DOI: http://dx.doi.org/10.26483/ijarcs.v16i3.7284

ISSN No. 0976-5697

Volume 16, No. 3, May-June 2025

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

SKIN LESION CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Saurav Patel
Bioinformatics Centre
Savitribai Phule Pune University
Pune, India

Smita Saxena*
Bioinformatics Centre
Savitribai Phule Pune University
Pune, India

Abstract: Skin cancer is a common cancer with a large count of patients diagnosed annually. Early diagnosis and correct classification of skin lesions may increase the likelihood of cure before it turns malignant and the cancer metastasises. In this study we have trained the system using the International Skin Imaging Collaboration (ISIC) curated datasets of a huge number of gold-standard lesion diagnosed training images from patients with different skin disease conditions. Convolutional neural networks are used in this study because of their superior performance in medical imaging. Various architecture models and data augmentation strategies are investigated to alleviate dataset imbalances and improve model resilience. The comparative study of model performance demonstrates the superiority of InceptionV3 model in terms of validation accuracy and computational efficiency.

Keywords: skin lesion; skin cancer; medical imaging; image classification; deep learning; convolutional neural network

I. Introduction

Skin cancer is one of the prevalent types of cancer worldwide, affecting millions of individuals annually. According to the World Health Organization (WHO), skin cancer accounted for approximately 1.5 million new cases in 2022, making it a significant global health concern [1]. The disease is broadly classified into melanoma and non-melanoma skin cancers, with melanoma being more aggressive. The incidences of skin cancer continue to rise, particularly in regions with high sun exposure [2]. Given the increasing burden of skin cancer, early detection and preventive measures are crucial for reducing morbidity and mortality. The advances in medical imaging techniques and Deep Learning algorithms have significantly improved dermatology-specific medical image classification, particularly in identifying skin cancers from dermoscopic or macroscopic images [3]. Skin lesion classification is a complex task that requires the analysis of various features and patterns within the images. Convolutional Neural Network (CNN) has demonstrated superior performance in image analysis tasks, including skin lesion classification, owing to their ability to automatically learn and extract relevant features from images. CNN can effectively capture spatial relationships within the image data, enabling accurate differentiation between benign and malignant skin lesions. Compared to traditional machine learning algorithms, CNNs have shown higher accuracy and generalizability in skin lesion classification tasks [4]. However, there are several challenges to automated skin cancer classification, such as the imbalance in the distribution of skin disease photographs used for training, the robustness and cross-domain flexibility of the model, and limited data availability. This study was undertaken to create a reliable and efficient method for diagnosing skin cancer using a variety of pre-trained CNN application models and data augmentation techniques. The aim was to enhance skin cancer detection accuracy, support therapeutic decision-making, and provide better access to specialized medical knowledge.

II. Materials and Methods

The dataset used for this work was obtained from the International Skin Imaging Collaboration (ISIC), which is a partnership between academia and industry aimed at advancing the application of digital skin imaging to reduce melanoma mortality [5]. This image collection represents a valuable resource for researchers and developers in the field of dermatology and machine learning. This dataset includes various skin conditions and diseases, including Actinic keratosis, Basal cell carcinoma, Dermatofibroma, Melanoma, Nevus, Pigmented benign keratosis, Seborrheic keratosis, Squamous cell carcinoma, and Vascular lesion. Such a diverse and large dataset of skin lesion images is crucial for the development and training of computer-aided diagnosis (CAD) systems [6]. Accurate classification of skin lesions as malignant or benign by CAD systems can help improve the accuracy and efficiency of clinical diagnoses and ultimately reduce morbidity and mortality associated with skin cancer. The dataset employed in this study includes 60,507 images of both benign and malignant oncological diseases collected from various ISIC challenges held between 2016 and 2020 [7]. The ISIC metadata provided information about the diagnosis and clinical characteristics of the skin lesions depicted in the images, including disease type, anatomical site, and clinical subtype. Table 1 shows the skin lesion categories and number of images available for each category of skin lesions [5].

