

Skin Cancer Detection Using Deep Learning Methods

THESIS

*Submitted towards the partial fulfilment of
the requirements of the award of the degree of*

Master of Technology In Data Science

Submitted by

SHIVASHISH KAUSHIK

2K24/DSC/10

Under the guidance of

Dr. Sonika Dahiya

Assistant Professor

Department of Software Engineering



DELHI TECHNOLOGICAL UNIVERSITY

(FORMERLY Delhi College of Engineering)

Bawana Road, New Delhi-110042

May 2026

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, New Delhi-110042

CANDIDATE'S DECLARATION

I, SHIVASHISH KAUSHIK, Roll No - 2K24/DSC/10, students of M.Tech (Data Science), hereby declare that the Thesis titled — “**Skin Cancer Detection Using Deep Learning Methods**” which is submitted by me to the Department of Software Engineering, Delhi Technological University, New Delhi in fulfillment of the requirement of the award of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associationship, Fellowship or other similar title or recognition.

Place: Delhi

Date:

SHIVASHISH KAUSHIK

2K24/DSC/10

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly New Delhi College of Engineering)

Bawana Road, New Delhi-110042

CERTIFICATE

I hereby declare that the Project titled **Skin Cancer Detection Using Deep Learning Methods** which is submitted by SHIVASHISH KAUSHIK - 2K24/DSC/10, Department of Software Engineering, New Delhi Technological University, Delhi in partial fulfillment of the degree of Master of Technology is a record of the project work carried out by the students under my supervision. To the best of my knowledge, this work has not been submitted in any part or fulfillment for any Degree or Diploma to this University or elsewhere.

Place : Delhi

Date :

Dr. SONIKA DAHIYA

SUPERVISOR

(Assistant Professor, SE, DTU)

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to Dr. Sonika Dahiya, Assistant Professor, Department of Software Engineering, Delhi Technological University, for her continuous guidance, support, and mentorship throughout the preparation of this thesis. Her insights and expertise were instrumental in shaping the direction of my research and explaining all the tasks to be accomplished, emphasising the importance of this work as well as its relevance to real-world applications.

I would also like to thank all the teachers and students of the Department of Software Engineering, Delhi Technological University, for providing the necessary resources, continuous support, and valuable suggestions that facilitated the completion of this research. Lastly, I would like to thank all my family and friends for providing continuous support and motivation during the process of this study.

SHIVASHISH KAUSHIK

2nd Year

2K24/DSC/10

Department of Software Engineering

Delhi Technological University

ABSTRACT

Skin cancer is one of the most common and dangerous types of cancer across the world, with around 1.5 million new cases being diagnosed every year. Traditional methods of diagnosis work effectively but suffer from problems related to inter-observer variability and specialist availability. This thesis presents two complementary contributions. First, a systematic review of eighteen deep learning studies published between 2020 and 2025 is conducted, covering convolutional neural networks, Vision Transformers, ensemble methods, and hybrid models on benchmark datasets including HAM10000 and the ISIC archive. Key research gaps identified include the absence of reproducible Vision Transformer benchmarks on standard dermoscopy datasets. Secondly, Low-Rank Adaptation (LoRA) is performed on ViT-B/16 for multiclass dermoscopic classification on the HAM10000 dataset by fine-tuning only 1.36% of all the model’s parameters. The proposed LoRA-ViT network provides a test-set macro F1 score of 0.7222, accuracy of 78.04%, and macro AUC-ROC of 0.9577, surpassing MobileNetV3 and approaching ResNet-50 with significantly lower computational overhead. The best choice, according to the rank ablation experiment, is $r = 32$. The presented experiments show that parameter-efficient fine-tuning of Vision Transformers can be considered a good method for detecting skin cancer from dermoscopic images under limited resources.

LIST OF PUBLICATIONS

1. Shivashish Kaushik & Sonika Dahiya (2026). *Advances in Automated Skin Cancer Detection Using Deep Learning Methods*. Proceedings of ComSIA 2026. [Springer LNNS] [Accepted]
2. Shivashish Kaushik & Sonika Dahiya (2026). *LoRA-ViT: Parameter-Efficient Fine-Tuning for Skin Cancer Detection*. Proceedings of InCODE 2026. [IEEE] [Submitted]

Contents

Candidate's Declaration	i
Certificate	ii
Acknowledgement	iii
Abstract	iv
LIST OF PUBLICATIONS	v
Table of Contents	vii
List of Tables	viii
List of Figures	ix
List of Abbreviations	x
CHAPTER 1: INTRODUCTION	1
1.1 Motivation	1
1.2 Problem Statement	1
1.3 Overview of Proposed Work	2
1.4 Objectives of the Study	2
CHAPTER 2: RELATED WORK	4
2.1 Deep Learning Architectures for Skin Cancer Detection	4
2.2 Vision Transformers in Medical Imaging	5
2.3 Existing Review Papers	5
CHAPTER 3: METHODOLOGY	7
3.1 Dataset	7
3.2 Data Partitioning and Class Imbalance Handling	8

3.3	Image Preprocessing and Augmentation	8
3.4	Proposed LoRA-ViT Architecture	10
3.5	Baseline Models	11
3.6	Training Configuration	11
3.7	Evaluation Metrics	13
CHAPTER 4: EXPERIMENTAL SETUP & RESULTS		15
4.1	Implementation Details	15
4.2	Evaluation Metrics	16
4.3	Baseline Model Results	16
4.4	LoRA-ViT Results	18
4.5	Comparison with Baseline Models	19
4.6	Ablation Study	20
4.7	Discussion	22
CHAPTER 5: CONCLUSION AND FUTURE WORK		24
5.1	Conclusion	24
5.2	Future Work	24

List of Tables

Table 2.1 : Comparative Performance of Skin Cancer Detection Methods from Systematic Review	6
Table 3.1 : LoRA Configuration for ViT-B/16 Fine-Tuning	10
Table 3.2 : Training Configuration Applied to All Models	12
Table 3.3 : Evaluation Metrics Used for Model Performance Assessment	14
Table 4.1 : Implementation Environment and Software Configuration	16
Table 4.2 : ResNet-50 Training Summary	17
Table 4.3 : MobileNetV3-Large Training Summary	17
Table 4.4 : Test Set Performance Comparison on HAM10000 (1,503 Test Samples)	19
Table 4.5 : LoRA Rank Ablation Study on HAM10000 Validation Set (20 Epochs)	21
Table 4.6 : Target Module Ablation Study on HAM10000 Validation Set ($r = 32, 10$ Epochs)	22

