American Journal of Psychiatric Rehabilitation

Expert Opinion Article

AI-Powered Skin Cancer Detection Using A CNN-Transformer Hybrid Model

T. Maheshselvi[1*], V.Bharathiraja[2], R.Bragadheesh[3], S.Harishvijayabaskaran[4]

[1*]Department of Computer Science and Engineering, University College of Engineering, Thirukkuvalai (A Constituent College of Anna University, Chennai) Nagapattinam, India thanumaheshselvi@gmail.com
[2]Department of Computer Science and Engineering, University College of Engineering, Thirukkuvalai (A Constituent College of Anna University, Chennai) Nagapattinam, India bharathiraj7890@gmail.com
[3]Department of Computer Science and Engineering, University College of Engineering, Thirukkuvalai (A Constituent College of Anna University, Chennai) Nagapattinam, India r.bragadheesh834@gmail.com
[4]Department of Computer Science and Engineering, University College of Engineering, Thirukkuvalai (A Constituent College of Anna University, Chennai) Nagapattinam, India Harish2003dpm@gmail.com

Abstract
This study details the development and implementation of a deep-learning-based automated skin analysis skin scan application. A hybrid model combining EfficientNet and Vision Transformer (ViT) was proposed to increase the accuracy of lesion classification. While EfficientNet retrieves fine-grained spatial characteristics, ViT gathers global contextual links and performs better in terms of accuracy than ViTs and solo CNNs, especially when it comes to identifying difficult cases. The model was trained and validated using the HAM10000 dataset. This method reduces dependency on conventional dermatological treatments by providing an easy-to-use and accessible AI-driven tool for early skin cancer detection.

Keywords — Skin Cancer Detection, EfficientNet, Vision Transformer, Hybrid Deep Learning, Medical Image Analysis.

Introduction

Skin cancer is a malignant tumor that usually develops in the exposed body regions. It is the aberrant proliferation of skin cells caused by damaged DNA that has not been repaired. The exact etiology of this illness remains unknown. According to the World Health Organization (WHO), the number of instances of malignant skin cancer has dramatically increased over the last ten years. Early detection of skin cancer is essential because it allows symptom classification and allows professionals to choose the best course of action for the patient [1]. Melanoma and non-melanoma are the two main types of skin cancers that may be separated. If detected early, non-melanoma skin cancer may be treated with ease [12]. Although melanoma accounts for only 4% of all skin cancer cases, it is the most dangerous type and is responsible for 80% of skin cancer fatalities [13]. Existing methods, such as manual inspection based on ABCDE criteria, are subjective and incorrect [1] because dermatologists have different levels of experience and malignant skin lesions may not appear in a straight line. Thus, we have contributed to the development of a reliable model that can identify and categorize skin lesions in real time. Patient survival is positively correlated with early detection of malignant skin lesions. When skin cancer is detected early, significant medical cost savings are achieved. Early diagnosis can provide significant economic benefits to the country through shortened treatment times and lower treatment costs.

Furthermore, enhancing patients' quality of life through early detection and care would have a direct positive impact on society's welfare. The findings of this study make a substantial contribution to the scientific literature by shedding light on the potential applications of deep learning and computer vision methods in the field of medical image processing. Furthermore, the findings of this study will be useful to engineers and researchers in the domains of deep learning and computer vision. Finally, a wider audience will gain from the creation and application of such cutting-edge technologies as they will increase public access to healthcare services.

Literature Review

To detect skin cancer, several cutting-edge methods search for an accurate deep-learning model from a number of pretrained models. For instance, in [2], the DenseNet169 architecture was trained on the HAM10000 dataset, which has seven types of skin lesions, and it was able to obtain 92.25% accuracy and 93.59% recall (sensitivity). To categorize seven distinct forms of skin lesions, Sulthana et al. [3] presented S-MobileNet, an end-to-end deep convolutional neural-network-based skin lesion classification framework. The 10,000 dermatoscopic images from various populations in the HAM10000 dataset were used to train the algorithm. Their suggested S-MobileNet, which includes the Mish activation function, performed better than other models and showed promise in the classification of skin lesion photos.

