VISVESVARAYA TECHNOLOGICAL UNIVERSITY

“Jnana Sangama”, Belagavi - 590018

A MINI-PROJECT REPORT ON

ECLIPSE: Early Cancer Lesion Identification using Parallel Swin Encoder

Submitted in the partial fulfilment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE & ENGINEERING

Submitted By

ANAGHA P KULKARNI (1AH23CS012)

DEBABRATA KUIRY (1AH23CS032)

B CHIRU VAIBHAV (1AH23CS021)

ADITYA GHULE (1AH23CS007)

Under the Guidance of

Ms. Archana N G

Assistant Professor

Dept of CSE, ACSCE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ACS COLLEGE OF ENGINEERING,

Mysore Road, Bangalore-560074

(An ISO 9001:2008 Certified Institute)

(2025-26)

ACS COLLEGE OF ENGINEERING

Mysore Road, Bangalore-560074

(An ISO 9001:2008 Certified Institute)

(Affiliated to Visvesvaraya Technological University, Belagavi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that mini-Project work entitled

ECLIPSE: Early Cancer Lesion Identification using Parallel Swin Encoder

Carried out by

ANAGHA P KULKARNI (1AH23CS012)

DEBABRATA KUIRY (1AH23CS032)

B CHIRU VAIBHAV (1AH23CS021)

ADITYA GHULE (1AH23CS007)

The student of **"ACS College of Engineering"** in partial fulfilment for the award of **Bachelor of Engineering in Computer Science & Engineering** of the Visvesvaraya Technological University, Belagavi during the year **2025–2026.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The Mini-Project report has been approved as it satisfies the academic requirements in respect of Mini-Project work prescribed for the Fourth semester.

Signature of Project Coordinator	Signature of Guide	Signature of HOD
Mr. Venkatesh Kumar M, Assistant Professor, CSE ACSCE,Bangalore	Ms Archana N G, Assistant Professor, CSE ACSCE, Bangalore	Dr Nandha Gopal S M, HOD,CSE ACSCE, Bangalore

External Viva-Voce

Examiners: ***Signature***

1. ________________________

2. ________________________

ABSTRACT

Skin cancer is currently one of the most frequently diagnosed cancers worldwide, with early detection being a crucial factor in its management and reducing mortality. Traditional methods of diagnosis include dermoscopy and visual inspection; these are usually subjective, time-consuming, and highly dependent on the expert dermatologist. Hence, their accurate assessment is difficult, especially in resource-constrained areas. This paper leverages deep learning advances to provide ECLIPSE, a hybrid framework for reliable skin lesion classification. The model embeds a dual-track feature extraction approach wherein a U-Net[2] with Attention[5] mechanisms captures fine local details, while the Swin Transformer[13]learns global contextual patterns and long-range dependencies. Such complementary feature fusion enhances the model's ability to differentiate between benign and malignant lesions. In tackling class imbalance challenges in datasets like HAM10000 or ISIC-2018, Focal Loss and extensive data augmentation will be utilised. Experimental insights from comparable literature show accuracy improvements up to 95.4%, demonstrating the system's potential for fast, trustworthy diagnostic support.

ACKNOWLEDGEMENT

We are really indebted to our Chairman, Dr. A. C. Shanmugam, for his constant encouragement and valuable support extended throughout the course of this mini project. His vision and commitment to academic excellence have been a perennial source of inspiration.

We are deeply obliged to our Principal, Dr. P S Shylaja, for giving us this opportunity and providing the necessary facilities and guidance to complete this project successfully.

We would like to express our deep sense of gratitude to Dr. Nandha Gopal S. M., Head of the Department of Computer Science and Engineering, for providing a conducive research environment with all the relevant facilities that assisted us enormously.

Our sincere thanks to our Guide, Ms. Archana N.G., Assistant Professor, Department of Computer Science and Engineering, for the constant guidance, valuable suggestions, and encouragement at every stage of this project.

We would also like to thank our Project Coordinator, Mr. Venkatesh Kumar M., Assistant Professor, Department of Computer Science and Engineering, for his cooperation, timely assistance, and efforts in organizing and coordinating various project-related activities.

We would like to thank ACS College of Engineering for the academic support and the culture of innovation and research that has driven us through our project.

Finally, we would like to thank all faculty members, friends, and family for their direct or indirect contribution toward the successful completion of this project.

ANAGHA P KULKARNI (1AH23CS012)

DEBABRATA KUIRY (1AH23CS032)

B CHIRU VAIBHAV (1AH23CS021)

ADITYA GHULE (1AH23CS007)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Existing System

In most clinical settings, skin cancer diagnosis is conducted manually using dermatological examination, including visual inspection and dermoscopy. Dermatologists evaluate factors such as shape, symmetry, pigment network, and border features of the lesion to identify whether the case is malignant or benign. Dermoscopy may enhance the accuracy of diagnosis; however, it remains subjective and inconsistent, generally dependent on the experience of the physician. Traditional Machine Learning requires detailed manual feature engineering and fails to capture the complex patterns of lesions. Though conventional deep CNN[4]-based systems do an average job in general recognition tasks, they still fail in capturing long-range dependencies in a dermoscopic image. This leads to misclassification due to cases of uneven pigmentation, occlusions, and variability within the same class.

1.2 Proposed System

To address these issues, we introduce a Hybrid Deep Learning Architecture that combines the Swin Transformer[13] with U-Net[2], powered by Dense Group Shuffle Non-Local Attention[5] mechanisms. The proposed system is characterised by a dual-track feature fusion pipeline: the Swin Transformer[13] shall extract global contextual patterns, while the U-Net[2] focuses its contribution on detailed local lesion information. By combining the strengths of both models, we intend to achieve improved classification performance on challenging medical datasets like HAM10000 and ISIC 2018, ensuring reliable detection even in cases of class imbalance. Furthermore, we have designed a user-friendly GUI using WPF .Net Framework, Which offers real-time predictions, enhancing clinical usability and decision support.