List of Figures

Figure 3.1 :	7
Figure 3.2 : Augmented training samples from HAM10000. Geometric transforms including rotation, flipping, and affine translation are combined with colour jitter to increase training data diversity and improve model generalisation.	9
Figure 4.1 : LoRA-ViT training and validation curves across 25 epochs. The green dashed line marks the best validation checkpoint at Epoch 19 (Val F1 = 0.6881, Test F1 = 0.7222).	18
Figure 4.2 : Per-class ROC curves for LoRA-ViT on the HAM10000 test set.	19
Figure 4.3 : Presents a visual comparison of F1-macro scores across all three models..	20
Figure 4.4 : LoRA rank ablation results on HAM10000 validation set. Left: Val F1 vs rank showing peak at $r = 32$.	21

List of Abbreviations

ABBREVIATION	DEFINITION
CNN	Convolutional Neural Network
ViT	Vision Transformer
LoRA	Low-Rank Adaptation
PEFT	Parameter-Efficient Fine-Tuning
DL	Deep Learning
ML	Machine Learning
SVM	Support Vector Machine
GAN	Generative Adversarial Network
AUC	Area Under the ROC Curve
ROC	Receiver Operating Characteristic
BCC	Basal Cell Carcinoma
SCC	Squamous Cell Carcinoma
MEL	Melanoma
ISIC	International Skin Imaging Collaboration
GPU	Graphics Processing Unit
SE	Squeeze-Excitation
MCC	Matthews Correlation Coefficient

Chapter 1

INTRODUCTION

1.1 Motivation

Increasing prevalence of skin cancer around the world and issues related to its diagnosis, a lot of attention has been paid to deep learning-based diagnostic systems. Deep learning techniques have been successfully used for skin cancer detection in an automated fashion. CNNs have performed well on this problem, and some of the most recent architectures include ResNet-50, DenseNet, VGG-16, and Inception-V3, all of which have shown accuracy levels ranging from 91% to 96% [4,5,7]. Further improvements have been made by using ensemble techniques and hybrid CNN-ML techniques, where the maximum achieved F1-score was 0.93 [4,5,7]. Ensemble approaches and hybrid CNN-ML models have further pushed performance, with the highest 0.93 F1-score reported through a Squeeze-Excitation DenseNet fused with a metaheuristic ensemble [10]. Despite these advances, Vision Transformers (ViTs) [13], which capture global patch-level dependencies that CNNs cannot model due to their local receptive fields, have received comparatively limited attention in this domain [2]. Another problem which has not been considered yet is computational accessibility. In order to adjust all the parameters of extensive pre-trained models like ViT-B/16., one needs to update all 86 million parameters which are computationally costly because it needs a powerful GPU-based system which cannot be afforded by most research and clinical organizations. Low-Rank Adaptation (LoRA) [14] was designed for Large Language Models in order to overcome this problem using low-rank adapter matrices in the frozen pre-trained weights.

1.2 Problem Statement

Deep learning-based skin cancer detection has advances a lot, but there are still significant obstacles to overcome. Current research mostly uses CNN-based architectures trained on dermoscopic benchmark datasets like the ISIC archive and HAM10000 [19]. Nevertheless, the research community lacks a repeatable benchmark for ViT performance in dermatology because Vision Transformers, which show theoretical advantages in capturing long-range spatial relationships in lesion images [2], have not been rigorously benchmarked on these standard

evaluation splits. The high computational cost of fine-tuning big pretrained vision models is a closely related issue. Complete fine-tuning methods necessitate updating every model parameter, which makes experimentation resource-intensive and unfeasible without advanced GPU infrastructure. Despite the impressive performance of parameter-efficient fine-tuning methods like LoRA [14] on natural language processing and general vision tasks in recent years, their use on dermo- scopic image classification has yet to be investigated. In addition, class imbalance remains a constant problem in skin cancer datasets. The HAM10000 dataset shows that Melanocytic Nevi accounts for 66.9% of all images, whereas Dermatofibroma comprises just 1.1%, yielding an imbalance ratio of 58:1 [19]. However, despite the acknowledged problem of imbalance in most papers, different approaches to dealing with this issue, namely different dataset splits, metrics, and class definitions, were used. As a result, there is a need for a standardized experimental setup where LoRA fine-tuning of ViT is tested against CNN baselines

1.3 Overview of Proposed Work

This thesis contributes to the above problems through two novel research studies. First, A comparison table that integrates accuracy, F1 score, AUC, sensitivity, and specificity in all reviewed papers is constructed. Five important research gaps are found during the review, among which the primary one is the lack of reproducible benchmarks for ViTs on standard dermoscopic datasets. The primary gap directly addressed by the second contribution. We propose and implement the application of Low-Rank Adaptation to ViT-B/16 [13] for multi-class dermoscopic skin lesion classification on HAM10000 [19]. Proposed LoRA-ViT framework fine-tunes only 1.36% of total model parameters by placing trainable low-rank adapter matrices into the Query and Value attention projections across all twelve transformer encoder blocks, while keeping the original pretrained weights frozen. AdamW optimizer used to train model with a OneCycleLR scheduler incorporating a 10% linear warmup, and is evaluated against ResNet-50 and MobileNetV3 baselines trained under identical experimental conditions [18]. A systematic rank ablation study across $r \in \{4, 8, 16, 32, 64\}$ and a target module ablation study are further conducted to empirically identify the optimal LoRA configuration for this task.

1.4 Objectives of the Study

The objectives of this research are as follows:

1. To identify critical research gaps in the existing literature, with particular focus on the

absence of Vision Transformer benchmarks on standard dermoscopic datasets and the lack of parameter-efficient fine-tuning approaches in dermatological AI.

2. To implement and evaluate a parameter-efficient LoRA-based fine-tuning framework for ViT-B/16 on the HAM10000 dermoscopic dataset, establishing the first reproducible LoRA-ViT benchmark in skin cancer detection.
3. To compare the proposed LoRA-ViT model against ResNet-50 and MobileNetV3 baselines under identical experimental conditions using standardized evaluation metrics including accuracy, F1-macro, and AUC-ROC.
4. To conduct a systematic ablation study across LoRA ranks and target module configurations in order to identify the Pareto-optimal trade-off between parameter efficiency and classification performance.
5. To demonstrate that parameter-efficient fine-tuning of Vision Transformers

Chapter 2

RELATED WORK

2.1 Deep Learning Architectures for Skin Cancer Detection

This landmark study established deep learning as a clinically viable approach and motivated a subsequent wave of research exploring progressively more sophisticated architectures.

Convolutional Neural Networks

CNNs have become the prevailing architectural choice for dermoscopic image classification owing to their capability of learning visual features hierarchically. Lower layers learn basic features like edges and color gradients, whereas higher layers learn sophisticated task-specific features like lesion texture, irregular borders, and pigment network patterns [21]. VGGNet was proposed by Simonyan and Zisserman [22], showing that the depth of the network with small 3×3 convolutions is an important aspect of classification success. VGG-16 and VGG-19 have been widely adopted as baseline architectures in dermatology. [10] extended this concept with a Squeeze-Excitation DenseNet (SE-DenseNet), incorporating channel attention mechanisms that adaptively recalibrate feature responses. Fused with a metaheuristic ensemble of multiple deep models.

