Skin Lesion Classification with MobileViTv2 and Dual-Weighted Learning

Kshitiz Regmi*
Department of Computer Science
Georgia State University
Atlanta, USA
kregmi3@student.gsu.edu

Ping Xu†
Department of Computer Science
Georgia State University
Atlanta, USA
pxu4@gsu.edu

Abstract

In real-world clinical settings, dermatology datasets are often highly imbalanced, with a small number of rare but clinically critical disease classes overshadowed by a large volume of common cases. This imbalance poses a significant challenge for standard deep learning training strategies, which tend to be biased toward majority classes and consequently underperform on minority conditions. A similar issue is observed in skin lesion classification, a medical computer vision task where accurate recognition of rare lesions is essential, as misclassification can lead to delayed diagnosis or inappropriate treatment. Given the potentially severe clinical consequences of model errors, addressing data imbalance is crucial for developing reliable and clinically applicable skin lesion classification systems. In this work, we construct a unified 14-class dataset by merging HAM10000 and MSLDv2.0, covering benign, malignant, and infectious skin conditions. We train a compact MobileViTv2 model with approximately 4M parameters and optimize for weighted F1-score. To address class imbalance, we propose a dual-weighted learning strategy that combines inverse square-root sampling with Effective Number of Samples (ENS) loss reweighting. On the held-out test set, the proposed approach achieves a weighted F1-score of 0.9263 and a weighted AUC of 0.9803. Comparative analysis with recent large-model benchmarks suggests that lightweight architectures can remain competitive when imbalance-aware training is applied.

1 Introduction

Early and accurate detection of skin lesions is critical for improving patient outcomes, particularly for malignant diseases such as melanoma. Machine learning systems have shown strong potential in assisting clinicians, yet their effectiveness in practice is limited by the properties of real-world dermatology data. In particular, datasets are highly imbalanced, with common benign lesions dominating training while rare but clinically critical classes remain underrepresented.

Recent work demonstrates that large convolutional and transformer-based models can achieve strong performance on skin lesion classification tasks. However, these models often require substantial computational resources, limiting their deployability in resource-constrained clinical settings. Lightweight models provide an attractive alternative due to their efficiency, but their performance under severe class imbalance remains underexplored.

In this work, we study whether lightweight models can achieve competitive performance when training explicitly accounts for class imbalance, highlighting the trade-offs between model capacity, efficiency, and reliability on rare disease classes.

*Email: kregmi3@student.gsu.edu
†Supervisor.

Preprint. Under review.

2 Dataset

We construct a 14-class skin lesion dataset by merging HAM10000 and MSLDv2.0 Tschandl et al. [2018], Ali et al. [2023]. The merged dataset includes dermatoscopic cancers and infectious skin diseases: Actinic keratoses, Basal cell carcinoma, Benign keratosis-like lesions, Chickenpox, Cowpox, Dermatofibroma, Healthy, HFMD, Measles, Melanocytic nevi, Melanoma, Monkeypox, Squamous cell carcinoma, and Vascular lesions.

For convenience, the combined dataset was downloaded from Kaggle (`kaggle datasets download -d ahmedxc4/skin-ds`), while all experiments cite the original dataset sources.

2.1 Class Imbalance

The merged dataset is severely imbalanced. Melanocytic nevi accounts for over 35% of training samples, while Dermatofibroma and Vascular lesions each represent less than 1%. This imbalance leads to poor batch-level exposure and gradient dominance when standard cross-entropy loss is used.

Table 1: Training data distribution after dataset merging.

Class	Count	Percent
Melanocytic nevi	10300	35.1%
Melanoma	3617	12.3%
Monkeypox	3408	11.6%
Basal cell carcinoma	2658	9.1%
Benign keratosis-like lesions	2099	7.2%
HFMD	1932	6.6%
Healthy	1368	4.7%
Chickenpox	900	3.1%
Cowpox	792	2.7%
Actinic keratoses	693	2.4%
Measles	660	2.3%
Squamous cell carcinoma	502	1.7%
Vascular lesions	202	0.7%
Dermatofibroma	191	0.7%

3 Related Work

Skin lesion classification has received significant attention due to its importance in early detection of skin cancer. Public benchmarks such as HAM10000 and ISIC datasets have enabled large-scale evaluation of automated diagnostic systems, but they present substantial challenges, including visual similarity between lesion types and severe class imbalance Tschandl et al. [2018]. As a result, model performance on rare malignant classes remains a critical concern.

Early approaches primarily relied on convolutional neural networks (CNNs) trained using transfer learning. By fine-tuning ImageNet-pretrained models such as VGG, ResNet, and Inception architectures, researchers achieved strong accuracy on dermoscopic image classification tasks Mahbod et al. [2019]. However, CNN-only models often focus on local texture patterns and may struggle to capture global lesion structure, especially when lesions exhibit subtle differences.

To improve robustness, ensemble-based methods have been widely explored. Recent work demonstrates that combining multiple CNN backbones can improve classification performance by leveraging complementary feature representations across architectures Thwin and Park [2024]. These ensemble models typically employ oversampling techniques to mitigate dataset imbalance and report notable gains in overall accuracy on ISIC and HAM10000 datasets. While effective, such approaches require multiple high-capacity networks, which increases computational cost and limits suitability for deployment in resource-constrained or mobile clinical settings.

More recently, hybrid CNN–Transformer architectures have been proposed to address limitations of pure CNN models. These architectures aim to capture both local visual features and global contextual

information, which is particularly important for dermoscopic analysis. MobileViTv2 represents a lightweight hybrid design that integrates efficient convolutional blocks with mobile-friendly self-attention, enabling strong representation learning with significantly fewer parameters Mehta and Rastegari [2022]. Although larger hybrid models can achieve strong benchmark performance, they often rely on substantially higher computational budgets Aruk et al. [2025].

