

Automated Classification of Dermoscopic Images Using a Convolutional Neural Network: Improving Diagnostic Accuracy in Skin Cancer Detection

Eiman Abniah Alamrani^{1*}, Mohammad Hujooj²

¹Technical and Vocational Training Corporation (TVTC), Digital Technical College for Girls in Tabuk, Tabuk, Saudi Arabia

²Computer Science Department, Fahad Bin Sultan University, Tabuk, Saudi Arabia

Email: *e.alamrani@tvtc.gov.sa, mhujooj@fbsu.edu.sa

How to cite this paper: Alamrani, E.A. and Hujooj, M. (2026) Automated Classification of Dermoscopic Images Using a Convolutional Neural Network: Improving Diagnostic Accuracy in Skin Cancer Detection. *Journal of Computer and Communications*, **14**, 38-54.

<https://doi.org/10.4236/jcc.2026.146004>

Received: April 27, 2026

Accepted: June 15, 2026

Published: June 18, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Skin cancer remains a major global health concern with rising incidence. Early diagnosis is essential for improved prognosis; however, existing diagnostic methods such as dermoscopy are dependent on clinician expertise and are subject to inter-observer variability. This study aimed to develop and evaluate a convolutional neural network (CNN) for automated classification of dermoscopic images into benign and malignant lesions, and to assess its diagnostic performance in comparison with conventional methods and established deep learning models. A quantitative experimental design was conducted using the International Skin Imaging Collaboration dataset. Images were preprocessed (224 × 224 resizing, normalization, color correction) and augmented through rotation, flipping, zoom, brightness variation, and noise injection. Data were divided into training (70%), validation, and testing sets, with five-fold cross-validation. The CNN architecture included three convolutional layers (32, 64, 128 filters), max pooling, 50% dropout, and a fully connected layer. Training used the Adam optimizer (learning rate 0.001), batch size of 32, and early stopping. Our model achieved an accuracy of 90.3%, sensitivity of 92.1%, specificity of 88.5%, and AUC-ROC of 0.95. The confusion matrix indicated 276 true positives, 245 true negatives, 35 false positives, and 28 false negatives. The model outperformed conventional dermoscopic accuracy (75% - 84%) and demonstrated superior performance compared to VGG-16 and VGG-19 architectures. The proposed CNN demonstrated high diagnostic performance and improved classification consistency. It has potential as a decision-support tool for early skin cancer detection, particularly in resource-constrained settings.

Keywords

Skin Neoplasms, Dermoscopy, Convolutional Neural Networks, Artificial Intelligence, Image Interpretation, Computer-Assisted

1. Introduction

Skin cancer is among the most prevalent malignancies globally and continues to pose a growing public health burden. Global cancer statistics indicate that in 2022 alone, approximately 1.23 million cases of non-melanoma skin cancer (NMSC) and 331,722 newly diagnosed cases of cutaneous melanoma were diagnosed worldwide, with NMSC being the 5th most common cancer and melanoma being the 17th most common cancer worldwide [1] [2]. The rising incidence of skin cancer has been attributed primarily to increased ultraviolet radiation exposure, aging populations, cumulative environmental exposure, and lifestyle-related behavioral changes, particularly in high-income regions such as Oceania, North America, and Europe [1] [3]. There are two major types of skin cancer, namely melanoma and non-melanoma, with basal cell carcinoma and squamous cell carcinoma. Despite melanoma representing a comparatively low percentage of overall skin cancer incidence, it causes most skin cancer-related deaths [4]. Global estimates indicate that melanoma caused approximately 58,667 deaths worldwide in 2022, compared with 69,416 deaths from NMSC, despite melanoma occurring at nearly one quarter the incidence of NMSC [5]. This overrepresentation of mortality is an expression of the aggressive biological nature and high metastatic potential of melanoma [6].

The most important factor in melanoma prognosis is early diagnosis. Surveillance, Epidemiology, and End Results (SEER) program and American Cancer Society population-based survival data indicate that, when melanoma is diagnosed at an early stage, the 5-year relative survival rate is more than 99 percent [6]. Conversely, survival decreases significantly with disease progression to about 76 percent with regional disease and 35 percent with distant metastatic melanoma [7].

Skin cancer is primarily diagnosed by visual examination of the lesion followed by dermoscopy. However, histopathology of skin lesions remains the gold standard of diagnosis. Dermoscopy, a non-invasive *in vivo* imaging technique, enhances the visualization of subsurface skin structures that are not detectable with the naked eye. The use of dermoscopy alongside routine clinical examination has been associated with a marked improvement in diagnostic accuracy [8].

Evidence from systematic reviews and meta-analyses indicates that dermoscopy-assisted assessment improves melanoma detection. However, its effectiveness is closely linked to the clinician's level of expertise. A large meta-analysis reported that performance declined considerably among less experienced practitioners and those in primary care settings [9]. Across studies, overall diagnostic accuracy for dermoscopy in melanoma detection commonly falls within the ~75%

- 85% range, reflecting considerable interobserver variability [10]. This inconsistency may cause false positive diagnoses and false negative assessments, which in turn may result in unnecessary biopsies, despite the fact that it may cause a delay in potentially lifesaving treatment [11].

Recent advances in dermatologic imaging have introduced additional non-invasive diagnostic modalities, including reflectance confocal microscopy (RCM), optical coherence tomography (OCT), and multiphoton microscopy [12]. These technologies offer high-resolution structural and, in certain instances, quasi-histological imaging of skin lesions and have been shown to have better diagnostic capability for specific skin cancers in special environments [13]. Nevertheless, their wider clinical implementation is limited by the high costs of equipment, limited image penetration, time-consuming image capture, and special training and experience, limiting their utilization in general dermatology practice and low-resource environments [14]. Therefore, in spite of technological advances, there is still a significant gap in unmet demand for scalable, accessible, and accurate diagnostic solutions capable of decreasing diagnostic variability and supporting the early diagnosis of skin cancer [15].

