International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 04 | Apr 2025 www.irjet.net p-ISSN: 2395-0072

Radiant Dermat: Skin dermatology tool using Computer Vision and ML

Sonal Chaudhari[1], Aaryaman Ahirao[2], Jivhesh Chaudhari[3], Vinay Kumkar[4], Aditya Kelaskar[5]

[1]*Assistant Professor, Dept. of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, India*
[2345]*Dept. of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, India*

Abstract - *The project focuses on the development of an Android-based application for disease detection using real-time image capture and image uploads. The app employs a machine learning model that utilizes advanced technologies such as deep learning, vision transformers, Swin Transformer technology, image segmentation, an encoder-decoder unit, and a SoftMax probabilistic classifier.*

Key Words: Aritificial Intelligence, Deep learning, Convolutional Neural Network.

1.INTRODUCTION

Radiant Dermat is an advanced skin diagnosis tool that leverages the power of machine learning and computer vision to provide accurate and personalized skin assessments. This innovative platform aims to revolutionize the field of dermatology by offering accessible and efficient skin healthcare solutions. Using cutting-edge image analysis and deep learning algorithms, Radiant Dermat can identify skin lesions and provide customized recommendations to users. This project aims to achieve the highest level of accuracy and reliability for the end users. By harnessing the power of machine learning (ML) and computer vision, this innovative platform offers accurate, personalized skin assessments, making expert-level dermatological evaluations accessible to anyone, anywhere.

2. LITEARTURE SURVEY

2.1 INFLUENCE OF AI IN DERMATOLOGY

The application of AI within dermatology is experiencing a period of rapid growth, extending across a broad spectrum of concerns from common skin conditions like acne and eczema to the more critical areas of skin aging and skin cancer detection [1] . This technological advancement addresses a crucial public health issue by enhancing diagnostic accuracy, efficiency, and accessibility [2] . AI is transforming dermatology by automating many aspects of diagnostic processes into digital systems, thereby reducing manual administrative tasks and allowing healthcare professionals to focus on more critical medical decisions [2] . This shift is particularly evident in early skin cancer detection, where AI-based systems leverage medical data and analytics to improve diagnostic processes and simplify medical services . Furthermore, AI serves as a valuable tool that augments the capabilities of dermatologists, particularly in the early detection of melanoma, which can significantly improve patient survival rates [2] . The capabilities of AI in this domain include providing diagnostic support, facilitating medical interventions, and aiding in the development of proactive health strategies . Many modern hospitals are increasingly incorporating AI technologies to reduce operational costs and enhance diagnostic precision [2] . These systems can rapidly analyze large volumes of medical data, identifying patterns and anomalies that may elude human detection, thus streamlining workflows and ensuring more accurate diagnoses.

2.2 MOBILE APPLICATIONS FOR SKIN LESION DETECTION

The development and interest in smartphone applications for skin lesion diagnosis and triage are on the rise [1] . The widespread use of smartphones has made it feasible to develop AI models that can run directly on patients' devices, offering a convenient and private means for initial skin assessments [3] . This on-device processing of AI models addresses data privacy concerns by ensuring that sensitive medical data remains on the user's device, rather than being transmitted to external servers for analysis [3] . However, while some mobile skin lesion analysis applications demonstrate promising accuracy in controlled settings, many face limitations in terms of sensitivity and specificity when evaluated in real-world scenarios [4] . Clinical validation of these apps remains a significant challenge, with studies revealing variable and often low accuracy, recommending caution against their sole use for diagnostic purposes [5]. For instance, a 2020 systematic review of several melanoma detection apps found poor study design and a high risk of bias, with many apps failing to accurately identify melanoma cases [5]. Conversely, some applications, like Dermalyser, have shown high diagnostic accuracy in prospective clinical trials, indicating the potential of AI in this domain when rigorously validated [6] . The performance of AI-powered skin lesion detection apps is also significantly influenced by the quality of the input images . Standardized image capture and high-quality images are crucial for ensuring the reliability of AI analysis, as variations in lighting, focus, and angle can affect the accuracy of the results [1].