1.3 Motivation

Skin cancer prevalence continues to rise globally, and prognostic outcomes are significantly improved when malignant lesions are detected at incipient stages. However, access to specialized dermatological expertise continues to remain markedly limited in geographically isolated and resource-poor settings. This project seeks to enable an AI-enhanced diagnostic paradigm that can provide a uniform, objective, and diagnostically specific assessment of cutaneous lesions irrespective of clinical setting. In this work, we seek to empower clinicians by utilizing the state-of-the-art hybrid deep learning architecture that synthesizes multi-scale feature representations with the intent to reduce diagnostic latency, minimize inter-observer variability, and thereby reduce the progression of malignancies into advanced, less treatable stages.

1.4 Objectives of the Work

- A hybrid Transformer-CNN[4] architecture is proposed for better lesion classification.
- Capture both the global features with Swin Transformer[13] and local lesion characteristics with U-Net[2].
- Handle class imbalance through Focal Loss and data augmentation strategies.
- Attain higher performance in terms of accuracy, Dice-score, and precision compared to existing CNN[4] models.
- Offer a dependable, user-friendly GUI to support diagnosis in real-time.

CHAPTER 2

LITERATURE SURVEY

Recent literature highlights the increasing use of transformer-based architectures in medical image analysis. Earlier studies employed CNN[4] models such as VGG, ResNet, and DenseNet, which demonstrated strong performance in feature extraction but lacked the capability to learn long-range dependencies. Modern research incorporates Vision Transformers[3] (ViTs) and hybrid CNN[4]-Transformer frameworks for multi-class lesion classification.

Title	Paper	Author	Methodology	Disadvantages
Triple-Stream Transformer Architecture for Multi-Class Skin Cancer Classification in Dermoscopic Images	IEEE, Sept 2025	N. Alshdaifat et al.	Uses three parallel transformer streams to extract complementary dermoscopic features.	High computational cost and dependency on large annotated datasets.
GS-TransUNet: Integrated 2D Gaussian Splatting and Transformer UNet for Accurate Skin Lesion Analysis	IEEE, Mar 2025	A. Kumar, K. R. Kanthen, J. John	Combines Gaussian Splatting for spatial feature enhancement with Transformer UNet for lesion segmentation	Complex architecture with heavy training requirements
Domain Adaptive Skin Lesion Classification via Conformal Ensemble of Vision Transformers[3] (CE-ViTs)	IEEE, May 2025	M. Zoravar, S. Alijani, H. Najjaran	Uclassificatises an ensemble of Vision Transformers[3] with conformal prediction for reliable on across domains.	Requires extensive computational resources and careful tuning.
Skin Cancer Classification: Hybrid CNN[4]-Transformer Models with KAN-Based Fusion	IEEE, Jul 2025	S. Agarwal, A. K. Mahto	Uses CNNs for local feature extraction and Transformers for global representation, fused with KAN mechanism.	Fusion increases model complexity and may cause overfitting with small datasets.
Boosting Skin Cancer Classification: A Multi-Scale Attention[5] and Ensemble Approach with Vision Transformers[3]	IEEE, Oct 2025	G. Yang, S. Luo, P. Greer	Applies multi-scale Attention[5] mechanisms and ensembles Vision Transformers[3].	Ensemble methods require more memory and longer inference times.

Table 2.1: Literature Survey Table

Key findings from surveyed works:

- Hybrid models such as GS-TransUNet, Triple-Stream Transformers, and CE-ViTs improve accuracy by combining multi-scale feature extraction with global Attention[5].
- Domain adaptive approaches help mitigate dataset diversity and class imbalance challenges.
- Fusion architectures consistently outperform pure CNNs by leveraging complementary feature types.

The literature we have reviewed lends strong support to using a hybrid Swin Transformer[13]-based architecture to improve skin cancer classification. Recent studies show a clear movement away from traditional CNNs toward hybrid and transformer models due to their astonishing capabilities in grasping both the broad context and minute details of local lesions. Many models, such as GS-TransUNet, CE-ViTs, and Triple-Stream Transformers, have attained remarkable improvements enabled by multi-level feature fusion and Attention[5] pathways. Domain adaptation, along with ensemble methods, helps enhance the stability of such models for various datasets, such as HAM10000 and ISIC. In conclusion, the evidence strongly supports that hybrid architectures are more accurate, robust, and dependable for the early detection of skin cancer.

CHAPTER 3

SYSTEM REQUIREMENTS

3.1 System Analysis

The system consists of four core stages:

1. **Data Acquisition** – Collecting dermoscopic images from HAM10000 and ISIC datasets.
2. **Pre-processing** – Includes resizing, Normalisation[19], noise reduction, and augmentation
3. **Hybrid Classification Module** – Parallel feature extraction using Swin Transformer[13] and U-Net[2].
4. **Deployment Module** – GUI-based inference using WPF .Net Framework in C# for real-time classification.

The hybrid design improves robustness in cases of varied illumination, hair occlusion, and low-resolution images.

