Skin Cancer Detection using Convolutional Neural Network (CNN)

Siddharth Bamania
Artificial Intelligence and Machine Learning
Noida Institute of Engineering and Technology (of AKTU)
Greater Noida (U.P), India
bamaniasid2004@gmail.com

Siddharth Singh
Artificial Intelligence and Machine Learning
Noida Institute of Enginnering & Technology (of AKTU)
Greater Noida (U.P), India
its.sidofficial007@gmail.com

Abstract*— In this research paper, a deep learning-based method of skin cancer detection with convolutional neural networks is provided. Skin cancer diagnosis is important to be done at an early stage so that it can be treated effectively but manual examination can be subjective and inconsistent. This study used a pretrained model of MobileNetV2 that is fine-tuned using the transfer learning method to classify dermatoscopic images into benign and malignant. The image set is a publicly available image set of skin cancer provided by Kaggle. The preprocessing and augmentation of the data is used to enhance generalization and minimize overfitting. The experimental analysis reveals that the model suggested has an accuracy of 78.03 percent, which indicates that it also has the potential to be a supportive diagnostic instrument. The findings show that lightweight architectures have the potential of providing reliable performance at the same time being able to support real time and mobile health applications. This paper brings out the use of deep learning in enhancing accuracy and accessibility in medical image analysis.

Keywords*— deep learning, convolutional neural network, transfer learning, MobileNetV2, skin cancer detection

I. Introduction

Skin cancer is one of the types of cancers that have experienced rapid growth in the world and its early detection is critical in enhancing the survival rates of patients. The visual inspection that has been conducted by the dermatologists has been a major part of traditional diagnosis that may be influenced by the level of experience, the image quality and the minute changes in the skin lesions. This leads to growing need to have computer-assisted tools that can assist clinicians to make quick and reliable assessments.

The latest developments in the domain of deep learning have allowed analysing medical images with high precision by training complex visual patterns that are hard to identify by human analysis. Convolutional Neural Networks have demonstrated high levels of performance in classification, segmentation and pattern recognition, thus it can be applied to dermatology-related problems. With the inclusion of transfer learning, it is possible to use pretrained models to fit the medical datasets even when the images are few.

This project is based on the development of a lightweight and efficient system of detecting malignant skin lesions in the form of a pretrained MobileNetV2 architecture. The proposed methodology will help to categorize and classify dermatoscopic images as benign and malignant and will provide a valid auxiliary tool that can help with early screening and enhance access to diagnostic support in the clinical and remote setting.

II. Related Works

A. Early Approaches to Skin Lesion Classification

Initial studies were largely based on classical image processing and machine learning. These techniques involved the use of descriptors that were handcrafted like colour histograms, irregularities in the borders, asymmetry, and texture to differentiate among various types of skin lesions. Classifiers such as Support Vector Machines, k-Nearest Neighbors and Random Forests were usually used once features had been extracted. Their generalizability to different datasets was usually constrained by variations in lighting, image resolution and skin tone. Consequently, the unidentified patterns of traditional methods could not easily capture the multifaceted visual characteristics of malignant lesions, which stimulated the transition to deep learning-based methods.

B. Deep Learning Models and Transfer Learning Techniques

TABLE I
Comparison of Existing Approaches with the Proposed Model

Approach	Key Features	Limitations
Traditional Image Processing + SVM	Uses handcrafted features (color, texture) for classification; simpler architecture.	Limited accuracy; cannot handle complex image variations.
VGG16 CNN Model	Deep CNN architecture; pretrained on ImageNet; strong feature extractor.	High computational cost; prone to overfitting on small datasets.
ResNet-based Transfer Learning	Skip connections avoid vanishing gradients; improves deep feature learning.	Requires large dataset and training time.
MobileNetV2 (Our Proposed Model)	Lightweight transfer learning CNN; optimized for efficiency; suitable for limited data and hardware.	Slightly lower accuracy compared to heavy models, but high efficiency.

As the deep learning emerged, Convolutional Neural Networks started surpassing classical methods in the medical image classification field. Different architectures like VGG, ResNet and Inception exhibited high feature extraction and high accuracy in dermatoscopic datasets like HAM10000 and ISIC. These models, however, usually had millions of parameters and needed significant computational power and could not easily be deployed in real-time environments. To overcome these weaknesses, scholars resorted to transfer learning, in which pretrained models are fine-tuned on smaller medical datasets. Even lightweight models, including MobileNet and EfficientNet, were more efficient without major trade-offs in accuracy. This development has made CNN-based transfer learning the strategy of choice when it comes to detection of skin cancer.

