OPTIMIZATION OF DEEP LEARNING BASED LIGHT WEIGHT SKIN DISEASE CLASSIFICATION MODEL USING TRANSFER LEARNING AND MODEL QUANTIZATION FOR RESOURCE-CONSTRAINED ENVIRONMENTS

W. Azeem[1], M. U. Ahmad[2], T. Ahamd[3], A. Farooq[3], A. I. Mirza[2] and I. Shahzadi[2]

[1]Department of Informatics and Systems, School of Systems and Technology, University of Management and Technology (UMT), Lahore
[2]Institute of Data Science, University of Engineering and Technology (UET), Lahore
[3]Department of Computer Science, University of Engineering and Technology (UET), Lahore
Corresponding Author Email: amjad.farooq@uet.edu.pk

ABSTRACT: Skin disease classification using dermoscopic images is one of challenging diagnostic process because of its visual similarities among various disease conditions, imbalance classes of diseased images, and the unavailability of annotated data. Many of Deep learning models have demonstrated encouraging detection performance. However, the most of Deep Learning Models requires high computational and memory requirements, that makes their deployment difficult in resource-constrained environments remains. In this study, we investigate an efficient convolutional neural network-based framework for multi-class skin disease classification using MobileNetV2. This technique is specially used for the modelling of systems for resource-constrained environments. The proposed technique will progressively enhance the performance of the model in three stages. The modelling will start from the baseline in initial stage and then considering data augmentation and unfreezing the layers. In middle stage partial fine-tuning, and class weighting to address class imbalance is considered for further improvement. Additionally, post-training quantization is used to reduce model size and enhance deployment potential with a little trade-off in performance. Experimental results show the improvements are moderately enhanced for performance macros across our three training stages. Quantization were achieved for a significant memory reduction while maintaining competitive performance of the model. This study highlights the suitability of the proposed approach for mobile based clinical applications.

Keywords: Skin disease classification, Transfer learning, MobileNetV2, Model quantization, Deep learning, Medical image analysis

(Received 06.10.2025 Accepted 20.12.2025)

INTRODUCTION

Skin diseases are among the most prevalent health conditions in Pakistan as well as around the world. Skin disease can affect individuals of all age groups. Early detection and accurate diagnosis are very important. Particularly some of most common and severe conditions such as Ance, Psoriasis, Eczema, SkinCancer, Vitiligo, where delayed detection can lead to life-threatening consequences. Most of the traditional diagnosis procedures rely on expert dermatological assessment, which is a time-consuming and costly process. Expert dermatologists may be inaccessible in low-resource settings and in rural areas.

Recent research in deep learning, and advanced convolutional neural networks (CNNs)[1-2], have significantly improved and eased the automatic medical image analysis mechanism. However, designing and training efficient deep learning networks from scratch needs a large amount of annotated datasets. It also requires high computational resources and memory, which may be limited in medical environments. The transfer learning technique is an effective way to extract meaningful features from limited datasets [3].

For real world applications many of deep learning models are impractical for their deployment in resource-constrained environments. Their deployment is difficult due to large model size and high inference latency. It can be made deployable if the model sized is reduced. This motivates the researchers to explore lightweight architectures and model compression techniques that may be suitable for resource-constrained environments. In this study, we focus on MobileNetV2, which is a lightweight CNN based optimization technique for mobile environments. We investigate its effectiveness for skin disease classification in multiple stages. We further applied post-training quantization to enhance its deployment efficiency without significantly compromising on models performance.

Related Work: Deep learning-based skin disease classification techniques have been widely explored in

recent years. Some of the past researchers have reported the utilization of handcrafted features and use of conventional classifiers. With the advancement of CNNs many of versatile architectures have been reported in recent literature. The CNN based architectures such as VGGNet, ResNet, DenseNet, and many other have achieved strong performance on dermoscopic datasets. However, these models often require substantial computational resources [4].

The advanced techniques enable us to use pre-trained models to adjust to medical imaging tasks with limited data. Although it may raise the risk of overfitting, several studies show that fine-tuning deeper layers enhances task-specific representation. In skin disease datasets, class imbalance remains an issue that results in predictions which can be skewed toward major classes having more data for tuning.

Using a custom dataset, Aishwarya and Gomathy classified ten different types of skin diseases using ResNet architectures. They reported training and testing accuracies above 98% and demonstrated the model's ability to capture discriminative lesion patterns because of its deep residual connections. [5]

M. Akter *et al* have shown ResNet variants such as ResNet50 and ResNet101 to achieve competitive performance when distinguishing between malignant and benign skin lesions. These variants have shown outperforming stats compared to traditional CNN frameworks [6].

Dermatological datasets often have a significant imbalance between multiple classes. The imbalance can degrade the performance of classification models. Many studies have shown that models trained without compensation mechanisms can give a skewed response in favor of the major classes. To resolve this issue, many strategies have been reported in literature like class weighting, data augmentation, and adapted loss functions [7].

