

SKIN DISEASE CLASSIFICATION USING DEEP LEARNING

Keethika S, Sanjai R, Sathya S, Pooja K

Department of AI, Kongu Engineering College, Perundhurai, Erode.

keethika.ai@kongu.edu

{sanjai.22aid, sathyaas.22aid, poojak.22aid}@kongu.edu

ABSTRACT:

Skin disease is a rapidly growing illness where abnormal skin cells grow uncontrollably due to UV radiation exposure. So, it is very essential for the early detection of the disease. It helps to keep cancer from growing and spreading, lowering healthcare expenses and problems. Early detection also reduces scarring and increases quality of life. The studies focus on enhancing skin cancer detection using various deep learning models. The CNN Hybrid model leverages the strengths of both ResNet50 and EfficientNetB0 architectures, aiming to improve feature extraction and classification performance. The CNN Transformer model integrates Convolutional Neural Networks (CNNs) with Transformer blocks to capture both local and global features, enhancing classification accuracy. The CNN model serves as a baseline, utilizing a straightforward architecture to classify different types of skin cancer, providing a basis for comparison with the more complex hybrid models.

Keywords: CNN, RNN, Transformer Precision, Recall, F1 Score, Skin diseases, ResNet50 and EfficientNetB0.

1. INTRODUCTION

Skin disease classification using deep learning is an emerging and transformative

field in medical diagnostics. Leveraging advanced neural networks, particularly convolutional neural networks (CNNs), this approach aims to enhance the accuracy and efficiency of detecting skin cancer from medical images. Deep learning models are trained on vast datasets comprising dermoscopic images of various skin conditions, learning to differentiate between benign and malignant lesions with high precision. These models can analyze subtle patterns and features in the images that may be imperceptible to human eyes, thus providing a powerful tool for early diagnosis. The integration of deep learning in skin cancer prediction not only augments the capabilities of dermatologists but also facilitates timely intervention, improving patient outcomes. Moreover, the use of such AI-driven techniques holds the promise of making high-quality diagnostic services more accessible, especially in underserved regions where specialist healthcare is limited. As research continues to advance, the accuracy, reliability, and practical applications of these predictive models are expected to further evolve, paving the way for more sophisticated and user-friendly diagnostic tools.

2. LITERATURE REVIEW

Research work [1] proposes a solution to the problem of higher medical costs behind diagnosis, lower accuracy rate in detection and portability problem of the manual detection system. In this system, dermoscopic

images are classified to predict skin cancer using a multi-layered CNN approach with multiple regularization techniques named dropout and batch normalization. As a result, the system has provided an accuracy of 93.58% which is higher than most other conventional approaches.

The classification results of the [2] proposed methods are compared with state-of-the-art approaches while utilizing a reduced number of factors/feature vectors. The experiments exhibit that the classification accuracy is about 89.29% using the proposed method. The outcome of the experimental investigation shows that our method has higher accuracy in comparison with state-of-the-art techniques in literature.

The diagnosing methodology of [3] uses concept of image processing and deep learning. Through using different tactics of image augmentation, the number of images has also been enriched. Finally, the transfer learning approach is used to further improve the accuracy of the classification tasks. Approximately 0.76 weighted average precision, 0.78 weighted average recall, 0.76 weighted average f1-score, and 79.45 percent accuracy are shown by the proposed CNN method[4].

The data pre-processing techniques like sampling, dull razor and segmentation using autoencoder and decoder is employed. Transfer learning techniques like DenseNet169 and Resnet 50 were used to train the model to obtain the result [5].

3. DATASET DESCRIPTION

The dataset includes training and the test data with nine different classes namely actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, vascular

lesion which is used for training and evaluation of the model.