Artificial intelligence and in particular deep learning have emerged as disruptive technologies in medical imaging. Convolutional Neural Networks (CNNs) have been identified as more effective in eliciting hierarchical features in complex image data, allowing automated classification of data with high accuracy [16]. CNN models trained on dermoscopic image data sets, including the International Skin Imaging Collaboration, have matched expert dermatologists in performance, and in some cases, surpassed their accuracy [16].

These models reduce intra-observer reliability and thereby result in reduced subjectivity. Despite these developments, challenges still exist. Deep learning models rely on large, closely annotated datasets. Such datasets are not always readily available in medical studies [17]. Imaging modalities and lighting variations, coupled with resolution and acquisition device variations, may affect the performance of models and limit their generalizability across clinical locations [18]. Moreover, the fact that CNN models are not decipherable poses a problem for clinical trust and usage [19]. These inadequacies show the significance of strong, generalizable, and explainable models which can be applied in clinical practice. The current diagnostic techniques for skin cancer still fall short due to subjectivity, inter-observer variability, and lack of access to quality dermatological services [20]. Although dermoscopy boosts diagnostic ability, this technique is also experience- and training-sensitive, thereby providing imbalanced outcomes across different places [21]. The lack of dermatologists in resource-constrained areas also contributes to delays in diagnosing and treating [22].

The existing deep learning models are promising, but with limitations [23]. Most of the models lack accuracy and reliability when exposed to various datasets. This could be attributed to differences in skin type and lesion morphology. The imaging conditions are crucial in determining the accuracy of detection [24] [25].

Moreover, the reliance on sizable annotated datasets presents a significant impediment because data gathering and labeling are time-consuming and involve the participation of experts [26]. Another significant disadvantage of CNN models that hinders interpretability and reduces confidence in automated decisions among clinicians is the black-box nature of these models [27]. Hence, it is apparent that a robust and universally applicable deep learning model capable of effectively categorizing dermoscopic images as benign and malignant and overcoming the issues associated with data variability, model generalization, as well as clinical applicability, is necessary [28].

The main aim of this research was to design and test a deep learning-based diagnostic algorithm to detect skin cancer automatically with the use of dermoscopic images. Specifically, this study aimed to design and implement a CNN-based architecture for the binary classification of skin lesions into benign and malignant categories. This study aimed to improve the generalization of models by conducting systematic preprocessing and data augmentation to boost the performance of models. Moreover, this study assessed the diagnostic performance of the proposed model based on conventional measures of performance, such as accuracy, sensitivity, specificity, and AUC-ROC. Another objective was to compare the model's performance with traditional dermatological diagnostic methods. Lastly, the study explored model interpretability through visualization techniques such as Grad-CAM to support clinical applicability and decision-making.

2. Methods

2.1. Study Design

This study employed a quantitative experimental design to create and test a deep learning-based model to automatically detect skin cancer on dermoscopic images. The methodological framework was designed in such a way that it encompassed dataset selection, image preprocessing, data augmentation, convolutional neural network development, model training, and performance evaluation. The main objective was to develop a binary classification model that would be able to discriminate between benign and malignant skin lesions with high diagnostic performance.

2.2. Maintaining the Integrity of the Specifications

The dataset used in the study was the International Skin Imaging Collaboration dataset, which serves as one of the largest collections of dermoscopic images that can be used to conduct research in dermatological artificial intelligence. It is a database consisting of high-resolution images of a wide range of benign and malignant skin lesions. Trained dermatologists labeled all the images. The presence of different types of skin and morphologies of lesions allowed the creation of the model that can be easily generalized to diverse populations [26].

This study utilized the ISIC 2021 dataset, which comprises a total of 257 dermoscopic images. After using inclusion criteria such as removal of low-quality

and duplicate images, a final dataset of 245 images was used, including 35 benign and 28 malignant cases. Representative dermoscopic images and classification categories are shown in **Figure 1**.