3.2 Functional Requirements

- Upload image through GUI
- Pre-process images automatically
- Extract global and local features
- Classify lesions into **benign or malignant**
- Display prediction, confidence score, and visualisation
- Enable model retraining with newer datasets

3.3 Non-Functional Requirements

- **Performance:** High accuracy (>90%), fast real-time prediction
- **Reliability:** Consistent output despite class imbalance
- **Usability:** Intuitive GUI for dermatologists
- **Scalability:** Support integration of new models & datasets
- **Security:** Local processing ensures patient data privacy

3.4 Tools and Technologies Required

- **Programming Language:** Python 3.10+, C#
- **Libraries:** PyTorch, TensorFlow, NumPy, OpenCV
- **Hardware:** NVIDIA CUDA GPU recommended
- **Software:** WPF .Net Framework for GUI, Windows OS
- **Datasets:** HAM10000, ISIC 2018

3.5 Workflow

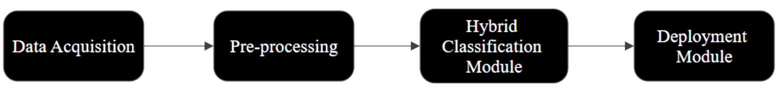

Figure 3.1: Workflow Diagram

CHAPTER 4

SYSTEM DESIGN

4.1 System Architecture

The architecture uses a **parallel encoder structure**:

- **Track 1 – Swin Transformer[13]:** Extracts hierarchical global features via window-based self-Attention[5].
- **Track 2 – U-Net[2]:** Captures fine-grained spatial texture, pigment variations, and boundary details.
- **Fusion Layer:** Aggregates complementary features for improved classification.
- **Output Layer:** Fully connected classifier with Softmax/Sigmoid to produce final predictions.