2.3 DEEP LEARNING METHODOLOGIES

Various deep learning architectures, particularly Convolutional Neural Networks (CNNs), are fundamental to the development of AI-powered skin lesion classification systems [7] . These networks excel at learning hierarchical representations from image data, making them highly effective for tasks like identifying and classifying skin lesions [8, 9] . Transfer learning, a technique that involves using pre-trained models on large datasets (like ImageNet) and fine-tuning them for specific tasks with smaller datasets, has proven valuable in improving the accuracy of skin lesion classifiers, especially given the limited size of many medical image datasets [10]. A common challenge in training deep learning models for skin lesion classification is data imbalance, where the number of images for different types of lesions can vary significantly . To address this, data augmentation techniques, such as image rotation, flipping, and zooming, are often employed to artificially increase the size and diversity of the training data, thereby improving the robustness and generalization of the models [8] . Furthermore, there is a growing recognition of the importance of explainable AI (XAI) in dermatology [11] . XAI methods aim to provide transparency into the decision-making processes of AI models, allowing clinicians to understand and validate the AI's diagnoses, which is crucial for building trust and facilitating the integration of AI into clinical practice [12] .

2.4 HYBDRID MODELS: APPLICATION AND PERFORMANCE OF U-NET AND VISION TRANSFORMERS IN MEDICAL IMAGE ANALYSIS

The U-Net architecture has become a cornerstone in the field of medical image segmentation [13] , widely utilized for its effectiveness in pixel-level classification of lesions and other anatomical structures . Its design features an encoder path that progressively extracts features from the input image while reducing its spatial dimensions, and a decoder path that reconstructs the segmented image at the original resolution. Crucially, U-Net incorporates skip connections that directly link corresponding layers in the encoder and decoder, enabling the decoder to access fine-grained details captured by the encoder, which is essential for precise localization of abnormalities in medical images [13]. Vision Transformers, including architectures like the SWIN Transformer, represent a significant advancement in image processing by effectively capturing global contextual information, a capability that traditional CNNs often lack due to their focus on local features [14, 15] .The SWIN Transformer employs a mechanism called shifted windows, which allows for efficient processing of high-resolution images by confining the attention computation to within local windows while also enabling cross-window connections to model global relationships [15] . This ability to understand the broader context of an image, coupled with the detailed local processing of CNNs, makes Vision Transformers highly effective for both image classification and segmentation tasks in medical imaging. Recognizing the complementary strengths of U-Net for precise localization and Vision Transformers for global context understanding, there is a growing trend in medical image analysis to combine these architectures into hybrid models [16] . These hybrid approaches aim to leverage the benefits of both types of models to achieve improved performance in medical image segmentation and classification. Many such models have demonstrated state-of-the-art results, achieving high segmentation accuracy, often measured by metrics like the Dice score and Intersection over Union (IoU), as well as superior classification accuracy [17, 18]. Radiant Dermat's architecture adopts a specific hybrid approach where the feature maps extracted by the U-Net encoder are directly fed into the SWIN Transformer, bypassing the decoder phase of U-Net. This design presents a novel strategy that could potentially offer computational efficiency by omitting the U-Net decoder while still capitalizing on the robust feature extraction capabilities of both U-Net and SWIN Transformer. The effectiveness of this particular architectural choice would need to be evaluated against the performance and efficiency of other hybrid models reported in the literature, such as those that integrate Transformers at different stages of the U-Net or use them in parallel pathways .

3. PROPOSED SOLUTION

3.1 OVERVIEW

Radiant Dermat distinguishes itself through several key features and potential innovations. The core diagnostic capability is driven by a novel hybrid model architecture that integrates U-Net and a SWIN Transformer for skin lesion detection and classification. This combination aims to achieve high accuracy by leveraging the segmentation prowess of U-Net and the global context understanding of SWIN. A notable feature is the weekly updating of new diseases on the server, complete with images and descriptions, which serves to continuously improve the user's knowledge about various skin conditions. The integration of Gemini 1.5 Pro provides users with an advanced chatbot capable of answering health-related inquiries and potentially offering personalized information. Furthermore, Radiant Dermat includes a comprehensive user profile feature that allows users to upload their past medical history. This capability to integrate user medical history could potentially enhance the personalization and accuracy of skin assessments, although it necessitates stringent privacy and security measures . The application also offers standard functionalities such as photo upload/capture of skin lesions, result display with detailed information about the diagnosed condition and the ability to store this

information in a history log, and a feature to view the user's test history. The technical architecture, utilizing Jetpack Compose for the frontend and Spring Boot for the backend, along with Firebase for authentication, provides a solid foundation for a scalable and maintainable application. The direct feeding of U-Net encoder features into the SWIN Transformer, skipping the U-Net decoder, could represent a potential innovation in terms of computational efficiency or model performance, warranting further investigation and comparison with existing hybrid models.