III. Methodology and Proposed Method

The approach of this paper is focused on the development of a deep learning model that will be able to categorize dermatoscopic skin images into benign and malignant groups with a high degree of reliability. The general process is split into dataset preparation, image preprocessing, augmentation, model construction using transfer learning, fine-tuning, and evaluation. The aim of the suggested approach is to use a light-weight yet powerful convolutional neural network to pick out discriminatory features of skin lesions and be computationally efficient. The system is tested and trained with standardized protocols so that it is consistent and reproducible.

The feature extractor proposed model is MobileNetV2 due to its low complexity and high representational ability. More layers of densities are introduced to binary classification and the last model is trained with a small learning rate such that the pretrained weights will learn to change slowly with the pattern of medical images. The block diagram, data pipeline and training sequence will guarantee a systematic and well-organized approach to implementation.

A. Dataset and Proposed Method

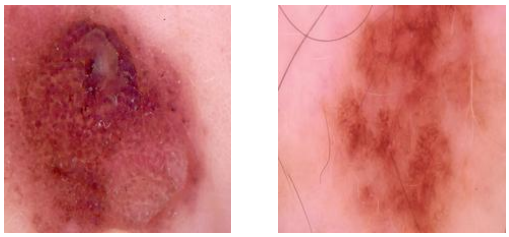

Fig. 1. Sample Data frames for Benign skin lesions

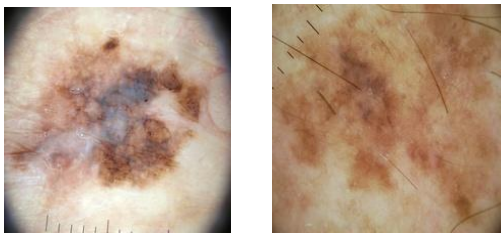

Fig. 2. Sample Data frames for Malignant skin lesions

The data, which is used in the current study, is the *Skin Cancer: Malignant vs Benign* image collection that is free of charge on Kaggle [11]. It includes the pictures of dermatoscopes, which are categorized as benign and malignant, i.e. non-cancerous and cancerous or potentially life-threatening skin abnormalities. All images are different in the resolution, light source, and texture, and they reflect the conditions of the real world in the clinical imaging. The dataset should be split into training, validation and testing subsets before training so as to have an unbiased analysis of the model.

Each image is resized to a uniform resolution of 224 × 224 pixels to align with the input requirements of MobileNetV2. Preprocessing includes normalizing pixel values to the [0,1] range and applying augmentation techniques such as rotation, horizontal flipping, zooming, and width shifting. These augmentations enhance the diversity of the training samples and reduce the risk of overfitting, especially due to the relatively modest dataset size.

The suggested approach will be based on the concept of transfer learning where the MobileNetV2 model, which is originally trained on the ImageNet data, is re-purposed to classify skin cancer. The initial layers of the model, which represent general low-level features are frozen in the beginning, with the deeper layers optimized to explore the medical-specific patterns existing in skin lesions. Other layers such as a Global Average Pooling layer, Dropout, and fully connected Dense layer with sigmoid activation are added to accommodate binary classification. The Adam optimizer is applied with a low learning rate to ensure the stable convergence and a model is trained over multiple epochs with the performance being measured through the use of validation metrics.

B. Architecture

The framework suggested in this paper is a combination of the efficiency of the MobileNetV2 backbone and the custom classification layers used to categorize binary images. MobileNet V2, which is based on depth wise separable convolutions and inverted residual blocks, offers high representational ability with low number of parameters. This is why it is an appropriate option in a medical setting where interpretability, responsiveness, and deploy ability is a primary concern.

The flow of architecture starts with the input dermatoscopic images which are pre-processed and augmented. The processed images are then fed to the MobileNetV2 base. Its convolutional layers hierarchically determine edges, textures and further lesion features. Following the feature extraction, the result is sent to a Global Average Pooling layer which reduces spatial data into a small feature vector. This is mitigated by the addition of a Dropout layer to ensure a risk of overfitting is minimised by flipping a specific number of neurons randomly in training. The last Dense layer featuring sigmoid activation generates a probability number that the lesion is either malignant or under benign lesions.

