ORIGINAL ARTICLE

Deep Learning-Based Binary Classification of Skin Lesions for Melanoma Detection Using Transfer Learning

Jawad Hussain Awan[1], Sadia Noreen[1*], Shazma Tahseen[2]

[1]*FEST, Iqra University Main Campus Karachi, Sindh, Pakistan*
[2]*Dr. A. H. S. Bukhari Postgraduate Centre of Information & Communication Technology, University of Sindh, Jamshoro*
Correspondence: *sadianoreen560@gmail.com*
doi: 10.22442/jlumhs.2026.01460

ABSTRACT

OBJECTIVE: To create and test a binary classification deep learning-based melanoma detector using dermoscopic skin lesions as inputs and classifying the lesions as either benign or malignant (melanoma).

METHODOLOGY: A pre-trained InceptionV3 architecture was used to create a convolutional neural network (CNN) model on the basis of transfer learning. This model was trained using a publicly accessible dataset of skin cancer images in the form of dermoscopic images that were classified into nevus, seborrheic keratosis and melanoma. The benign lesions (nevus and seborrheic keratosis) were combined as one category whereas melanoma was categorized as the malignant category. The images were down-sampled to 299 x 299 pixels and normalised before being trained. This data was separated into training, validation and testing sets. RMSProp was used to optimize the model on binary cross-entropy loss.

RESULTS: The trained model performed on the test dataset with good and balanced results. It had an overall accuracy of 84.67 and this showed good quality skin lesion classification. The F1 of 0.81 indicates a good compromise between recall and precision. In addition, the ROC AUC of 0.88 supports the fact that the model has a great capacity of differentiating malignant and benign lesions in different thresholds.

CONCLUSION: The paper shows that transfer learning and deep learning have the potential to be used to classify binary skin cancer successfully i-e: melanoma, especially in dermatological environments.

KEYWORDS: Skin, Cancer, Melanoma, Detection, CNN, Medical Image and Classification

INTRODUCTION

Skin cancer is described as the unregulated proliferation of malignant cells that form the epidermis which is the outer layer of skin[1]. It is mainly induced by DNA damage that is not repaired causing mutations that cause rapid cell multiplication and the formation of tumors. Worldwide, one out of eight hundred thousand new cases and more than two thousand one hundred deaths of skin cancer is recorded[2]. The major risk factors are the long-term exposure to bad ultraviolet (UV) rays of sunlight and the use of the UV tanning beds[3].

Skin cancer has a geographical variation. The highest prevalence is reported in regions with the northern latitudes as opposed to the southern regions, which may indicate a northsouth gradient[4]. The causes of this difference are the light pigmentation of the skin, the exposure to the elevation, and the presence of more occupational activities that involve the outdoor[5].

In modern clinical practice, skin lesions are mostly diagnosed by the unaided eye, with the help of dermoscopy[6]. Dermoscopy is a technique that is used to view skin lesions with a special instrument that has a magnifying lens and polarized light, whereby, more details are viewed in the lesions in terms of their structure[7]. Dermoscopic images are normally taken with the help of high-resolution digital single-lens reflex (DSLR) cameras or smartphone-based accessories. Various factors influence diagnostic decisions; they include history of the patient, ethnicity, lifestyle and exposure to sun[8].

The growing amount of large publicly available dermoscopic images datasets has made the field of AI-based skin lesion analysis advance rapidly. The computer-aided diagnostic (CAD) systems that are advocated with the help of AI are changing the area of medical imaging by enhancing the accuracy and consistency of the diagnosis[9]. Deep learning methods have been shown to be very effective in the detection of abnormalities in different medical imaging modalities such as breast cancer, brain tumors, lung cancer, and skin lesions[10].

Melanoma is a cancerous growth that starts as a result of melanocytes which are cells that produce melanin. These lesions are frequently similar to benign moles and may occur in areas of the body that are exposed to the sun and non-exposed areas[11]. Early diagnosis of melanoma dramatically increases the survival of the patient but the ability to differentiate benign and malign lesions by visual means is still difficult because of the slight morphological resemblance[12]. The variability of diagnostic results can be dependent on clinical experience, workload, and complexity of lesions.