Custom and Lightweight Architectures

The model shows that task-specific architecture design can yield comparable results without requiring the high computational cost associated with big pre-trained networks. Howard et al. [27] introduced MobileNetV3, lightweight architecture that can be deployed on mobile and edge platforms by leveraging depthwise separable convolutions to minimize computation. Li et al. [11] evaluated MobileNetV3 alongside DenseNet on ISIC 2017 and 2019 datasets, reporting 92.7% accuracy with sensitivity of 0.86, demonstrating that lightweight architectures can achieve clinically relevant performance while remaining deployable hardware. Musthafa et al. [6] proposed a checkpoint-based optimised CNN incorporating periodic model saving during training to recover the best generalising model state. Evaluated on HAM10000 and ISIC 2019, this approach achieved 94.3% accuracy with precision of 0.91, demonstrating that train-

ing strategy optimisation can improve performance independently of architectural changes.

2.2 Vision Transformers in Medical Imaging

This global attention capability offers a theoretical advantage for dermoscopic classification, involve relationships between spatially distributed image regions. Yang et al. [2] investigated ViT-based architectures specifically for skin cancer and reported superior classification on lesion patterns requiring holistic spatial reasoning. Their study, however, did not include controlled benchmarking on standard HAM10000 or ISIC evaluation protocols, leaving the quantitative performance of ViTs on established dermoscopy splits uncharacterised. More broadly, transformer-based models have produced strong results across medical imaging domains including radiology, computational pathology, and ophthalmology. Chen et al. [29]. Nevertheless, ViTs typically require substantially larger training corpora than CNNs to reach comparable accuracy, and their full fine-tuning demands considerable GPU memory, limiting their accessibility for many research groups.

2.3 Existing Review Papers

Their review evaluated studies using dermoscopic benchmark datasets including HAM10000 and ISIC, reporting that CNN-based approaches consistently achieve high accuracy on dermoscopic images. However, their review was confined entirely to dermoscopic datasets, with no evaluation of studies using clinical photographs, and did not consider Vision Transformer architectures which emerged as a significant research direction from 2021 onwards.

Table 2.1: Comparative Performance of Skin Cancer Detection Methods from Systematic Review

Ref.	Method	Dataset	Accuracy	Key Result
[3]	Hybrid CNN + SVM	ISIC, HAM10000	93.5%	High accuracy with AUC of 0.89
[4]	CNN (ResNet-50, DenseNet)	ISIC 2017, 2018	92%	Strong sensitivity of 0.88
[5]	CNN (SkinNet-14)	DermIS, ISIC 2018	90.2%	Good balance of accuracy and training time
[6]	Optimized CNN (Checkpoint-based)	HAM10000, ISIC 2019	94.3%	High precision of 0.91
[7]	CNN (VGG-16, VGG-19, ResNet, DenseNet)	ISIC 2016, 2017	91%	Good recall of 0.89
[8]	CNN (Custom Deep Neural Network)	ISIC 2017, 2018	93%	Strong accuracy with AUC of 0.88
[9]	CNN (InSiNet Architecture)	ISIC 2019, HAM10000	95%	High precision of 0.92
[10]	Hybrid CNN + Meta-heuristic Ensemble	ISIC 2018, 2019	96%	Highest F1-score of 0.93
[11]	CNN (MobileNetV3, DenseNet)	ISIC 2017, 2019	92.7%	Good sensitivity of 0.86
[30]	CNN (ResNet-50 with Regularizer)	ISIC 2016	94.5%	High accuracy with AUC of 0.92
[31]	Hybrid CNN + ANN (Image + Tabular Fusion)	ISIC + Structured	68.7%	Moderate accuracy combining modalities
[32]	Ensemble CNN (VGG, ResNet)	ISIC 2018, HAM10000	95.1%	Strong recall of 0.92
[33]	CNN (Custom Architecture)	HAM10000, ISIC 2019	91.5%	Reliable classification results
[12]	CNN (VGG-16, DenseNet, ResNet)	ISIC 2017, 2018	93%	Consistent sensitivity of 0.89

Chapter 3

METHODOLOGY

The overall pipeline encompasses dataset preparation, class imbalance handling, image pre-processing and augmentation, Vision Transformer architecture with LoRA adaptation, baseline model training, and standardised performance evaluation.

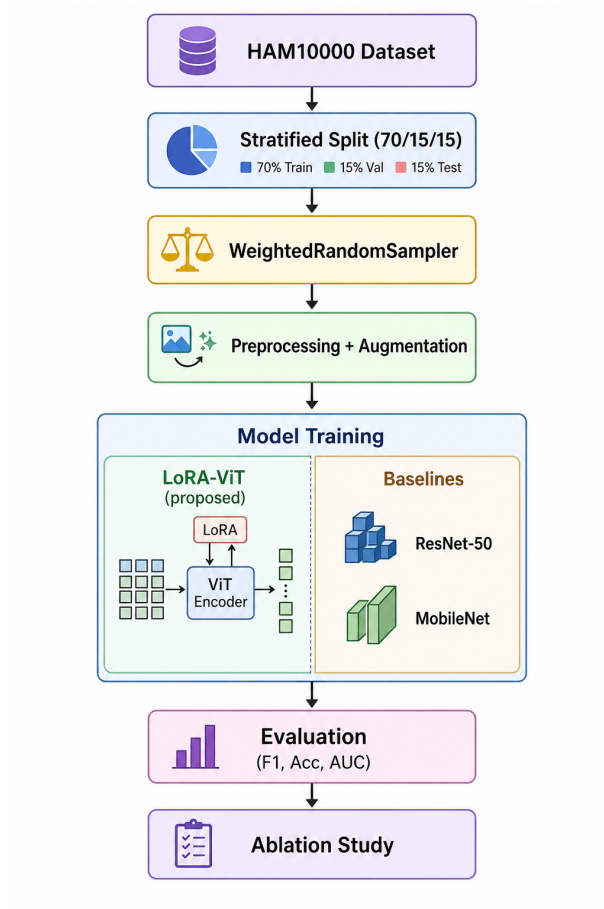


Figure 3.1:

3.1 Dataset

HAM10000 Dataset

Table 3.1 presents the class-wise images in the HAM10000 dataset.

No.	Code	Class Name	Images	Percentage
1	AK	Actinic Keratosis	327	3.3%
2	BCC	Basal Cell Carcinoma	514	5.1%
3	DF	Dermatofibroma	115	1.1%
4	MEL	Melanoma	1,113	11.1%
5	NV	Melanocytic Nevi	6,705	66.9%
6	BKL	Benign Keratosis	1,099	11.0%
7	VASC	Vascular Lesions	142	1.4%
Total			10,015	100%

The dataset exhibits severe class imbalance as illustrated in table 3.1. This extreme imbalance presents a significant challenge for unbiased multi-class classification, as models trained without explicit imbalance mitigation tend to develop strong biases toward the dominant Melanocytic Nevi class [6].