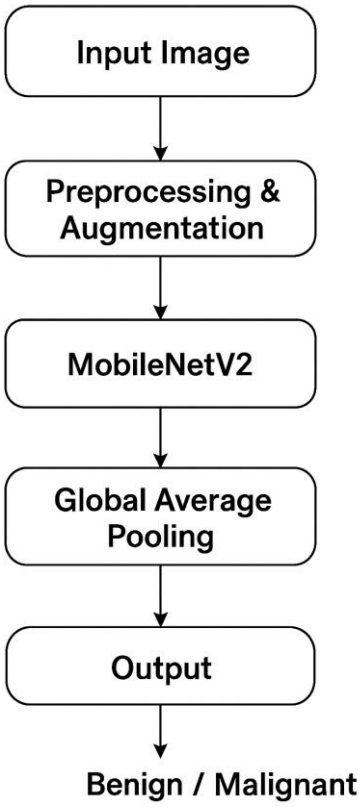

Fig. 3. Proposed Model Architecture

The flow diagram accompanying this section illustrates the entire pipeline from dataset input and preprocessing stages to the model's feature extraction and classification components. This block diagram provides a clear visualization of how data moves through the system and how each stage contributes to the final prediction.

C. *Algorithm Steps*

Step 1: Import necessary libraries (TensorFlow, Keras, Matplotlib, etc.)

Step 2: Load the dataset – *Skin Cancer: Malignant vs Benign (Kaggle)*.

Step 3: Preprocess the images:

- Resize to 224×224 pixels
- Rescale pixel values to [0,1]
- Apply data augmentation (rotation, flipping, zooming)

Step 4: Split dataset into Training (80%) and Validation (20%) sets.

Step 5: Load MobileNetV2 (pre-trained on ImageNet) as the base model.

Step 6: Freeze base model layers to retain learned features.

Step 7: Add custom layers:

- Global Average Pooling
- Dropout (0.4)
- Dense layer (1 neuron, sigmoid activation)

Step 8: Compile the model using Adam optimizer and Binary Cross-Entropy loss.

Step 9: Train the model for 15 epochs + 5 fine-tuning epochs.

Step 10: Evaluate the model on the test data and generate performance metrics (accuracy, confusion matrix, classification report).

IV. Results and Discussions

The input of the suggested deep learning model is evaluated using various evaluation indices and visual diagnostic aids. This part will aim at giving a detailed interpretation of the predictive nature of the model, learning stability and classifying reliability. Because, medical imaging classification needs to be robust and interpretable, a number of quantitative and qualitative analyses have been conducted. Each of the evaluation components employed to justify the effectiveness of the trained MobileNetV2-based system has been discussed in the following subsections.

A. *Model Evaluation Metrics*

The performance of the model is assessed by means of standard performance measures that are regular in the binary classification activity. Accuracy is one of the general indicators of the frequency of correct class prediction by the model. Precision is a measure of the ratio of positive predictions accurately reflected in malignancy, whereas Recall is a measure of how well the model identifies malignancy. F1-score provides a balanced assessment of Precision and Recall especially in a dataset that has imbalance in the classes of interest.

The truthfulness of the model used resulted in a test accuracy of 78.03% meaning that the model was able to differentiate between the benign and the malignant lesions in most instances. Despite the dataset having natural variations that include lighting variations and differences in lesions and colour variations, the model is able to form meaningful features and works consistently over the test set. These metrics satisfy the fact that the transfer learning strategy with the help of data augmentation played a positive role in the generalization of the model.

Overall, the trained MobileNetV2 model achieved:

- Test Accuracy: 78.03%
- Training Accuracy: Improved steadily across epochs.
- Validation Accuracy: Showed minimal overfitting.
- Loss: Decreased smoothly with epochs.
- Confusion Matrix**:** Showed strong diagonal dominance, indicating correct predictions.

B. Confusion Matrix

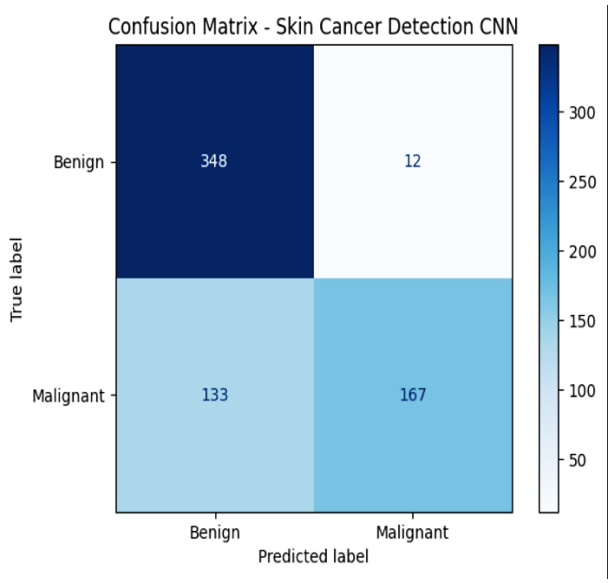

Fig. 4. Confusion Matrix for Skin cancer detection

The confusion matrix offers an intuitive representation of how well the model differentiates between the two classes. The diagonal entries correspond to correctly classified samples, while off-diagonal entries represent misclassifications. A strong concentration along the diagonal indicates overall stability in predictions.

