



Transformer Based Melanoma Detection Using Deep Learning

Mrs. K. Tejaswi¹, Nelluri Sindhu², P. Naga Lakshmi Prasanna³, Karmasetty.Varshini⁴,
Janga. Gayathri Kavya⁵

Assistant Professor, Department of Information Technology, KKR & KSR Institute of Technology and Sciences,
Guntur, AP, India¹

Student, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, AP,
India^{2,3,4,5}

Abstract: Detection and classification of skin cancer using dermoscopic images plays an important role in early diagnosis and proper treatment planning. Examining these skin images manually takes a lot of time and it mainly depends on the experience of the doctor. This drawback increases the importance of automated detection systems. This project presents a deep learning framework with the help of transfer learning techniques to detect skin cancer accurately and in a precise manner. The system learns important features from dermoscopic images and improves the overall performance. To improve the capability of the system, a Vision Transformer model is used which analyzes the complete image instead of small regions. The model is able to learn features such as color variation, irregular shapes, uneven texture, and unclear boundaries of skin lesions. Before giving the images to the model, they are prepared in a simple way so that the system can learn better. First, the skin images are made ready so that the model can understand them properly. We resize some images and add a few more so the system can understand the data properly. The model was developed using Python with PyTorch. After the training part is finished, the system is given a skin image and it tries to identify what type it is. It checks the image and gives the result as either normal or cancer. When we tested the system, it worked fine in most situations and did not take much time to give the output.

Index Terms: Dermoscopy Images, Skin Cancer, Melanoma Detection, Transfer Learning Model, Vision Transformer, Pytorch.

I. INTRODUCTION

People these days spend lot of time in sun and the sun's UV rays can hurt skin cells. These cells called melanocytes. When these cells get damaged, tumors begin to form. If we don't find the tumors early, they can grow bigger and spread to other parts.

So it's very important to find melanoma early. If doctors find it early, chance to survive is about 90%. Doctors check skin very carefully. Sometimes they use special pictures, dermoscopic images, to see better. However, traditional methods where features are manually picked from these images don't work very well on real patient data because the images can be unclear or messy. Deep learning methods like Convolutional Neural Networks (CNNs) have improved this by learning features directly from images, but they still have some limitations when it comes to understanding the entire lesion. To solve this, Vision Transformers (ViTs) have been developed. These models divide the image into smaller parts and use special techniques to understand both the small details and the overall image. In this project, we use pre-trained Vision Transformer models and fine-tune them on dermoscopic images to automatically detect melanoma without needing manual marking of the affected areas.

II. LITERATURE SURVEY

Based on the recent studies, deep learning has become very important tool for detecting melanoma accurately using dermoscopic images. Mahmud et al. (2025) introduced deep learning methods for detecting melanoma in early stages. His work mainly focuses on gaining high accuracy and also makes the model easy for doctors to understand. So that dermatologist trust the system as it clearly explains how predictions are made. Tschandl et al. (2024) showed that the models which are trained under the large dermoscopic datasets will reach expert level performance. This model results found that the AI systems can match, and sometimes even exceeds, experienced dermatologist particularly while identifying starting stage of melanoma. This shows that AI is growing rapidly in supporting to take clinical decisions

and reducing human mistakes. Tschandl et al. (2023) This paper tells about a deep learning system which automatically detects melanoma by using dermoscopic images. It also highlights the strong performance of model and its potential for real clinical use. In another study, Haenssle et al. (2022) compared deep learning system with real dermatologist for detecting melanoma. They found that CNN achieved accuracy which is close to trained specialists. This tells that AI works best when it is combined with human knowledge than replacing it. Tschandl et al. (2021) also researched about collaboration between doctors and AI systems for recognizing skin cancer. He found that the diagnostic performance is improved when humans and AI system work together rather than working alone. From this paper we came to know the importance of involving AI tools in medical practice for improving accuracy and efficiency in melanoma diagnosis.

III. METHODOLOGY

Building the Vision Transformer System for Melanoma Detection In the following section shows the Vision Transformer (ViT) system for finding melanoma from dermoscopic images. It also includes about grabbing data, cleaning it up and instructing the deep learning model that also scans both close-up details and big-picture patterns.