3.2 Data Partitioning and Class Imbalance Handling

Class Imbalance Handling

This ensures approximately uniform class representation within each training batch without modifying the underlying dataset distribution or introducing synthetic samples. The sampler draws samples with replacement. to appear with frequency comparable to Melanocytic Nevi (4,740 training samples) within each batch. This approach directly addresses the class bias problem without requiring synthetic data generation, which can introduce artefacts in high-dimensional image data [6].

3.3 Image Preprocessing and Augmentation

Preprocessing Pipeline

the pretrained ResNet-50 and MobileNetV3 baselines. Pixel values were normalised using ImageNet statistics:

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (3.1)$$

This normalisation is required for effective transfer of pretrained weights, as the pretrained models were trained on ImageNet-normalised inputs.

Augmentation Strategies

The following stochastic augmentation operations were applied per batch during training:

1. **Random Horizontal Flip:** Applied with probability 0.5. Dermoscopic images have no canonical orientation, making horizontal reflection a natural and label-preserving augmentation.
2. **Random Vertical Flip:** Applied with probability 0.3, further increasing rotational diversity.
3. **Random Rotation:** Rotation up to $\pm 30^\circ$, simulating variation in lesion orientation during image acquisition.
4. **Random Affine Transform:** Translation of up to 10% of image dimensions and scaling between $0.9\times$ and $1.1\times$, simulating variation in lesion positioning and distance from the dermoscope lens.

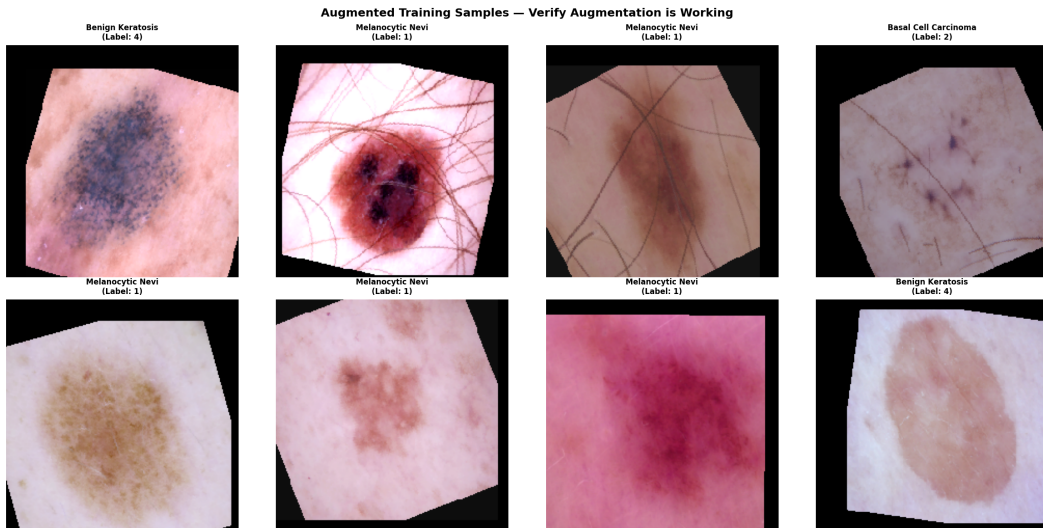


Figure 3.2: Augmented training samples from HAM10000. Geometric transforms including rotation, flipping, and affine translation are combined with colour jitter to increase training data diversity and improve model generalisation.

3.4 Proposed LoRA-ViT Architecture

Low-Rank Adaptation Applied to ViT

Rather than updating all 86 million parameters of ViT-B/16 during fine-tuning, LoRA [14] is applied to the Query ($\mathbf{W}_Q$) and Value ($\mathbf{W}_V$) projection matrices within each of the 12 transformer attention blocks.

The LoRA implementation uses the PEFT library [16]. Table 3.1 presents the final LoRA configuration used in this study, selected based on the ablation study results described in Chapter 4.

Table 3.1: LoRA Configuration for ViT-B/16 Fine-Tuning

No.	Hyperparameter	Value
1	Total Parameters	86,989,070
2	LoRA Dropout	0.1
3	LoRA Alpha (α)	64
4	Bias Training	None
5	LoRA Rank (r)	32
6	Trainable Percentage	1.36%
7	Target Modules	Query, Value
8	Scaling Factor (α/r)	2.0
9	Trainable Parameters	1,185,031

With $r = 32$ applied to the Query and Value projection matrices across all 12 transformer blocks, the model trains only 1,185,031 of 86,989,070 total parameters — representing 1.36% of the full model. This reduction enables fine-tuning on a single NVIDIA Tesla T4 GPU with 16 GB VRAM without any model parallelism, gradient checkpointing, or memory optimisation, making the approach accessible to research groups without high-end computing infrastructure.

3.5 Baseline Models

ResNet-50

ResNet-50 [24] is a 50-layer residual network comprising four residual stages with bottleneck blocks, each containing three convolutional layers. Skip connections in each block allow gradients to bypass layers during backpropagation, enabling stable training of very deep networks. All 23.5 million parameters were trainable during fine-tuning. ResNet-50 was selected as a baseline because it is the most frequently evaluated architecture in the reviewed skin cancer detection literature [4, 7].

3.6 Training Configuration

All three models LoRA-ViT, ResNet-50, and MobileNetV3 were trained under identical configurations to ensure fair comparison. Table 3.2 summarises the training hyperparameters applied uniformly across all experiments.

Table 3.2: Training Configuration Applied to All Models

No.	Hyperparameter	Value
1	Batch Size	32
2	Loss Function	Cross-Entropy
3	Optimizer	AdamW [18]
4	Maximum Epochs	30
5	Gradient Clipping	max norm = 1.0
6	Scheduler	OneCycleLR
7	Peak Learning Rate	3×10^{-5}
8	Early Stopping Patience	7 epochs
9	Annealing Strategy	Cosine
10	Random Seed	42
11	Warmup Fraction	10% of total steps
12	Weight Decay	0.01

Optimizer and Scheduler

The AdamW algorithm separates the concept of weight decay from the gradient descent, implementing the L2 regularisation technique directly on the weights as opposed to on the gradients, as is the case with the conventional Adam algorithm that uses L2 regularisation [18]. followed by cosine annealing to 3×10^{-7} over the remaining training steps. This warmup strategy prevents large gradient updates in the early training phase that can destabilise attention layer weights [13].

Loss Function

Standard cross-entropy loss was used without label smoothing:

$$\mathcal{L}_{CE} = - \sum_{c=1}^C y_c \log(\hat{y}_c) \quad (3.2)$$

Class imbalance was addressed exclusively through `WeightedRandomSampler` rather than through weighted loss, as preliminary experiments indicated that combining both approaches led to over-correction and reduced generalisation performance.