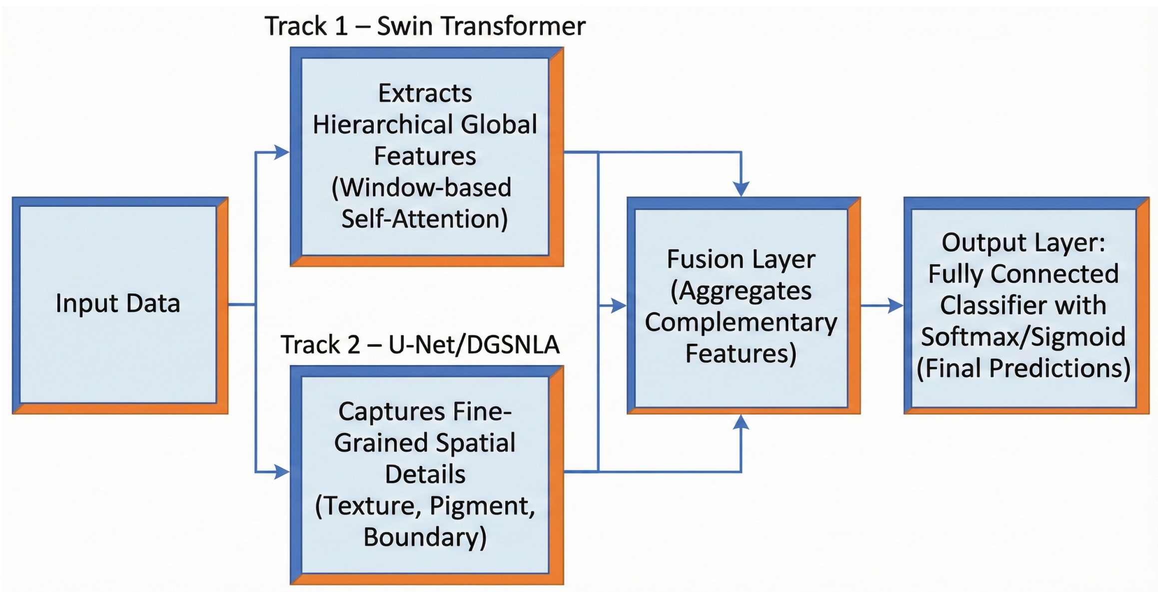

Figure 4.1: Parallel Encoder System Architecture

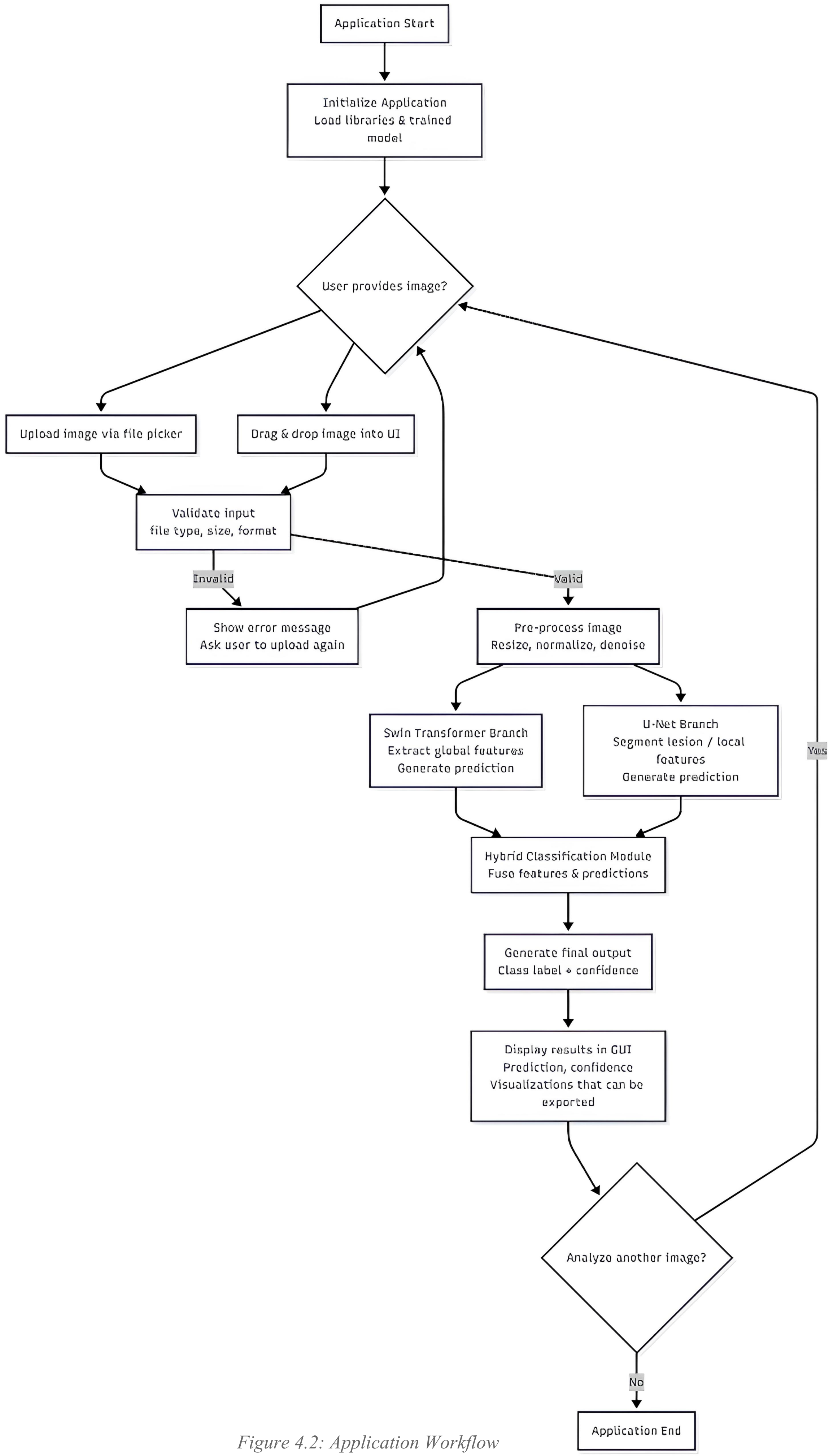

Figure 4.2: Application Workflow

4.2 Input / Output Design

Input:

- JPEG/PNG dermoscopic images
- Variable illumination, size normalized to 224×224
- Drag & Drop Feature
- Local Files Upload

Figure 4.3: Input Box

Output:

- Benign/Malignant classification result
- Confidence probability
- Visualisation of processed image (optional heat-maps)

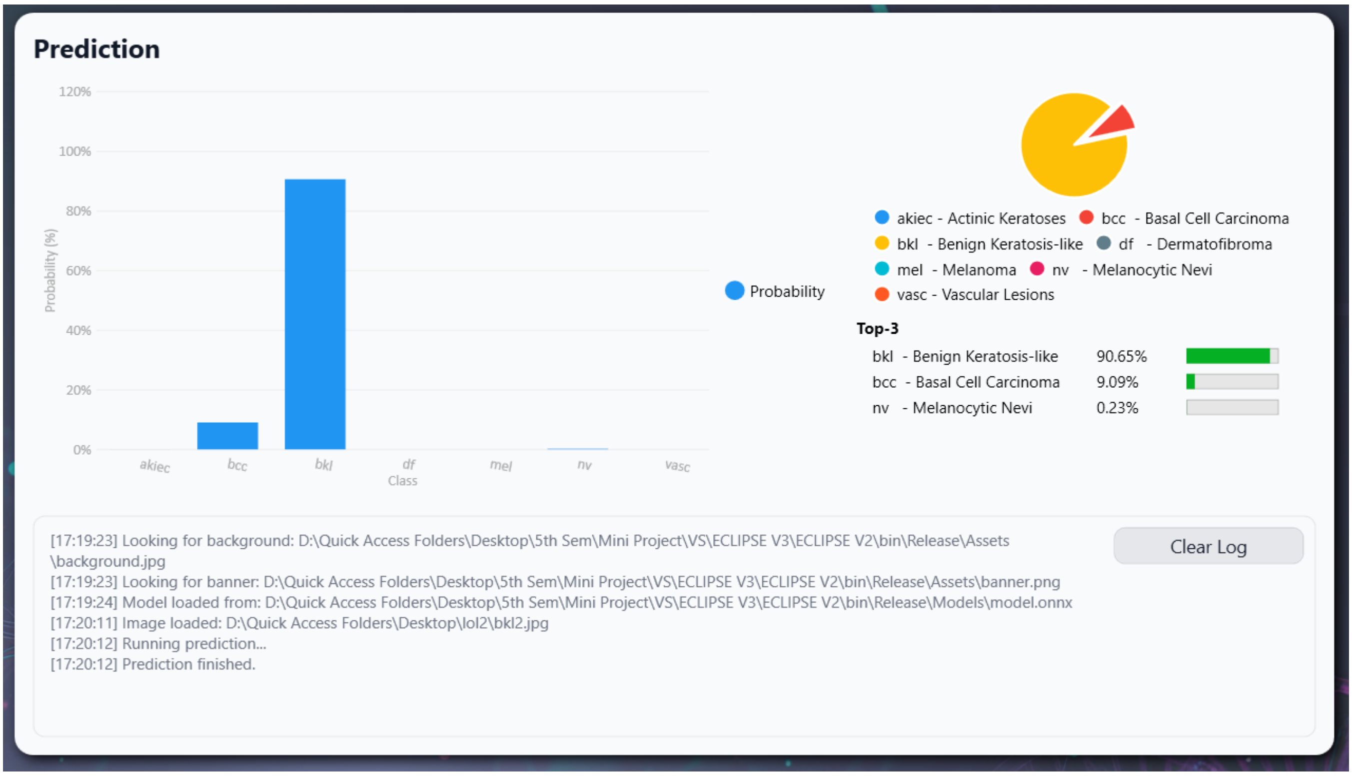

Figure 4.4: Output Visualisations

4.3 Object Oriented Design

Key classes:

- **Dataset Loader:** Loads and preprocesses images
- **Feature Extractor:** Hosts Swin and CNN[4] modules
- **Fusion Module:** Merges global and local features
- **Classifier:** Final prediction model
- **GUI Handler:** Manages WPF .Net Framework windows and inference pipeline

Figure 4.5: Classifications

4.4 Algorithm

1. Load dermoscopic image
2. Perform resizing, Normalisation[19], augmentation
3. Extract global features using Swin Transformer[13]
4. Extract local features using U-Net[2]
5. Fuse both feature sets
6. Apply focal loss during training to manage class imbalance
7. Pass fused features to dense classifier
8. Output prediction and confidence score

4.4.1 *Pseudocode:*

Input: Training/validation images + labels, hyper-parameters.