Howard *et al*., discussed that lightweight CNN architectures, such as MobileNet, have been widely adopted for skin lesion classification due to their low computational cost. Through the use of depth-separable convolutions, these models offer a good balance between accuracy and efficiency, making them suitable for embedded and mobile digital health applications [8].

Pan *et al* reported that Transfer learning is widely used to improve model performance on limited medical datasets. By initializing networks with weights pre-trained on large datasets and then progressively refining them, models can learn more robust representations. This approach reduces overfitting and improves overall classification performance [9].

Recently MobileNet architectures have been widely reported for skin disease classification. MobileNet got attention of many of researchers due to its lightweight structure and efficient deployment enabling capacity on resource-constrained devices. Chaturvedi *et al.* proposed a seven-way automated multi-class skin cancer classification system. They used a pre-trained MobileNet model on the HAM10000 dataset. The reported model achieved high weighted average precision, recall, and F1-score values about 0.83 [10].

Further work on MobileNet-based classification has explored combining the MobileNet backbone with advanced feature mechanisms. MobileNet-V2 with attention blocks and atrous spatial pyramid pooling (ASPP) has shown improved discrimination of contextual information among skin disease datasets. The results have shown the validation for possible utility of MobileNet in both generalization and mobile deployment of skin disease detection systems [11]. for deployment on resource-constrained devices model quantization has been discussed for making lightweight architectures using MobileNet. These model can be more suitable for real world applications deployment because of reduced model size and computational cost. For example, the Quantization Friendly MobileNet (QF-MobileNet) framework proposes architectural optimizations that address redundancy and quantization loss inherent in baseline MobileNet models. The Quantization results in significant reductions of tunable parameters, as well as it also reduces inference time with minimal degradation in accuracy after quantization. This approach makes the architecture more efficient for embedded applications on smartphones and low hardware devices [12].

METHODOLOGY

The figure 1 shows the research methodology framework for this research work. Publically available skin disease dataset was acquired from Kaggle. The dataset was quite with total of 21 disease classes, so we planned to train the model on limited classes of diseases. Then top five disease classes were selected for model training. The model MobileNet V3 was adopted for training, as this is the lightweight model that can be quantized and performs well under the constrained hardware resources. The training was performed in three stages.

Dataset Description and Preprocessing: The experiments are conducted on a publicly available skin disease image dataset containing multiple dermatological conditions. The dataset consists of RGB images categorized into distinct disease classes. There were 21 diseased classes and 1 normal class, in original dataset. However, classes were widely imbalanced. For this research work we decided to select to choose five disease classes which were nearly balance and as well as these were most important diseased for which early detection is favorable. Additionally, one normal class were also considered. So the total of 6 classes were used to train the

model. Figure 2 shows sample images of five selected diseased classes.

The dataset is organized into training, validation, and test splits as per most of the practice reported in research. Then, for preprocessing Images are resized to a fixed resolution that is compatible with MobileNetV2 input requirements. For this study we did not perform manual segmentation to allow the model learn discriminative features directly from raw images.

