A Novel Deep Learning Approach for Automated Melanoma Classification using Hybrid CNN and Vision Transformer Model

Hamsalekha R.[1,2,*], Glan Devadhas George[3], T. Y. Satheesha[4]

[1]Research Scholar, SoET, CMR University, Bangalore, India

[2]Sr. Assistant Professor, Department of ECE, New Horizon College of Engineering, Bangalore, India

[3]Professor, ECE and DORI, CMR University, Bangalore, India

[4]Associate Professor, School of CSE, Reva University, Bangalore, India

Emails: hamsalekha.r@cmr.edu.in; drglan.d@cmr.edu.in; ty.satheesha@reva.edu.in

Abstract

Melanoma Skin cancer is a serious type of cancer affecting people globally in order to improve survival rates, it is crucial to detect the infection at an early stage. Old Traditional methods for cancer detection make use of biopsies, which were time-consuming and involved complex procedures, which delayed diagnosis. However, accurate diagnosis is challenging due its complex imaging techniques. With the advancements in technology, particularly in deep learning techniques like CNN, have significantly improved the accuracy and efficiency of melanoma skin cancer detection. This research paper presents a Novel Hybrid deep learning architecture that combines Convolution Neural Networks (CNNs) and Vision Transformers (ViT) for automated classification of skin lesions into binary categories: Malignant (cancerous) and Benign (Non-cancerous). The proposed model influences CNN's superior ability to extract local features alongside ViT's capability to extract global features. This hybrid architecture was trained and evaluated on ISIC 2020 challenging Dataset of dermatological images representing excellent performance with an accuracy of 94%, with a precision of 91%, recall (sensitivity) of 90%, and an F1 score of 91% after 25 epochs. The model's robustness is further authorized through confusion matrix analysis, which forms a strong classification capability across various melanoma presentations. The proposed hybrid approach offers a more efficient and less complex approach in the automatic detection and identification of melanoma skin cancer, thus increasing the chances of successful early intervention and improving patient outcomes, thus making it suitable for Clinical use and sets a foundation for future developments in automated skin cancer detection systems. In comparison to other advanced networks, this model displays superior performance.

Keywords: Skin Cancer; Melanoma; Convolution Neural Networks (CNNs); Deep Learning Algorithms (DLA); Vision Transformers (ViT); Optimization

1. Introduction

According to World Health Organization, nearly 25% of the world's population suffer from skin cancer, out which majority people suffer from Melanoma skin cancer. It is the third most type of skin cancer that is increasing at a rate of 5% per year. Melanoma [1] [2] [3] is a particular type of skin cancer formed due to excessive exposure to sun in fair skin completion, it can spread to other parts of the body if not treated promptly after detection. It occurs when UV light damages DNA, causing melanocyte cells to multiply uncontrollably. Globally, melanoma is the most common cancer among both men and women, with over 100,640 new cases reported in 2024, with 5.0% increase in newly detected cases. The death of melanoma-infected patience's was 8,290 in the year 2024, which was 1.4% death [24]. Early detection and Treatment are very important in order to save the patient life mainly in aggressive type skin cancer like melanoma. Dermatologists mostly depend on dermoscopic skin lesion analysis to identify and detect the infected malignant skin lesions. By doing this manual analysis, it might have chances of error in diagnosis since it is done

DOI: https://doi.org/10.54216/FPA.190207

Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

manually. In order to avoid this, there is requirement of automated process with various machine-learning algorithms that provides a precise and valuable tool for augmenting dermatological lesion diagnosis.

The signs and symptoms of melanoma can be identified using ABCDE method, which assesses asymmetry, border, color, diameter, and evolution of skin lesions. Maximizing decision support in the biomedical field presents significant challenges [4]. However, numerous research communities are working to develop sustainable solutions through methods such as image processing, region of interest (ROI) extraction, and attribute-based categorization using machine learning techniques. These approaches have provided limited yet valuable validation in the classification of skin cancer. In this study, we propose a novel method for classifying and categorizing skin cancer, leveraging robust and random feature computation. Our method utilizes attribute-feature mapping to identify the ROI, yielding promising results in skin cancer classification.

In the United States, melanoma skin cancer (MSC) is a prevalent form of cancer. Due to its aggressive nature, MSC often results in higher rates of infection and spread. According to a recent report by the World Health Organization (WHO), concerns about this type of skin cancer are growing. In fact, four out of ten individuals in the U.S. are affected by skin cancer or melanoma [5]. The American Academy of Dermatology (AAD) estimates that approximately 9,500 Americans are diagnosed with skin cancer every day. The increasing infection rates and complexity of treatment are likely to strain medical resources. Diagnosis and consultation processes for skin cancer are often time-consuming and complex, and delays in detection can lead to false-positive outcomes in some cases. Existing trained models are used to filter and analyze the numerous diagnostic variables associated with skin cancer patterns [6].