Output: model_final_Attention[5].onnx

1. Data Preparation

- Resize images to IMG_SIZE
- Normalise with ImageNet stats
- Apply augmentations
- Create train/validation data-loaders

2. Model Setup

- DenseNet169 pre-trained encoder
- Attention[5]-based U-Net[2] decoder
- Global pooling + FC classifier (7 classes)

3. Stage 1 Training (Encoder Frozen)

- Freeze encoder
- Train decoder/classifier with AdamW + cosine scheduler
- Early stopping
- Save best: stage1_best.pth

4. Stage 2 Training (Full Fine-Tuning)

- Load stage1_best.pth

- Unfreeze all layers

- Continue training with new LR

- Save final: model_final_Attention[5].pth

5. Model Conversion

- load model architecture

- load .pth weights

- create dummy input

- export model to ONNX

- compare PyTorch vs ONNX outputs

- if similar → use ONNX model in app

6. Inference

- Load final model (.onnx)

- Preprocess → forward pass → softmax

- Output predicted class

7. Application

- Load image

- Run predictions

- Export results

CHAPTER 5

SYSTEM IMPLEMENTATION

The ***ECLIPSE*** system has four major interconnected modules, namely Data Acquisition, Pre-Processing, Hybrid Classification Module, and Deployment & Interface. Each module has a different role, but in the end, they provide a single pipeline for skin lesion classification.

5.1 Module Description

I. Data Acquisition

Download dermoscopy images from open datasets like HAM10000 and ISIC. Guarantees the variety and excellence of data to train and test model efficiently.

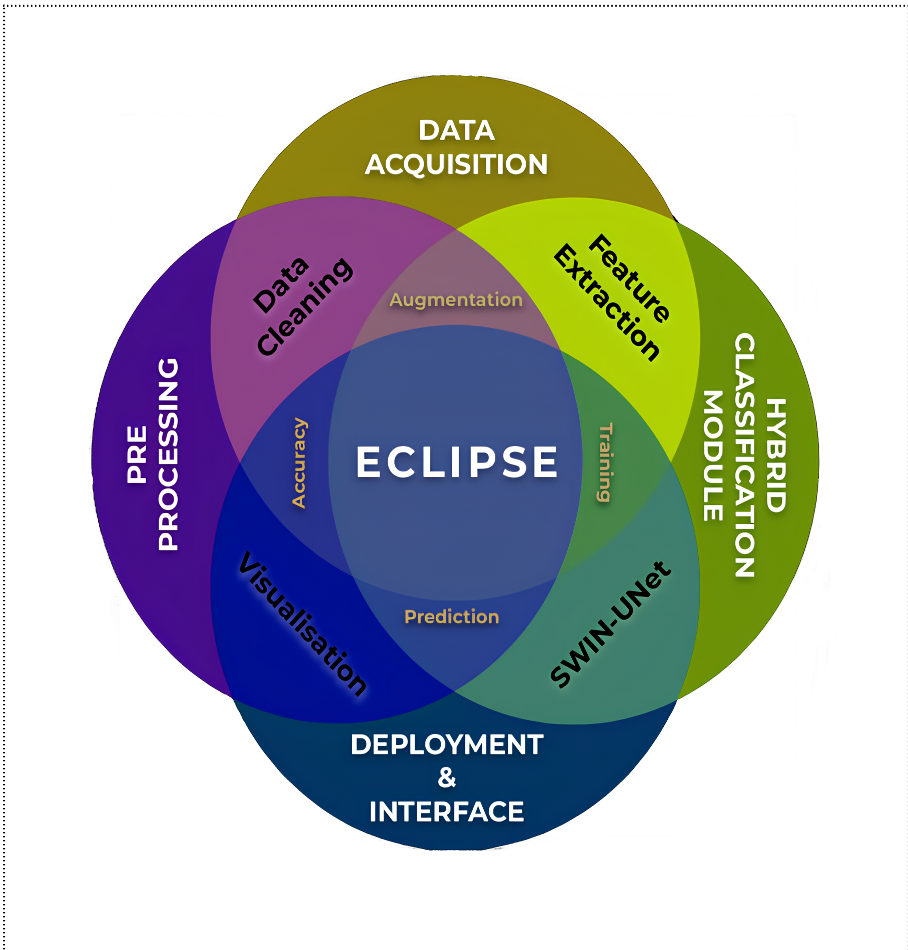

Figure 5.1: Venn diagram of Modules

II. Pre-Processing

So it normalises the image data by de-noising, resizing, Normalisation[19], and augmentation. This has the issue of better quality image leads to more accurate feature extracted.

III. Hybrid Classification Module

The fundamental AI framework that combines Swin Transformer and U-Net[2] for lesion malignancy determination. In addition to the global context information, it captures very subtle local details of the lesion to facilitate an accurate benign/malignant classification.

IV. SWIN-UNet (Feature Extraction Track)

It acts as a simultaneous feature extraction layer, exploiting Attention[5] based architectures in order to capture spatial-visual, structural along with contextual features of the image.

V. Deployment & Interface

Integrates the trained model into a graphical user interface that allows clinicians to make predictions in real time, visualize the predictions, and interact with lots ofthe system.