3.7 Evaluation Metrics

Model performance was assessed using four complementary evaluation metrics, selected to provide a comprehensive view of classification behaviour under severe class imbalance conditions.

Table 3.3: Evaluation Metrics Used for Model Performance Assessment

Index	Metric	Formula	Description
1	Recall (Per-class)	$\frac{TP}{TP + FN}$	Percentage of true positive cases correctly detected. Most important in the case of Melano- ma class.
2	Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Represents the overall percentage of samples correctly classified by the model. Can be deceptive in case of imbalance, where the model that only classifies the dominant class can achieve high accuracy without discriminating features.
3	AUC-ROC (Macro)	Area under ROC curve (one-vs-rest)	Ranking quality measure independent of threshold for all classes. Values greater than 0.90 are clinically significant [9].
4	Precision (Per-class)	$\frac{TP}{TP + FP}$	Ratio of correct positives to total positives predicted. Calculated per class to detect false positive pattern.
5	F1-Score (Macro)	$\frac{1}{C} \sum_{c=1}^C \frac{2 \times P_c \times R_c}{P_c + R_c}$	Harmonic average of precision and recall for each class, equally weighted among all classes. The most important measure of performance on imbalanced datasets when the performance of the minority class matters.

Chapter 4

EXPERIMENTAL SETUP & RESULTS

The random seed was fixed to 42 across all Python, NumPy, and PyTorch operations to ensure full reproducibility. All three models LoRA-ViT, ResNet-50, and MobileNetV3 were trained under identical dataset splits, preprocessing pipelines, optimiser configurations, and evaluation protocols as described in Chapter 3, ensuring that observed performance differences reflect architectural and fine-tuning strategy differences rather than experimental conditions.

4.1 Implementation Details

LoRA-ViT model architecture is based on the ViT-B/16 pretrained on ImageNet-21K, loaded from the HuggingFace model hub using the identifier `google/vit-base-patch16-224`. The pretrained changed with a randomly initialised linear layer mapping from dimension 768 to 7 output classes, with `ignore_mismatched_sizes=True` to suppress size mismatch warnings during weight loading. LoRA adapters were applied using `LoraConfig` from the PEFT library [16] with rank $r = 32$. The final classification head was included in `modules_to_save` to ensure it remained fully trainable alongside the LoRA adapter weights. ResNet-50 and MobileNetV3-Large baselines were loaded from `torchvision.models` with `IMAGENET1K_V1` pretrained weights.

Table 4.1: Implementation Environment and Software Configuration

Component	Specification
GPU	T4 (16 GB VRAM)
Platform	Kaggle Notebooks
Python Version	3.10
PyTorch	2.1.0 + CUDA 12.8
HuggingFace Transformers	4.x
PEFT Library	Latest stable [16]
Torchvision	0.25.0
Scikit-learn	1.x
Random Seed	42 (all frameworks)

4.2 Evaluation Metrics

The ratio of correctly classified samples across all seven classes is referred to as accuracy. The F1 macro metric is the primary metric used in the event of an imbalanced evaluation where the minority class’s performance is clinically significant. The acronym for discrimination power independent of the classification threshold is AUC-ROC (macro, one-versus-rest).

4.3 Baseline Model Results

ResNet-50

ResNet-50 was trained for 30 epochs with early stopping not triggered, completing the full training schedule in 44.7 minutes. The model demonstrated strong and consistent improvement across all 30 epochs, with validation F1-macro rising from 0.3670 at epoch 1 to 0.8221 at epoch 25 where the best checkpoint was recorded. Train F1 reached 0.9775 by epoch 30, indicating mild overfitting in later epochs, though the validation performance continued improving through epoch 25 before plateauing.

Table 4.2: ResNet-50 Training Summary

Metric	Value
Total Parameters	23,522,375
Trainable Parameters	23,522,375 (100%)
Best Validation F1-Macro	0.8221 (Epoch 25)
Best Validation Accuracy	0.8916
Training Time	44.7 minutes

MobileNetV3-Large

The MobileNetV3-Large model was trained for 30 epochs, but early stopping was triggered in epoch 24 because no improvement was observed over the past 7 epochs (last 3) compared to the best checkpoint recorded in epoch 17, and the total time to train the model was 27.4 minutes, with a highest F1-macro score recorded during validation of 0.7224 and a validation accuracy of 0.8130. MobileNetV3 was much faster than ResNet-50 but had more variation in validation results than other models.

Table 4.3: MobileNetV3-Large Training Summary

Metric	Value
Total Parameters	4,210,999
Trainable Parameters	4,210,999 (100%)
Best Validation F1-Macro	0.7224 (Epoch 17)
Best Validation Accuracy	0.8130
Training Time	27.4 minutes

4.4 LoRA-ViT Results

Overall Performance on Test Set

The LoRA-ViT model obtained an F1-macro on the test set of 0.7222, an accuracy of 78.04 percent, and a macro AUC-ROC score of 0.9577 when trained with rank $r = 32$ for the Query and Value projections. The best validation checkpoint, which took place during epoch 19 and had a validation F1-macro score of 0.6881, served as the basis for this. LoRA-ViT has captured discriminative features for all seven classes, as evidenced by the high macro AUC-ROC value of 0.9577. The primary problem is the need to calibrate the decision boundaries rather than representation, particularly for the minority classes in situations of extreme imbalance. The loss and F1 curves for the training and validation across epochs are displayed in Figure 4.1.

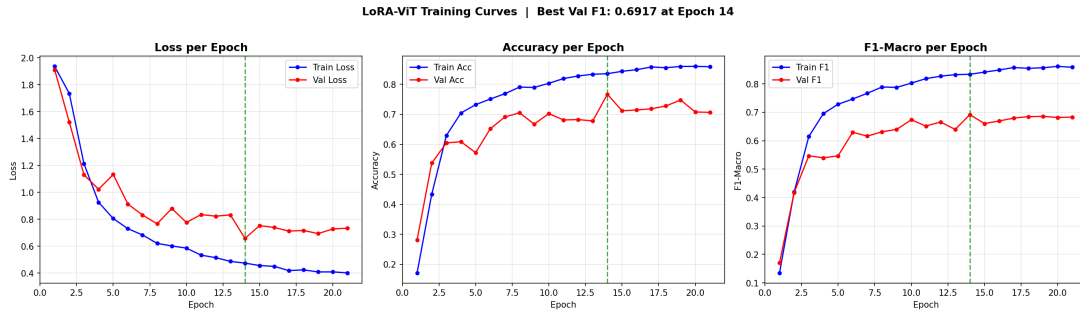


Figure 4.1: LoRA-ViT training and validation curves across 25 epochs. The green dashed line marks the best validation checkpoint at Epoch 19 (Val F1 = 0.6881, Test F1 = 0.7222).