In a thorough assessment of CNNs for melanoma detection, Haenssle et al. [4] achieved good sensitivity and specificity. Their results demonstrated that CNNs may improve dermatologists' diagnostic skills, especially when it comes to identifying high-risk lesions that require further investigation. Support Vector Machines (SVM) and Probabilistic Neural Networks (PNN) are used to detect melanoma. The results showed that the PNN is more effective than the SVM in identifying skin damage caused by melanoma [5]. U.O. Dorj et al. [6] used an ECOC SVM classifier in combination with the pre-trained AlexNet TLM for skin cancer stratification.

A comparison of many pre-trained CNN models and self-built models was reported by M. R. Hasan et al. [7] in order to categorize a dataset of 6594 skin cancer photos into two groups: benign and malignant. With the VGG16 model, the highest classification accuracy was 93.18%. Skin cancer detection has significantly increased owing to the use of hybrid models and attention strategies. By adding attention modules to the CNN designs, Li et al. [8] enabled the model to concentrate on the most important areas of the image. The forecast accuracy and interpretability were enhanced using this approach. Other studies, such as Liu et al. [9], examined hybrid models that combined CNNs with different machine learning techniques, such as support vector machines (SVMs), to enhance classification performance.

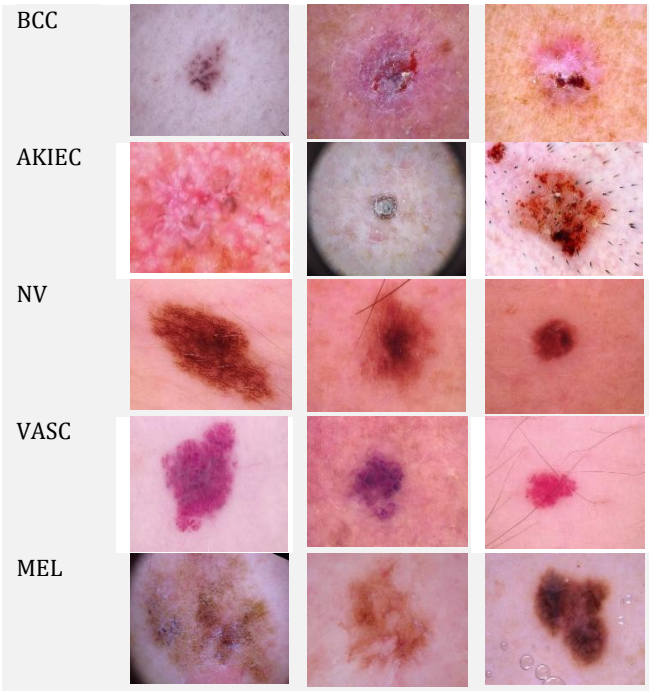
Studies in this field have advanced greatly because of the HAM10000 dataset, which was produced by Tschandl et al. [10]. A total of 10,015 microscopic images of pigmented tumors collected from various sources constituted the dataset. Consequently, the taught models are more dependable and practical. Numerous researchers have developed and validated machine-learning models for skin lesion classification using this dataset, demonstrating improved diagnostic reliability and accuracy. Transfer learning was used by Pham et al. [11] to refine pre-trained models, such as ResNet and InceptionV3, on the HAM10000 dataset. Their findings demonstrated how well transfer learning works to achieve excellent classification accuracy with comparatively little training data.

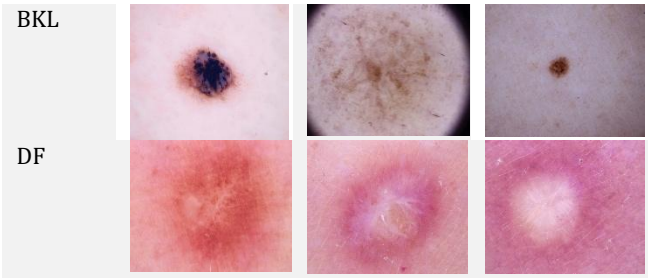
Implementation

A. Dataset

The proposed hybrid model was developed and evaluated using the HAM10000 dataset as an input dataset. The CNN classified 10015 images into seven different classes of skin cancer: Vascular Lesions (Angiomas, Angiokeratomas, Pyogenic Granulomas, & Hemorrhage) (VASC), dermatoma (DF), melanoma (MEL), Melanocytic Nevi (NV), benign keratosis-like lesions (Solar Lentigines/Seborrheic Keratoses & Lichen-Planus like Keratoses) (BKL), and Actinic Keratoses & Intraepithelial Carcinoma/Bowen's Disease (AKIEC). The classes categorized in this study are listed in Table 1, along with examples of each category.