3.2 DATA AND PREPROCESSING

The dataset utilized in this study is the HAM10000 ("Human Against Machine with 10000 training images") dataset [Reference to HAM10000 paper if available]. This publicly available dataset comprises a collection of 10,015 dermatoscopic images of pigmented skin lesions. These images represent seven distinct diagnostic categories of skin diseases, making it a suitable benchmark for multi-class skin lesion classification tasks. The seven classes included in the dataset are:

1. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec)

2. Basal cell carcinoma (bcc)

3. Benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratosis) (bkl)

4. Dermatofibroma (df)

5. Melanoma (mel)

6. Nevus (nv)

7. Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage) (vasc)

The original images in the HAM10000 dataset have a native resolution of approximately 600 x 450 pixels with three color channels (RGB). The preprocessing pipeline aimed to enhance the quality and standardize the input images before feeding them to the UNetSwinClassifier model. Initially, Digital Hair Removal (DHR) was performed to mitigate the obscuring effects of hair in dermatoscopic images. This involved applying the Black Hat transformation, a morphological operation, to highlight the hair structures, followed by an impainting technique to seamlessly fill in the hair regions, allowing the model to focus on the underlying skin lesion. Subsequently, a Median Filter was employed to reduce noise present in the images while preserving essential textural details and edges, thereby improving the signal-to-noise ratio. Finally, to ensure uniformity and compatibility with the model's input requirements, all preprocessed images were resized from their original dimensions of approximately 600x450 pixels to a fixed size of 224x224 pixels using bilinear interpolation, which helps to maintain image integrity during the scaling process.

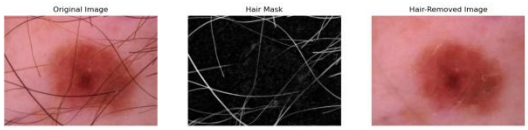

Fig -1: Results of Digital Hair Removal (DHR) Applied on an image from the dataset using Black-Hat transformation and impainting.

3.3 WORKING MODEL

This section details the architecture of the proposed deep learning model for skin disease classification. The model, named UNetSwinClassifier, leverages the strengths of both convolutional neural networks (CNNs) for local feature extraction and Transformer networks for capturing global dependencies. The architecture comprises four main components: a U-Net style encoder for hierarchical feature extraction, a bottleneck convolutional layer for channel reduction, a Swin Transformer module for learning long-range relationships, and a final classification head.