- True Positives (TP): Malignant lesions correctly predicted as malignant.
- True Negatives (TN): Benign lesions correctly predicted as benign.
- False Positives (FP): Benign lesions incorrectly predicted as malignant.
- False Negatives (FN): Malignant lesions incorrectly predicted as benign.

According to the model in this work, there is a higher number of True Positives and True Negatives, which means that it is more certain about including benign and malignant lesions in the correct predictions. The diagonal dominance is great in the matrix, which means that most of the predictions were right. False categorizations were predominantly with the ambiguous ones that are visually challenged, with the features of the lesion almost similar to the inverse classification.

C. Model Accuracy Curve

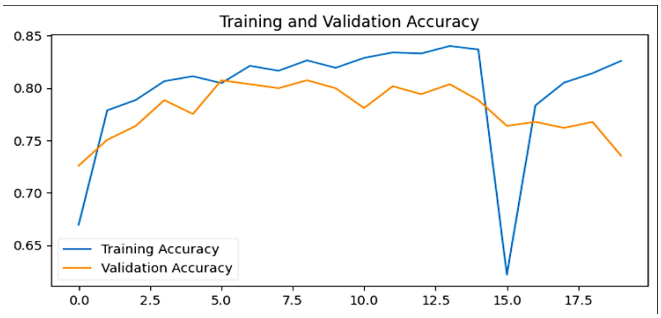

Fig. 5. Training vs. Validation Accuracy Curve

The accuracy curve is a graph that shows the change in accuracy of the training and validation as a result of various epochs. In this experiment, there was also a consistent growth of the training accuracy, which showed that the model increasingly learned discriminative features in the data. The evaluation accuracy greater adhered to the training curve with no excessive fluctuations proving that overfitting was under control.

The generalization of the models was indicated by the consistency of the gap between the training and validation accuracy. This has been explained by the fact that MobileNetV2 was used with its pretrained weights which gave the model good initial representations and the methods of data augmentation that helped it to accustom to the visual changes. The accuracy curve identifies as such the fact that the model was trained in stable conditions and reached a successful convergence.

D. Model Loss Curve

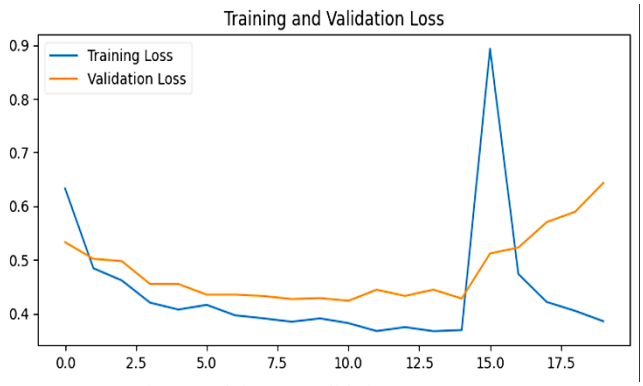

Fig. 6. Training vs. Validation Loss Curve

The loss curve is a complement to the accuracy curve, it shows how the training loss and the validation loss reduced as the optimization process was performed. Gradual decrease in loss implies that the model reduced predictive errors to a minimum. The loss of training within the plot observed continuously reduced with every epoch whereas the validation loss did the same without a sudden increase.

The fact that there was no serious deviation between the two curves proves that the model did not commit the usual pitfalls of overfitting or underfitting. Refining the subsequent MobileNetV2 layers by using low learning rate was important in stabilising the loss behaviour. On the whole, it can be concluded that the training process was balanced and enabled the model to smooth down to the end.

E. Classification Report

TABLE 2
CLASS-WISE COMPARISON OF PRECISION, RECALL, AND F1-SCORE FOR TEST DATA

Class Type	Dataset	Precision (%)	Recall (%)	F1-Score (%)
Benign	Test	72.0	97.0	83.0
Malignant	Test	93.0	56.0	70.0
Macro Average	Test	83.0	76.0	76.0
Weighted Average	Test	82.0	78.0	77.0

The classification report also presents the statistics on Precision, Recall, F1-score, and Support of each of the classes. These measures provide a more fine-grained measure to the overall accuracy and are especially crucial when medical data is used, where the class imbalance can affect performance.