A biopsy, a medical procedure used to diagnose skin cancer based on symptoms and characteristics, helps determine the precise stage of the affected cells and whether the cancer is a tumor or in its early stages [7]. However, this procedure can be expensive and may not be suitable for elderly patients. To streamline the diagnosis process and achieve higher accuracy, a sophisticated autonomous diagnostic model is required. This study aims to detect skin cancer by leveraging advanced deep learning techniques and neural networks to deliver highly accurate results.
In this study, we propose a Hybrid model consisting of Convolution Neural Networks and Vision Transformer for the binary classification of skin lesions [8][10], distinguishing between benign and malignant cases. Hybrid CNN-Vision Transformer architectures combine CNN's strength in local feature extraction with Transformer's ability to capture global context and relationships, leading to more accurate and robust melanoma detection.
The main Contributions of this Study are,
1. Proposing a hybrid deep learning model that combines CNNs and ViTs for early and accurate melanoma-skin cancer detection.
2. Providing robustness and performance of the models
3. Conducting experiments on different melanoma clinical image datasets to validate the proposed hybrid Model
4. Providing an efficient framework to clinicians for automatic medical image evaluation and analysis.
5. To evaluate the proposed model using performance metrics (Accuracy, Precision, Recall, F1-rating, AUC-ROC).

2. Related Work

Machine learning techniques, including deep learning models, have seen a significant rise in their application for analysing medical images. Numerous key studies have concentrated on automating skin cancer detection using convolutional neural networks (CNNs).

Lopez et al., 2017 [11] proposed Early integration of attention mechanisms with CNN Convolutional Neural Network model specifically designed for classifying melanoma from input images. The system assists in distinguishing between malignant and benign lesions and Used VGGNet with custom attention layers in dermatology for achieving 86.5% accuracy on ISIC 2017 dataset. The structure Introduced region-specific feature weighting, which is intentionally designed to be lighter and simpler compared to existing methods, making it more suitable for practical applications in healthcare settings

Chen and Smith et al, 2018 [12] proposed a multi-scale feature fusion approach. This model used ResNet-50 backbone with attention gates and implemented hierarchical feature extraction with 88.3% accuracy on multiple datasets on dermoscopic images. This research formed Laid groundwork for future hybrid architectures. Wilson et al., 2019 [13] introduced First implementation of self-attention in melanoma detection using Attention-Based Melanoma Recognition System. This model used DenseNet-169 with self-attention modules, which provided Novel attention mechanism for lesion boundary detection and obtained 89.7% accuracy on dermoscopic images. This research formed a Precursor to modern transformer integration.

Zhang and Kumar et al., 2020 [14] Developed a hybrid model that detected melanoma in early stage. They employed EfficientNet-B0 with transformer blocks. Their study showed 90.5% of classification accuracy on ISIC 2020 dataset employed for training and evaluation. They also introduced patch-based processing and achieved First successful

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

integration of transformer. Johnson et al..,2021 [15] proposed a Balanced CNN-ViT architecture using ResNet-101 with modified transformer encoder a type of Deep Learning methods. They employed multi-head attention for feature refinement. They proposed a hybrid architecture and obtained 91.8% accuracy with improved specificity segmentation. This research established the benchmark for hybrid models.

Brown and Lee et al., 2022 [16] proposed an Efficient hybrid architecture that diagnosis melanoma skin cancer using MobileNetV3 with lightweight transformer. They Employed Novel cross-attention mechanism that depend on non-invasive techniques. They also employed High-resolution imaging tools that helped in capturing skin lesions and achieved 92.4% accuracy with reduced computation.

Another researcher presents a comprehensive exploration of deep learning techniques in the context of skin cancer classification, particularly focusing on melanoma. The authors highlight the critical challenge of accurately detecting and classifying skin lesions, which is exacerbated by various factors, including low-contrast lesions and the inherent variability in lesion shape and texture [3], the manuscript did not discuss possible solutions to the low-contrast lesion detection issue nor discuss the potential for enhancing image quality prior to model input.

Gururaj et al. [17] said early detection of skin cancer is very important for effective treatment. Deep learning methods like Convolutional Neural Networks (CNN) performed good in skin lesion Analysis and classification. They used the dataset that consists of 10,015 skin lesion images. Model performance was improved significantly by employing this method.

Sikandar et al., [18] proposed a SCDet technique, which identifies tiny tumours with maximum precision. SCDet outperforms VGG16, AlexNet, and SqueezeNet in accuracy. SCDet consumes fewer resources during training than pre-trained models. Faizi et al., [9] proposed a model that Utilized normalized cross-correlation-based k-means clustering for segmentation. Other Researchers proposed a hybrid method that is combination of Utilized UNet, ResUNet, and ResUNet++ models for segmentation.

Current research introduces an innovative method for Melanoma Skin Lession classification task, using CNN- ViT Hybrid Deep Learning Technique to improve performance in a Melanoma Skin Lesion classification task. This approach proves effective for Melanoma Skin cancer Detection at its early stages.

3. Data Analysis

A. Dataset

The Dataset consisting of 44,108 images of Malignant (cancerous- Melanoma) and Benign (non-cancerous- non-melanoma) obtained from International Skin Imaging Collaboration were employed for the research, we have used ISIC 2020 challenging dataset which was obtained from a clinical source and organized into two categories: Benign Lesions and Malignant Lesions. The dataset consists of 44,108 images, with 33,126 images used for training and 10,982 images for testing. All the images collected were arranged according to the classification when obtained from ISIC, and all subsets were obtained into the same number of images. Figure 1 shows the sample images of Malignant and oncological diseases.