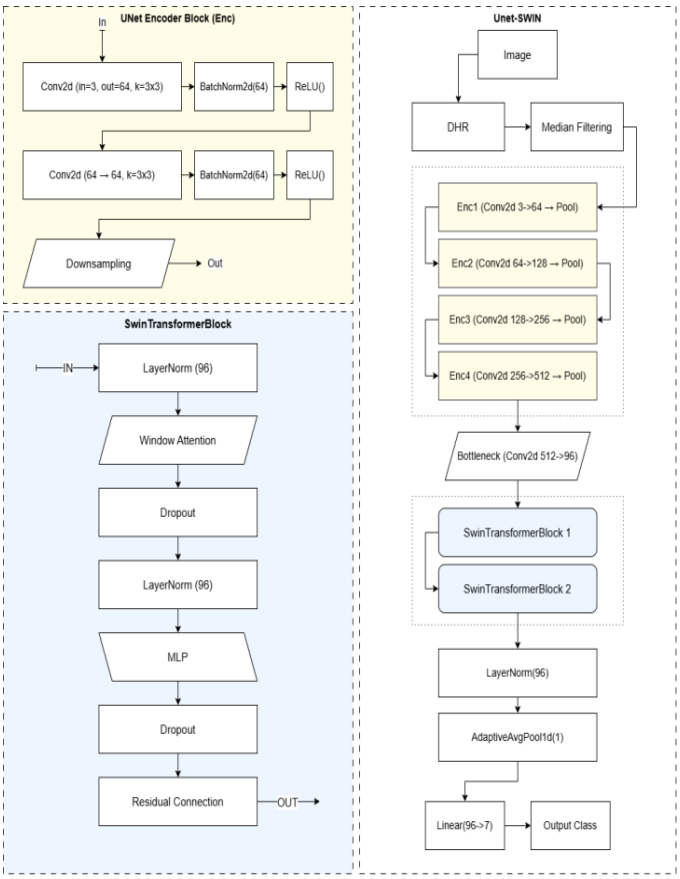

Fig-2: Model breakdown

- U-Net Style Encoder - The initial stage of our architecture employs a U-Net inspired encoder to extract hierarchical features from the input skin lesion images. The encoder consists of four sequential UNetEncoderBlock layers. Each UNetEncoderBlock is designed to perform feature extraction and downsampling.

- Double Convolutional Layer - Each UNetEncoderBlock begins with a double convolutional layer. This module comprises two consecutive convolutional layers, each followed by Batch Normalization and a Rectified Linear Unit (ReLU) activation function. The convolutional operation can be mathematically represented as:

$$(f * g)(x, y) = \sum_i \sum_j f(i, j) \cdot g(x - i, y - j)$$

Fig-3: Formula I

where f is the input feature map and g is the convolutional kernel. In our case, we utilize 2D convolutional layers with a kernel size of 3x3 and a stride of 1, maintaining the spatial dimensions through padding.Batch Normalization (BN) is applied after each convolutional layer to stabilize learning by normalizing the activations of the previous layer. For a batch of activations B={x1,...,xm}, BN computes the mean μB and variance σB2 and normalized output yi is then given by:

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \qquad y_i = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta$$

Fig-4: Formula II

where γ and β are learnable scale and shift parameters, and ϵ is a small constant for numerical stability.The ReLU activation function, applied element-wise, introduces non-linearity.

- Max pooling layer - Following the double convolutional layer in each UNetEncoderBlock, a Max Pooling layer is used for spatial downsampling. Specifically, we employ a 2x2 Max Pooling operation with a stride of 2, which reduces the height and width of the feature maps by a factor of 2. The Max Pooling operation selects the maximum value within each 2x2 window. The input image of size 224x224 undergoes this encoding process, resulting in feature maps of decreasing spatial dimensions and increasing channel counts (64, 128, 256, and 512 channels after each of the four encoder blocks). The final output of the encoder has a spatial dimension of 14x14.

- Bottleneck Convolutional Layer - After the U-Net style encoder, a bottleneck convolutional layer is introduced. This layer consists of a single 2D convolutional layer with a kernel size of 1x1 and a stride of 1. The purpose of this layer is to reduce the number of feature channels from 512 (the output of the final encoder block) to a lower dimension of 96. This reduction helps in managing the computational complexity of the subsequent Swin Transformer module. The 1x1 convolution performs a linear combination of the input channels at each spatial location:

$$Output_{i,j,k} = \sum_{c=1}^{C_{in}} Input_{i,j,c} \cdot Kernel_{c,k}$$

Fig-5: Formula III

- Swin Transformer module - To capture global contextual information crucial for accurate skin disease classification, we integrate a Swin Transformer module. The Swin Transformer operates on the feature maps produced by the bottleneck layer, which have a spatial resolution of 14x14 and 96 feature channels. This module consists of two sequential SwinTransformerBlock layers.
- Swin Transformer block - Each SwinTransformerBlock implements a shifted window-based multi-head self-attention mechanism. The block comprises several sub-layers:

1. Layer Normalization (LN): Applied before the attention mechanism and the MLP to stabilize training
2. Window-based Multi-Head Self-Attention (W-MSA) / Shifted Window-based Multi-Head Self-Attention (SW-MSA): The core of the Swin Transformer block is the attention mechanism. The input feature map is divided into non-overlapping windows of size 7x7. Within each window, self-attention is computed. For subsequent blocks, a shifted window partitioning scheme is employed to enable cross-window connections, enhancing the ability to model global relationships.
3. DropPath: A form of stochastic depth that randomly drops entire residual connections during training to improve robustness.
4. **MLP:** A Multi-Layer Perceptron with two linear layers and a GELU (Gaussian Error Linear Unit) activation function in between. The first linear layer expands

the feature dimension by a ratio of 4 (as defined by swin_mlp_ratio), and the second linear layer projects it back to the original dimension. Dropout is applied after the GELU activation.