Convolutional Neural Network (CNN) is a type of artificial neural network that is commonly used in image and video recognition tasks. CNNs are particularly effective in image recognition tasks because they can learn to identify local patterns and structures within an image such as edges, corners and other features regardless of their position within the image [8]. This is achieved by using numerous layers in the architecture of CNN. Convolutional Layer applies a set of filters or kernels (Horizontal Edge filter, Vertical Edge filter etc.) to the input data. Each filter is convolved with the input data to generate a feature map that focuses on specific features. Padding is a technique used in CNN to add extra pixels around the border of the input data before convolution. This can be useful for preserving the spatial resolution of the feature maps produced by the convolutional layers. When a convolutional filter is applied to the input data, it slides over the data and produces an output feature map that is smaller than the input data. Without padding, the filter will not be able to fully cover the edge pixels of the input data, resulting in a loss of spatial resolution in the output feature maps. In the case of 'Valid' padding, no padding is added to the input data and the filter is only applied to the parts of the input that completely overlap with the filter. For 'Same' padding, the input data is zero-padded to ensure that the output feature maps have the same spatial dimensions. Pooling layers are used to minimize the size of feature maps. It reduces the number of parameters to learn and the amount of computation required in the network. The pooling layer summarizes the features present in the region of the feature map generated by the convolutional layers. The flatten layer is generally added after the convolutional and pooling layers to convert the high-dimensional feature maps produced by the convolutional and pooling layers into a one-dimensional vector that can be fed into a fully connected neural network layer. The flatten layer reshapes the feature maps into a single long vector by stacking the values of each feature map together. Fully connected layers are typically used at the end of the network to make predictions based on the high-level features extracted by the earlier convolutional and pooling layers. The fully connected layer takes the flattened feature vector produced by the previous layer and applies a set of weights to each element of the vector. The weights are learned during the training process and are used to make predictions based on the features extracted by the earlier layers. The output of the fully connected layer is typically passed through a non-linear activation function, such as a ReLU (Rectified Linear Unit), to introduce non-linearity into the network. This helps in improving the network's capability to recognize and learn the patterns from input data. Different models have different number of layers and number of trainable parameters. In this case, if the CNN is being used to classify skin lesion images into 8 different classes, the fully connected layer would have 8 neurons, one for each class.

Table I. Number of images present per class in ISIC dataset

Type of Lesion	***Image Count***	***Sample Image***
Actinic Keratosis	1,064	
Basal Cell Carcinoma	3,323	
Benign Keratosis	2,624	
Melanoma	5,333	
Nevus	18,068	
Seborrheic Keratosis	1,761	
Squamous Cell Carcinoma	628	
Unknown	27,706	
Total	60,507	

Keras Applications are designed to facilitate the use of deep learning algorithms in real-world applications, such as medical imaging, by providing pre-trained models that can be utilized to solve specific tasks [9]. These models are based on cutting-edge architectures and have been trained on large-scale datasets, yielding improved results. Keras Applications were utilized as the backbone of the skin cancer diagnosis algorithm, enabling the development of a robust and trustworthy method for identifying skin cancer [10]. The pre-trained weights provided by these models allowed for faster training and fine-tuning, while the use of transfer learning enabled the model to learn from the ISIC dataset and achieve high accuracy. The study employed InceptionV3, InceptionResNetV2, ResNet50V2 and MobileNetV2 models to classify skin lesion images [11-14].

The first step of the process involved downloading an image dataset of benign and malignant oncological conditions from ISIC and sorting them according to ISIC classification into 7 different classes and 1 unknown class using metadata [5-6]. The images in different classes were found to be unequally divided. 6000 images per class were used for the final dataset. Data augmentation techniques such as cropping, rotating, resizing, translating, and flipping were used to address the class imbalance using the ImageDataGenerator tool in Python Keras module. Reshaping the images was necessary to address the issue of size disparity, which was also achieved using the ImageDataGenerator tool. After balancing and reshaping the images, the final dataset of 48000 images was divided into 80% training and 20% test data. Pre-trained models of Keras were trained on the training data. The trained models were validated using the test data to understand prediction trends and monitor accuracy. Reliable models were saved following training and validation, and these saved models were then used to make predictions.

III. Results and Discussion

The study utilized TensorFlow's Keras API, a high-level deep learning framework designed for building and training neural networks efficiently. Keras simplifies model development by providing pre-built layers, optimizers, and data preprocessing tools, making it suitable for skin lesion classification task. The performance of four deep learning models- InceptionV3, InceptionResNetV2, ResNet50V2 and MobileNetV2 was evaluated using key metrics such as accuracy, precision, recall, F1-score and AUC-ROC. These metrics are mathematically evaluated using the following formulas:

$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$ (1)
$\text{Precision} = TP / (TP+FP)$ (2)
$\text{Recall}=TP/ (TP+FN)$ (3)
$F1 = 2\times(\text{Precision}\times\text{Recall})/ (\text{Precision}+\text{Recall})$ (4)
$\text{Loss} = -\sum y\log\hat{y}$ (5)

where, TP-True Positive; TN- True Negative; FP- False Positive and FN- False Negative.