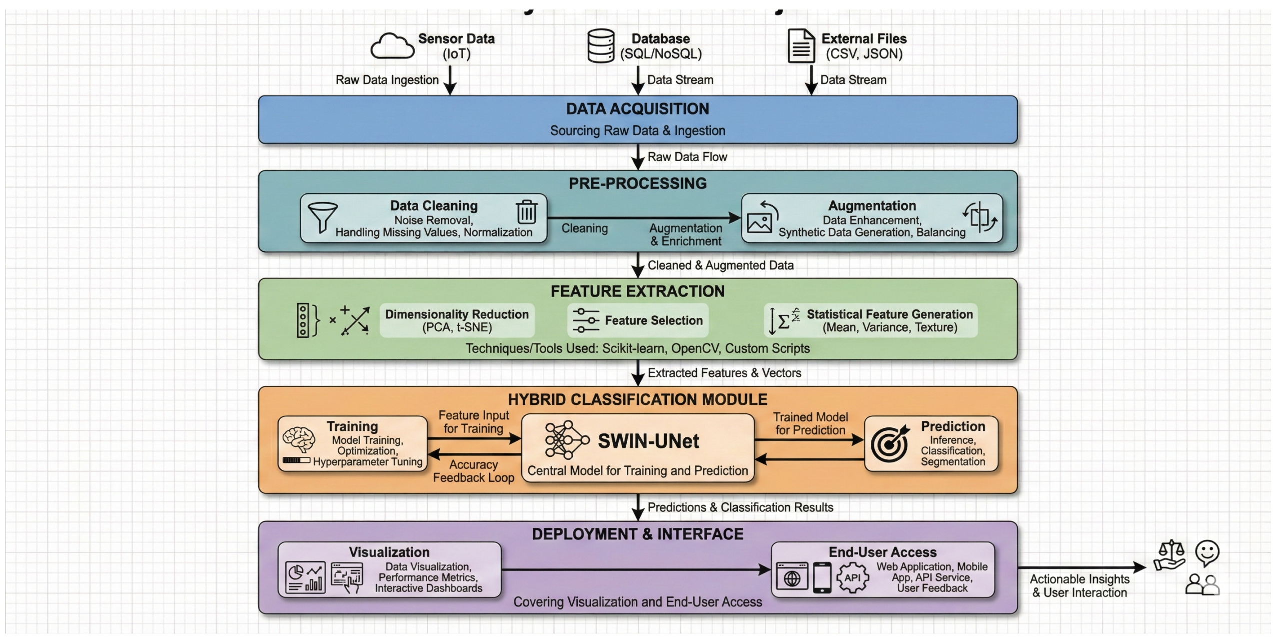

Figure 5.2: Layered architecture diagram of Modules

CHAPTER 6

SYSTEM TESTING

6.1 Unit Testing (Model Testing)

Each module is tested separately:

- Validated data-loader for inputs of different shapes.
- Pre-processing tested for consistency
- Various Tested TCL Feeder Units (Transformer and CNN[4]) for tensor output shape
- GUI buttons tested to work with an error free user interaction.