The Swin Transformer module processes the input feature map of size 14x14, maintaining the spatial resolution while refining the feature representations through the attention mechanism. The output of the Swin Transformer module has a shape of (Batch Size, 14*14, 96).

- Final clasification layers - The final stage of the architecture is responsible for mapping the learned features to the classification output. This stage consists of the following layers:

1. Layer Normalization: Another Layer Normalization layer is applied to the output of the Swin Transformer module.
2. Adaptive Average Pooling: An Adaptive Average Pooling layer with an output size of 1 is used to reduce the spatial dimensions of the feature map. The input tensor of shape (Batch Size, 14*14, 96) is reshaped to (Batch Size, 96, 14*14) before being passed to this layer, resulting in an output of shape (Batch Size, 96, 1). This effectively performs global average pooling across the spatial dimensions
3. Linear Layer (Classification Head): A final Linear layer with an input size of 96 and an output size of 7 (corresponding to the number of skin disease classes) is used to perform the classification. This layer outputs the logits for each class.

The overall forward pass of the UNetSwinClassifier takes an input image of size (Batch Size, 3, 224, 224) and produces an output tensor of shape (Batch Size, 7) representing the classification logits for the seven different skin disease classes. These logits are then typically passed through a Softmax function to obtain class probabilities for prediction. This detailed description provides a comprehensive overview of the proposed UNetSwinClassifier architecture, outlining the function and mathematical formulations of its key components. This methodology aims to effectively leverage both local and global feature extraction for accurate skin disease classification.

4. PERFORMANCE METRICS

- Accuracy: Percentage of correctly detected diseases.
- Latency: Time taken from image capture/upload to result display.
- Precision and Recall: Used to evaluate the model's ability to detect the correct disease without false positives or negatives.
- Scalability: Ability to handle multiple users and real-time requests without a significant drop in performance.

Table-1: Performance Metrics

Model name	Dataset	Task	Accuracy(%)	Sensitivity(%)	F1-Score(%)	Publication
Swin Transformer Large	Dermoscopic	Classification	93.82	93.5		2023
SkinSwinViT	ISIC2018	Classification	97.88	99.36	97.79	2024
LUSwin	OCT	Denoising				2023
TESL-NET	ISIC 2016	Segmentation				2024
TESL-NET	ISIC 2017	Segmentation				2024
TESL-NET	ISIC 2018	Segmentation				2024
ScaleFusionNet	ISIC 2016	Segmentation				2025
ScaleFusionNet	ISIC 2018	Segmentation				2025
Hybrid(U-Net + Inception-ResNet-v2 + ViT)	ISIC 2020	Classification	98.65	98.03		2024
Hybrid(U-Net + MobileNet-V3)	HAM-10000	Classification	98.86	97.32	97.32	2023
Hybrid(Swin Transformer + DGSNLA)	HAM-10000	Classification	94.21		96.64	2024
Ours	HAM-10000	Classification	83.86	83.92	83.75	2025

5. FUTURE WORK

1. Integration of a Disease Progression Tracking System: Future work could focus on developing a system to track the progression of skin diseases over time. This would involve incorporating longitudinal data analysis capabilities, potentially allowing users to monitor changes in their condition based on sequential image analysis and other relevant information.

2. Enhancing Model Accuracy and Robustness: Further research will aim to improve the classification accuracy and robustness of the proposed model. This could involve exploring advanced architectures, incorporating attention mechanisms more deeply, leveraging larger and more diverse datasets, and investigating techniques for handling class imbalance and improving generalization to unseen data.

6. REFERENCES

[1] Gabriela Lladó Grove, Gorm Reedtz, Brian Vangsgaard, Hassan Eskandarani, Merete Haedersdal, Flemming Andersen, and Peter Bjerring. 2024. “Artificial Intelligence Smartphone Application for Detection of Simulated Skin Changes: An In Vivo Pilot Study.” Skin Research and Technology 30(10): e70056. DOI: 10.1111/srt.70056.

[2] Kavita Behara, Ernest Bhero, and John Terhile Agee.2024. "AI in Dermatology: A Comprehensive Review into Skin Cancer Detection." PeerJ Computer Science 10: e2530. DOI: 10.7717/peerj-cs.2530.