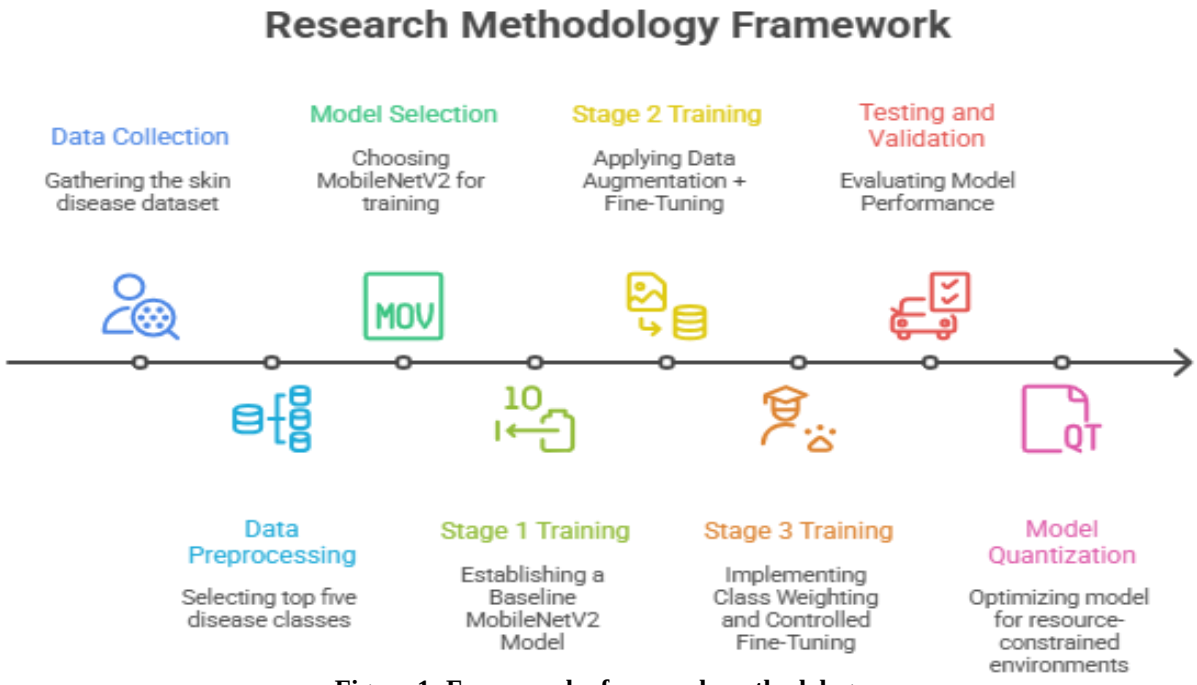

Figure 1: Framework of research methodology

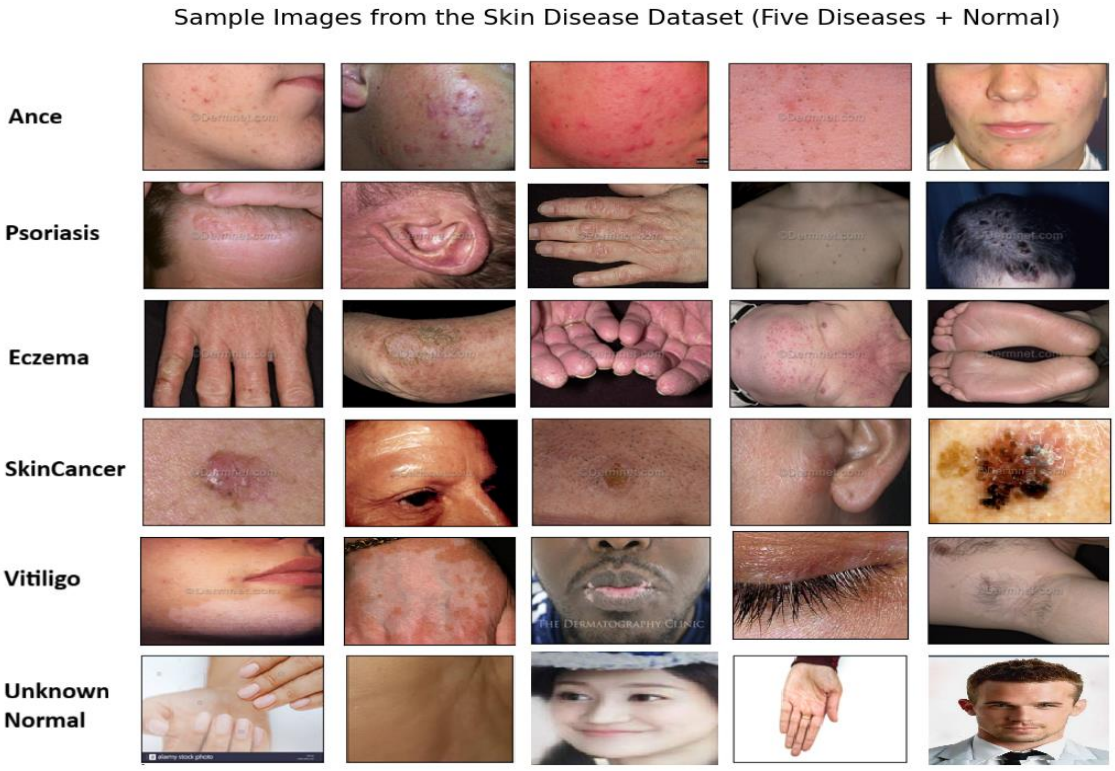

Figure 2: Sample Images of Selected Disease Classes

Model Architecture and Training Strategy: MobileNetV2 is selected as the base architecture due to its lightweight design and efficiency. The pre-trained weights on ImageNet are utilized to initialize the model. It is designed for efficient on-device vision applications. MobileNetV2 architecture consists of 53 convolutional layers. Its Initial layers perform a standard 3x3 convolution with 32 filters and stride. The its Bottleneck Layers have a series of 19 residual bottleneck blocks with varying expansion factors, output channels, and strides. The final Layers of MobileNetV2 have a 1x1 convolution (1280 filters), a Global Average Pooling layer, and a final fully connected layer for classification. MobileNetV2's lower layers are generally used to detect simple features like edges and corners. Whereas the model's upper layers are designed to foresee the complex features and patterns.