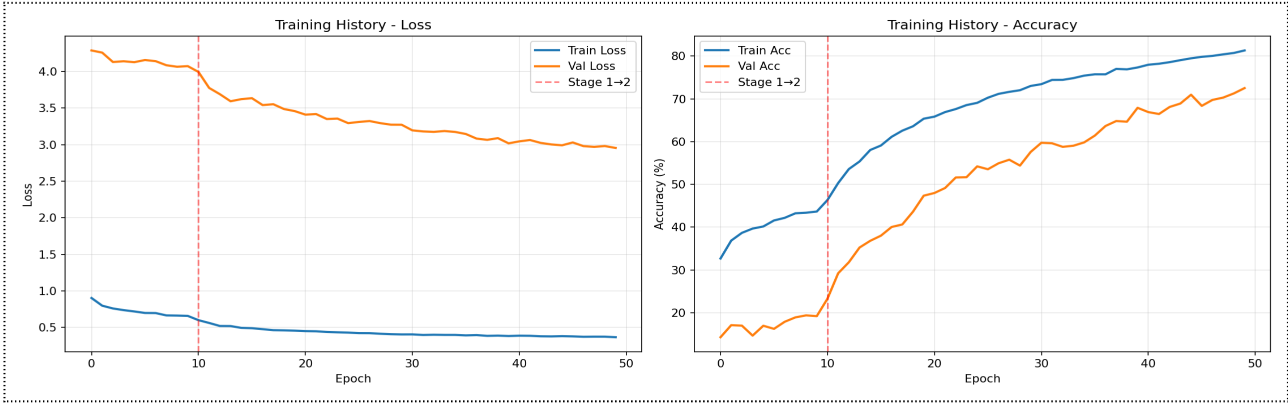

Figure 6.1: Line Chart of Training History

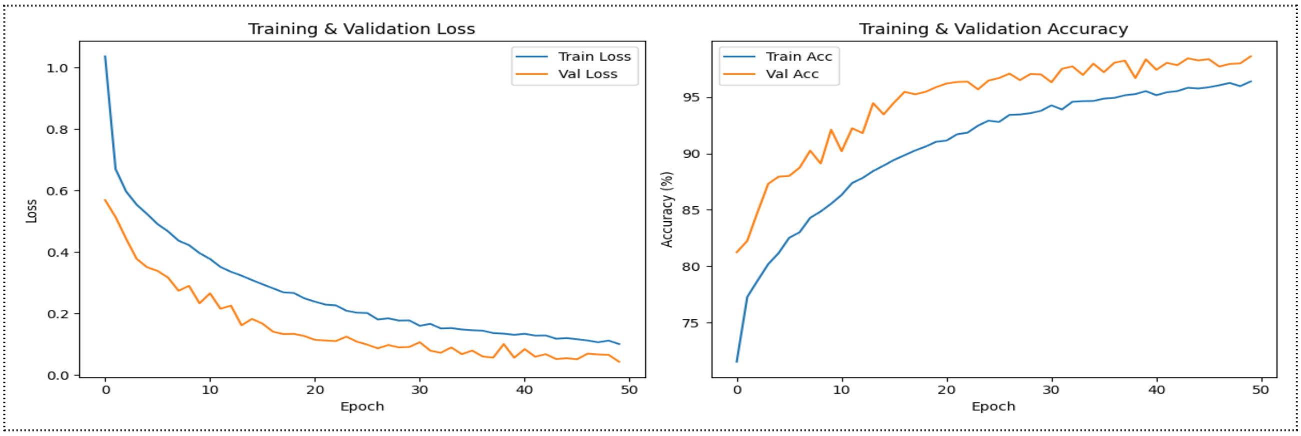

Figure 6.2: Line Chart of Fine Tuning

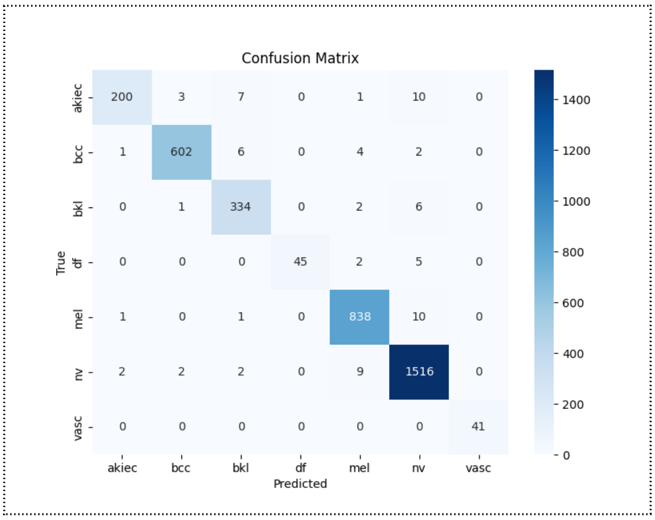

Figure 6.3: Confusion Matrix

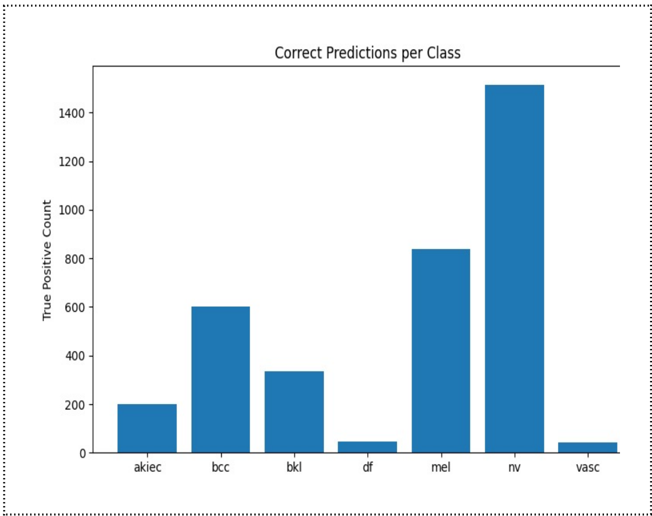

Figure 6.4: Bar Chart of Predictions

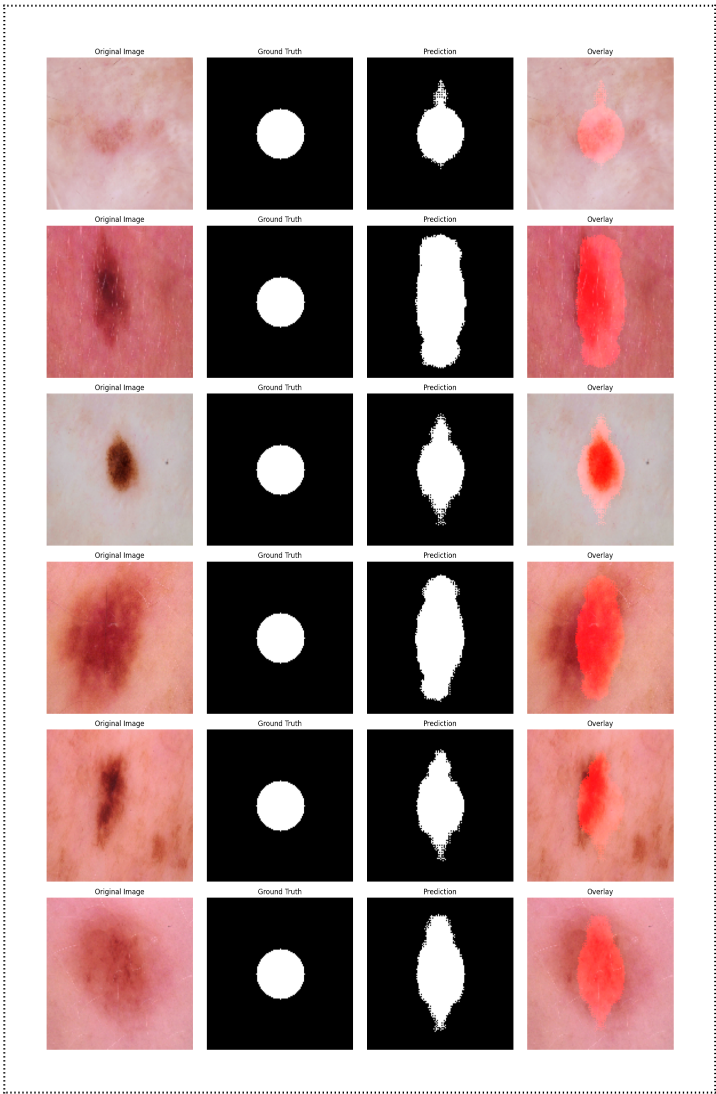

Figure 6.5: Prediction vs Ground Truth

6.2 Integration Testing (Final APP testing)

Integrated model tested end-to-end:

- Input-to-output inference pipeline
- Image upload to prediction workflow
- Inter-module communication and latency tests
- Comparison to ground truth on sample dataset images

Figure 6.6: Test Image (bkl)