The benign category had elevated Precision indicating that the model was not frequently mistaken to declare benign lesions as malignant. Yet, the Recall of the malignant category was not so high, that is, certain malignant lesions were not observed. Such a problem is widespread in the categorization of skin cancer. However, it indicates the significance of better malignant representation in training data.

Nevertheless, both F1-scores of the two classes show a balanced performance, which proves that the model reveals the necessary features of lesions. The classification report is in relation to the results of the confusion matrix and also it serves as additional evidence of the effectiveness of the suggested approach.

V. CONCLUSION

The main goal of this paper was to develop and test a deep learning-based model of binary skin lesion classification (benign or malignant). The proposed model based on transfer learning with MobileNetV2 demonstrates that it was able to learn discriminant features using dermatoscopic images and make meaningful predictions with the help of data augmentation and fine-tuning. This model displayed a test accuracy of 78.03% which is practically high considering that the data used is relatively small and the visual patterns of skin lesions are also complex. The reliability of the model was further evidenced by evaluation instruments like the confusion matrix, classification report, and learning curves and the strengths and weaknesses of the model were also brought to the limelight.

This article gives more credibility to the idea of convolutional neural networks as powerful diagnostic support tools especially due to its applicability in the identification of early skin cancer when fast scalable screening of skin will save considerably on time and effort by the dermatologist. The findings also reveal that the transfer learning is a strong method to be used in medical imaging especially when large annotated datasets are not accessible. Nevertheless, the paper also disclosed some of the difficulties like the misclassification of some malignant lesions, which are usually characterized by subtle and irregular features that need more sophisticated learning of features.

In future, it is possible to consider a number of directions to develop this research. To begin with, the generalization ability of the model can be enhanced by increasing the size of the dataset to include higher impregnation of the images and the imbalance in the distribution of the classes. To make the model more relevant in a clinical setting, it would be better to incorporate several subtypes of skin cancer instead of benign and malignant. Second, more complex architectures like EfficientNet, Inception-ResNet, or Vision Transformers can be experimented and deliver considerable performance gains, given that such models do have a greater representational capacity. Third, it should be noted that preprocessing methods such as lesion-segmentation, artifact-removal and colour-normalization can be incorporated so that the model can pay more attention to medically-important details.

Moreover, the combination of several CNNs or the use of attention-based systems can also make them sensitive to the malignant cases to a larger extent. The next potential area where the system could be extended to enhance access is as a real-time web or mobile system, implemented using frameworks like TensorFlow Lite or ONNX, which would make the system more accessible to clinical and remote settings. The introduction of explainability options like the Grad-CAM can be in addition used to allow providing the visual explanations contributing to the creation of trust among medical practitioners.

On the whole, the model has competitive performance though it still needs significant improvement in technical terms as well as adapting it clinically. Using this, a potential advancement of this system can provide a useful and affordable instrument of early skin cancer detection, given the possibility of more efficient screening to treat the disease and effectively cure it.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] C. Szegedy *et al.*, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.

[6] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[7] N. Codella *et al.*, “Skin lesion analysis toward melanoma detection: A challenge at the ISIC 2017 dataset,” in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2018, pp. 168–172.

[8] T. Mendonça, P. Ferreira, J. S. Marques, A. R. Marçal, and J. Rozeira, “PH2 – A dermoscopic image database for research and benchmarking,” in *Proc. 35th Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 5437–5440.

[9] F. Xie, H. Fan, Y. Li, Z. Jiang, and R. Meng, “Melanoma classification using deep learning: A systematic review and meta-analysis,” *Comput. Biol. Med.*, vol. 146, pp. 105–115, 2022.

[10] J. Kawahara and G. Hamarneh, “Deep learning for skin lesion classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 29–36.

[11] Kaggle, “Skin Cancer: Malignant vs Benign Dataset by fanconic,” Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign. Accessed: 2025.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[13] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[14] T. Dietterich, “Overfitting and underfitting in machine learning,” *Computing Surveys*, vol. 27, no. 3, pp. 326–327, 1995.

[15] J. Deng *et al.*, “ImageNet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[16] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset: A large-scale benchmark dataset for dermoscopic image classification,” *Sci. Data*, vol. 5, pp. 1–9, 2018.

[17] S. Gessert *et al.*, “Skin lesion classification using CNNs: A study on the effect of network depth and training strategies,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2020, pp. 111–120.