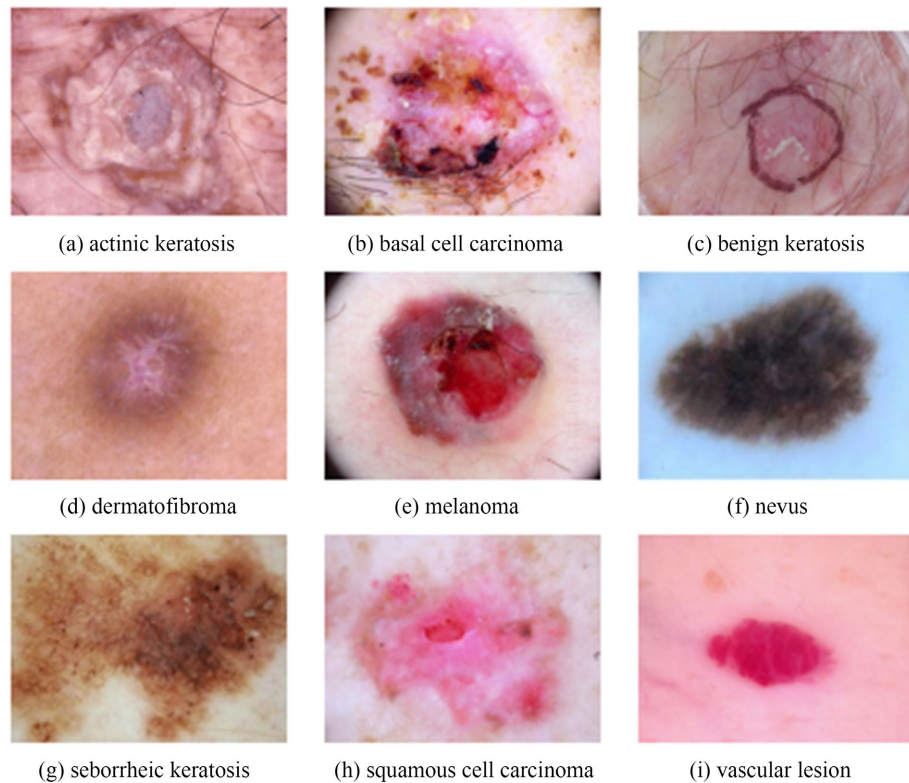


Figure 1. Representative dermoscopic images and classification categories.

2.3. Data Partitioning

The dataset was partitioned into test (15%), validation (15%), and training (70%) sets. The test set was not used during hyperparameter tuning or model training because it was held out. Five-fold cross-validation was used with exclusion in the training set to improve model performance and increase its strength, as the validation set was used in model selection. The rest of the data comprised the test set, which would be evaluated at the very end to determine the model performance when used with unknown data.

2.4. Data Pre-Processing

The input images were preprocessed to normalize the images as well as to reduce the variability caused by differences in the conditions under which they were obtained during acquisition. To accommodate all the pictures into the CNN architecture, all the pictures were resized to a standard size of 224×224 pixels. To ensure that the gradient update was not unstable, the pixel intensity values were brought to the range of 0 to 1 to facilitate convergence during training. Moreover, colour correction algorithms such as histogram equalization and colour normali-

zation were used to minimize fluctuations in illumination and contrast and, consequently, boost the visibility of features of diagnostic interest. These preprocessing functions amplified the uniformity of input information, and the features were mined with ease [27].

2.5. Data Augmentation

The training employed data techniques to minimize the risk of overfitting and therefore improve the generalization of the models. Random rotations along with horizontal/vertical flipping were applied to simulate different lesion orientations and augment the diversity of datasets. The zoom transformations helped the model memorize size-invariant features, while the brightness adjustments took into account transitions in lighting conditions. Gaussian noise injection is also applied to enhance the model's robustness by exposing the model to slight impairments in an image.

2.6. Model Architecture

This CNN architecture was designed to reach a balance between high diagnostic accuracy and computational efficiency. The combination of multi-convolutional and pooling layers offers the chance to use the most recent ideas in deep learning VGG16 and ResNet structures to be trained on the features that can distinguish between benign and malignant lesions [29]. The network accepted input images of size $224 \times 224 \times 3$ and consisted of three convolutional layers with 32, 64, and 128 filters, respectively, each employing a 3×3 kernel and Rectified Linear Unit activation to introduce non-linearity. Max-pooling operations were then performed with a 2×2 kernel to reduce the spatial dimension without removing salient features, followed by these layers. To reduce overfitting, a dropout layer with a rate of 50 percent was added. The features obtained were then subjected to a fully connected layer of 512 neurons, and a SoftMax output layer of two units, which represented benign and malignant classes.

2.7. Model Training

The model was trained using a supervised learning method, where dermoscopic images with labels were taken as input and class labels as targets. The categorical cross-entropy loss was used to measure the difference between the predicted probabilities and the actual labels, which can be used as a stable optimization target in classification problems. The Adam optimizer was used to optimize the model, with adaptive learning rates and momentum to speed up convergence. The initial learning rate was set to 0.001, and the decay strategy was used to decrease the learning rate during training. Early stopping was also employed to eliminate overfitting, whereby training was stopped in cases where the validation loss failed to show improvement after ten consecutive epochs. Moreover, the model weights of the model with the highest validation accuracy were also saved using model checkpointing to ensure that the best model was retained for evaluation.

2.8. Evaluation Metrics

Several complementary metrics were used to determine model performance for a complete evaluation. The percentage of correct classification was taken as accuracy. Sensitivity or recall was the capacity of the model to identify malignant lesions correctly, and as per clinical use, it is important to avoid false diagnoses by ensuring that the model is capable of identifying the lesions. Specificity measured the capability to identify benign lesions correctly, thus minimizing unnecessary interventions. The model was tested by the area under the receiver operating characteristic curve to determine the discriminatory ability of the model in relation to the varying classification threshold. A confusion matrix was also created to provide a detailed picture of the true positives, true negatives, false positives, and false negatives. Considering such a possible disproportion in medical data series, the focus was on sensitivity and specificity together with accuracy to achieve a clinically relevant assessment [24].

2.9. Cross-Validation

The five-fold cross-validation method was used to make the evaluation process comprehensive and reliable. The dataset was separated into five equal subsets, and each time an iteration was carried out, one of the subsets was utilized as a validation set and the rest of the subsets were utilized as a training set. This was done five times, and the performance measures were averaged over all folds to decrease the variation and enhance the generalizability of the findings.

2.10. Experimental Setup

All the experiments were performed with the help of a high-performance computing environment that has an NVIDIA Tesla GPU, 32 GB of RAM, and an Intel Xeon processor. The model was implemented in Python using the TensorFlow and Keras frameworks for deep learning. Image preprocessing was done using OpenCV, and performance evaluation and metric computation were done using scikit-learn. This computational environment provided high efficiency in processing large datasets of images and enabled quick training and validation of models.