Initially MobileNetV2 applies a single filter per input channel and all m channels are convolved independently with its own k x k kernel for depth wise filtering. Model architecture can be represented as shown in figure 3:

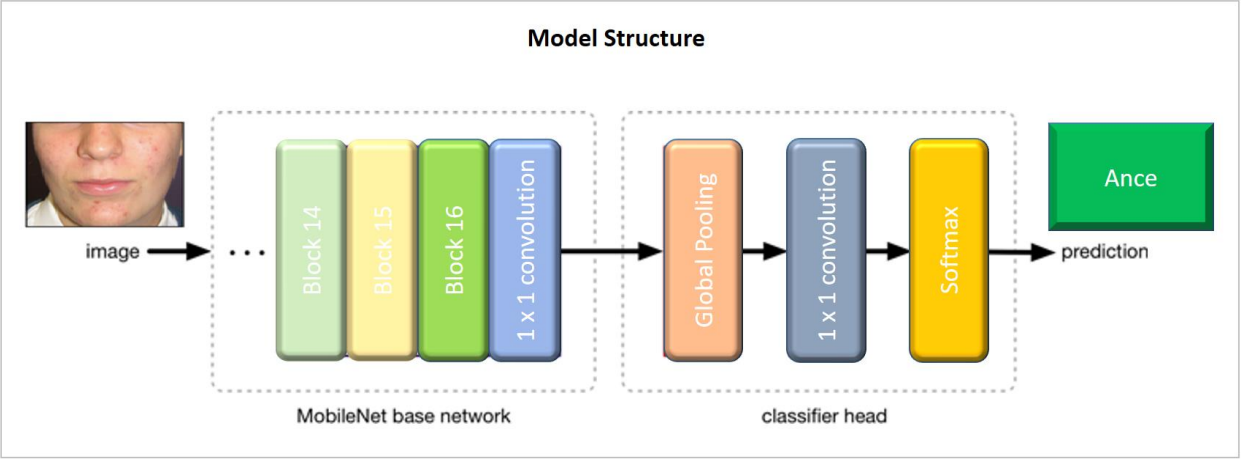

Figure 3: Model Structure Architecture of Proposed Methodology

Its mathematical expressions can be presented as:

$$O_{dw}(i,j,m) = \sum_{k,l} W_{k,l,m} \cdot F_{i+k,j+l,m} \quad (1)$$

Then a 1x1 convolution is used to combine all the depth wise outputs into new features, that can be represented as:

$$O_{pw}(i,j,n) = \sum_{m} W_{n,m} \cdot O_{dw}(i,j,m) \quad (2)$$

To systematically enhance the performance, a training process has been conducted in three progressive stages, as shown in Table 1:

Table 1: Description of Stage wise Training Strategy

Stage	Description
Stage 1	Baseline MobileNetV2
Stage 2	+ Data Augmentation + Partial Fine-Tuning
Stage 3	+ Class Weighting + Controlled Fine-Tuning

Stage 1: Baseline Model: The MobileNetV2 backbone is frozen, and only the classification head is trained. The freezing is done to avoid model overfitting and making it faster to safely learn underlying patterns by using pre-learned knowledge. The stage 1 training was performed for 12 epochs, whereas subsequent fine-tuning stages e.g., 2 and 3, were trained for fewer epochs to prevent overfitting and preserve learned representations

Stage 2: Data Augmentation and Partial Fine-Tuning: In our second stage of model training we used data augmentation techniques, like rotation, flipping, and zooming of images so that generalization can be enhance for better results. In this stage the upper layers of the backbone were also made unfrozen so that the task-specific features may also be adopted and learned. Stage 2 training was carried out for 10 epochs.

Stage 3: Class Weighting and Fine-Tuning: In our Stage 3 we introduced class weights so that imbalance of classes may be adjusted. In this process minority classes having less data are emphasized during the training process. Fine-tuning is also performed using a low learning rate, so that a stabilize optimization can be

achieved. Stage 3 training was terminated earlier at 8th epoch.

Finally, to improve the deployment efficiency of model, post-training quantization technique is applied to convert the trained model from 32-bit floating-point precision to 8-bit integer representation. This process reduces the model size and memory utilization, to make its deployment convenient for the edge and mobile based devices.

RESULTS AND DISCUSSION

Stage-Wise Training Performance Analysis: Figure 4 shows the epoch-wise comparison of training accuracies, loss, precision, and recall for our trained model during the three training stages. At Stage-1 the MobileNetV2 backbone was trained in which some of the convolutional layers were left frozen to retain generic features of our pre trained model. No additional optimization strategy was considered in stage 1, the model shows a steady increase in performance with running epochs. In this stage, the convergence was slower and final training accuracy remains lower compared to later stages. The baseline representation of skin disease patterns was established in this stage.

In Stage-2, we noticed an improvement in all metrics after data augmentation and partial unfreezing of layers were introduced. Training accuracy, precision, and recall increased, whereas the loss decreases rapidly on running early epochs. The decrease in loss represents fast convergence and improved feature learning. These results confirmed that the data augmentation enhanced the model's robustness changes in skin lesion appearance. The process of unfreezing the layers enables better adaptation of high-level features in dataset.

Stage-3, which incorporates class weighting and fine-tuning, achieves the highest overall performance, demonstrating the most stable and consistent learning trends. The further improved precision and recall in this stage 3 were observed due to better handling of imbalanced classes and minority disease categories. The class weighting and fine-tuning significantly improvised the results.