TABLE I. IMAGE DATASET SAMPLE





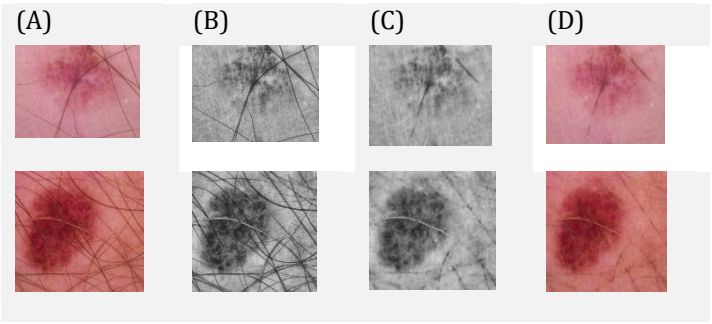
B. Preprocessing

The images were meticulously prepped before being fed into the model to preserve consistency and enhance the performance. This required implementing several picture modifications, standardizing the pixel values, and modifying their size to increase the dependability of the model.

1. Hari remove

Table 2 lists the image processing pipeline employed as the initial stage of our hair removal approach. Section A shows the original dermoscopic images of body hair, each of which represents a distinct type of lesion, such as carcinoma or melanoma. Section B presents the grayscale conversion of these images, which are required for morphological processes. The morphological modification of BlackHat is utilized in Section C to highlight the hair detection procedure. A hair mask was then made using binary thresholding. After hair removal, which was achieved by inpainting using the telea algorithm and optional smoothing, the finished images are displayed in Section D. This indicates that it is possible to successfully repair the affected areas while preserving lesion characteristics.

TABLE II. HAIR REMOVE

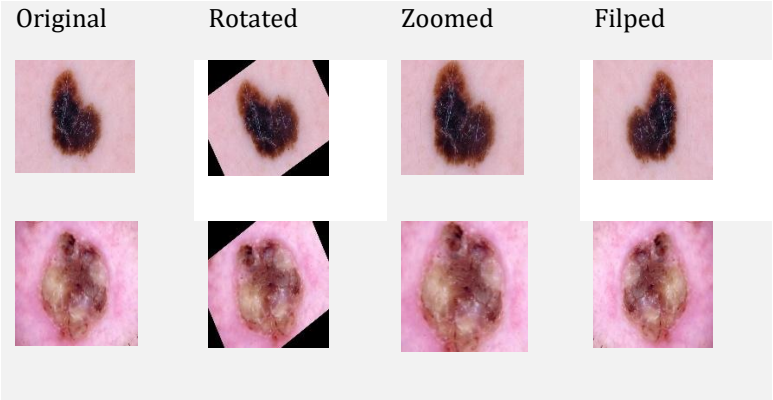


2. Data Augumendation

There are class imbalances in the HAM10000 dataset, which consists of the input photographs; that is, there are more images for certain classes and fewer for the other classes. Therefore, this class imbalance was eliminated by data augmentation [13]. Data augmentation was used to increase the size and diversity of the training dataset for computer vision and deep learning. Rotation, scaling, cropping, flipping, and addition of noise or blur are some of the transformations and changes made to the original dataset to produce new training data. By preventing overfitting and increasing the ability of the deep learning model to generalize to new, untested data, data augmentation improves the model's performance and accuracy. By creating more diverse training data, data augmentation may help alleviate dataset imbalances, such as unequal class distributions.

There are class imbalances in the HAM10000 dataset, which consists of the input photographs; that is, there are more images for certain classes and fewer for the other classes. Therefore, this class imbalance was eliminated by data augmentation [13]. Data augmentation was used to increase the size and diversity of the training dataset for computer vision and deep learning. Rotation, scaling, cropping, flipping, and addition of noise or blur are some of the transformations and changes made to the original dataset to produce new training data. By preventing overfitting and increasing the ability of the deep learning model to generalize to new, untested data, data augmentation improves the model's performance and accuracy. By creating more diverse training data, data augmentation may help alleviate dataset imbalances such as unequal class distributions.

TABLE III. DATA AUGMENTATION RESULTS



As shown in Table 3, the input dataset underwent data augmentation using the rotation, zooming, and flipping procedures. Consequently, there were fewer differences among the classes in the initial dataset.