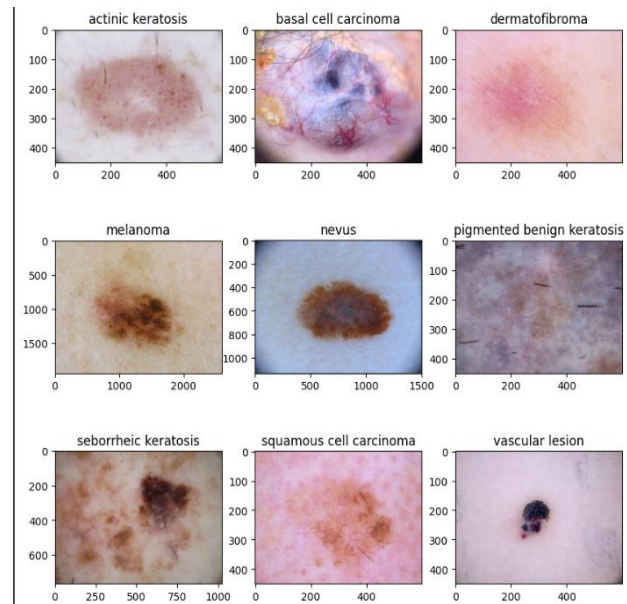


Fig 1. Types of skin lesions

Dataset:

<https://www.kaggle.com/code/sharanharsoor/skin-cancer-detection/input>

4. PROPOSED METHODOLOGY

The CNN model utilizes a straightforward architecture with multiple convolutional and pooling layers followed by dense layers for classification, serving as a baseline. The CNN Transformer model integrates CNNs with Transformer blocks, where CNN-extracted features are fed into Transformer blocks applying multi-head attention mechanisms for improved classification. The CNN Hybrid model combines ResNet50 and EfficientNetB0 architectures, freezing their convolutional bases and concatenating features, followed by global average pooling and dense layers for final classification. Each approach aims to optimize feature extraction and classification performance for skin cancer detection.

4.1 Deep Convolutional Neural Network

Deep Convolutional Neural Networks (CNNs) have revolutionized the field of image recognition and classification. These networks are particularly adept at automatically and adaptively learning spatial hierarchies of features from input images. A CNN typically consists of multiple layers, including convolutional layers, pooling layers, fully connected layers, and dropout layers, each playing a crucial role in feature extraction and classification. The architecture described here employs a sequential model starting with a rescaling layer to normalize the input data followed by a series of convolutional and max-pooling layers which progressively capture spatial hierarchies and abstract features from the input images. Dropout layers are interspersed to prevent overfitting by randomly disabling certain neurons during training. Finally, the network is flattened and passed through fully connected layers to produce the output probabilities for classification.

The model trained on the given dataset, which includes 9 classes, demonstrated a progressive improvement in both training and validation accuracy over the epochs. Initially, the model starts with an accuracy of 21.37% on the training set and 21.70% on the validation set. As the training progresses, significant improvements are observed, with the final training accuracy reaching 85.10% and the validation accuracy at 50.11%. The model's loss metric follows a similar trend, decreasing from an initial loss of 2.0509 to 0.3716 on the training set and showing fluctuating but overall decreasing patterns on the validation set. The convolutional layers with increasing filter sizes (16, 32, 64, 128) capture diverse features, while the dropout and max-pooling layers effectively handle

overfitting and dimensionality reduction. This architecture ensures a balance between model complexity and generalization, achieving a robust performance on the image classification task.