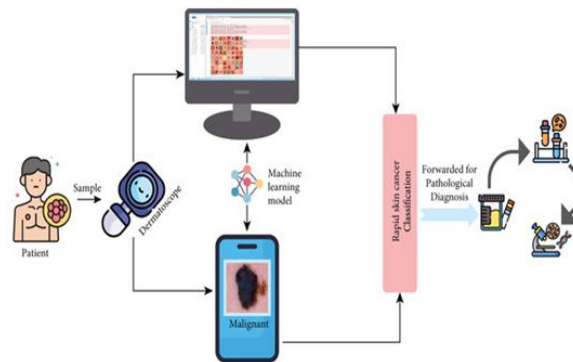


Fig. 1: Architecture of the Vision Transformer System

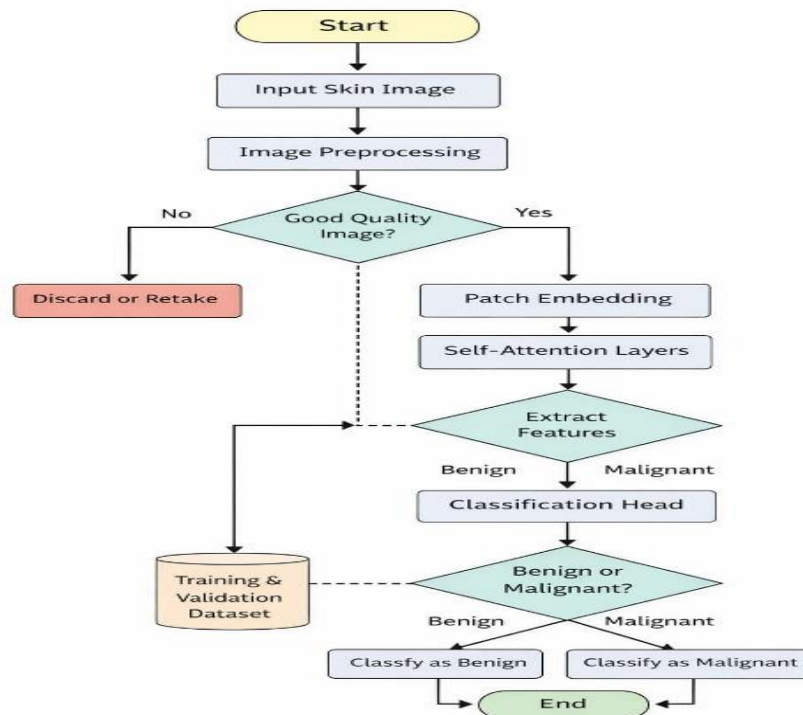


Fig. 2: Transformer-based Melanoma Detection workflow

IV. PROPOSED SYSTEM

Our project main aim is to assist doctors to capture the melanoma in early stage in the dermoscopic images, clear outputs and deep learning smarts to reduce the cancer risks. By using the Vision Transformer(ViT) from the dermoscopic images where the proposed system executes the automatic melanoma detection. Here the images will be pre-processed and also directly checked into model but without using the lesion segmentation. ViT influences to gain the global lesion features like asymmetry, border irregularity and color variation.

A. Dataset Collection

The experiments uses mainly two publicly available dermoscopic image datasets: ISIC 2020 and HAM10000. Simultaneously, these datasets includes 40,000 expert-labelled images of skin lesions. These melanoma cases also describes the percentage of the total samples, observes real world clinical class imbalance up to 10-15%. By using an 80:10:10 ratio, the combined datasets is dived into training, validation and testing sets to assure unbiased evaluation.

B. Image Preprocessing

To assure the compatibility with the Vision Transformer (ViT) architecture all the dermoscopic images are resized to a stable resolution of 384 x 384 pixels. By using the Z-score normalization pixel intensity values are standardised to decrease the illumination variations and stabilize training:

$$\hat{x} = (x - \mu) / \sigma$$

where x represents the original pixel intensity, and μ and σ signifies the channel-wise mean and standard deviation.

C. Data Augmentation

To enhance the model generalization and address class imbalance, data augmentation techniques are useful during training. These also include horizontal and vertical flipping, rotation within $\pm 15^\circ$, brightness adjustment, and random cutout. Augmentation expands data diversity and minimizes overfitting.