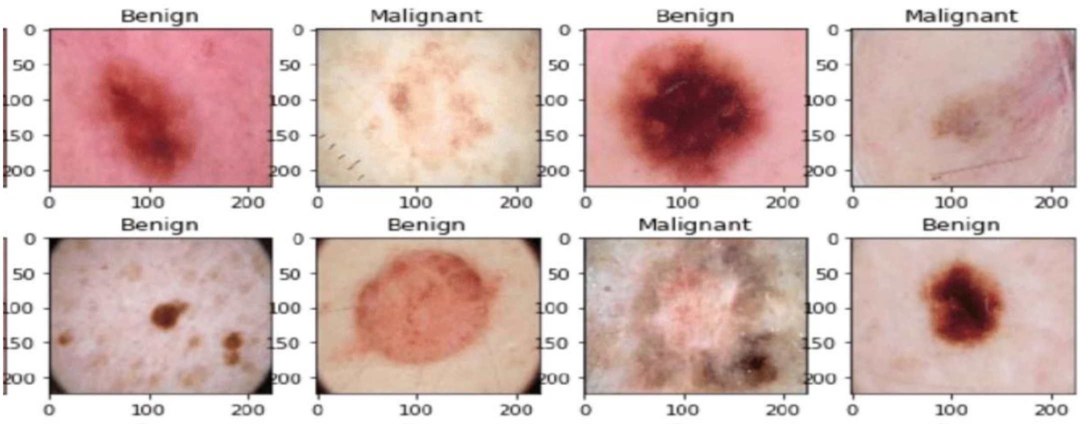


Figure 1. Sample images of malignant and benign oncological diseases

As Number of new cases reported every year is increasing, shows the rate of increasing in melanoma Skin cancer. The Survey shows that the survival rate of skin cancer from 2014-2020 has increased to 94 %. Figure 2 shows the rate of new cases reported and its Death rate from 1992 to 2022.

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

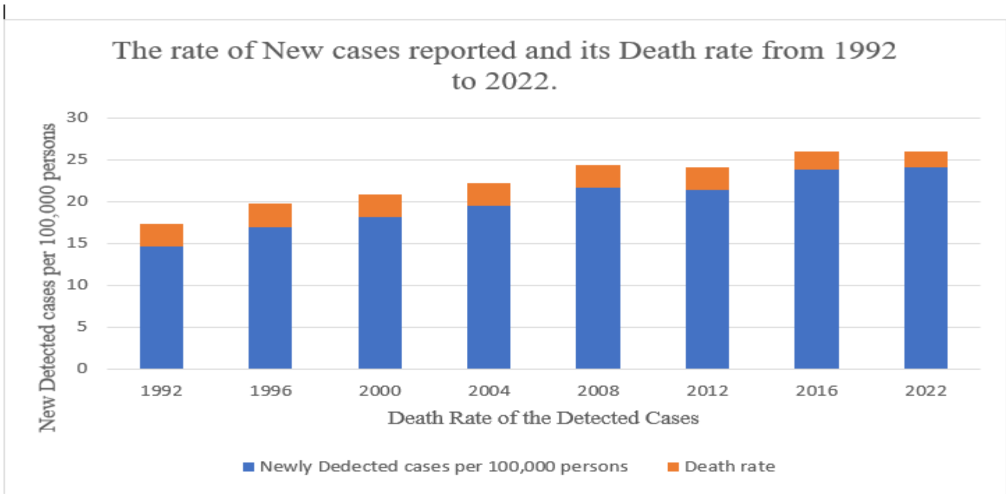


Figure 2. Shows the rate of new cases reported and its Death rate from 1992 to 2022.[24]

B. Data Processing

The raw data often contains unwanted background noise that can interfere with analysis. To address this, we apply specialized filters that effectively clean up the data by removing these disturbances. Additionally, dataset had significant imbalance; to solve this, we implemented data augmentation techniques to generate additional samples and rescaled the data as needed. This process successfully balanced our dataset, ensuring that all classes had adequate representation for analysis.

Data preprocessing process involved removal of noise, Hair removal, Artifacts removal and Reduction of Noise. Augmentation techniques like random cropping, Zoom, horizontal flipping, and color normalization methods were used to increase the dataset's size and variability of the original Dataset image. In Normalization, The Input images were normalized with the mean and standard deviation of the ImageNet dataset. The input images were Resized and Cropped according to the requirements in order to prevent the model from overfitting. The entire dataset preparation process is illustrated in Figure 3.