3. Normalization

Normalization is the process of scaling data, in this case, to a range of 0–1. The confidence score for each bounding box is calculated and normalized to facilitate object detection. For instance, in Fig. 1.

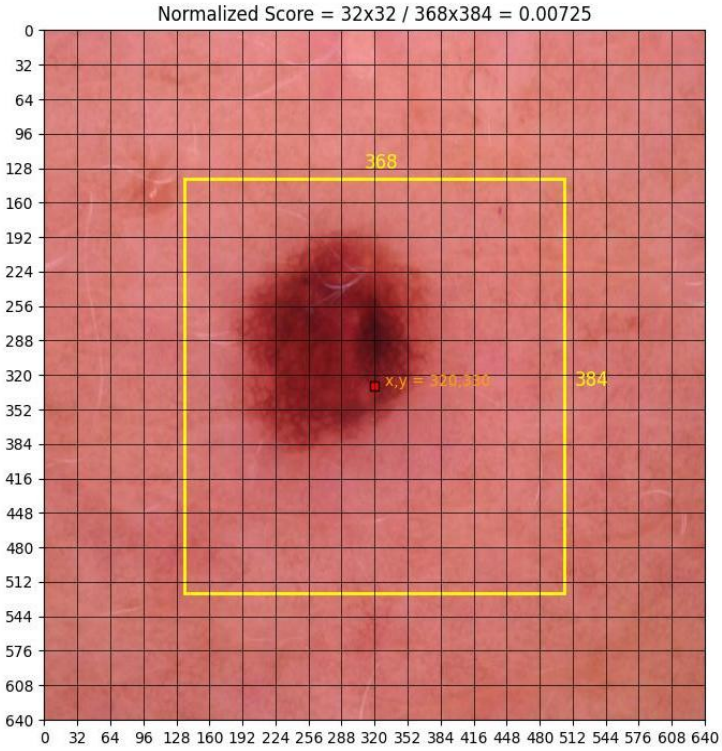


Fig. 1. Bounding Box

C. Segmentation

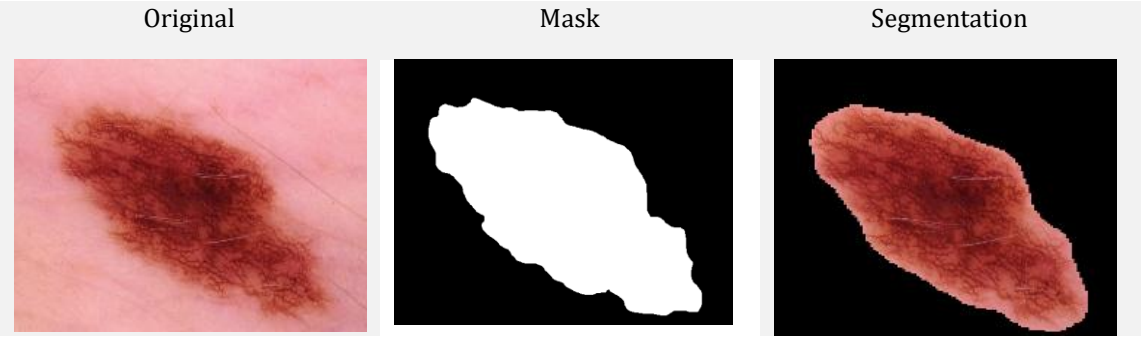
Segmentation assists in identifying particular regions of interest in a given image; for example, identifying cancerous regions in diagnostic images. Segmentation does this by examining every pixel and determining whether it is a part of the desired area. This generates a segmentation mask, which is mathematically defined as:

to 1.

$$\text{Mask}(X) = \sigma(\text{Conv}_{1\times1}(D_0))$$

X is the input image in Equation (1), and mask (X) is the output map indicating the probability that each pixel is part of the target. D_0 is downsampled by passing it through a 1×1 convolution layer to decrease the depth of channels to one, then a sigmoid activation function σ transforms the outputs to probability values ranging from 0

TABLE IV. SEGMENTATION RESULT



The decoder output D_0 is constructed in multiple stages and formulated as

$D_0 = D(B(E(X)))$

Equation (2) deconstructs the U-Net structure, as follows: Encoder E(X) takes the input deep features. The features are fed into bottleneck layer B, which stores the essential information. Subsequently, decoder D progressively upsamples the information, leveraging the outputs of the encoder as skip connections to retrieve the spatial details. The output D_0 is ultimately the input to Equation (1), which generates the segmentation mask.