The current developments in the field of deep learning, specifically convolutional neural networks (CNNs), have demonstrated a significant potential in the analysis of medical images[13]. CNNs are hierarchical visual features (edges, textures, shapes, etc.) that are naturally learned, which means that no feature engineering is required and the classification can be robust[14]. Transfer learning has become a powerful approach in medical imaging tasks with small labeled data to be able to use learned representations trained on large-scale datasets like ImageNet and fine-tune them to domain-specific data sets such as melanoma detection[15].

The proposed research presents an approach to binary classification of melanoma with the help of deep learning and dermoscopic images. An InceptionV3 architecture is used to identify malignant melanoma and benign skin lesions in pre-trained architecture. This research aims to generate a CNN-based skin lesion binary classification model, measure its performance based on clinically relevant metrics and determine its possible use as a decision-support instrument in the screening of melanoma. It now explicitly explains: (1) the clinical burden of melanoma, (2) diagnostic variability among dermatologists, (3) limitations in resource-constrained settings, and (4) the need for AI-assisted early detection systems using transfer learning.

METHODOLOGY

Dataset and Preprocessing

In this research, a publicly accessible dataset on skin cancer that was available in the Udacity Dermatologist AI repository was used. The dataset preprocessing, model development, training, and evaluation were conducted between March and September 2025. The computational work and analysis conducted at Iqra University Main Campus, Karachi, Sindh, Pakistan, and University of Sindh, Jamshoro, Pakistan. The data set is made up of dermoscopic images belonging to three clinical types, namely, nevus, seborrheic keratosis, and melanoma from the ISIC dataset (one benign and one malignant example) as shown in **Figure 1**. In this research, the pictures of nevus and seborrheic keratosis were classified as benign, and the pictures of melanoma were classified as malignant, which formed a binary classification task.

The data was split into training, validation and testing databases to make sure that performance is evaluated impartially. CSV metadata files which were generated programmatically were used to store image file paths and matching labels. The images have been resized to 299 x 299 pixels to fit the inputs of the InceptionV3 mode and normalized to between 0 and 1 of intensity range. In order to enhance training efficiency and consistency, the data pipeline on TensorFlow was used, with the addition of caching, shuffling, batching, and prefetching.

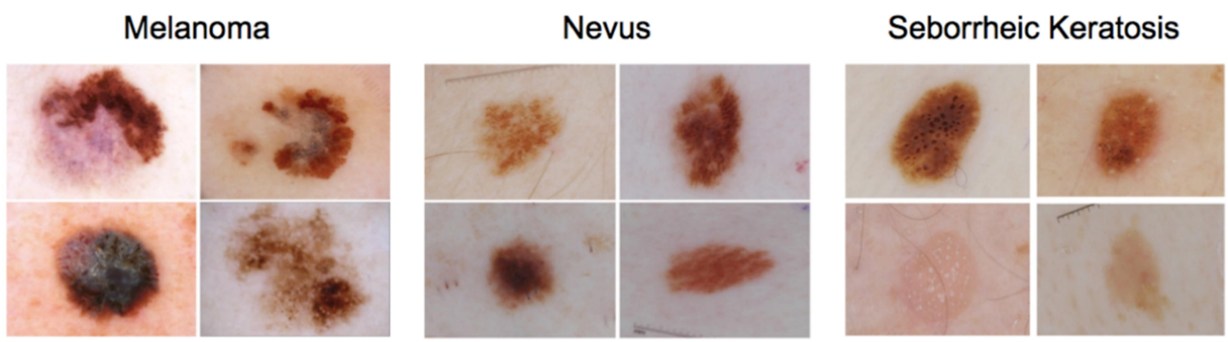

Figure 1: Three clinical types data of nevus, seborrheic keratosis, and melanoma

Table I: Dataset

Dataset Split	Number of Images
Training	1400
Validation	300
Testing	300
Total	2000

In **Table I**, The ISIC dataset was separated into training, validation and testing subsets. Model optimization was done with the training set, hyperparameter tuning and checkpointing were done with the validation set, and the test set was used to evaluate final performance.

Model Architecture

The implementation of a deep learning model based on transfer learning was through the InceptionV3 architecture that was pre-trained on the ImageNet dataset. Freezing all

convolutional layers to create pre-trained network made it a fixed feature extractor. The method allows the generalization of acquired generic visual representations and minimizes the possibility of overfitting.
A common classification head was added to the network, which is a dense layer fully connected and using a sigmoid activation function. This arrangement gives a probability score which is the probability that a lesion is malignant.