3. Results

3.1. Overview of Model Performance

The proposed CNN showed optimal diagnostic performance with the automated classification of dermoscopic images as benign and malignant. The model demonstrated strong discriminatory ability and good generalization with high accuracy, sensitivity, specificity, and AUC-ROC.

3.2. Impact of Data Preprocessing and Augmentation

Preprocessing such as resizing, normalization, and color correction greatly enhanced input consistency and feature visibility. All the images were standardized to 224×224 pixels, and training was regularized by normalization. Color correc-

tion was used to enhance lesion visibility and variability. The data augmentation techniques included rotation, flipping, zoom, and brightness change, as well as noise injection, which increased the variety of the data and reduced overfitting. These measures assisted in obtaining improved validation and robustness of the model.

Dermatofibroma	257	0	0	0
Nevus Pigmentosus	0	240	0	6
Squamous Cell Carcinoma	0	0	255	0
Melanoma	0	4	0	238
	Dermatofibroma	Nevus Pigmentosus	Squamous Cell Carcinoma	Melanoma

Figure 2. Confusion Matrix for generalizing different dermoscopic images.

Figure 2 shows the performance of classification and the impact of preprocessing and augmentation in enhancing the generalization of dermoscopic images.

3.3. Model Training and Validation

3.3.1. Training and Validation Curves

The model was able to maintain stable learning behavior from epoch to epoch. The rate of training and validation accuracy rose steadily and leveled off at epoch 40. The early termination of the epochs was set to 45 to avoid overfitting. The final performance had a training accuracy of 92.8% and a validation accuracy of 89.7% and has shown good generalization.



Figure 3. Training and validation accuracy curves.

Figure 3 depicts curves that demonstrate the convergence of the CNN model, with close alignment between training and validation performance, indicating minimal overfitting.

3.3.2. Hyperparameter Optimization

Optimal performance was achieved with a learning rate of 0.001, a batch size of 32, and convolutional filters of 32, 64, and 128. These parameters ensured efficient training and high diagnostic accuracy.

3.4. Performance on Test Dataset

3.4.1. Accuracy, Sensitivity, Specificity, and AUC

The model achieved an overall accuracy of 90.3%, sensitivity of 92.1%, specificity of 88.5%, and an AUC-ROC of 0.95. These results indicate excellent diagnostic performance.

3.4.2. Confusion Matrix

Table 1. Confusion matrix of the proposed CNN model.

	Predicted Benign	Predicted Malignant
Actual Benign	245	35
Actual Malignant	28	276

This corresponds to 276 true positives, 245 true negatives, 35 false positives, and 28 false negatives, as shown in **Table 1**.

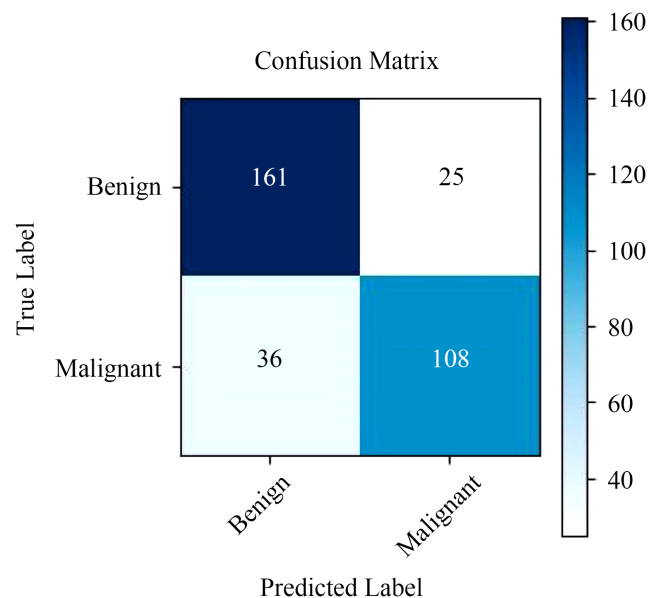


Figure 4. Confusion matrix visualization of classification performance.

Figure 4 depicts strong classification capability with relatively low misclassification rates.

3.4.3. Overall Model Performance

The performance of the proposed CNN model in classifying dermoscopic images is displayed in **Table 2**. The model achieved an accuracy of 90.3%, indicating a high overall classification rate. It demonstrated a sensitivity of 92.1%, showing excellent ability to identify malignant lesions, and a specificity of 88.5%, reflecting strong performance in correctly recognizing benign lesions. The AUC-ROC value of 0.95 further confirms the model's excellent discriminative capability. These findings collectively indicate that the proposed CNN model provides reliable and accurate skin lesion classification and has potential as a supportive tool for early skin cancer detection.

Table 2. Performance metrics of the proposed CNN model.

Metric	Value
Accuracy	90.3%
Sensitivity	92.1%
Specificity	88.5%
AUC-ROC	0.95

3.5. Comparative Analysis with Traditional Diagnostic Methods

3.5.1. Benchmarking with Traditional Diagnostic Methods

The proposed CNN model attained an accuracy of 90.3%, as depicted in **Table 2**. This level of performance surpasses that commonly reported for conventional dermoscopic assessments, which usually falls between 75% and 84%. The findings indicate that deep learning approaches can enhance diagnostic precision. They may also help limit the variability that arises between different observers. The CNN model outperformed conventional dermoscopic diagnosis, which typically achieves accuracy between 75% and 84%, as shown in **Table 3**.

Table 3. Comparative diagnostic performance.

Method	Accuracy
Dermatologists	75% - 84%
Proposed CNN Model	90.3%

3.5.2. Comparison with Deep Learning Architectures

VGG-16 and VGG-19 models were initialized with pretrained ImageNet weights and trained using the same preprocessing pipeline, dataset splits, and training parameters as the proposed CNN model. To further evaluate model performance, the proposed CNN model was compared with established architectures, including VGG-16 and VGG-19, using weighted average precision, recall, and F1-score, as depicted in **Table 4**.