Accuracy measures the proportion of correctly classified samples. Precision indicates how many predicted positives are actually correct. Recall (Sensitivity) shows how well the model identifies actual positives. F1 Score combines precision and recall using the harmonic mean. It balances precision and recall. It ensures that both precision and recall must be high for the F1 score to be high. AUC-ROC (Area Under Curve - Receiver Operating Characteristic) evaluates the model's ability to distinguish between classes. It is computed from the ROC curve, which plots True Positive Rate (TPR) vs. False Positive Rate (FPR). Model Loss quantifies the error during training for classification. In (5), y is the true label and ŷ is the predicted probability.

The results obtained from these models and the run time of the respective models and some sample predictions are shown and discussed below. Table 2 indicates the training accuracy, validation accuracy and the code runtime between defining the class weights, combining base model with top layers, training, saving the best model after performance evaluation and fitting. The codes were executed on Anaconda distribution v22 (Python 3.9) on i5 processor desktop with 16GB RAM and Windows 10 Operating System.

Table II. Model accuracy and runtime

Model Name	*Accuracy*	*Validation Accuracy*	*Runtime (hr:min:sec)*
InceptionV3	94.39%	59.04%	18:07:38
InceptionResNetV2	93.97%	57.80%	28:06:44
ResNet50V2	95.91%	56.47%	24:55:37
MobileNetV2	95.20%	48.70%	2:35:32

The MobileNetV2 model had least runtime but the validation accuracy was poor. InceptionV3 model gave comparable accuracy with less runtime than the remaining two models.

Fig. 1 is a combined plot of the InceptionV3 model indicating various performance metrics progression over the epochs during training and validation. The model training accuracy reaches 0.94 while the loss reduces to 0. AUC reaches a value of 0.99. Precision and recall values are 0.97 indicating good performance of the model. F1-score is also very high at 0.97.