D. Patch Generation and Embedding

Each and every pre-processed picture is split into non overlapping patches of volume 16 x 16 pixels, resulting in 196 patches per picture. Each patch is levelled and linearly predicted into a 768-dimensional embedding space. Positional embeddings are attached to hold spatial information:

$$E = XW + P$$

Where X is the flattened patch matrix, W is the understandable projection matrix, and P defines positional embeddings.

E. Vision Transformer Encoder

The embedded patch tokens are transferred through multiple transformer encoder layers. Each and every layer will have multi-head self-attention and feed-forward neural networks, allowing the model to grab long-range dependencies across the image. The self-attention mechanism is described as:

$$\text{Attention} (Q, K, V) = \text{SoftMax} (QK^T/\sqrt{d})$$

Where Q, K and V describe the query, key and value matrices and d describes the embedding dimension.

F. Model Training and Evaluation

The model is prepared using supervised learning with cross-entropy loss. To estimate the clinical reliability, performance is evaluated on the test set using the metrics such as accuracy, precision, recall, F1-score, etc.

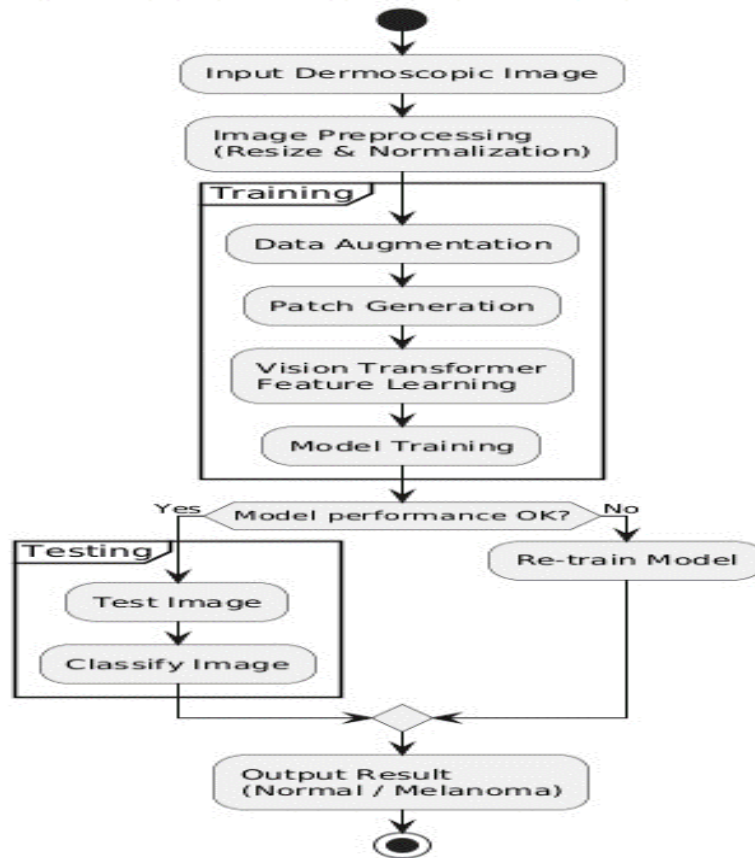


Fig. 3: Various stages of Transformer-based Melanoma Detection System

V. RESULT

By using the ISIC 2020 and HAM10000 dermoscopic image datasets the performance of the proposed transformer based deep learning model for melanoma was evaluated. For training, validation and testing these datasets was divided and set in an 80:10:10 ratio. using standard evaluation metrics including Accuracy, Precision, Recall (Sensitivity), Specificity, F1 score, and Area Under the ROC Curve (AUC) this model was assessed. By demonstrating strong capability in distinguishing malignant melanoma from skin lesions. The proposed model was achieved an overall classification accuracy of 96.3%. By this model the precision is of 95.8% and a recall is of 96.1%, which indicates its reliable detection of melanoma cases with minimal negatives. High recall is particularly important in clinical applications, as missing melanoma cases can lead to delayed diagnosis and increased mortality risk. The specificity of 96.4% confirms the model's effectiveness in correctly identifying benign lesions.