[3] Iren Valova, Peter Dinh, and Natacha Gueorguieva. 2023. "Mobile Application for Skin Cancer Classification Using Deep Learning." Journal of Machine Intelligence and Data Science (JMIDS) 4: 15–26. DOI: 10.11159/jmids.2023.003.

[4] Sangers, T., Reeder, S., van der Vet, S., Jhingoer, S., Mooyaart, A., Siegel, D. M., Nijsten, T., & Wakkee, M. (2022). Validation of a Market-Approved Artificial Intelligence Mobile Health App for Skin Cancer Screening: A Prospective Multicenter Diagnostic Accuracy Study. *Dermatology*, *238*(4), 649–656.

[5] Sun, M. D., Kentley, J., Mehta, P., Dusza, S., Halpern, A., & Rotemberg, V. (n.d.). Accuracy of commercially available smartphone applications for the detection of melanoma. Br J Dermatol., 186(4), 744–746.

[6] Panagiotis Papachristou, My Söderholm, Jon Pallon, Marina Taloyan, Sam Polesie, John Paoli, Chris D Anderson, and Magnus Falk. 2024. "Evaluation of an artificial intelligence-based decision support for the detection of cutaneous melanoma in primary care: a prospective real-life clinical trial". British Journal of Dermatology 191(1) DOI 10.1093/bjd/ljae021

[7] Abdulrahman Takiddin, Jens Schneider, Yin Yang, Alaa Abd-Alrazaq, and Mowafa Househ. 2021. "Artificial Intelligence for Skin Cancer Detection: Scoping Review." Journal of Medical Internet Research 23(11): e22934. DOI: 10.2196/22934.

[8] Yinhao Wu, Bin Chen, An Zeng, Dan Pan, Ruixuan Wang, and Shen Zhao. 2022. "Skin Cancer Classification With Deep Learning: A Systematic Review." Frontiers in Oncology 12: 893972. DOI: 10.3389/fonc.2022.893972.

[9] Mohamed A. Kassem, Khalid M. Hosny, Robertas Damasevicius, and Mohamed Meselhy Eltoukhy. 2021. "Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review." Diagnostics 11(8): 1390. DOI: 10.3390/diagnostics11081390.

[10] MdSirajul Islam and Sanjeev Panta. 2024. "Skin Cancer Images Classification using Transfer Learning Techniques." arXiv:2406.12954v1.

[11] Muhammad Bilal Jan, Muhammad Rashid, Raja Vavekanand, and Vijay Singh. 2025. "Integrating Explainable AI for Skin Lesion Classifications: A Systematic Literature Review." Studies in Medical and Health Sciences 2(1): 1–14. DOI: 10.48185/smhs.v2i1.1422.

[12] K. Srilatha and N. Satheesh Kumar. 2024. "Explainable AI Framework for Skin Cancer Classification, and Melanoma Segmentation." ARPN Journal of Engineering and Applied Sciences 19(15): 1819–6608.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In Medical Image Computing and Computer-Assisted Intervention (MICCAI), edited by Nassir Navab, Joerg Hornegger, William M. Wells, and Alejandro F. Frangi, 234–41. Cham: Springer International Publishing. arXiv:1505.04597v1.

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." In International Conference on Learning Representations. arXiv:2010.11929v2.

[15] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." arXiv:2103.14030v2.

[16] Pengsong Jiang, Wufeng Liu, Feihu Wang, and Renjie Wei. 2025. "Hybrid U-Net Model with Visual Transformers for Enhanced Multi-Organ Medical Image Segmentation." *Information* 16(2): 111. DOI: 10.3390/info16020111.

[17] Ovais Iqbal Shah, Danish Raza Rizvi, and Aqib Nazir Mir. MAPUNetR: A Hybrid Vision Transformer and U-Net Architecture for Efficient and Interpretable Medical Image Segmentation. Department of Computer Engineering Jamia Millia Islamia (A Central University) New Delhi.ment of Computer Engineering Jamia Millia Islamia (A Central University) New Delhi.

[18] Li, X., Zhu, W., Dong, X., Dumitrascu, O. M., & Wang, Y. (2024). EVIT-UNET: U-NET LIKE EFFICIENT VISION TRANSFORMER FOR MEDICAL IMAGE SEGMENTATION ON MOBILE AND EDGE DEVICES (arXiv:2410.15036v1).