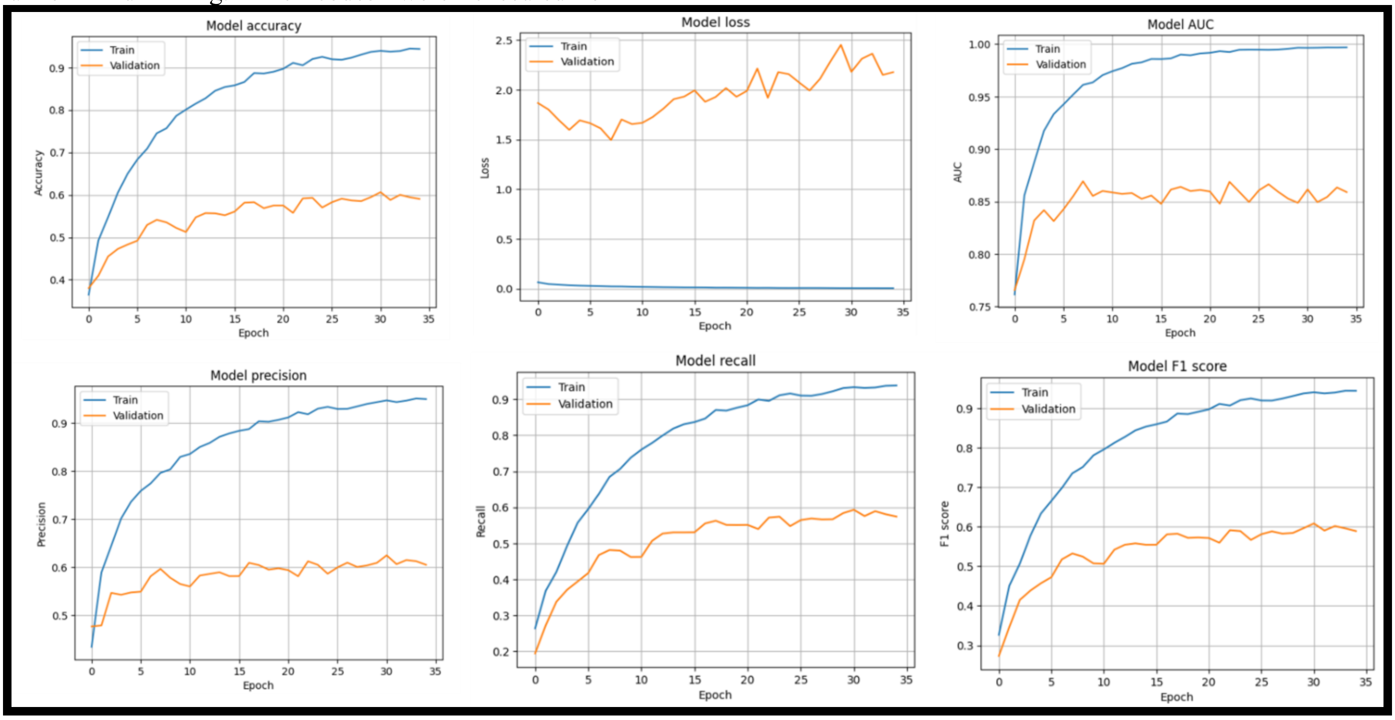
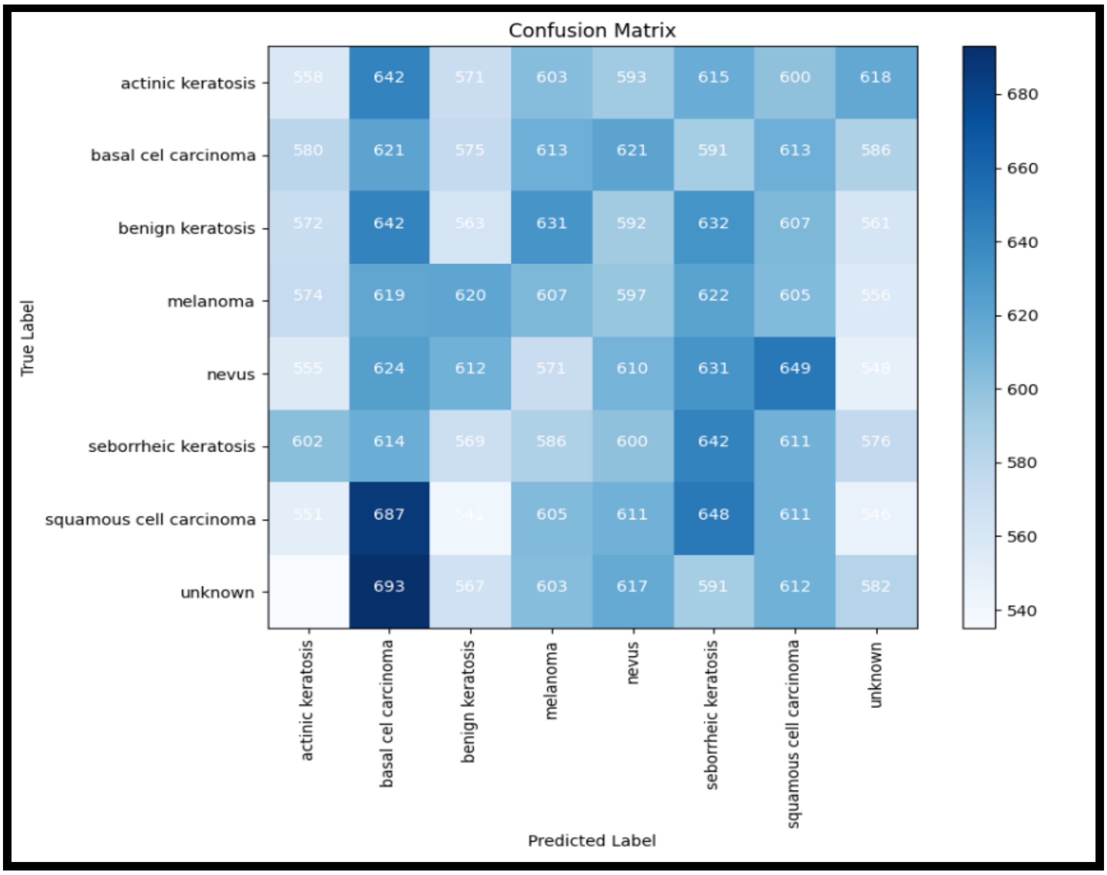


Figure 1: Training and validation of InceptionV3 model: Accuracy, Loss, AUC, Precision, Recall and F1-score

Fig. 2 is the confusion matrix that shows the number of predicted vs. true label for all the skin lesion images belonging to the eight different classes.