D. Hybrid Model

Our framework combines the complementary strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to classify cancer presence and stage or type from diagnostic images. The input image is preprocessed and resized to dimensions of (3, 300, 300) to fit EfficientNet-B3's optimal dimensions. EfficientNet-B3 was used as a feature extractor with compound-scaled convolutional layers, squeeze-and-excitation blocks, and Swish activations to yield fine-grained local features such as cell boundaries, tissue textures, and micro-level irregularities. The resultant feature map of size (1536, 10, 10) was flattened to 100 patch tokens and transposed to a pre-training compatible sequence format.

TABLE V. MODEL SUMMARY

Stage	Layer / Module	Input Shape	Output Shape
1	Input Layer	(3, 300, 300)	(3, 300, 300)
2	Stem Convolution	(3, 300, 300)	(40, 150, 150)
3	MBConv1	(40, 150, 150)	(24, 75, 75)
4	MBConv6 (×2)	(24, 75, 75)	(40, 38, 38)
5	MBConv6 (×2)	(40, 38, 38)	(80, 19, 19)
6	MBConv6 (×3)	(80, 19, 19)	(112, 19, 19)
7	MBConv6 (×3)	(112, 19, 19)	(192, 10, 10)
8	MBConv6 (×4)	(192, 10, 10)	(320, 10, 10)
9	Conv Head	(320, 10, 10)	(1536, 10, 10)
10	Flatten Spatial Grid	(1536, 10, 10)	(1536, 100)
11	Transpose for Transformer	(1536, 100)	(100, 1536)
12	Vision Transformer Encoder ×6	(100, 1536)	(100, 1536)
13	Transpose Back	(100, 1536)	(1536, 100)
14	Adaptive Average Pooling (1D)	(1536, 100)	(1536,)
15	Dense Layer 1	(1536,)	(256,)
16	Dense Layer 2	(256,)	(128,)
17	Dense Layer 3	(128,)	(64,)
18	Dense Layer 4	(64,)	(32,)
19	Output Layer	(32,)	(7,)
20	Final Prediction	(7,)	Cancer Class Label

The ViT encoder uses multihead self-attention to capture global relations from different regions in the image to help the network understand large-scale structural patterns such as tumor shape, spread, and clustering. The resulting token sequence was pooled into a fixed-size embedding vector, which was then fed through the fully connected layers to yield a seven-class softmax output. The hybrid architecture effectively leverages the local sensitivity of CNNs with the global awareness of transformers to make it highly appropriate for cancer detection, cancer staging, cancer type, and clinical decision-making, particularly in cases with limited annotation.

Results

The performance of the proposed hybrid model was evaluated using the parameters obtained from the Confusion Matrix following 25 epochs of implementation. The performance of the proposed hybrid model is assessed by how well it can categorize the input images from the HAM10000 dataset into seven types of skin cancer.

A. Loss Analysis

The training and testing losses of the proposed hybrid model are illustrated in Fig. 2.

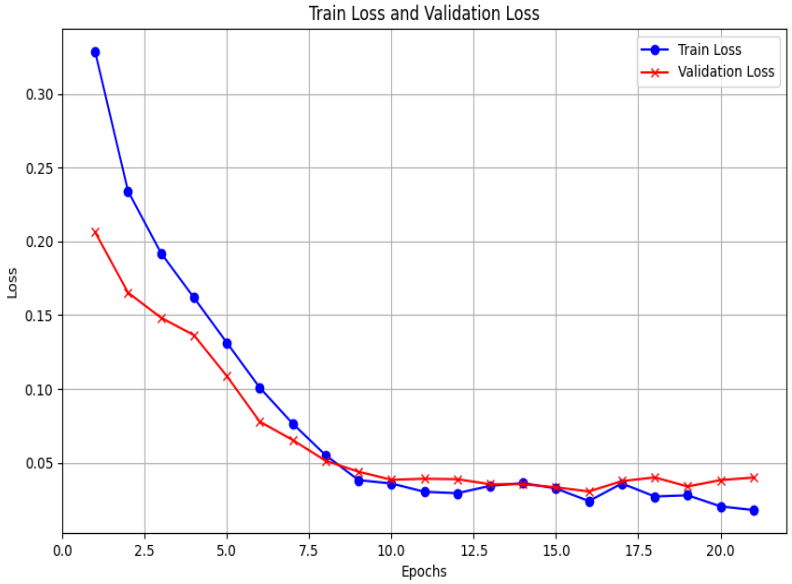


Fig.2. Loss Analysis

The model showed effective progress as the training and validation losses steadily dropped over the epochs. Both losses stabilized at low values after roughly ten epochs, suggesting a good generalization without overfitting.