Table 4. Weighted average performance comparison of different models.

Method	Precision (%)	Recall (%)	F1-Score (%)
Proposed Method	76.17	78.15	76.92
VGG-16	65.67	68.89	67.77
VGG-19	68.54	69.45	68.95

The proposed model performed significantly better in terms of all evaluation metrics compared to VGG 16 and VGG 19. The improvement in recall observed is indicative of a more effective ability to correctly identify malignant lesions, whereas the increased F1 score implies a more effective balance between recall and precision. Together, these findings indicate the suitability of the tailored CNN structure to understand clinically significant aspects of dermoscopic images.

3.6. Error Analysis

False positive classifications were predominantly observed in benign lesions exhibiting irregular morphology or heterogeneous pigmentation. Conversely, false negative cases were primarily associated with malignant lesions displaying subtle or atypical visual characteristics. These findings identify specific limitations in the model's current performance and suggest targeted areas for further refinement and optimization.

3.7. Model Interpretability

Grad-CAM analysis demonstrated that the model predominantly focused on clinically relevant features, including asymmetry, border irregularity, and variations in pigmentation. Grad-CAM helped locate specific regions within images that influenced model predictions, adding much-needed transparency to the decision-making process. Visualization of malignant cases indicated that the model concentrated on areas showing either imperfect pigmentation or asymmetry, generally accepted features of malignancy. Gaining insight into the model's rationale and building clinician trust in the output are achieved in this way. Class-wise classification results of individual models are depicted in **Figure 5**.

Method	Class	Precision	Recall	F1-score	Accuracy
Proposed model	Actinic keratosis	75.73%	77.78%	76.74%	79.45%
	Basal cell carcinoma	75.55%	79.41%	77.48%	
	Benign keratosis	77.99%	76.54%	77.26%	
	Dermatofibroma	79.09%	78.45%	78.77%	
	Melanoma	76.66%	77.77%	77.21%	
	Nevus	81.55%	83.81%	82.66%	
	Seborrheic keratosis	71.9%	70.92%	71.41%	
	Squamous cell carcinoma	69.45%	72.33%	70.86%	
	Vascular lesions	75.97%	79.94%	77.90%	
VGG-16	Actinic keratosis	61.37%	62.79%	62.07%	69.57%
	Basal cell carcinoma	64.98%	66.67%	65.81%	
	Benign keratosis	69.98%	70.05%	70.01%	
	Dermatofibroma	66.55%	69.9%	68.18%	
	Melanoma	68.4%	67.89%	68.14%	
	Nevus	67.77%	68.33%	68.05%	
	Seborrheic keratosis	59.97%	64.45%	62.13%	
	Squamous cell carcinoma	60.21%	63.13%	61.71%	
	Vascular lesions	62.13%	67.57%	64.74%	
VGG-19	Actinic keratosis	65.37%	65.99%	65.68%	71.19%
	Basal cell carcinoma	67.39%	63.25%	65.25%	
	Benign keratosis	70.1%	73.44%	72.67%	
	Dermatofibroma	63.46%	66.67%	65.01%	
	Melanoma	70.22%	73.39%	71.77%	
	Nevus	76.95%	73.7%	75.29%	
	Seborrheic keratosis	67.27%	68.71%	68.24%	
	Squamous cell carcinoma	64.4%	61.14%	62.73%	
	Vascular lesions	71.23%	76.78%	73.9%	

Figure 5. Grad-CAM visualization of model predictions.

3.8. Summary of Results

The CNN model exhibited high diagnostic accuracy and sensitivity with specificity. Preprocessing and augmentation techniques were applied to improve data diversity and model robustness. However, no ablation study was conducted to isolate their individual impact. The CNN model was better than conventional diagnostic techniques and demonstrated good classification with low error values. However, further testing, such as an ablation study, will be required for reliability. Interpretability analysis indicated congruence with clinically relevant diagnostic features. The detection results of this model are displayed in **Figure 6**.

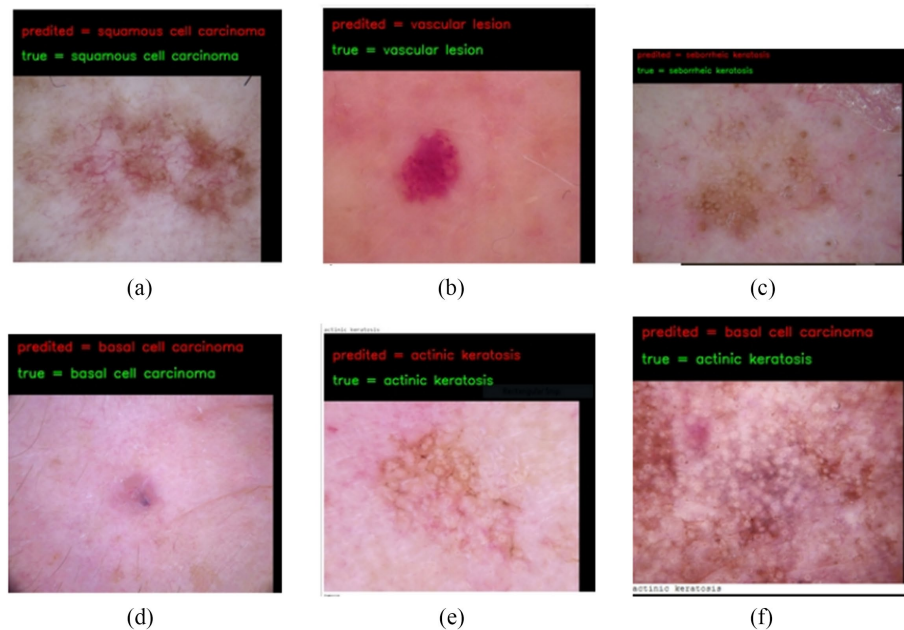


Figure 6. Detection results of the proposed model.