The progressive stages wise improvement in training results validates the effectiveness of the proposed training strategy.

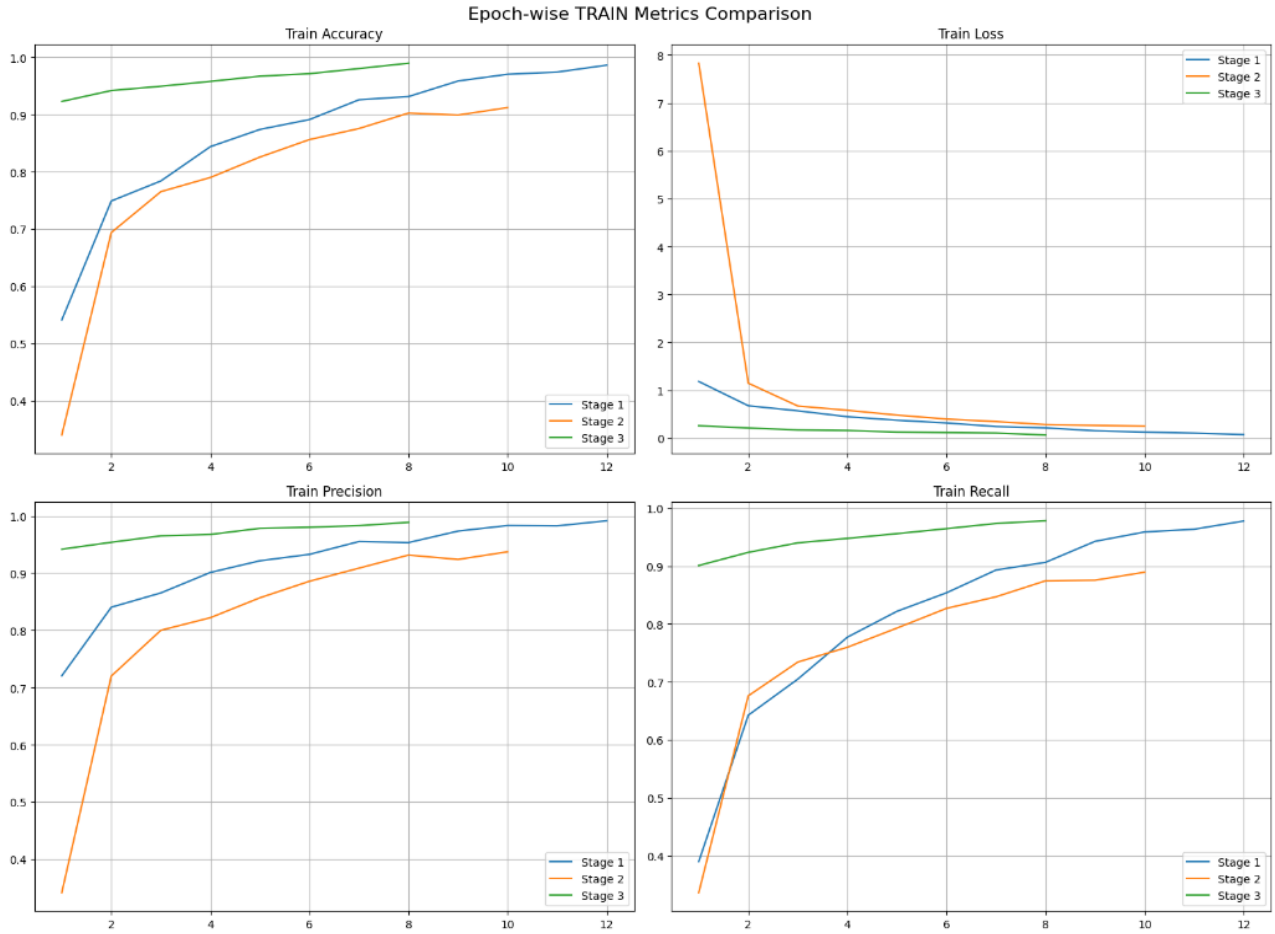

Figure 4: Epoch wise Training Results for all Stages

Stage-Wise Validation Performance Analysis: The figure 5 shows the validation results, a progressive impact of our proposed stage-wise training on model generalization can be visualize from these graphs. The model achieved a moderately stable validation accuracy and precision during stage 1 at baseline learning and frozen feature extractors. Then a higher validation loss and fluctuations were observed in beginning of Stage-2 training due to layer unfreezing of layers and increased model flexibility. However, after training few epochs, progressive improvement was achieved in validation accuracy, precision, and recall. The progressive improvement suggests the better adaptation of task-specific skin disease features. The Stage-3 shows a balanced validation performance, with consistently higher recall. This reflects the effectiveness of class weighting and fine-tuning to address the imbalance among classes for better recognition. The overall validation curves of all three stages confirm that the stage-wise training strategy has enhanced the generalization and also controlled the overfitting. This has led to a more robust and clinically enhanced classification model.