B. Accuracy Analysis

Figure 3 highlights the accuracy achieved during the Training and Testing of the proposed hybrid model.

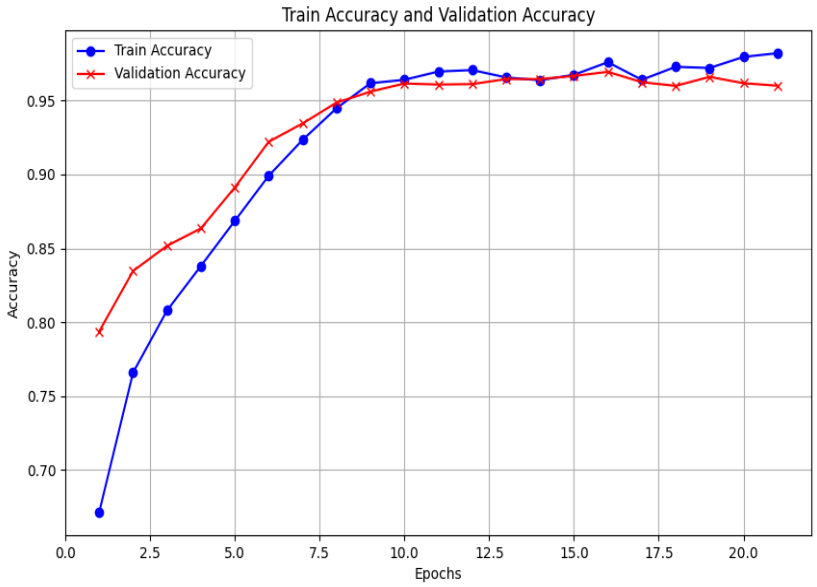


Fig.3 Accuracy Analysis

As the number of epochs increased, the training and validation accuracies gradually increased, until they exceeded 97%. Strong and reliable model performance is indicated by the close alignment of the training and validation accuracies.

C. Confusion Matrix

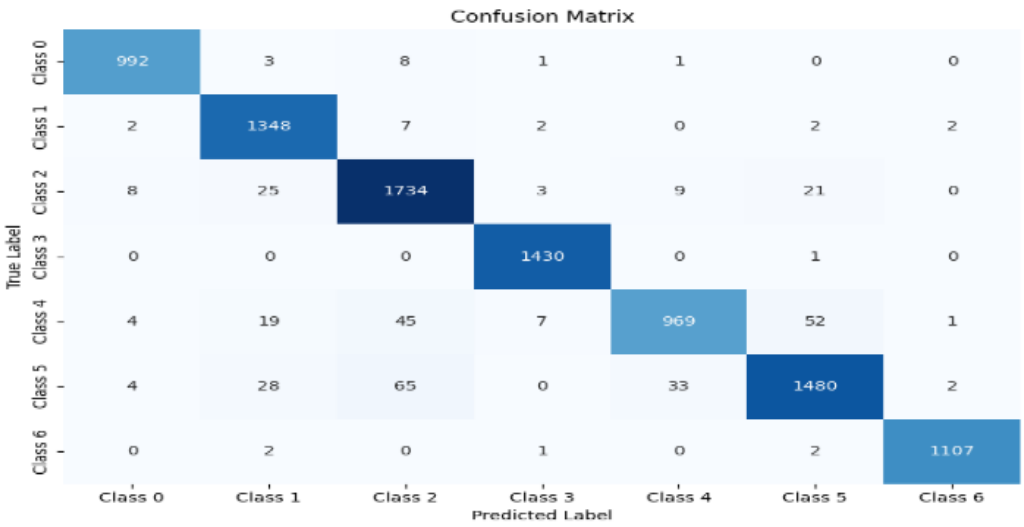


Fig. 4 Confusion Matrix

The outcome was remarkable, with a superb classification in all categories. The ability of the model to attain such high accuracy makes it valuable for clinical purposes by facilitating early and precise diagnosis.