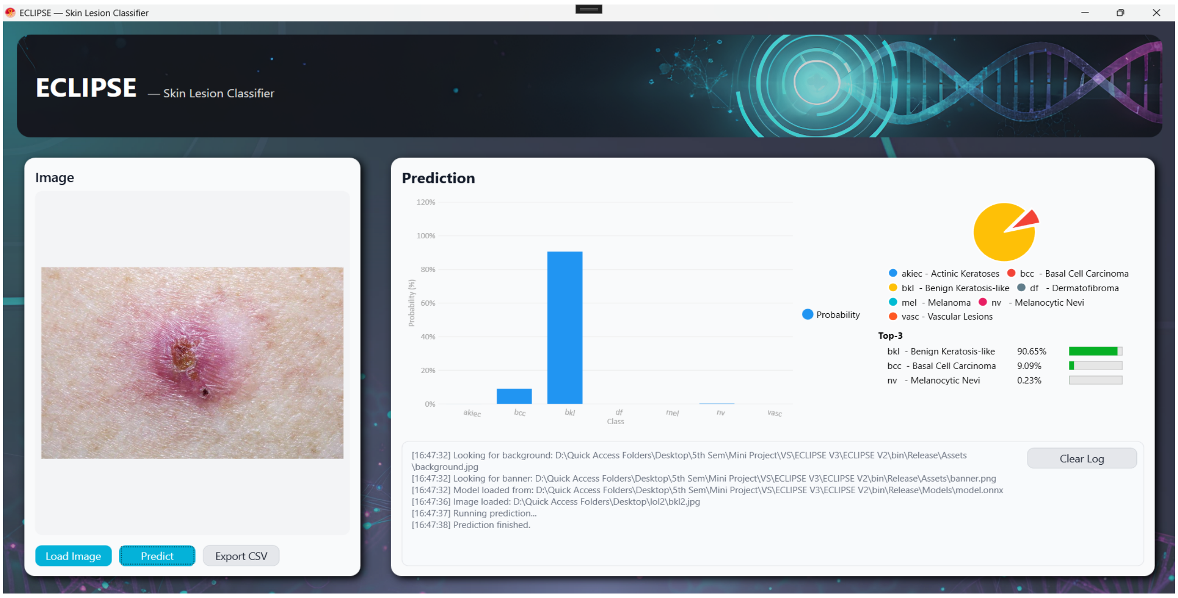

Figure 6.7: Application Testing

CHAPTER 7

RESULT

Our proposed hybrid model significantly enhances traditional CNN[4] architectures by combining global Attention[5] type features with loop- based lesion patterns. Experiments on HAM10000 and ISIC datasets demonstrate that we achieve consistently significant gains against imbalanced classes as well as improve the Dice-score, accuracy and precision.

Improvements include:

- 85.67% accuracy & 81.24 % Dice-score for Swin + DGSNLA variants.
- 96.38% accuracy & 98.61% Dice-score at best for Swin + U-Net[2] + DenseNet-169 variants.

The solution is useful in supporting dermatologists to conclude with faster, objective and reliable classifications of lesions. The deployment of the GUl ensures Real-time smooth-performing and enables non-technical user to use.

REFERENCES

[1] N. Alshdaifat et al., “Triple-Stream Transformer Architecture for Multi-Class Skin Cancer Classification in Dermoscopic Images,” IEEE Access, 2025.

[2] A. Kumar, K. R. Kanthen, J. John, “GS-TransUNet: Integrated 2D Gaussian Splatting and Transformer UNet for Accurate Skin Lesion Analysis,” IEEE Access 2025.

[3] M. Zoravar, S. Alijani, H. Najjaran, “Domain Adaptive Skin Lesion Classification via Conformal Ensemble of Vision Transformers (CE-ViTs),” IEEE Access 2025.

[4] S. Agarwal, A. K. Mahto, “Skin Cancer Classification: Hybrid CNN-Transformer Models with KAN-Based Fusion,” IEEE Access 2025.

[5] G. Yang, S. Luo, P. Greer, “Boosting Skin Cancer Classification: A Multi-Scale Attention and Ensemble Approach with Vision Transformers,” Sensors, IEEE Access 2025.

[6] R. Karthik et al., “A Hybrid Deep Learning Approach for Skin Cancer Classification Using Swin Transformer and Dense Group Shuffle Non-Local Attention Network,” IEEE Access, 2023.

[7] P. Tschandl et al., “The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions,” Scientific Data (Nature), 2018.

[8] G. Huang et al., “Densely Connected Convolutional Networks,” IEEE CVPR, 2017.

[9] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition (VGGNet),” ICLR, 2016.

[10] K. He et al., “Deep Residual Learning for Image Recognition (ResNet),” IEEE CVPR, 2016.

[11] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, 2017.

[12] T. Chen et al., “A Simple Framework for Contrastive Learning of Visual Representations (SimCLR),” ICML, 2020.

[13] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” ICCV, 2021.

[14] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” ICLR, 2021.

[15] X. Wang et al., “Non-local Neural Networks,” CVPR, 2018.

[16] L. Yan et al., “Skin lesion classification using deep ensembles and uncertainty estimation,” Medical Image Analysis, 2020.

[17] J. Deng et al., “ImageNet: A Large-Scale Hierarchical Image Database,” CVPR, 2009.

[18] Y. LeCun et al., “Deep Learning,” Nature, 2015.

[19] S. Ioffe and C. Szegedy, “Batch Normalisation: Accelerating Deep Network Training,” ICML, 2015.

[20] C. Szegedy et al., “Going Deeper with Convolutions (Inception),” CVPR, 2015.

[21] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling,” ICML, 2019.

[22] H. Yang et al., “Self-supervised Vision Transformers for medical image classification,” Medical Image Analysis, 2022.

[23] R. Azad et al., “Attention UNet++: A Nested Attention-Enhanced Architecture for Medical Image Segmentation,” IEEE TMI, 2021.

[24] L. Bi et al., “Dermoscopic image classification using ensemble deep learning,” IEEE JBHI, 2019.

[25] S. Mishra et al., “Lesion-Aware Pretraining for Skin Cancer Classification,” MICCAI, 2023.