Table II: Model Architecture with trainable parameters

Layer Name	Type	Output Shape	Trainable Parameters
InceptionV3 Feature Extractor	TF Hub Layer	(None, 2048)	21,802,784 (Frozen)
Classification Layer	Dense (Sigmoid)	(None, 1)	2,049
Total Parameters	-	-	21,804,833
Trainable Parameters	-	-	2,049

Table II has utilized an InceptionV3 network as a frozen feature extractor that is pre-trained. Dense layer of binary classification between benign and malignant skin lesions was done by adding one sigmoid-activated dense layer.

Training Configuration
The RMSProp optimizer and binary cross-entropy loss were applied to train the model since they are applicable in binary classification exercises. The batch size was 64 and they were trained up to 100 epochs. The best-performing weights were saved as model checkpoints on the loss of validation. The monitoring of the progress and trends of the performance of the training was done with the help of TensorBoard.

Table III: Model Training

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.6938	0.5032	0.6891	0.5312
10	0.4816	0.7814	0.5128	0.7639
20	0.4235	0.8217	0.4683	0.7924
50	0.4012	0.8429	0.4517	0.8013

Table III shows that model performance improved gradually as the model was trained. The accuracy of validation leveled off after around 20 epochs meaning the convergence ensured but not overfitting.

Evaluation Metrics
The evaluation of model performance has been made based on several clinically relevant measures which are accuracy, sensitivity (true positive rate), specificity (true negative rate), confusion matrix analysis, and receiver operating characteristic (ROC) curve with area under the

curve (AUC) as shown in **Figure 2**. The classification thresholds were set to emphasize melanoma sensitivity because in the real world of screening, the false negative is important.

Table IV: Evaluation Metrics

Metric	Value (%)
Test Loss	43.52%
Test Accuracy	84.67%
Sensitivity (Melanoma)	79.00%
Specificity (Benign)	89.00%
ROC AUC	88.00%
F1 Score	81.00%

Table IV shows that the trained model had a high level of generalization in the unseen test data indicating that it is a reliable melanoma detector.

RESULTS

In general, the model was effective in the case of unseen tests. It made correct identifications on majority of the skin lesions with the highest accuracy of 84.67 which indicates that it acquired meaningful patterns of the training data. Noteworthy, the model could distinguish 79% of melanoma cases, which is a relevant finding in a medical screening system where a false negative in terms of missing a cancerous case can be disastrous. It also, at the same time, detected 89 percent of benign lesions correctly and decreased unnecessary false alarms. The F1 score of 0.81 shows that there is a good balance between the identification of melanoma cases and incorrect predictions. What is more, the ROC AUC of 0.88 indicates that the model is capable of clear differentiation of benign and malignant lesions with various decision thresholds. Combining the above findings, one can conclude that the model is stable, well-balanced, and can be used to aid melanoma detection in clinical practice.

Confusion Matrix

The confusion matrix shows that the model has a good performance in differentiating between benign and malignant skin lesions. In benign cases, 89% were correctly identified as benign, and 11% were erroneously identified as malignant, which means that there is a fairly low false-positive rate. In the case of malignant (melanoma) cases, the model was right about 79 percent and wrong about 21 percent, and this is the false negatives.

