek MM (2022) Accurate skin cancer diagnosis based on convolutional neural networks. Indones J Electr Eng Comput Sci 25(3):1429–1441
17. Stafford H et al (2023) Non-melanoma skin cancer detection in the age of advanced technology: a review. Cancers (Basel) 15(12):3094
18. Patil H (2024) Frontier machine learning techniques for melanoma skin cancer identification and categorization: a thorough review. Oral Oncol Rep 100217
19 Han M, Zhao S, Yin H, Hu G, Ghadimi N (2024) Timely detection of skin cancer: an AI-based approach on the basis of the integration of Echo State Network and adapted seasons optimization algorithm. Biomed Signal Process Control 94:106324. https://doi.org/10.1016/j.bspc.2024.106324
20. Malibari AA et al (2022) Optimal deep neural network-driven computer aided diagnosis model for skin cancer. Comput Electr Eng 103:108318
21. Saeed M, Naseer A, Masood H, Rehman SU, Gruhn V (2023) The power of generative ai to augment for enhanced skin cancer classification: a deep learning approach. IEEE Access 11:130330–130344
22. Supriyanto C, Salam A, Zeniarja J, Wijaya A (2023) Two-stage input-space image augmentation and interpretable technique for accurate and explainable skin cancer diagnosis. Computation 11(12):246
23. Majtner T, Yildirim-Yayilgan S, Hardeberg JY (2019) Optimised deep learning features for improved melanoma detection. Multimed Tools Appl 78(9):11883–11903. https://doi.org/10.1007/s11042-018-6734-6
24. Teodoro AAM et al (2023) A skin cancer classification approach using gan and roi-based attention mechanism. J Signal Process Syst 95(2–3):211–224

25. El-Shafai W, El-Fattah IA, Taha TE (2023) Deep learning-based hair removal for improved diagnostics of skin diseases. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-16646-6
26. Ajmal M et al (2023) BF2SkNet: best deep learning features fusion-assisted framework for multiclass skin lesion classification. Neural Comput Appl 35(30):22115–22131. https://doi.org/10.1007/s00521-022-08084-6
27. Lembhe A, Motarwar P, Patil R, Elias S (2022) Enhancement in skin cancer detection using image super resolution and convolutional neural network. Procedia Comput Sci 218:164–173. https://doi.org/10.1016/j.procs.2022.12.412
28. Esteva A et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
29. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G (2018) Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE J Biomed Heal informatics 23(2):538–546