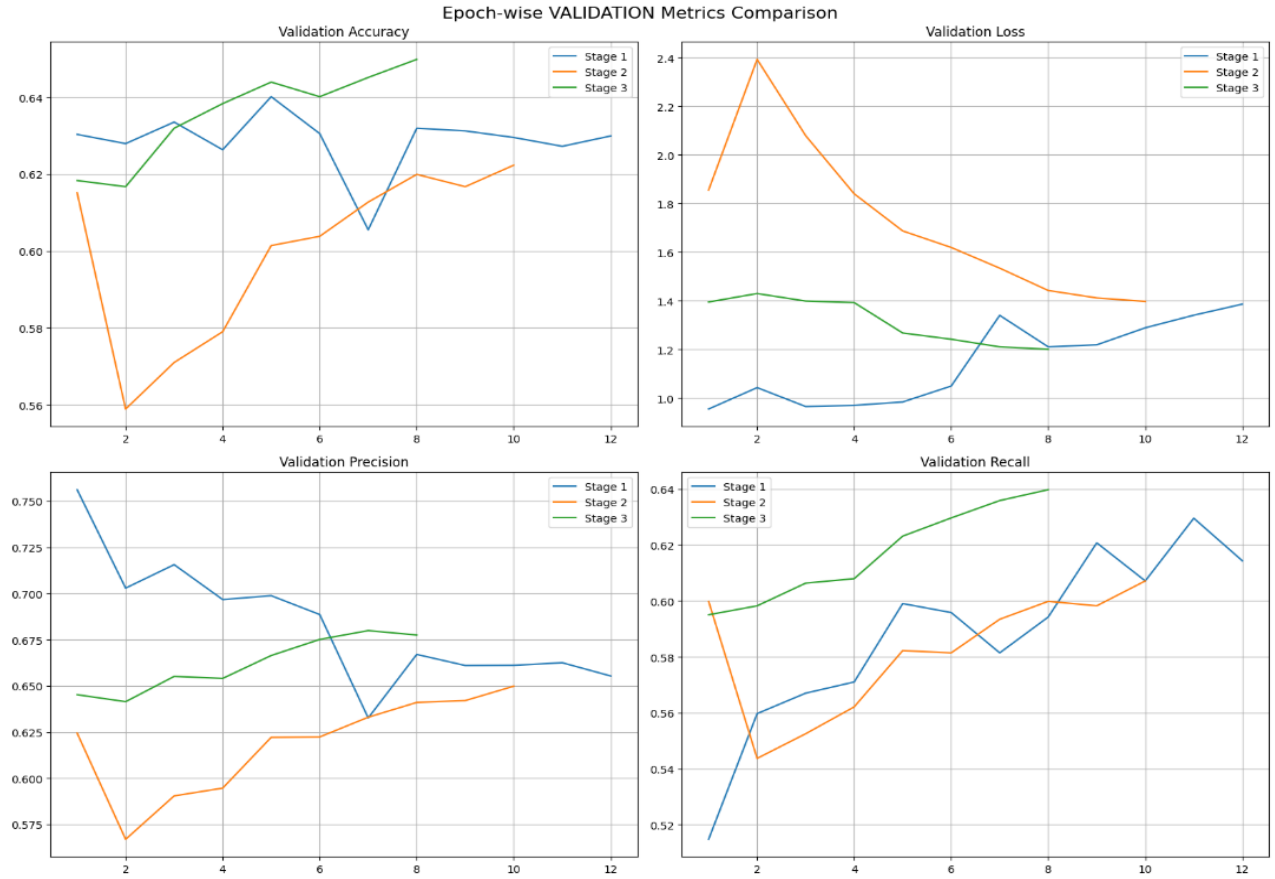

Figure 5: Epoch wise Validation Results for all Stages

Confusion Matrix Analysis: Confusion matrix shown in figure 6, reveal a reduced misclassification among visually similar diseases as training progresses. Notably, recall for clinically critical classes improves after applying class weighting, highlighting the effectiveness of imbalance handling strategies.

Figure 7 shows progressive improvements across training stages. The macro-averaged accuracy, precision, recall, and F1-score increase consistently, indicating improved robustness across classes.

Statistical Significance Analysis: Finally, paired t-tests were conducted on epoch-wise validation metrics, to validate that the observed performance improvements across training stages of our model were statistically correct. The table show that the transition from Stage-1 to Stage-2 and then from Stage 2 to Stage-3 yields statistically significant gains ($p < 0.05$). This confirms the effectiveness of progressive fine-tuning of our model in three stages. Furthermore, bootstrap-based confidence interval of CI= 95% shows a consistent mean improvement in validation accuracy from Stage-1 to Stage-3. These stage wise findings demonstrate that our proposed strategy leads to a reliable and robust performance improvement rather than random fluctuations.