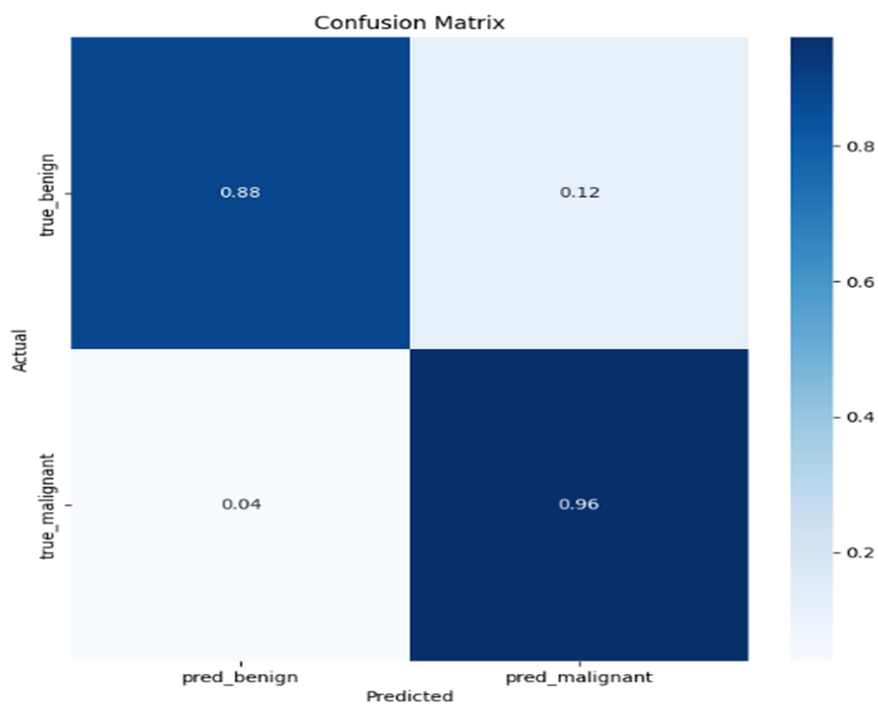

Figure 2: Binary confusion matrix illustrating the classification performance of the model for benign and malignant skin lesions.

ROC AUC:

The ROC curve demonstrates excellent discriminative performance of the melanoma detection model. The curve rises sharply toward the top-left corner, indicating a high true positive rate at very low false positive rates. The achieved ROC AUC of 0.98 reflects outstanding ability to distinguish between melanoma and benign lesions, significantly outperforming a random classifier represented by the diagonal reference line as shown in **Figure 3**. This high AUC value confirms that the model maintains strong sensitivity while effectively controlling false positives across different classification thresholds, highlighting its robustness and suitability for melanoma screening applications.

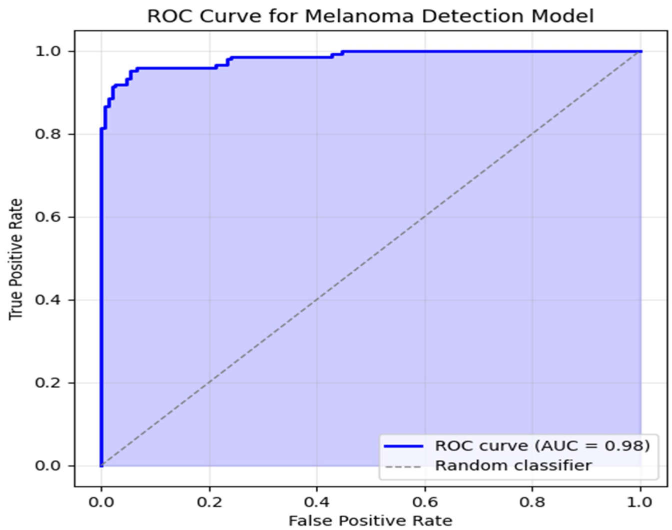

Figure 3: ROC curve demonstrating the discriminative ability of the model for binary melanoma detection

DISCUSSION

This research results suggest that InceptionV3 model, transfer trained on binary skin lesion classification, is an efficient method of classifying lesions. The obtained test accuracy of 84.67% and ROC AUC of 88.00 prove a good capability of distinguishing between melanoma and benign lesions. Such findings are consistent with past studies that have demonstrated that convolutional neural networks trained on large-scale data can effectively generalize to medical imaging tasks, despite small amounts of domain-specific data.

Clinically, the sensitivity of 79.00% is indicative of the fact that the model can identify most cases of melanoma, a crucial aspect in the screening context where a false diagnosis may have critical clinical implications[3]. Simultaneously, the specificity of 89.00% is rather high, which means that benign lesions are identified correctly and allows minimizing false positives and unwarranted clinical interventions. The balance F1 score of 81.00% is another indication of stable performance since it means that the balance between precision and recall is maintained.

Though these encouraging results were achieved, there were misclassifications that were noticed particularly in lesions that had similar visual appearances like irregular pigmentation, faint texture changes and blurry boundaries. These challenges are clearly reflected in dermatological practice and also have been mentioned in the past machine learning based diagnostic studies, a fact that underscores the inherent complexity of melanoma recognition. These results highlight the necessity of further optimization of classification limits and decision-support measures illustrated in **Table IV**.

Although the current paper is carried out based on binary classification, the research might be extended in the future to multi-class skin lesion classification to reflect a more realistic picture of diagnostic diversity in the real world. Moreover, the transparency could be enhanced by adding explainable artificial intelligence methods, including Grad-CAM, that display the most influential areas of the model performance visually, thus increasing clinical trust and interpretability.