Such consistent performance further makes the measurement of performance indicators such as accuracy, precision, and recall. The confusion matrix generated by the developed model is shown in Figure 4.

To assess the efficacy of the model, we calculated the recall, accuracy, precision, and F1 scores using a confusion matrix. The results are shown in Table VI.

TABLE VI. CLASS WISE PERFORMANCE METRICS

Class	Accuracy	Precision	Recall	F1 Score
Actinic Keratoses	0.987065	0.982178	0.987065	0.984615
Basal Cell Carcinoma	0.988995	0.945965	0.988995	0.967001
Benign Keratos	0.963333	0.932760	0.963333	0.947800
Dermatofibroma	0.999301	0.990305	0.999301	0.994783
Melanocytic Nevi	0.918114	0.949936	0.928114	0.933754
Melanoma	0.883318	0.957510	0.883318	0.918919
Vascular Lesions	0.995504	0.995504	0.995504	0.995504

This table shows the Accuracy, Precision, Recall and F1 scores of the different types of cancer detected by our model.

Conclusion

EfficientNet and Vision Transformer hybrid models have been employed within this project for the detection of skin cancer. It was employed in the publicly released HAM10000 dataset. The use of a hybrid model was found to work effectively in the detection of small, medium, and large-sized lesions and also improved early diagnosis of skin cancer. It was tested using parameters including accuracy, precision, recall, and the F1 score that yielded high levels of achievement. The built model guarantees dermatologists an F1 score of 96.31%, precision of 96.48%, recall of 96.22%, and accuracy of 96.22%. The model's performance was improved in addition to durability using enhanced deep-learning strategies that include Vision Transformer and EfficientNet. It was deployed for use using a web app that serves as a beneficial means through which the diagnosis of cancer of the skin might be performed on a self-based system. In future work, the app's classification should include working together side by side with medical staff that would supply insightful opinions and provide extra authentication to support the extensive use of the built app. Incorporating features that include real-time analysis as well as study material would improve the

use experience and provide a way for users to monitor their own skin more effectively.

References

[1] O. Abuzaghleh, B. D. Barkana, et M. Faezipour, "Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention," IEEE J. Transl. Eng. Heal. Med., vol. 3, pp. 1-12, 2015. [DOI:10.1109/JTEHM.2015.2419612].

[2] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, "Deep learning methods for accurate skin cancer recognition and mobile application," Electronics, vol. 11, no. 9, p. 1294, 2022.

[3] R. Sulthana, V. Chamola, Z. Hussain, F. Albalwy, and A. Hussain, "A novel end-to-end deep convolutional neural network based skin lesion classification framework," Expert Systems with Applications, vol. 246, p. 123056, 2024

[4] H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk, and L. Uhlmann, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," Annals of oncology, 29(8), pp.1836-1842, 2018.

[5] Barman M, Choudhury JP, Biswas S. Automated detection of melanoma skin disease using classification algorithm. In: Dasgupta K, Mukhopadhyay S, Mandal JK, Dutta P, eds. Computational Intelligence in Communications and Business Analytics. CICBA 2023. Communications in Computer and Information Science. Vol 1955. Springer; 2024: 153-164. doi:10.1007/978-3-031-48876-4_14.

[6] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," Multimed. Tools Appl., vol. 77, no. 8, pp. 9909–9924, 2018.

[7] M. R. Hasan, M. I. Fatemi, M. Monirujjaman Khan, M. Kaur, and A. Zaguia, "Comparative analysis of skin cancer (benign vs. Malignant) detection using convolutional neural networks," J. Healthc. Eng., vol. 2021, p. 5895156, 2021.

[8] X. Li, C. Wang, L. Zhang, X. Gao, and Y. Liu, "Attention-based deep ensemble model for skin lesion classification," IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3693–3704, 2020.

[9] Y. Liu, A. Jain, and e. a. Eng, Christina, "A deep learning system for differential diagnosis of skin diseases," Nature Medicine, vol. 26, no. 6, pp. 900–908, 2019.

[10] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, p. 180161, 2018.

[11] T. C. Pham, C. M. Luong, and V. C. Hoang, "A comprehensive study on classification of skin lesions using convolutional neural networks," IEEE Access, vol. 9, pp. 39843–39851, 2021.

[12] Leffell, D. J., & Brash, D. E. (1996). Sunlight and Skin Cancer. Scientific American, 275(1), 52–59.