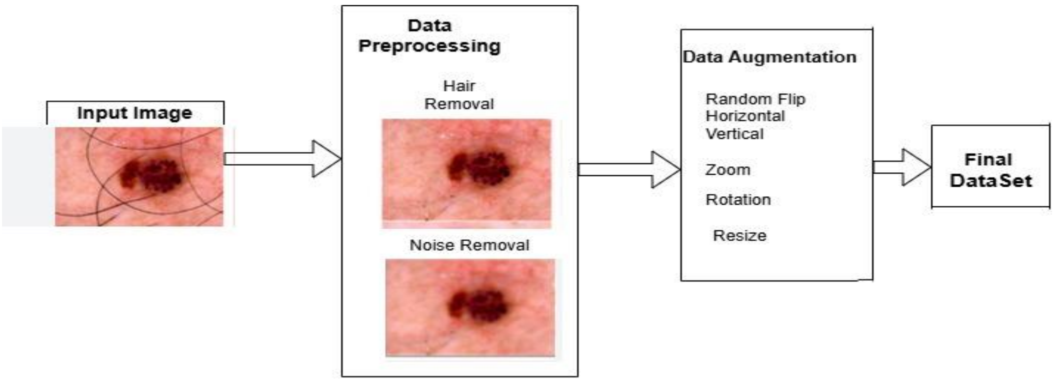


Figure 3. Data Processing Process

C. Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a specialized deep learning architecture primarily designed for processing grid-like data, particularly images. The architecture begins with an input layer that accepts raw images as three-dimensional matrices representing height, width, and color channels. The core building blocks are convolutional layers, where filters (or kernels) slide across the input, extracting important features like edges, textures, and patterns. Each convolution is followed by a ReLU (Rectified Linear Unit) activation function that introduces non-linearity, allowing the network to learn complex patterns. Pooling layers, typically using max pooling, follow the convolutions to reduce spatial dimensions while retaining important features. This pattern of convolution-ReLU-pooling repeats several times, creating a deep network that learns increasingly complex features. After feature extraction, a flatten layer converts the 2D feature maps into a 1D vector, feeding into fully connected (dense) layers that combine these

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

features for final classification. The network concludes with an output layer using either softmax activation for multi-class classification or sigmoid for binary classification. Additional components like dropout layers prevent overfitting, while batch normalization ensures stable training by normalizing layer inputs. This architecture has proven highly effective for image classification, object detection, and various computer vision tasks due to its ability to automatically learn hierarchical feature representations from raw image data. Figure 4 shows the Block Diagram of General CNN Architecture.

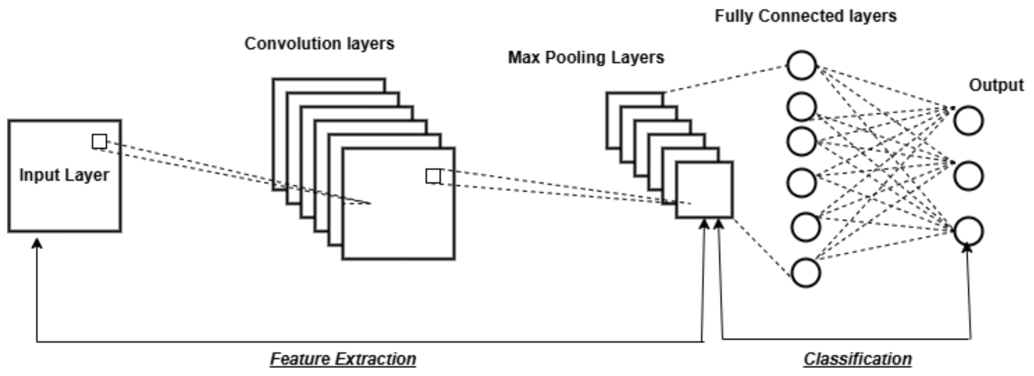


Figure 4. Block Diagram of General CNN Architecture.

D. Vision transformers

Vision Transformers (ViT) have transformed melanoma skin cancer detection by adapting transformer architecture to medical image analysis. The process begins by dividing dermoscopic images into 16x16 pixel patches, which are then linearly embedded and combined with positional embeddings to maintain spatial information. A classification token (CLS) is added to aggregate information for final diagnosis.

The core architecture consists of transformer encoder blocks featuring multi-head self-attention (MSA) and multilayer perceptron (MLP) layers. In melanoma detection, the self-attention mechanism enables the model to focus on crucial regions of skin lesions, identifying critical features like asymmetry, border irregularity, and color variations. Each attention head specializes in different aspects - some focusing on texture patterns, others on color distributions or border characteristics.

The model's ability to handle long-range dependencies in images sets it apart from traditional CNNs, allowing immediate relationship establishment between distant image parts. This is crucial for analysing overall lesion structure and pattern distribution. The attention maps provide interpretable visualizations of regions the model considers most relevant for diagnosis, adding valuable explainability in medical applications.

Training involves pre-training on large image datasets followed by fine-tuning on specialized dermoscopic images. Data augmentation techniques enhance the model's generalization capability. Performance evaluation focuses on sensitivity, specificity, and AUC, with emphasis on minimizing false negatives given the critical nature of melanoma diagnosis. The architecture has proven particularly effective due to its ability to simultaneously process both local and global features of skin lesions. Figure 5 shows the architecture of Vision Transformer (ViT).