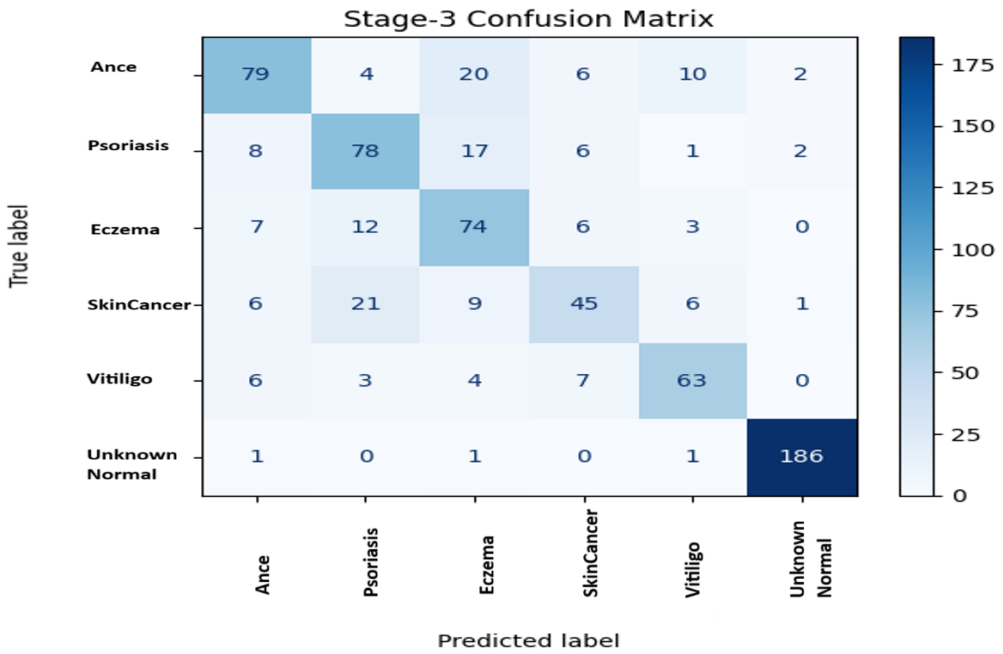

Figure 6: Confusion Matrix of Final Stage

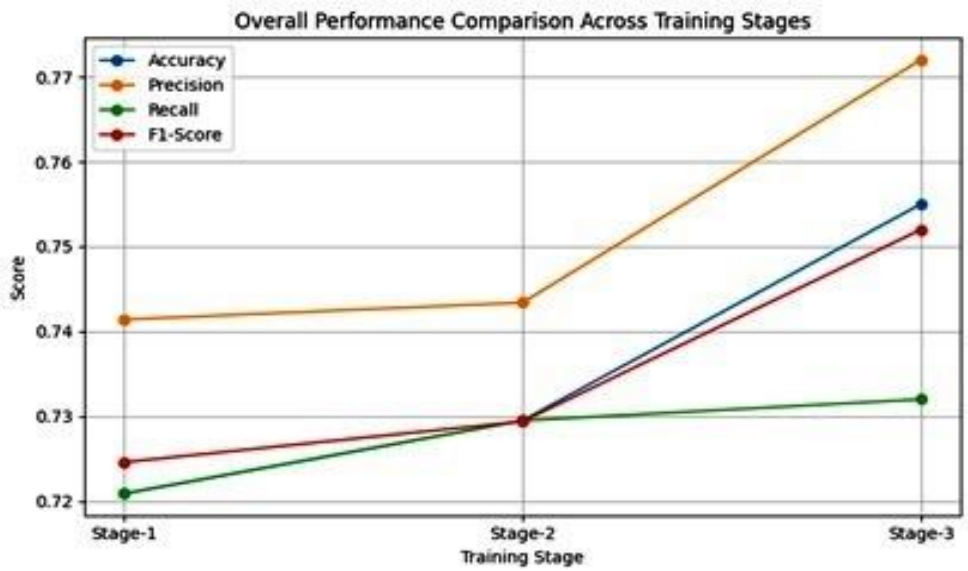

Figure 7: Overall Performance Comparison

Table 2: Statistical Significance Analysis:

Metric	Comparison	t-value	p-value	Significance
Val Accuracy	Stage 2 vs Stage 1	−3.5609	0.0092	Significant
	Stage 3 vs Stage 2	5.8462	0.0006	Highly Significant
Val Precision	Stage 2 vs Stage 1	−4.6456	0.0024	Significant
	Stage 3 vs Stage 2	8.3865	0.0001	Highly Significant

Quantized vs Non-Quantized Model Performance: The figure 8 show quantized model exhibits only a minor 3 to 5 percent reduction in performance compared to the non-quantized version, while achieving substantial model size reduction. This trade-off is acceptable for real-world deployment scenarios.