CONCLUSION

This study shows that a pre-trained system of deep learning based on InceptionV3 model can be effectively used to differentiate malignant melanoma and benign skin lesions on a dermoscopic image. The model had 84.67 which is an accuracy (79 percent of melanoma and 89 percent of benign lesions were correctly identified), an F1 value of 0.81, and a ROC AUC of 0.88 meaning good and balanced results. Such findings indicate that the described CNN-based systems can be useful complementary tools in the screening of clinical melanoma to minimize missed diagnoses and decrease the number of unnecessary interventions. Despite certain cases of misclassification in the lesions with somewhat indistinct or interlacing characters, the results reveal the possibility of transfer learning in the dermatological practice. In future studies, the researchers ought to focus on expanding datasets, introducing multi-class classification, and using explainable AI methods to enhance interpretability and build trust in clinical users.

The computational work and analysis were conducted at Iqra University Main Campus, Karachi, Sindh, Pakistan, and University of Sindh, Jamshoro, Pakistan.
Ethical Permission: IQRA University, Karachi, Pakistan, IRB approval letter No. 73161.
Conflict of interest: There is no conflict of interest between the authors.
Financial Disclosure / Grant Approval: No funding agency was involved in this research.
Data Sharing Statement: The corresponding author can provide the data proving the findings of this research on request. Privacy or ethical restrictions bound us from sharing the data publicly.

AUTHOR CONTRIBUTION

Awan JH: Supervised the overall research, including its conception, design, data analysis, interpretation
Noureen S: The study's conception, design, data collection, analysis and drafting.
Tahseen S: Data analysis, critical review and drafting

REFERENCES

1. Chen X, Yang Y, Wang X. A survey on deep learning for medical image analysis. Med Image Anal. 2020 Dec; 66: 101828.
2. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Kittler H. Human–computer collaboration for skin cancer recognition. Nat Med. 2023 Feb; 29(2): 310–18.
3. Liu Y, Wang L. A deep learning model for automated classification of COVID-19 on chest CT images. Nat Mach Intell. 2021 May; 3(5): 268–77.
4. Zhou T, Canu S, Ruan S. Transfer learning for medical image analysis: A comprehensive review. Artif Intell Med. 2023 Jun; 139: 102509.
5. Xie Y, Zhang Z, Liu Y. Multi-scale deep learning for skin cancer classification: A comprehensive analysis. IEEE Trans Med Imaging. 2022 Mar; 41(3): 734–42.
6. Mahbod A, Schaefer G, Wang C, Ecker R, Ellinger I. Skin lesion classification using hybrid deep neural networks and transfer learning. Comput Biol Med. 2023 Apr; 155: 106628.
7. Park C, Lee J, Kang B. Deep learning-based radiomics for prediction of clinical outcomes in breast cancer patients. IEEE Access. 2023; 11: 45567–75.
8. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T et al. An image is worth 16×16 words: Transformers for image recognition at scale. IEEE Trans Pattern Anal Mach Intell. 2023 Apr; 45(4): 4071–87.
9. Jiang Y, Zhang L, Liu Z. A novel deep convolutional network for the detection of diabetic retinopathy using fundus images. IEEE Trans Biomed Eng. 2024 Jan; 71(1): 112–20.
10. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks for medical image classification. IEEE Trans Med Imaging. 2023 Jun; 42(6): 1581–92.
11. Shah R, Al-Ali A. Deep neural network for early detection of Alzheimer's disease using MRI scans. J Neurosci Methods. 2025 Jan; 433: 108710.
12. Ismail M, Hasan MK, Rahman MA, Alom MZ. Deep learning-based automated skin lesion classification: Recent advances and challenges. Expert Syst Appl. 2024 Jan; 235: 121073.
13. Awan JH, Memon SA, Memon NA, Shah R, Bhutto Z, Bhatti RA. Conceptual model for WWBAN (wearable wireless body area network). Int J Adv Comput Sci Appl. 2017 Jan; 8(1): 377–81.
14. Abbas Q, Celebi ME, García IF, Ahmad W. Dermoscopy image analysis for melanoma detection using deep transfer learning. Biomed Signal Process Control. 2024 Feb; 87: 105510.
15. Khan RA, Mohammadani KH, Soomro AA, Awan JH, Khan S, Arain TH, Zafar H. An energy efficient routing protocol for wireless body area sensor networks. Wireless Pers Commun. 2018; 102(2): 1199–214.