Similar calculations for performance metrics were carried out for the other models namely, InceptionResNetV2, ResNet50V2 and MobileNetV2. The performance matrics plots are not shown and discussed here in the text to avoid redundancy. However, the codes, results and plots can be shared upon request.



True Label \ Predicted Label	actinic keratosis	basal cel carcinoma	benign keratosis	melanoma	nevus	seborrheic keratosis	squamous cell carcinoma	unknown
actinic keratosis	558	642	571	603	593	615	600	618
basal cel carcinoma	580	621	575	613	621	591	613	586
benign keratosis	572	642	563	631	592	632	607	561
melanoma	574	619	620	607	597	622	605	556
nevus	555	624	612	571	610	631	649	548
seborrheic keratosis	602	614	569	586	600	642	611	576
squamous cell carcinoma	551	687	541	605	611	648	611	546
unknown	535	693	567	603	617	591	612	582

Figure 2: Confusion matrix for the InceptionV3 model

IV. Conclusion

The performance of four deep learning models InceptionV3, InceptionResNetV2, ResNet50V2 and MobileNetV2 was evaluated for skin lesion classification, focusing on key metrics. Among the tested models, InceptionV3 model achieved the highest classification accuracy of 94% indicating superior feature extraction capabilities. ResNet50V2 model gave higher accuracy but lower training accuracy with a higher runtime. InceptionResNet50V2 model had lower accuracy with highest runtime. Meanwhile, MobileNetV2, known for its lightweight architecture, provided faster inference in the shortest time but the training accuracy turned out to be the lowest. This model appears to be a promising candidate for real-time clinical applications. The study showed a significant improvement in skin lesion type detection precision, indicating the potential of deep learning algorithms in improving diagnosis and treatment decision-making. Early diagnosis of malignant lesions may be helpful in cure and preventing further stages of the disease. While deep learning models enhance early skin cancer detection, challenges such as data variability, misclassification, and computational complexity remain. Future research should focus on optimizing model efficiency, expanding diverse datasets, increasing the number of training images for different lesion types and further improvement in diagnostic reliability. The advancements in deep learning techniques and high-quality medical imaging will support clinicians in more accurate decision-making benefitting the patient care. Additionally, expanding the collection of publicly available datasets will enable researchers to develop more robust models and improve generalizability across different populations and ethnic groups.

V. Acknowledgment

S.P. thanks Department of Biotechnology, Government of India for PG program support and studentship. S.S. and S.P. acknowledge Savitribai Phule Pune University for infrastructure support.

VI. References

[1] World Health Organization, World health statistics 2022: Monitoring health for the SDGs, sustainable development goals, 2022.

[2] International Agency for Research on Cancer, Global Cancer Observatory: Skin Cancer Statistics 2025, 2025.

[3] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017. doi: 10.1038/nature21056.

[4] P. Tschandl, N. Codella, B. N. Akay, G. Argenziano, R. P. Braun, H. Cabo, et al., "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," Lancet Oncol., vol. 20, no. 7, pp. 938–947, Jul. 2019. doi: 10.1016/S1470-2045(19)30333-X.

[5] International Skin Imaging Collaboration, SIIM-ISIC 2020 Challenge Dataset, 2020. doi: 10.34970/2020-ds01.

[6] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, et al., "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," Sci Data, vol. 8, p. 34, 2021. doi: 10.1038/s41597-021-00815-z.

[7] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," Med Image Anal., vol. 75, p. 102305, Jan. 2022. doi: 10.1016/j.media.2021.102305.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[9] F. Chollet, Keras, GitHub, 2015. [Online]. Available: https://github.com/fchollet/keras..

[10] A. S. Reddy and G. M. P, "A Comprehensive Review on Skin Cancer Detection Strategies using Deep Neural Networks," J. Comput. Sci., vol. 18, no. 10, pp. 940–954, 2022. doi: 10.3844/jcssp.2022.940.954.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2818–282, April 1955.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," Proc. AAAI Conf. Artif. Intell., vol. 31, no. 1, pp. 4278–4284, 2017.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," Proc. Eur. Conf. Comput. Vis. (ECCV), pp. 630–645, 2016.. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[14] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, et al., "Searching for MobileNetV3," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 1314–1324, Oct. 2019.