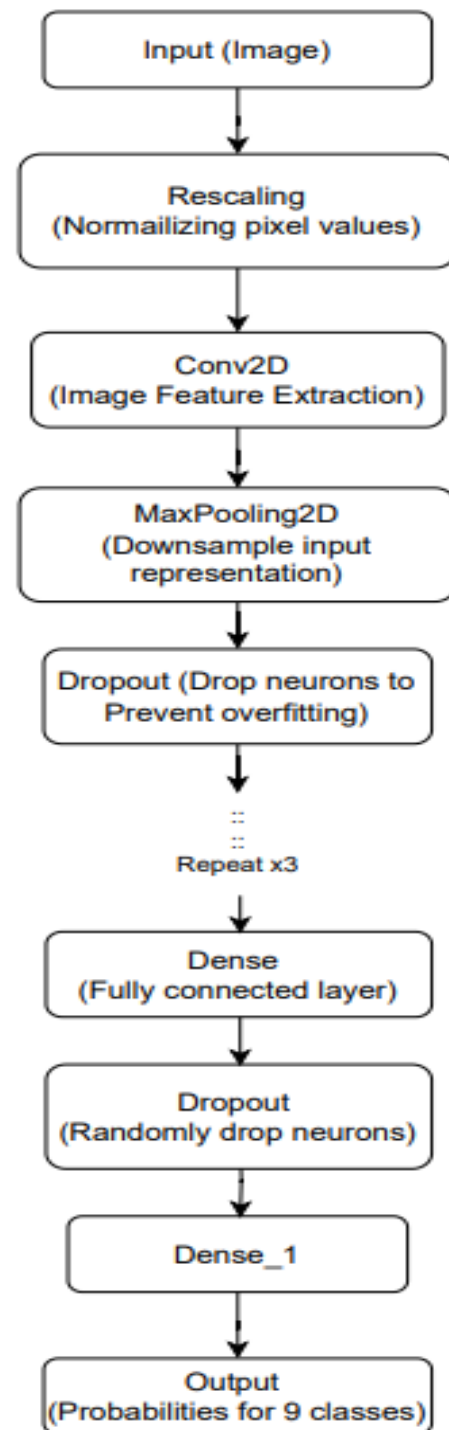


Figure 2: Flow diagram of Deep CNN

4.2 CNN-Transformer Model

The integration of CNN and Transformer architectures in the current deep learning model yielded promising improvements in both accuracy and loss metrics over a 30-epoch training period. The model starts with a series of convolutional layers to extract spatial features, followed by a transformer block to capture sequential dependencies and global context. Initially, the accuracy saw a significant increase from 19.59% to approximately 67.02% by epoch 17. Although the accuracy slightly declined towards the later epochs, peaking at 90.01%, the model maintained a higher overall performance compared to traditional deep CNNs. Similarly, the loss value decreased steadily from an initial 2.1224 to around 0.2344 by the final epoch, indicating a well-fitted model with reduced prediction errors.

This model comprises several hidden layers, each contributing uniquely to its performance. The initial layers consist of Conv2D and MaxPooling2D, designed to handle spatial feature extraction. With increasing complexity, the model includes dense layers followed by a reshaping layer to prepare the data for the transformer block. The Multi-Head Attention mechanism within the transformer block plays a critical role in capturing intricate patterns and dependencies across the entire input space. Subsequent layers include LayerNormalization, which stabilizes training and accelerates convergence, followed by an additional dense layer that refines the feature representation before the final classification. This combination of CNN and Transformer layers facilitates the extraction of both local and global features, enhancing the model's accuracy and robustness.

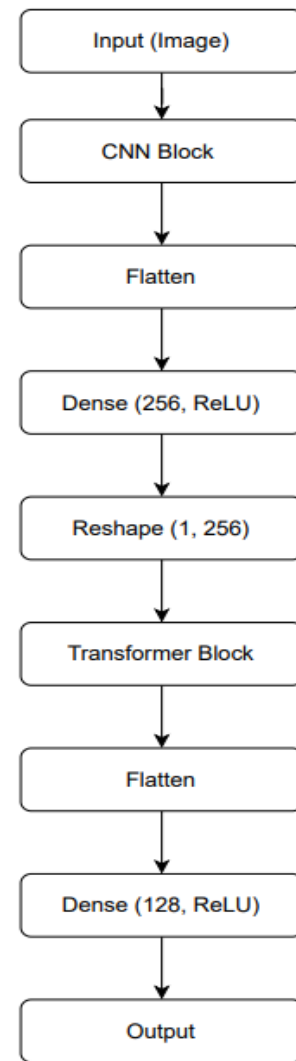


Figure 3: Flow diagram of CNN- transformer model

4.3 CNN Hybrid Model

The integration of ResNet50 and EfficientNetB0 models into our deep convolutional neural network has yielded promising results in improving the classification accuracy and reducing the loss values. The model architecture leverages the strengths of both pre-trained networks by concatenating their feature maps, followed by global average pooling and a dense layer with dropout for robust classification. During the 30 epochs of training, the model demonstrated

consistent improvement in accuracy, starting at 38.45% and reaching up to 85.99%. Concurrently, the loss values significantly decreased from 1.87 in the initial epoch to 0.35 by the final epoch, indicating effective learning and convergence.