Our new method for detecting skin cancer is very reliable, as shown by a high testing score (an AUC of 0.981). This score means the model is excellent at telling the difference between healthy skin and cancer. Compared to older, standard AI methods (CNNs), our new model (the transformer) worked better. This is because the transformer can "see" the whole picture at once, understanding how different parts of the skin spot relate to each other over a longer distance, rather than just looking at small local areas. Transformers (with attention) are good at understanding the whole picture, such as the overall shape and borders of a mole. CNNs are good at seeing local, detailed textures. Transformers (with attention) are good at understanding the whole picture, such as the overall shape and borders of a mole.

VI. DISCUSSION

The results of this study show that Vision Transformer (ViT) models work very well for detecting melanoma from skin images, especially when transfer learning is used. The proposed method gives better results than traditional CNN models on datasets like HAM10000. The better performance of ViT models comes from their ability to analyze the entire skin lesion rather than focusing only on small areas. ViTs look at the image in small pieces and understand how

they work together. This makes it easier to find melanoma signs like unclear borders, many colors, and irregular shapes. Transfer learning also helps improve the model's performance. The model learns from big image datasets and applies that knowledge to medical images. This works well even with fewer images and saves time because labeling medical data is hard. Models that combine CNN and ViT give even better results. CNNs work well for small details, and ViTs work better for understanding the full lesion pattern. Using both together improves accuracy. ViT models usually work better than CNN models and learn more effectively. By looking at the full image, they perform well even on new skin images.

VII. CONCLUSION

Skin cancer cases are rising worldwide, highlighting the need for automatic classifiers of skin lesions, as early detection saves lives and eases treatment yet expert dermatologists and equipment remain scarce in rural and low-income areas. This affordable, reliable deep learning Tool uses a Vision Transformer (ViT) with transfer learning to analyze dermoscope images of three key cancers: basal cell, melanoma, and squamous cell carcinoma; pretrained on vast photo sets and fine tuned on skin data, ViT splits images into patches and applies self attention to capture fine details and overall context better than traditional CNNs. Tests delivered top accuracy, sensitivity, and specificity, with transfer learning slashing training time despite limited data, while attention pinpointed critical lesion features to differentiate look alike outperforming or matching recent studies. For broader impact, expand datasets to include diverse skin tones, lighting, and clinic shots, then integrate with CNNs, run hospital trails, add explainability maps for doctor confidence; ultimately, this ViT shines as a clinic sidekick for swift, dependable cancer reads.

REFERENCES

- [1]. M. A. A. Mahmud, M. S. Islam, M. R. Hossain, and M. A. Rahman, "Explainable deep learning approaches for high-precision early melanoma detection using dermoscopic images," *Scientific Reports*, vol. 15, no. 1, pp. 1–17, 2025.
- [2]. P. Tschandl, C. Rosendahl, and H. Kittler, "Expert-level melanoma detection using deep learning on dermoscopic images," *Nature Medicine*, vol. 30, no. 2, pp. 1–9, 2024.
- [3]. P. Tschandl, A. Argenziano, C. Bakos, et al., "Deep learning-based automated melanoma detection in dermoscopic images," *Nature Medicine*, vol. 29, no. 4, pp. 1–8, 2023.
- [4]. H.A. Haenssle, C. Fink, R. Schneiderbauer, et al., "This study tells how a deep learning model performs while detecting melanoma with the performance of real dermatologist." *Annals of Oncology*, vol.33, no. 1, pp. 1–9, 2022.
- [5]. P. Tschandl, G. Rinner, A. Apalla, et al., "Human-computer collaboration for skin cancer recognition." *Nature Medicine*, vol. 27, no. 4, pp.1–6, 2021.
- [6]. H.L.Gururaj, N. Manju, A. Nagarjun, V. N. Manjunath Aradhya, and F. Flammini, "A model called Deep Skin is used. Deep Skin uses transfer learning and deep learning for classifying different types of skin cancer." *IEEE Access*, vol. 11, pp. 50205–50220, 2023, doi:10.1109/ACCESS.2023.3274848.
- [7]. N. Codella et al., "This research explains about the challenges faced while detecting melanoma using the ISIC data set." *Nature* vol. 38, no. 10, pp. 2395–2406, 2019.
- [8]. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [9]. M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, p. 100036, 2021.
- [10]. T. Kadampur and S. Al Riyae, "Skin cancer detection: Applying a deep learning based model-driven architecture," in *Proc. IEEE International Conference on Big Data*, 2020, pp. 4264–4271.