[13] Miller, A. J., & Mihm, M. C., Jr (2006). Melanoma. The New England journal of medicine, 355(1), 51–65.

[14] N. Deniz and C. Tastimur, "Skin Cancer Detection Based on YOLOv8 Through A Mobile Application," in Proc. 8th Int. Artif. Intell. Data Process. Symp. (IDAP'24), Malatya, Türkiye, Sep. 2024, pp. 1–6. doi: 10.1109/IDAP64064.2024.10711093.

[15] Z. Li, "A Skin Cancer Detection System Based on CNN with Hair Removal," in Proc. 2023 IEEE 3rd Int. Conf. Power, Electron. Comput. Appl. (ICPECA), Shenyang, China, Jan. 2023, pp. 1291–1294. doi: 10.1109/ICPECA56706.2023.10076164.

[16] R. Thinakaran, J. Somasekar, V. Neerugatti, and K. Ganga, "Advancements in Skin Cancer Detection: A Comprehensive Review of Convolutional Neural Network Approaches," in Proc. 2024 14th Int. Conf. Softw. Technol. Eng. (ICSTE), pp. 232–235. doi: 10.1109/ICSTE63875.2024.00047.

[17] S. M. Afifi, R. Kaur, H. GholamHosseini, O. Azzam, and R. Taha, "Deep Learning-Powered Mobile Application For Early Skin Cancer Detection," in Proc. 2024 11th Int. Conf. Soft Comput. & Mach. Intell. (ISCMI), pp. 318–321, 2024. doi: 10.1109/ISCMI63661.2024.10851700.

[18] M. A. Riyadi, A. Ayuningtias, and R. R. Isnanto, "Detection and Classification of Skin Cancer Using YOLOv8n," in Proc. 2024 11th Int. Conf. Electr. Eng., Comput. Sci. & Informatics (EECSI), Sep. 2024. doi: 10.1109/EECSI63442.2024.10776505.

[19] S. Sharma, S. Kaur, and N. Kaur, "Ensemble of CNN based Deep Learning Model for the Recognition of Skin Cancer Images," in Proc. 2024 IEEE 5th India Council Int. Subsections Conf. (INDISCON), 2024. doi: 10.1109/INDISCON62179.2024.10744259.

[20] M. Muskan, P. Venkateshwari, and D. Chandra Mohan, "Hierarchical Deep Learning Model for Skin Cancer Detection and Skin Disease Diagnosis," in Proc. 2024 3rd IEEE Delhi Section Flagship Conf. (DELCON), 2024. doi: 10.1109/DELCON64804.2024.10866731.

[21] K. Iqtidar, S. Aziz, A. Iqtidar, M. U. Khan, and W. Ali, "Image Pattern Analysis towards Classification of Skin Cancer through Dermoscopic Images," in Proc. 2020 First Int. Conf. Smart Syst. Emerging Technol. (SMARTTECH), 2020. doi: 10.1109/SMART-TECH49988.2020.00055.

[22] R. Pillai, N. Sharma, and R. Gupta, "Proposed Convolution Neural Network for Skin Cancer Diagnosis and Classification," in Proc. 2023 Int. Conf. Recent Advances in Electr., Electron. & Digital Healthcare Technol. (REEDCON), 2023. doi: 10.1109/REEDCON57544.2023.10151029.

[23] V. Aadiwal, B. Sharma, and D. P. Yadav, "Revolutionizing Dermatology: Novel Convolutional Neural Network Framework for Skin Cancer Detection," in Proc. 2024 3rd Int. Conf. Advancement in Technol. (ICONAT), Goa, India, Sep. 2024. doi: 10.1109/ICONAT61936.2024.10774995.

[24] D. C. Carvajal, B. M. Delgado, D. G. Ibarra, and L. C. Ariza, "Skin Cancer Classification in Dermatological Images Based on a Dense Hybrid Algorithm," in Proc. 2022 IEEE Int. Conf. Electronics, Electr. Eng. & Comput. (INTERCON), 2022. doi: 10.1109/INTERCON55795.2022.9870129.

[25] Madhan, S. and Kalaiselvan, A. (2024) 'Omics data classification using constitutive artificial neural network optimized with single candidate optimizer', *Network: Computation in Neural Systems*, 36(2), pp. 343–367. doi: 10.1080/0954898X.2024.2348726.