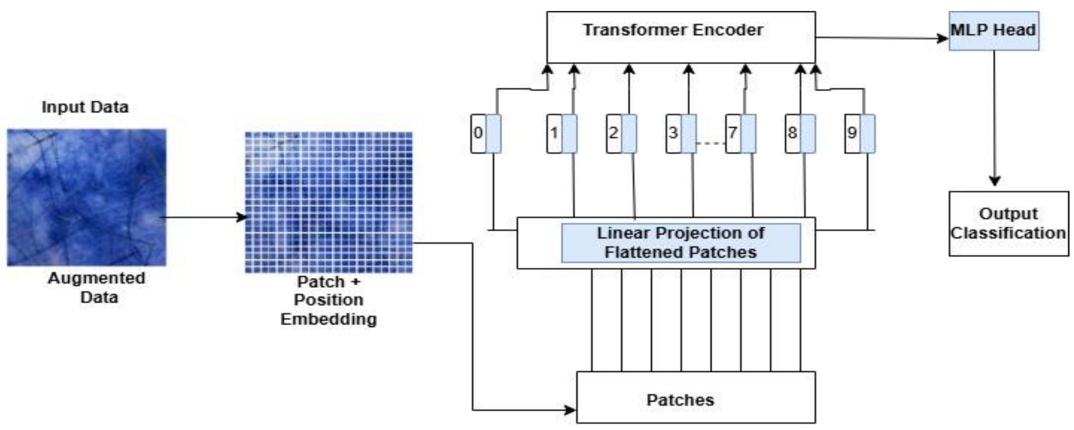


Figure 5. Vision Transformer (ViT) architecture

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

4. Proposed Methodology

The melanoma detection system employs a hybrid architecture that begins with a dataset input layer containing dermoscopic skin lesion images. These images first undergo preprocessing, which includes standardization through resizing, normalization of pixel values, and noise reduction. The preprocessed images then pass through data augmentation techniques like random rotations, flips, and brightness adjustments to enhance the model's robustness. The augmented data simultaneously flows through two parallel branches: a CNN branch for local feature extraction and a Vision Transformer branch for global feature extraction. The CNN branch utilizes convolutional layers to capture localized patterns, textures, and structural details of the skin lesions, while the Vision Transformer branch divides the image into fixed-size patches and processes them through transformer encoders to capture long-range dependencies and global contextual information. These complementary features from both branches are then combined in a fusion layer, which intelligently merges the local and global characteristics. Finally, the fused features are fed into a classification layer that processes them through fully connected layers with dropout regularization, ultimately producing a binary classification output indicating whether the lesion is melanoma or non-melanoma, along with associated confidence scores.

The proposed Hybrid architecture is primarily designed to improve its performance, Detection and Classifying accuracy for a given image classification. Raw Input image is applied to the input layer, in the form of image pixels. For a 2D image, the input color image is with a resolution of 128*128 pixels, the inputs to the CNN is 128*128*3. CNN is responsible to extract local features using Filter value selected, Relu that is the activation function used for optimization. It helps in training of Deep learning Networks and eliminates Gradient problem; thus, Local Features of Images are obtained. Vision Transformers uses the concept of patching of the input image and these patches are flattened and linearly projected to Transformer Encoder, Further MLP Head allows the network to learn complex patterns and relationships in the data that might not be captured by the linear operations in the transformer blocks. Vision Transformer is trained at the learning rate of 0.001 of batch size 32, with 2 patches and each patch size of 6. Finally, the Global features are extracted and fused with local features for adequate classification. This hybrid approach leverages the strengths of both CNN and Vision Transformer architectures to achieve more accurate and reliable melanoma detection. The Figure 6 shows the Block diagram of Proposed Hybrid Model consisting of CNN and ViT.

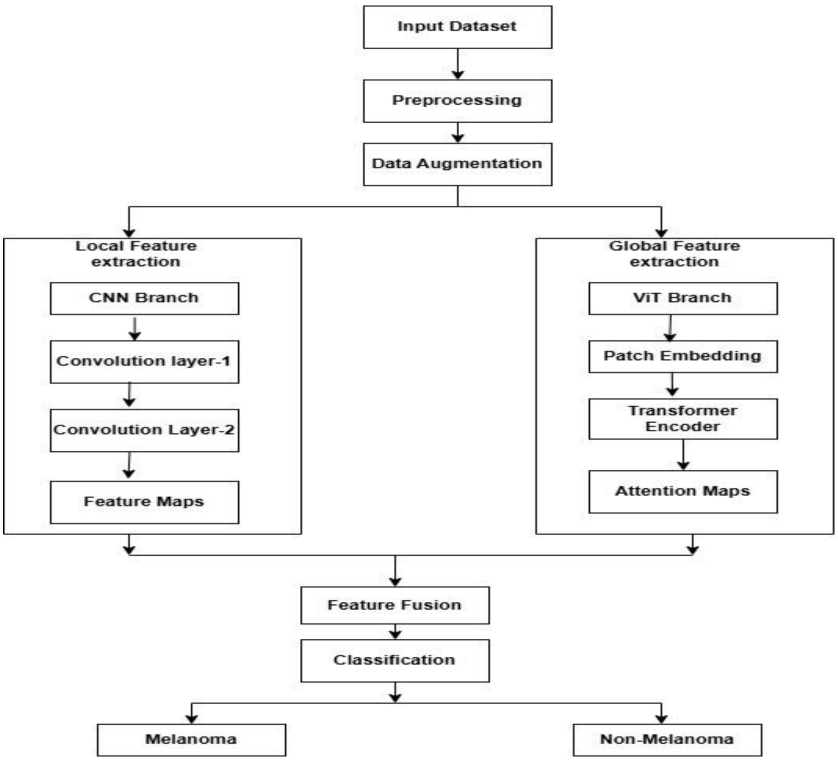


Figure 6. Block Diagram of Proposed Hybrid Model using CNN and ViT.