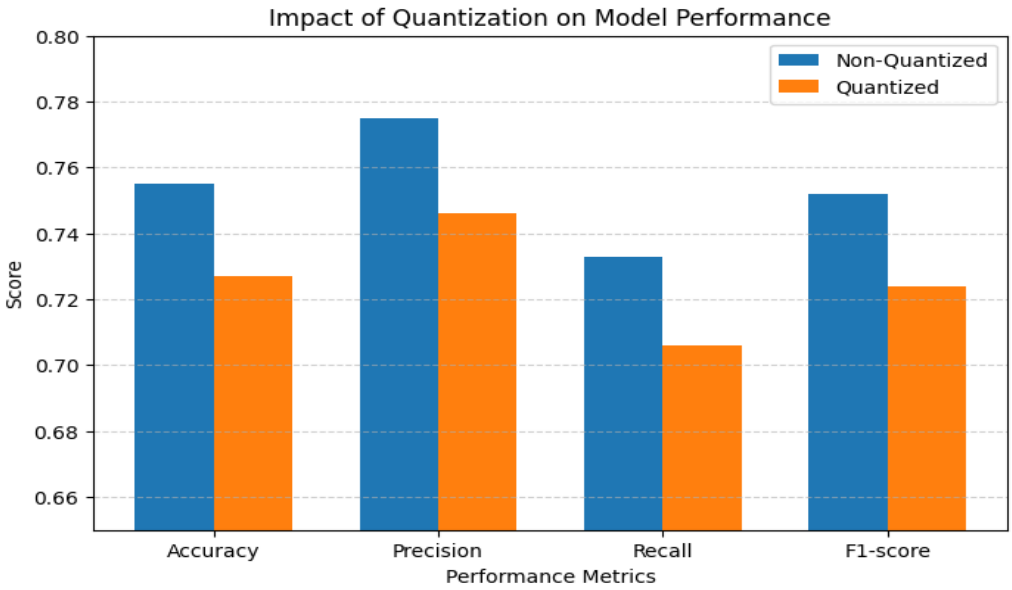

Figure 8: Impact of model quantization of performance metrics

Conclusion and Future Work: This paper presents an efficient skin disease classification framework using MobileNetV2, progressive fine-tuning, class weighting, and post-training quantization. The proposed stage wise approach achieved balanced performance improvements while significantly reducing model size, enabling deployment in resource-constrained environments.

Experimental results of model training in three stages have shown a consistent improvement in training and validation metrics. The validation accuracy, precision, recall, and F1-score were enhanced with strategical fine-tuning, class weighting and un freezing of layer during training stages. The significance analysis using paired t-tests and confidence intervals were also conducted, that confirmed the performance improvements, from Stage-1 to Stage-1 and then form Stage-2 to Stage-3, are not due to some random variation. This shows that a meaningful learning improvements were achieved with proposed methodology. Despite the lightweight nature of MobileNet-V2, the proposed approach achieved competitive and enhanced performance. So the proposed methodology makes it suitable for resource-constrained environments such as edge and mobile deployments.

Future work will focus on incorporating attention mechanisms, exploring lightweight segmentation-assisted pipelines, and validating the model on cross-dataset clinical benchmarks.

REFERENCES

[1]. Touvron, Hugo, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. "Training data-efficient image transformers & distillation through attention." In *International conference on machine learning*, pp. 10347-10357. PMLR, 2021.

[2]. Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In *International conference on machine learning*, pp. 6105-6114. PMLR, 2019.

[3]. Janoria, Honey, Jasmine Minj, and Pooja Patre. "Classification of skin disease from skin images using transfer learning technique." In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 888-895. IEEE, 2020.

[4]. Choy, Shern Ping, Byung Jin Kim, Alexandra Paolino, Wei Ren Tan, Sarah Man Lin Lim, Jessica Seo, Sze Ping Tan et al. "Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease." *NPJ Digital Medicine* 6, no. 1 (2023): 180.

[5]. V. Aishwarya and B. Gomathy, "Skin Disease Classification Using Deep Neural Networks With Resnet," *Journal of Pharmaceutical Negative Results*, vol. 13, no. S10, pp. 5067–5072, 2022.

[6]. M. Akter *et al.*, "Multi-class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network," *arXiv*, Mar. 2023.

[7]. Z. Yao, J. G. Chen, and M. Liu, "Deep learning for imbalanced medical image classification," *arXiv preprint arXiv:2102.01284*, 2021.

[8]. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision

applications," *arXiv preprint arXiv:1704.04861*, 2017.

[9]. S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[10]. S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin Lesion Analyser: An Efficient Seven-Way Multi-Class Skin Cancer Classification Using MobileNet," *arXiv*, July 2019.

[11]. B. et al., "MobileNet-V2: An Enhanced Skin Disease Classification by Attention and Multi-Scale Features," PubMed, 2025.

[12]. U. Kulkarni, M. S. M., S. V. Gurlahosur, and G. Bhogar, "Quantization Friendly MobileNet (QF-MobileNet) Architecture for Vision Based Applications on Embedded Platforms," *Neural Networks*, vol. 136, pp. 28–39, 2021.