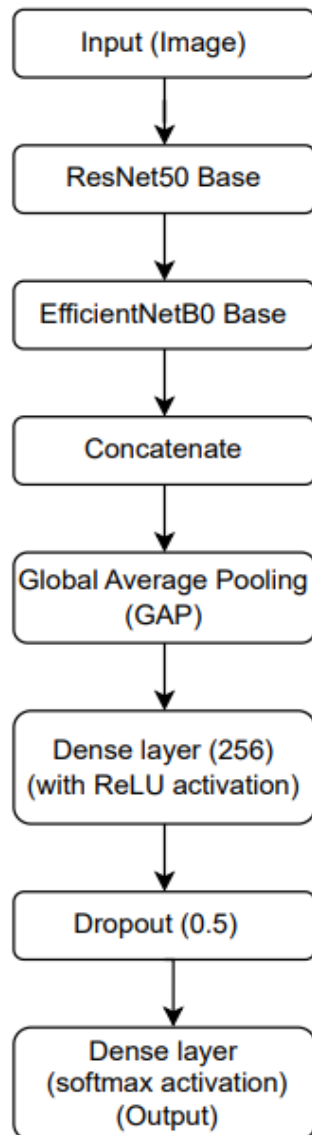


Figure 4: CNN Hybrid model

The inclusion of a 256-unit dense hidden layer with ReLU activation and a dropout rate of 0.5 contributed to enhancing the model's generalization capabilities, preventing overfitting while maintaining high accuracy.

This strategic architectural choice, combined with the powerful feature extraction capabilities of ResNet50 and EfficientNetB0, enabled the model to achieve a peak validation accuracy of 63.31%, with a steady validation loss around 1.29. Overall, the model's performance highlights the effectiveness of combining advanced convolutional architectures with a well-designed classification head for robust image classification tasks.

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 100, 100, 3)	0
resnet50 (Functional)	(None, 4, 4, 2048)	23,587,712
efficientnetb0 (Functional)	(None, 4, 4, 1280)	4,049,571
concatenate (Concatenate)	(None, 4, 4, 3328)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 3328)	0
dense (Dense)	(None, 256)	852,224
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 9)	2,313

Total params: 28,491,820 (108.69 MB)

Trainable params: 854,537 (3.26 MB)

Non-trainable params: 27,637,283 (105.43 MB)

Figure 5: Model Summary of CNN Hybrid model

5. PERFORMANCE EVALUATION

The performance evaluation of the models reveals varying degrees of effectiveness in skin disease classification. The CNN model's training accuracy provides a baseline for comparison, with its validation accuracy serving as a benchmark to evaluate the effectiveness of hybrid models and its validation loss used to assess improvements over the standard approach. The CNN Transformer model starts with a training accuracy of 19.59% in Epoch 1, reaching 90.01% by Epoch 30, while its validation accuracy improves from 21.70% in Epoch 1 to 58.17% in Epoch 24, with validation loss

showing a general decrease over epochs, indicating better model performance. The CNN Hybrid model's training accuracy increases from 38.45% in Epoch 1 to 85.99% by Epoch 30, with validation accuracy ranging from 49.44% in Epoch 1 to 63.31% in Epoch 17, and its validation loss fluctuating but generally trending downward, indicating improved performance.

6. CONCLUSION

The CNN-Transformer model shows improved classification performance for skin cancer detection compared to a standard CNN with an accuracy of 90.01% whereas the accuracy of standard CNN is 81.92%. The CNN Transformer model enhances the model's ability to capture both local and global features, leading to improved accuracy in skin cancer classification. The baseline CNN model demonstrates a fundamental approach to skin cancer classification. While effective, its performance is significantly enhanced by incorporating advanced techniques like hybrid models and Transformer blocks, as evidenced by the improved metrics in the more complex models.

References:

- [1] Hossin, M. A., Rupom, F. F., Mahi, H. R., Sarker, A., Ahsan, F., & Warech, S. (2020, October). Melanoma skin cancer detection using deep learning and advanced regularizer. In *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 89-94). IEEE.
- [2] Ashraf, R., Kiran, I., Mahmood, T., Butt, A. U. R., Razzaq, N., & Farooq, Z. (2020, November). An efficient technique for skin cancer classification using deep learning. In *2020 IEEE 23rd International Multitopic Conference (INMIC)* (pp. 1-5). IEEE.
- [3] Rezaoana, N., Hossain, M. S., & Andersson, K. (2020, December). Detection and classification of skin cancer by using a parallel CNN model. In *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)* (pp. 380-386). IEEE.
- [4] Gururaj, H. L., Manju, N., Nagarjun, A., Aradhya, V. N. M., & Flammini, F. (2023). DeepSkin: a deep learning approach for skin cancer classification. *IEEE Access*.
- [5] Malo, D. C., Rahman, M. M., Mahbub, J., & Khan, M. M. (2022, January). Skin cancer detection using convolutional neural networks. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0169-0176). IEEE.