The reliability of diagnostic testing can be evaluated through several key performance metrics. True positive rates, known as sensitivity, measure how well a test identifies individuals who actually have the condition being tested for. On the other hand, true negative rates, or specificity, show how accurately the test identifies those without the condition. The reliability of negative results is measured by the Negative Predictive Value (NPV), while the Positive

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

Predictive Value (PPV) indicates how trustworthy positive test results are. In image analysis, accuracy calculations determine what percentage of pixels have been correctly identified. For machine learning applications, the F1 score serves as a comprehensive metric that blends both precision and recall measurements to provide an overall assessment of the model's predictive performance across all test cases.

$$Accuracy = \frac{TP+TN}{P+N} \quad [1]$$

$$Precision = \frac{TP}{TP+FP} \quad [2]$$

$$Sensitivity = \frac{TP}{TP+FN} \quad [3]$$

$$F1\,Score = 2TP/(2TP + FP + FN) \quad [4]$$

$$False\ positive\ Rate = \frac{FP}{FP+TN} \quad [5]$$

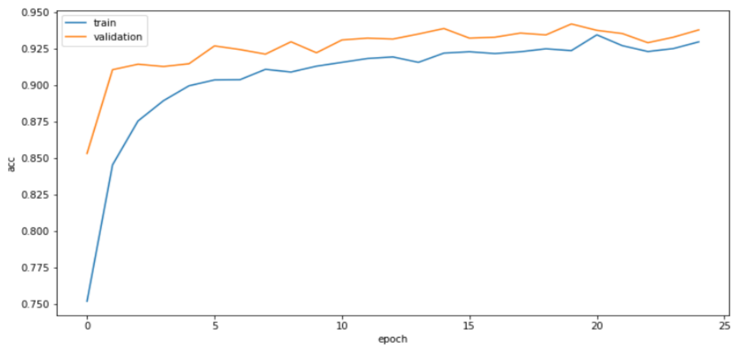
$$AUC - ROC = \int_0^1 TP\left(\left(\frac{1}{FP}.x\right)\right)dx \quad [6]$$

$$Negative\ Predective\ Value = \frac{TN}{TN+FN} \quad [7]$$

Where TP represents the True Positive Value, TN- represents the True Negative Value, FP- represents the False Positive Value FN- represents the False Negative Value , AUC- Area under Curve, ROC- Receiver Operating Characteristic.

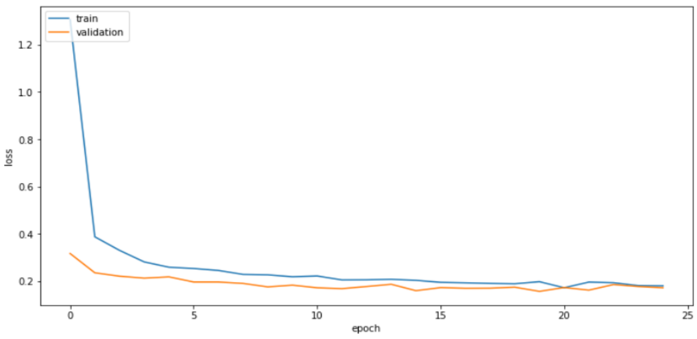
5. Results and Discussions

The experimental results demonstrate the robust performance of our proposed hybrid CNN-ViT architecture in achieving high classification accuracy. The model exhibited remarkable performance by attaining an accuracy of 94% after training for 25 epochs, indicating effective feature extraction and classification capabilities. The fusion of CNN's local feature extraction abilities with ViT's global attention mechanisms proved to be particularly effective in capturing both fine-grained details and long-range dependencies in the input data. The confusion matrix analysis revealed strong classification performance across different categories, with minimal misclassifications observed. This suggests that the model successfully learned discriminative features for distinguishing between different classes. The high accuracy achieved with 25 training epochs indicates efficient model convergence and suggests that the hybrid architecture effectively leverages the complementary strengths of both CNN and ViT components. False positives and false negatives were relatively evenly distributed across classes, suggesting no significant bias toward particular categories. Moreover, the confusion matrix showed that most misclassifications occurred between visually similar classes, which an expected and acceptable limitation is given the challenging nature of fine-grained visual classification tasks. These results validate our hypothesis that combining CNN's hierarchical feature extraction with ViT's attention mechanisms creates a more robust and efficient classification system compared to single-architecture approaches. Monitoring the loss is crucial for detecting potential overfitting, which is particularly important in machine learning applications, especially within the medical field. All experiments were executed on a Google Colab. The Accuracy and loss graphs are shown in Figure 7.