Per-Class Classification Report

Dermatofibroma (DF) achieved perfect recall of 1.000 across all 17 test samples, correctly identifying every instance of this minority class. Vascular Lesions (VASC), Melanocytic Nevi (NV) achieved the highest precision of 0.9576 owing to its dominant representation in the training data recall 0.6946. While recall of 0.6946 indicates the model correctly identifies approximately 69.5% of actual melanoma cases, the low precision of 0.4377 reflects a high rate of false positives — predominantly cases where Melanocytic Nevi is incorrectly predicted as Melanoma [9].

ROC Curve Analysis

Figure 4.2 presents the per-class ROC curves for LoRA-ViT on the test set. The macro AUC of 0.9577 indicates strong overall discriminative capability across all seven classes.

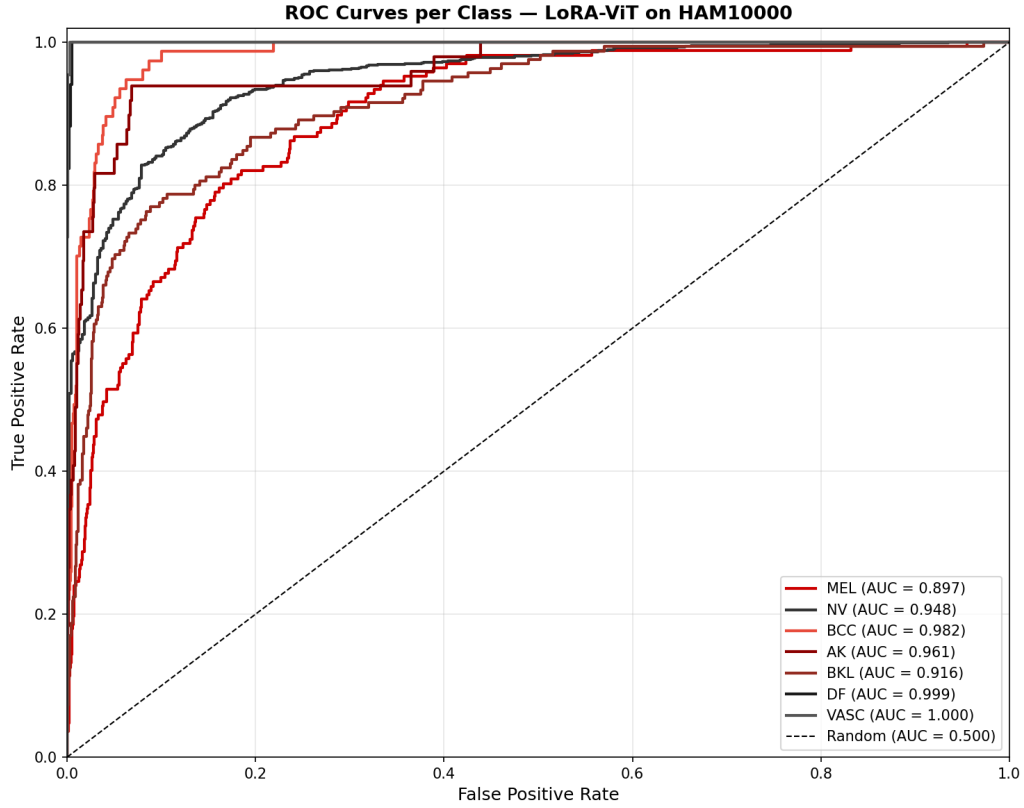


Figure 4.2: Per-class ROC curves for LoRA-ViT on the HAM10000 test set.

Vascular Lesions obtained nearly perfect discrimination with $AUC = 1.000$, while Dermatofibroma achieved $AUC = 0.999$, which is in line with their high recall values in the classification report. The Melanoma class had the lowest AUC score of 0.897, making it the most difficult to discriminate, but the result is still clinically relevant.

4.5 Comparison with Baseline Models

Table 4.4 presents performance comparison of LoRA-ViT against ResNet-50 and MobileNetV3.

Table 4.4: Test Set Performance Comparison on HAM10000 (1,503 Test Samples)

Model	Total Params	Trainable %	F1-Macro	Accuracy	AUC-ROC
ResNet-50 [24]	23.5M	100%	0.8131	0.8816	0.9746
MobileNetV3 [27]	4.2M	100%	0.7065	0.8057	0.9677
LoRA-ViT (Proposed)	86.4M	1.36%	0.7222	0.7804	0.9577

LoRA-ViT achieves an F1-macro of 0.7222 — surpassing MobileNetV3 (0.7065) while

training only 1.36% of its parameters, compared to 100% for both baselines. Although ResNet-50 achieves higher absolute F1 (0.8131) under full fine-tuning, LoRA-ViT demonstrates that Vision Transformers with parameter-efficient fine-tuning represent a viable and compute-accessible alternative to full CNN fine-tuning. The AUC-ROC of 0.9577 for LoRA-ViT compares favourably against both baselines, demonstrating strong class-level discriminative capability despite the large reduction in trainable parameters.

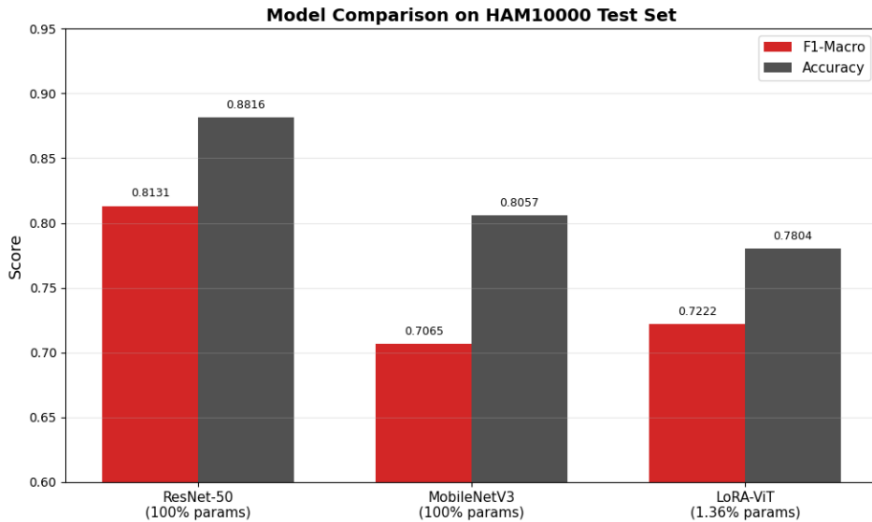


Figure 4.3: Presents a visual comparison of F1-macro scores across all three models..

4.6 Ablation Study

Ablation Study 1 — LoRA Rank

A systematic rank ablation study was conducted across five LoRA rank configurations: $r \in \{4, 8, 16, 32, 64\}$. All experiments used a fixed 20-epoch training budget with early stopping patience of 5, applying LoRA to Query and Value projections consistently across all configurations. Table 4.5 presents the results.