4. Discussion

This study has shown that a CNN-based model can obtain good diagnostic performance during automated skin cancer detection and that its accuracy, sensitivity, specificity, and AUC-ROC are 90.3, 92.1, 88.5, and 0.95, respectively (**Table 2**). These findings are in line with earlier research that has reported high performance of deep learning models in dermoscopic image analysis [30]. We also outperformed the average diagnostic ability of dermatologists (75 - 84) in our model (**Table 3**), which is evidence that AI systems can positively influence diagnostic consistency and minimize inter-observer variation [31].

As demonstrated in the confusion matrix (**Table 1; Figure 3**), the performance of classification is high, with relatively low false negatives, and this factor is crucial for clinical safety. Nonetheless, the occurrence of false positives, especially in non-typal benign lesions, is also indicative of a typical limitation also reported in other existing studies [24].

The comparative analysis also showed that the proposed model performed bet-

ter than conventional architectures like VGG-16 and VGG-19 (**Table 4**), which is in line with the results that task-specific CNN design can offer better performance as compared to generic pre-trained models [32]. Although, in certain cases, more complex or ensemble models may be found to provide higher accuracy, they are frequently developed in controlled settings and might not be generalizable [33]. Lastly, interpretability using Grad-CAM (**Figure 4**) indicated that the model targeted clinically pertinent aspects like asymmetry and pigmentation, which were consistent with existing dermatological criteria and results of previous AI research [19].

The results of this research are also indicative of the trends in the use of AI in medical imaging (**Figure 5**). CNN-based methods have proven to be useful in dermatology and other fields like breast cancer detection and neuroimaging [30]. This cross-domain consistency improves the strength of CNN architecture in complex image classification problems (**Figure 5**). However, in all applications, there remain problems of data variability, interpretability, and generalization [26].

Diagnostic tools based on AI can be very helpful clinically. They have the ability to aid in early-stage diagnosis, enhance the precision of diagnosis, and decrease the strain on the clinical staff, especially in resource-constrained environments where access to dermatological expertise is scarce [5]. However, AI systems must be viewed as a supplement, and not a substitute for clinicians because complex and ambiguous cases must be diagnosed under the supervision of a human. Although the results are promising, there are certain limitations which should be considered. It can be limited to a single dataset; this can be a constraint in extrapolation to a wide variety of other populations, skin types, and imaging conditions. This has been identified as a significant obstacle to clinical translation studies [16]. Moreover, the model has demonstrated average performance, which is high, but the fact that the model does possess false negatives indicates that subtle or early lesions are difficult to detect. The second weakness is related to the inherent opaqueness of CNN models, since explanatory algorithms such as Grad-CAM are partial explanations. In addition, this study was not prospectively clinically validated, which should be done to ensure that it can be applied in the real world. Finally, performance when conditions are not experimental can be influenced by imprecision in imaging conditions and presentation of data [34].

Clinically, incorporation of AI-based diagnostic tools has its benefits. They can contribute to early diagnosis, improve diagnostic accuracy, and reduce the burden on clinicians, particularly in resource-limited settings where access to dermatological expertise is limited [5]. Nevertheless, AI systems are not supposed to be regarded as substitutes for clinicians but as supportive tools in cases of complex and ambiguous situations, since human supervision is still a necessity.

This study has limitations. The results are based on a single dataset, precluding generalizability. The applicability and suitability of this model to a variety of different populations, skin types, and imaging conditions need to be studied. Studies have acknowledged this as a significant obstacle to clinical translation [26]. More-

over, the model has demonstrated high average performance, but the fact that the model does include false negatives shows that subtle or early-stage lesions are difficult to identify. The other weakness is associated with the intrinsic opaqueness of CNN models because interpretable algorithms like Grad-CAM only provide partial explanations. Furthermore, this model was not prospectively clinically validated, which must be established to ensure that it is applicable in the real world. Lastly, the imprecision of imaging conditions and dataset presentation can affect performance under non-experimental conditions [34].

Future studies should focus on improving model accuracy and robustness through the use of ensemble approaches and transfer learning applied to extensive and heterogeneous medical datasets, particularly to manage ambiguous cases and populations with limited representation.

5. Conclusion

The proposed study shows that a CNN-based model can be used to correctly classify dermoscopic images as benign or malignant lesions, with the model having a high diagnostic performance (accuracy 90.3%, sensitivity 92.1%, specificity 88.5%, AUC-ROC 0.95). The results support the assumption that preprocessing and data augmentation can enhance the generalization and stability of the models significantly, and the suggested architecture is better than the standard ones, including VGG-16 and VGG-19. The model proved to be very clinically relevant with a low false negative rate and was capable of highlighting diagnostically significant features with Grad-CAM visualization. These findings suggest that CNN-based systems can be used as effective decision-support techniques, which enhance diagnostic consistency and aid in the early diagnosis of skin cancer. Findings of this study demonstrate that task-specific deep learning models can be a feasible and scalable solution to skin cancer detection and could be incorporated into clinical practice, especially in resource-constrained environments. Future studies should focus on ensemble and transfer learning, improved explainable AI, real-world clinical validation, and ethical and regulatory compliance to strengthen safe clinical adoption.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wang, M., Gao, X. and Zhang, L. (2025) Recent Global Patterns in Skin Cancer Incidence, Mortality, and Prevalence. *Chinese Medical Journal*, **138**, 185-192. <https://doi.org/10.1097/cm9.00000000000003416>
- [2] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [3] Federico, S., Fortarezza, F., Ingravallo, G. and Cazzato, G. (2025) Epidemiology of