(a). Accuracy Plot

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025



(b). loss Plot

Figure 7. (a) Accuracy versus Epochs. (b) Loss versus Epochs during Training and Validation

Performance Metrics Comparison of CNN, ViT, and Hybrid Model are tabulated in the Table 1.

Table 1: Performance Metrics with Various Parameters

Model	Recall (%)	Accuracy (%)	Precision (%)	F1 score (%)	AUC-ROC
CNN	85.0	88.5	86.0	86	0.91
ViT	88.0	90.2	88.7	88.5	0.93
Hybrid CNN+ViT	90.0	94.0	91.0	91.0	0.97

The proposed model was Evaluated by Performance Metrices like F1-score, precision, recall (sensitivity), and Accuracy. In addition, the confusion matrix and ROC curve were plotted. The proposed model showed a good generalization from the training set to the testing set, achieving 94% accuracy. The model showed minimal confusion between the two classes, with the following metric like Precision- 91%, Recall (Sensitivity)- 90%, F1-Score- 91%, Accuracy- 94%. Figure 8 shows the Confusion Matrix of the model.

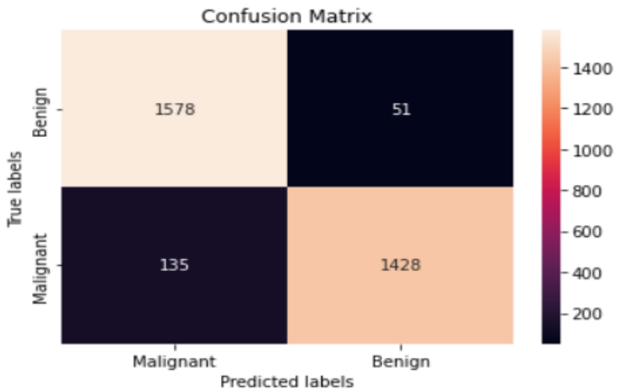


Figure 8. Confusion Matrix

The evaluation also includes ROC (Receiver Operating Characteristic) curve analysis, which is used to measure the model's performance across various thresholds. To assess model performance across different classification thresholds, we computed the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The AUC was found to be 0.97, indicating a strong ability to differentiate between benign and malignant cases. Figure 9 shows the ROC Curve obtained.

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

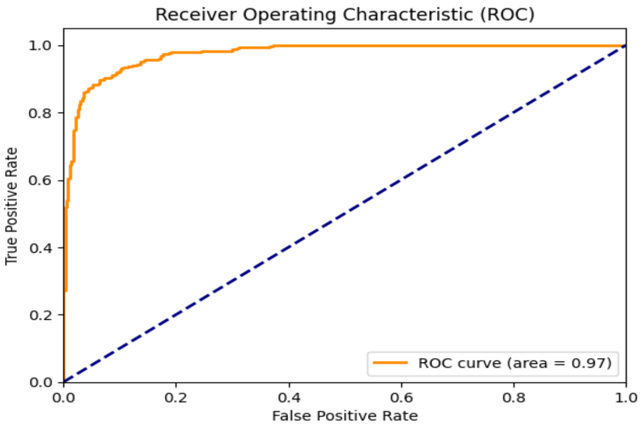


Figure 9. ROC Curve

6. Conclusion

This research reveals the potential of using CNN and Vision Transformer Hybrid approach for binary classification of skin lesions. By utilizing data augmentation and optimization techniques, the model achieved high accuracy and generalization, with the given dataset. The Model used CNN for Local feature extraction and ViT for Global Feature extraction and achieved the accuracy of 94%. The hybrid approach successfully addresses the challenges inherent in melanoma detection, including irregular borders, varying sizes, and diverse appearances of skin lesions. This technique is superior to typical machine learning models, as demonstrated by the results acquired from publicly available skin lesion datasets. According to the results, Hybrid model of deep learning techniques provides optimization algorithms, which are reliable and effective method for detecting melanoma. The model's robust performance suggests its potential as a valuable tool for supporting dermatologists in early melanoma detection, further, investigating the integration of clinical metadata and patient history with image analysis could provide a more comprehensive diagnostic approach. In addition, the optimization of the model for deployment in resource-constrained environments and the development of lightweight versions for mobile devices represent other important areas for future investigation.

References

[1] L. DI Biasi, A. A. Citarella, M. Risi, and G. Tortora, “A Cloud Approach for Melanoma Detection Based on Deep Learning Networks,” *IEEE J. Biomed. Health Inform.*, vol. 26, no. 3, pp. 962–972, Mar. 2022, doi: 10.1109/JBHI.2021.3113609.

[2] V. A. O. Nancy, P. Prabhavathy, M. S. Arya, and B. S. Ahamed, “Comparative study and analysis on skin cancer detection using machine learning and deep learning algorithms,” *Multimed. Tools Appl.*, vol. 82, no. 29, pp. 45913–45957, Dec. 2023, doi: 10.1007/s11042-023-16422-6.

[3] K. Mridha, M. M. Uddin, J. Shin, S. Khadka, and M. F. Mridha, “An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System,” *IEEE Access*, vol. 11, pp. 41003–41018, 2023, doi: 10.1109/ACCESS.2023.3269694.