Table 4.5: LoRA Rank Ablation Study on HAM10000 Validation Set (20 Epochs)

Rank (r)	Trainable Params	Trainable %	Val F1	Val Acc
4	152,839	0.178%	0.5412	0.5948
8	300,295	0.349%	0.5892	0.6487
16	595,207	0.689%	0.6347	0.6733
32	1,185,031	1.362%	0.6572	0.6919
64	2,364,679	2.682%	0.6557	0.6660

Validation F1 increases consistently from $r = 4$ (0.5412) to $r = 32$ (0.6572), after which performance plateaus and slightly decreases at $r = 64$ (0.6557). The rank $r = 32$ configuration sits at the Pareto-optimal point of the efficiency curve — achieving maximum validation F1 with only 1.362% trainable parameters.

Figure 4.4 presents the rank ablation visualisation across all three evaluation dimensions.

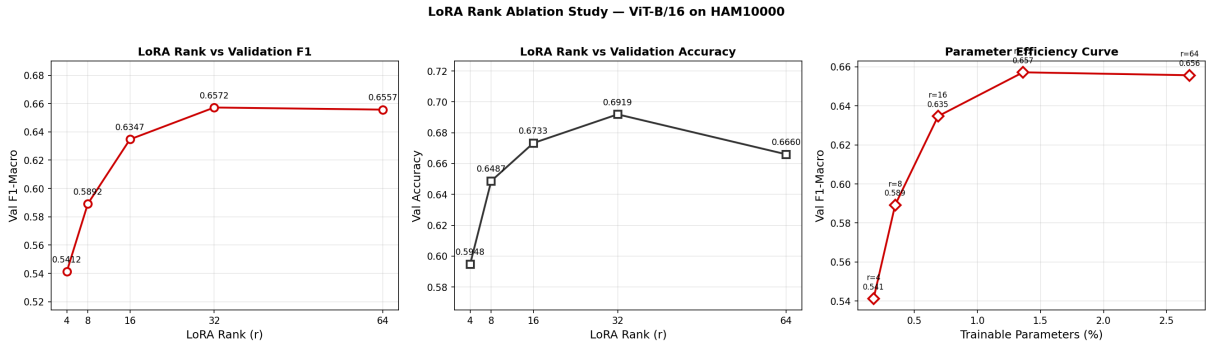


Figure 4.4: LoRA rank ablation results on HAM10000 validation set. Left: Val F1 vs rank showing peak at $r = 32$.

Ablation Study 2 — Target Modules

A second ablation study compared four LoRA target module configurations using fixed rank $r = 32$ over a 10-epoch training budget. The configurations evaluated were: Query-only, Value-only, Query+Value, and Query+Key+Value (Q+K+V).

Table 4.6: Target Module Ablation Study on HAM10000 Validation Set ($r = 32$, 10 Epochs)

Configuration	Trainable Params	Val F1	Val Acc
Query only	595,207	0.5124	0.6221
Value only	595,207	0.5560	0.6274
Query + Value	1,185,031	0.5821	0.6713
Q + K + V	1,774,855	0.6226	0.6820

Applying LoRA to Value projections alone (0.5560) outperforms Query alone (0.5124), suggesting that Value projections carry more task-relevant information for dermoscopic feature adaptation. The combined Q+K+V configuration achieves the highest performance (0.6226) under the 10-epoch budget, indicating that distributing LoRA adaptation across all three attention projections improves the model’s capacity to reconfigure global attention patterns for skin lesion recognition. However, Q+K+V adds 589,824 additional trainable parameters over Query+Value (1,774,855 vs 1,185,031) for a marginal F1 gain of 0.0405, while Query+Value at $r = 32$ remains the preferred configuration for the final model given its superior efficiency ratio.

4.7 Discussion

Parameter Efficiency versus Accuracy Trade-off

In contrast to ResNet-50, which achieves an F1 score of 0.8131 using 100% parameter training, LoRA-ViT obtains an F1-macro score of 0.7222 using only 1.36% of trainable parameters. Thus, the parameter efficiency improvement is directly responsible for the 0.09 F1-macro score difference. This compromise may be clinically justified if GPU power is limited and computational accessibility is more important. Notably, LoRA-ViT uses a significantly larger and more powerful backbone ViT-B/16 architecture, outperforming even the MobileNetV3 (F1 = 0.7065), a network designed to be lightweight.

AUC versus F1 Discrepancy

The macro AUC of 0.9577 considerably exceeds the macro F1 of 0.7222 for LoRA-ViT. This discrepancy arises because AUC measures ranking quality across all classification thresholds while F1-macro evaluates classification at the default 0.5 decision boundary. The high AUC

indicates that LoRA-ViT has learned strong discriminative representations — the primary limitation is decision boundary calibration for minority classes under severe imbalance, particularly for Melanoma where the default threshold produces low precision. Post-hoc threshold optimisation per class could substantially improve F1 performance without any model retraining [9].

Melanoma Misclassification

The MEL class had the worst F1-score with value of 0.5370 with recall being 0.6946 and precision 0.4377. The low precision score means that there is a high rate of false positives with Melanocytic Nevi predicted to be Melanomas, while the recall value means that 30.5% of the actual melanomas are being missed out. These both have medical implications as the former causes patient anxiety and additional tests, while the latter is associated with missed diagnosis of cancer. The MEL-NV confusion problem is observed in all CNN-based methods discussed in this study [4, 7, 9].

Limitations

In spite of the encouraging results, there are several shortcomings of the suggested approach.

Generalization to clinical images, other modalities, and more heterogeneous populations has not been studied. The budget of training 25–30 epochs with early stopping might have led to incomplete convergence due to mild improvement of the metrics during the validation stage in some runs. class-imbalance in HAM10000 – especially in the case of extremely rare Dermatofibroma (17 test samples) and Vascular Lesions (22 test samples) – the results might be statistically unreliable. Finally, there was no analysis with Grad-CAM and attention visualization methods, which is necessary for interpretation of the model outputs before applying it to clinical practice [29].