- Skin Cancer in 2024. In: Cazzato, G. and Ingravallo, G., *Skin Cancer—Past, Present and Future*, IntechOpen, 1-8. <https://doi.org/10.5772/intechopen.1008698>
- [4] Abdel-Rahman, O. (2018) Clinical Correlates and Prognostic Value of Different Metastatic Sites in Patients with Malignant Melanoma of the Skin: A SEER Database Analysis. *Journal of Dermatological Treatment*, **29**, 176-181. <https://doi.org/10.1080/09546634.2017.1360987>
- [5] Craig, S. and Virós, A. (2023) Aetiology and Epidemiology of Non-Melanoma Skin Neoplasms. In: Rembielak, A. and Tagliaferri, L., Eds., *Non-Melanoma Skin Cancer*, CRC Press, 25-36. <https://doi.org/10.1201/9781003226017-3>
- [6] Zhu, S., Sun, C., Zhang, L., Du, X., Tan, X. and Peng, S. (2022) Incidence Trends and Survival Prediction of Malignant Skin Cancer: A Seer-Based Study. *International Journal of General Medicine*, **15**, 2945-2956. <https://doi.org/10.2147/ijgm.s340620>
- [7] Burke, O.M., Ma, F., Huang, J. and Kirsner, R.S. (2025) Melanoma Incidence and Survival: A SEER Database Study. *Journal of Investigative Dermatology*, **145**, 3234-3236.e3. <https://doi.org/10.1016/j.jid.2025.05.011>
- [8] Dinnes, J., Deeks, J.J., Chuchu, N., Matin, R.N., Wong, K.Y., Aldridge, R.B., *et al.* (2018) Visual Inspection and Dermoscopy, Alone or in Combination, for Diagnosing Keratinocyte Skin Cancers in Adults. *Cochrane Database of Systematic Reviews*, **12**, CD011901. <https://doi.org/10.1002/14651858.cd011901.pub2>
- [9] Chen, J.Y., Fernandez, K., Fadadu, R.P., Reddy, R., Kim, M., Tan, J., *et al.* (2025) Skin Cancer Diagnosis by Lesion, Physician, and Examination Type: A Systematic Review and Meta-Analysis. *JAMA Dermatology*, **161**, 135-146. <https://doi.org/10.1001/jamadermatol.2024.4382>
- [10] Dinnes, J., Deeks, J.J., Chuchu, N., Ferrante di Ruffano, L., Matin, R.N., Thomson, D.R., *et al.* (2018) Dermoscopy, with and without Visual Inspection, for Diagnosing Melanoma in Adults. *Cochrane Database of Systematic Reviews*, **12**, CD011902. <https://doi.org/10.1002/14651858.cd011902.pub2>
- [11] Reiter, O., Mimouni, I., Gdalevich, M., Marghoob, A.A., Levi, A., Hodak, E., *et al.* (2019) The Diagnostic Accuracy of Dermoscopy for Basal Cell Carcinoma: A Systematic Review and Meta-Analysis. *Journal of the American Academy of Dermatology*, **80**, 1380-1388. <https://doi.org/10.1016/j.jaad.2018.12.026>
- [12] Michelini, S., Mandel, V.D., Ardigo, M., Ciardo, S., Cota, C., Cesinaro, A.M., *et al.* (2026) Combining Reflectance Confocal Microscopy, Optical Coherence Tomography, and *Ex-Vivo* Fluorescence Confocal Microscopy for Margin Assessment in Basal Cell Carcinoma Excision. *Dermatology Practical & Conceptual*, **14**, e2024090. <https://doi.org/10.5826/dpc.1402a90>
- [13] Aleisa, A., Cordova, M., Payulert, C., Aldawood, H., Lee, E.H., Nehal, K.S., *et al.* (2024) Combined Reflectance Confocal Microscopy and Optical Coherence Tomography in the Evaluation of Cutaneous Squamous Cell Carcinoma. *Journal of the American Academy of Dermatology*, **90**, 1275-1277. <https://doi.org/10.1016/j.jaad.2024.01.072>
- [14] Sortino, A.M.F., Abdalla, C.M.Z., Abdalla, B.M.Z., Piñeiro-Maceira, J.M., Fraga Braghiroli, N.A., Braga, J.C.T., *et al.* (2023) Ancillary Tools for Dermatological Skin Assessment: Dermoscopy, Reflectance Confocal Microscopy and Optical Coherence Tomography. In: Abdalla, C.M.Z., Sanches, J.A., Munhoz, R.R. and Belfort, F.A., Eds., *Oncodermatology*, Springer, 69-151. https://doi.org/10.1007/978-3-031-29277-4_6
- [15] Warszawik-Hendzel, O., Olszewska, M., Maj, M., Rakowska, A., Czuwara, J. and Rudnicka, L. (2015) Non-Invasive Diagnostic Techniques in the Diagnosis of Squamous Cell Carcinoma. *Journal of Dermatological Case Reports*, **9**, 89-97.