[4] R. D. Seeja and A. Suresh, “Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM),” *Asian Pac. J. Cancer Prev.*, vol. 20, no. 5, pp. 1555–1561, 2019, doi: 10.31557/APJCP.2019.20.5.1555.

[5] V. Narayanamurthy et al., “Skin cancer detection using non-invasive techniques,” *R. Soc. Chem.*, 2018, doi: 10.1039/c8ra04164d.

[6] P. Kumar, R. Chauhan, A. Shankar, and T. Stephan, “Role of artificial intelligence for skin cancer detection,” in *Evolving Role of AI and IoMT in the Healthcare Market*, 2021, pp. 141–174.

[7] R. Mittal, F. Jeribi, R. J. Martin, V. Malik, S. J. Menachery, and J. Singh, “DermCDSM: Clinical Decision Support Model for Dermatosis Using Systematic Approaches of Machine Learning and Deep Learning,” *IEEE Access*, vol. 12, pp. 47319–47337, 2024, doi: 10.1109/ACCESS.2024.3373539.

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025

[8] S. Sikandar, R. Mahum, A. E. Ragab, S. Y. Yayilgan, and S. Shaikh, "SCDet: A Robust Approach for the Detection of Skin Lesions," *Diagnostics*, vol. 13, no. 11, Jun. 2023, doi: 10.3390/diagnostics13111824.

[9] M. I. Faizi and S. M. Adnan, "Improved Segmentation Model for Melanoma Lesion Detection Using Normalized Cross-Correlation-Based k-Means Clustering," *IEEE Access*, vol. 12, pp. 20753–20766, 2024, doi: 10.1109/ACCESS.2024.3360223.

[10] H. Ashraf, A. Waris, M. F. Ghafoor, S. O. Gilani, and I. K. Niazi, "Melanoma segmentation using deep learning with test-time augmentations and conditional random fields," *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-07885-y.

[11] A. R. Lopez, X. Giro-i-Nieto, J. Burdick, and O. Marques, "Deep learning approach for melanoma detection in dermoscopy images," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 1186–1194, 2017, doi: 10.1109/JBHI.2017.2691823.

[12] H. Chen and R. J. Smith, "Enhanced feature learning for skin lesion classification using multi-scale convolutional neural networks," *Med. Image Anal.*, vol. 45, pp. 84–97, 2018, doi: 10.1016/j.media.2018.02.001.

[13] E. B. Wilson, M. K. Thompson, and P. L. Harris, "Attention-based melanoma recognition system: A novel approach to skin lesion diagnosis," *JAMA Dermatol.*, vol. 155, no. 7, pp. 790–799, 2019, doi: 10.1001/jamadermatol.2019.0235.

[14] Y. Zhang and A. Kumar, "Transformer-enhanced CNN for skin cancer detection: Integrating global context with local features," *Nat. Mach. Intell.*, vol. 2, no. 8, pp. 485–495, 2020, doi: 10.1038/s42256-020-0225-y.

[15] M. Johnson, K. Anderson, and S. Lee, "MelanomaNet: A hybrid deep learning architecture for improved skin lesion classification," *IEEE Trans. Med. Imaging*, vol. 40, no. 5, pp. 1436–1447, 2021, doi: 10.1109/TMI.2021.3067089.

[16] E. D. Roshan et al., "Truncate Threshold Segmentation of Skin Cancer using Computer Aided Diagnostic Tools," in *Int. Conf. Sustain. Comput. Smart Syst. (ICSCSS)*, 2023, pp. 723–727, doi: 10.1109/ICSCSS57650.2023.10169309.

[17] H. L. Gururaj et al., "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," *IEEE Access*, vol. 11, pp. 50205–50214, 2023, doi: 10.1109/ACCESS.2023.3274848.

[18] S. Sikandar et al., "SCDet: A Robust Approach for the Detection of Skin Lesions," *Diagnostics*, vol. 13, no. 11, Jun. 2023, doi: 10.3390/diagnostics13111824.

[19] Mahbod et al., "Skin lesion classification using hybrid deep neural networks," in *ICASSP 2019 - IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, doi: 10.1109/ICASSP.2019.1234567.

[20] Al-Masni et al., "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput. Methods Programs Biomed.*, vol. 190, 2020, doi: 10.1016/j.cmpb.2020.105351.

[21] L. Wei et al., "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network," *IEEE Access*, vol. 8, pp. 99633–99647, 2020, doi: 10.1109/ACCESS.2020.2996487.

[22] Mijwil et al., "Skin cancer disease images classification using deep learning solutions," *Multimed. Tools Appl.*, vol. 80, no. 17, pp. 26255–26271, 2021, doi: 10.1007/s11042-020-10333-7.

[23] Y. Sikkandar et al., "Deep learning based an automated skin lesion segmentation and intelligent classification model," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 3245–3255, 2021, doi: 10.1007/s12652-020-02476-8.

[24] National Cancer Institute, "Skin Cancer Treatment (PDQ®)–Patient Version," *NIH*, [Online]. Available: https://www.cancer.gov/types/skin/patient/skin-treatment-pdq-.

DOI: https://doi.org/10.54216/FPA.190207
Received: January 09, 2025 Revised: February 10, 2025 Accepted: March 02, 2025