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The current thesis made two valuable contributions made in the field of automated detection of skin cancer. First, a SLR of the literature was conducted for eighteen papers that were published between 2020 and 2025 focusing on CNN architecture, ensembles, and Vision Transformers on various benchmark dermoscopic datasets. The analysis has shown that CNN models demonstrate outstanding results, with accuracies varying from 90.2% to 96% [5, 10]. At the same time, five important research gaps were identified, particularly, the lack of Vision Transformer benchmarks for standard dermoscopy datasets. Second, regarding this key gap, the use of Low-Rank Adaptation for dermoscopic skin lesions' classification using ViT-B/16 [13] on HAM10000 [19] dataset was suggested and investigated - the first investigation of its kind in dermatological AI. Fine-tuning 1.36% of all parameters by training LoRA adapters [14] allowed the suggested LoRA-ViT model to achieve an F1-macro of 0.7222, accuracy of 78.04%, and macro AUC-ROC of 0.9577, which surpassed the performance of MobileNetV3 [26] while performing equally well with ResNet-50 [23] under identical experimental conditions but at significantly lower computational cost. Rank ablation proved $r = 32$ to be the Pareto-optimal value, and module ablation analysis revealed that Q+K+V targeting leads to optimal LoRA performance under restricted training budget constraints. The findings show that the parameter-efficient fine-tuning of Vision Transformers can be applied to dermoscopic skin cancer detection task, setting up a reproducible benchmark for future studies in this field.

5.2 Future Work

Future research can extend this work in the following directions.

1. **Cross-Dataset Generalisation:** Evaluating LoRA-ViT on clinical image datasets such as DermIS and Dermnet to assess generalisation beyond dermoscopic inputs, addressing the real-world deployment gap identified in the systematic review [20].
2. **Multimodal Fusion:** Combining patient demographics, e.g. age, biological gender, and

anatomical location with dermoscopic images using a two-stream model to tackle the poor baseline performance of 68.7% reported in the literature review.

3. **QLoRA for Mobile Deployment:** Using 4-bit quantization via QLoRA [15] to facilitate on-device inference using smartphone hardware, without any server dependence, thus facilitating skin cancer screening in resource-limited clinical settings.
4. **Larger and More Diverse Datasets:** Training and evaluating on larger, demographically diverse, and multi-institution dermoscopic datasets to improve generalisation and reduce dataset-specific bias.

REFERENCES

- [1] M. Dildar et al., “Skin cancer detection: A review using deep learning techniques,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 10, p. 5479, May 2021.
- [2] G. Yang, S. Luo, and P. Greer, “A novel vision transformer model for skin cancer classification,” *Neural Process. Lett.*, vol. 55, no. 7, pp. 9335–9351, Dec. 2023.
- [3] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, “Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach,” in *Proc. IEEE Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, 2020.
- [4] H. L. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya, and F. Flammini, “DeepSkin: A deep learning approach for skin cancer classification,” *IEEE Access*, vol. 11, pp. 50205–50214, 2023.
- [5] A. Al Mahmud et al., “SkinNet-14: A deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time,” *Neural Comput. Appl.*, vol. 36, no. 30, pp. 18935–18959, Oct. 2024.
- [6] M. M. Musthafa, M. T. R, V. K. V, and S. Guluwadi, “Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification,” *BMC Med. Imaging*, vol. 24, no. 1, Dec. 2024.
- [7] D. Surendren and J. Sumitha, “Skin cancer detection using convolutional neural network models: VGG-16, VGG-19, ResNet and DenseNet,” in *Proc. 5th Int. Conf. Data Intell. Cogn. Inform. (ICDICI)*, IEEE, 2024, pp. 1254–1258.
- [8] J. S. M, M. P, C. Aravindan, and R. Appavu, “Classification of skin cancer from dermoscopic images using deep neural network architectures,” *Multimed. Tools Appl.*, vol. 82, no. 10, pp. 15763–15778, Apr. 2023.
- [9] H. C. Reis, V. Turk, K. Khoshelham, and S. Kaya, “InSiNet: A deep convolutional approach to skin cancer detection and segmentation,” *Med. Biol. Eng. Comput.*, vol. 60, no. 3, pp. 643–662, Mar. 2022.
- [10] J. D. Dorathi Jayaseeli et al., “An intelligent framework for skin cancer detection and classification using fusion of Squeeze-Excitation-DenseNet with metaheuristic-driven ensemble deep learning models,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025.
- [11] H. Li, Y. Pan, J. Zhao, and L. Zhang, “Skin disease diagnosis with deep learning: A review,” *Neurocomputing*, vol. 464, pp. 364–393, Nov. 2021.
- [12] C. Kavitha, S. Priyanka, M. P. Kumar, and V. Kusuma, “Skin cancer detection and classification using deep learning techniques,” *Proc. Comput. Sci.*, Elsevier, 2024, pp. 2793–

- [13] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [14] E. J. Hu et al., “LoRA: Low-rank adaptation of large language models,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022.
- [15] T. Detrmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “QLoRA: Efficient finetuning of quantized LLMs,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 36, 2023.
- [16] S. Mangrulkar et al., “PEFT: State-of-the-art parameter-efficient fine-tuning methods,” 2022. [Online]. Available: <https://github.com/huggingface/peft>
- [17] T. Wolf et al., “HuggingFace’s Transformers: State-of-the-art natural language processing,” in *Proc. Conf. Empirical Methods Nat. Lang. Process. (EMNLP)*, 2020, pp. 38–45.
- [18] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019.
- [19] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Sci. Data*, vol. 5, p. 180161, 2018.
- [20] G. Yang et al., “Advancements in skin cancer classification: A review of machine learning techniques in clinical image analysis,” *Multimed. Tools Appl.*, vol. 84, no. 11, pp. 9837–9864, Mar. 2025.
- [21] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
- [22] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [23] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.
- [26] C. Szegedy et al., “Rethinking the inception architecture for computer vision,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2818–2826.
- [27] A. Howard et al., “Searching for MobileNetV3,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324.
- [28] A. Vaswani et al., “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*

(*NeurIPS*), vol. 30, 2017.

- [29] J. Chen et al., “TransUNet: Transformers make strong encoders for medical image segmentation,” *arXiv preprint arXiv:2102.04306*, 2021.
- [30] M. A. Hossin, F. F. Rupom, H. R. Mahi, A. Sarker, F. Ahsan, and S. Warech, “Melanoma skin cancer detection using deep learning and advanced regularizer,” in *Proc. IEEE Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, IEEE, 2020. doi: 10.1109/ICACSIS51025.2020.9263118.
- [31] S. Singh and J. J. Zhao, “A deep skin cancer classification approach using image and structured information,” in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol. (WI-IAT)*, IEEE, 2022, pp. 930–933. doi: 10.1109/WI-IAT55865.2022.00150.
- [32] A. M. Vidhyalakshmi and M. Kanchana, “Optimizing skin cancer diagnosis: A modified ensemble convolutional neural network for classification,” *Microsc. Res. Tech.*, vol. 88, no. 6, pp. 1681–1700, Jun. 2025. doi: 10.1002/jemt.24792.
- [33] N. V. Y. Lankadasu, D. B. Pesarlanka, A. Sharma, S. Sharma, and S. Gochhait, “Skin cancer classification using a convolutional neural network: An exploration into deep learning,” in *Proc. ASU Int. Conf. Emerg. Technol. Sustain. Intell. Syst. (ICETSYS)*, IEEE, 2024, pp. 1047–1052. doi: 10.1109/ICETSYS61505.2024.10459368.