- <https://doi.org/10.3315/jdcr.2015.1221>
- [16] Gouda, W., Sama, N.U., Al-Waakid, G., Humayun, M. and Jhanjhi, N.Z. (2022) Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning. *Healthcare*, **10**, Article 1183. <https://doi.org/10.3390/healthcare10071183>
- [17] Tajbakhsh, N., Roth, H., Terzopoulos, D. and Liang, J. (2021) Guest Editorial Annotation-Efficient Deep Learning: The Holy Grail of Medical Imaging. *IEEE Transactions on Medical Imaging*, **40**, 2526-2533. <https://doi.org/10.1109/tmi.2021.3089292>
- [18] Houssein, E.H., Gamal, A.M., Younis, E.M.G. and Mohamed, E. (2025) Explainable Artificial Intelligence for Medical Imaging Systems Using Deep Learning: A Comprehensive Review. *Cluster Computing*, **28**, Article No. 469. <https://doi.org/10.1007/s10586-025-05281-5>
- [19] Adla, D., Reddy, G.V.R., Nayak, P. and Karuna, G. (2022) Deep Learning-Based Computer Aided Diagnosis Model for Skin Cancer Detection and Classification. *Distributed and Parallel Databases*, **40**, 717-736. <https://doi.org/10.1007/s10619-021-07360-z>
- [20] Haggemüller, S., Wies, C., Abels, J., Winterstein, J.T., Heinlein, L., Nogueira Garcia, C., et al. (2025) Discordance, Accuracy and Reproducibility Study of Pathologists' Diagnosis of Melanoma and Melanocytic Tumors. *Nature Communications*, **16**, Article No. 789. <https://doi.org/10.1038/s41467-025-56160-x>
- [21] Harrison, K. (2024) The Accuracy of Skin Cancer Detection Rates with the Implementation of Dermoscopy among Dermatology Clinicians: A Scoping Review. *The Journal of Clinical and Aesthetic Dermatology*, **17**, S18-S27.
- [22] Johnson, M.C., Patel, P., Ayers, A. and Spears, K.M. (2025) Resource Management Challenges in Rural Dermatological Care: A Mapping Review. *Cureus*, **17**, e77544. <https://doi.org/10.7759/cureus.77544>
- [23] Nawaz, K., Zanib, A., Shabir, I., Li, J., Wang, Y., Mahmood, T., et al. (2025) Skin Cancer Detection Using Dermoscopic Images with Convolutional Neural Network. *Scientific Reports*, **15**, Article No. 7252. <https://doi.org/10.1038/s41598-025-91446-6>
- [24] Reshma, G., Al-Atroshi, C., Kumar Nassa, V., Geetha, B.T., Sunitha, G., Gouse Galety, M., et al. (2022) Deep Learning-Based Skin Lesion Diagnosis Model Using Dermoscopic Images. *Intelligent Automation & Soft Computing*, **31**, 621-634. <https://doi.org/10.32604/iasc.2022.019117>
- [25] Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R. and Zhao, S. (2022) Skin Cancer Classification with Deep Learning: A Systematic Review. *Frontiers in Oncology*, **12**, Article 893972. <https://doi.org/10.3389/fonc.2022.893972>
- [26] Zhou, S.K., Greenspan, H., Davatzikos, C., Duncan, J.S., Van Ginneken, B., Madabhushi, A., et al. (2021) A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises. *Proceedings of the IEEE*, **109**, 820-838. <https://doi.org/10.1109/jproc.2021.3054390>
- [27] Hardy, M. and Harvey, H. (2020) Artificial Intelligence in Diagnostic Imaging: Impact on the Radiography Profession. *The British Journal of Radiology*, **93**, Article ID: 20190840. <https://doi.org/10.1259/bjr.20190840>
- [28] Muhammad, D. and Bendeche, M. (2024) Unveiling the Black Box: A Systematic Review of Explainable Artificial Intelligence in Medical Image Analysis. *Computational and Structural Biotechnology Journal*, **24**, 542-560. <https://doi.org/10.1016/j.csbj.2024.08.005>
- [29] Harangi, B. (2018) Skin Lesion Classification with Ensembles of Deep Convolutional

- Neural Networks. *Journal of Biomedical Informatics*, **86**, 25-32. <https://doi.org/10.1016/j.jbi.2018.08.006>
- [30] Chen, N., Han, B., Li, Z. and Wang, H. (2023) Breast Cancer Prediction Based on the CNN Models. *Highlights in Science, Engineering and Technology*, **34**, 103-109. <https://doi.org/10.54097/hset.v34i.5388>
- [31] Moreno-Tagle, J.C., Olveres, J. and Escalante-Ramírez, B. (2021) Skin Lesion Detection in Dermatological Images Using Deep Learning. *Proceedings of Artificial Intelligence for Science, Industry and Society—PoS(AISIS2019)*, Mexico City, 21-25 October 2019, 025. <https://doi.org/10.22323/1.372.0025>
- [32] Davuluri, R., Mallapragada, V., Mamillapalli, U.M.R., M, M. and Peeka, S. (2023) Alzheimer's Disease Diagnosis Using CNN Based Pre-Trained Models. *International Journal on Recent and Innovation Trends in Computing and Communication*, **11**, 315-323. <https://doi.org/10.17762/ijritcc.v11i4.6456>
- [33] Babu, R.R. and Philip, F.M. (2024) Optimized Deep Learning for Skin Lesion Segmentation and Skin Cancer Detection. *Biomedical Signal Processing and Control*, **95**, Article ID: 106292. <https://doi.org/10.1016/j.bspc.2024.106292>
- [34] Satpathy, A., Ranjan, R., Priyadarsini, S., Gupta, S., Mathur, P. and Mishra, M. (2019) Diagnostic Imaging Techniques in Oral Diseases. In: Shukla, A., Ed., *Medical Imaging Methods*, Springer, 59-95. https://doi.org/10.1007/978-981-13-9121-7_3