Across datasets, class imbalance remains a dominant challenge. While many studies address imbalance through oversampling, this strategy alone does not fully resolve majority-class dominance during optimization. Loss reweighting methods such as class-balanced loss based on the Effective Number of Samples (ENS) directly address gradient imbalance by assigning higher importance to rare classes Cui et al. [2019]. Despite their effectiveness, such loss formulations are less commonly explored in combination with lightweight hybrid architectures.

Motivated by these observations, our work focuses on a compact MobileViTv2 backbone combined with a dual-weighted learning strategy. Rather than relying on large ensemble models, we aim to achieve competitive performance through principled imbalance handling at both the data sampling and loss optimization stages. This approach targets strong weighted F1-score on a diverse and highly imbalanced 14-class skin lesion dataset, while maintaining efficiency and deployability.

4 Method

4.1 Backbone Architecture

We adopt MobileViTv2 as the backbone network. The architecture integrates MobileNet-style convolutional blocks for local feature extraction with lightweight transformer blocks for global context modeling. This design enables efficient multi-scale representation learning with low parameter count.

4.2 Dual-Weighted Learning Strategy

To address severe class imbalance, we apply weighting at two stages. First, inverse square-root weighted sampling increases the probability of selecting rare classes in mini-batches. Second, ENS-based class weights are applied within cross-entropy loss to reduce majority-class gradient dominance. This dual-weighted approach improves both data exposure and optimization stability.

5 Evaluation Metrics

Due to the imbalanced nature of the dataset, accuracy alone is insufficient to assess model performance. We therefore use weighted F1-score as the primary evaluation metric. Weighted F1 computes the harmonic mean of precision and recall for each class and weights them by class support, ensuring that both frequent and rare classes contribute appropriately. We also report macro F1-score and weighted AUC, which are commonly used in medical image classification under class imbalance.

6 Experiments

All experiments are conducted using the MobileViT v2 (mobilevitv2_100) architecture with fixed train/validation/test splits and ImageFolder-based data loading. To address class imbalance, effective-number class weighting and a `WeightedRandomSampler` are applied.

Models are trained with the AdamW optimizer and cosine learning rate scheduling. Input images are resized to 256×256, gradient clipping is used to stabilize training, and performance is evaluated using accuracy, precision, recall, F1-score, AUC, and confusion matrices.

Training is performed for a maximum of 60 epochs with a batch size of 100 and an initial learning rate of 3×10^{-3}. Early stopping based on the validation weighted F1-score is employed to prevent overfitting. The best-performing model is automatically selected and saved using a centralized checkpointing mechanism, while full checkpoint-based training resumption is retained.

7 Results and Analysis

7.1 Overall Performance

Table 2: Final performance on the held-out test set.

Metric	Score
Weighted F1-score	0.9263
Test Accuracy	0.9276
Macro F1-score	0.9185
Weighted AUC	0.9803
Test Loss	0.6552

Overall, the model achieves strong and well-balanced performance, with a test accuracy of 92.8% and a weighted F1-score of 92.6%, indicating robust classification across classes.

7.2 Model Training and Loss

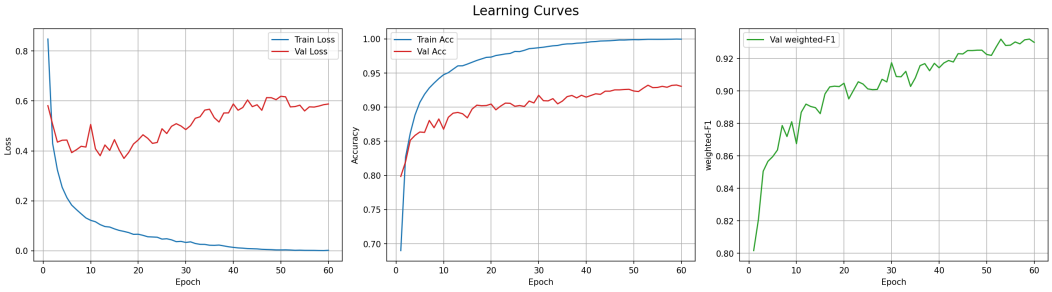

Figure 1: Training and validation loss, accuracy, and weighted F1-score over epochs.

The training curves indicate stable convergence throughout training. Validation weighted F1-score steadily increases and peaks above 0.92, while the gap between training and validation loss remains small, suggesting only mild overfitting and good generalization.

7.3 ENS Weight Analysis

The learned class weights, illustrated in Figure 2, demonstrate a precise inverse relationship to the training data distribution outlined in Table 1. This mechanism is critical for counteracting the gradient dominance of majority classes.

As shown in Figure 2, the highest loss weights are assigned to the most underrepresented classes: *Dermatofibroma* ($w = 2.76$) and *Vascular lesions* ($w = 2.63$). In contrast, majority classes such as *Melanocytic nevi* and *Monkeypox* are assigned significantly lower weights of 0.48 and 0.50, respectively.

This weighting scheme confirms that the Effective Number of Samples (ENS) formulation successfully rebalances the optimization landscape. By down-weighting the loss contribution of majority classes (approx. $0.5\times$), the model prevents frequent examples from overwhelming the gradient updates. Conversely, the high penalties assigned to minority classes (approx. $2.7\times$) force the optimization process to prioritize feature learning for rare conditions despite their scarcity. Crucially, the ENS formulation keeps these weights within a stable numerical range ($0.48 - 2.76$), avoiding the instability and exploding gradients often associated with simple inverse-frequency weighting, which would otherwise assign magnitude weights exceeding $50\times$ for the rarest classes.

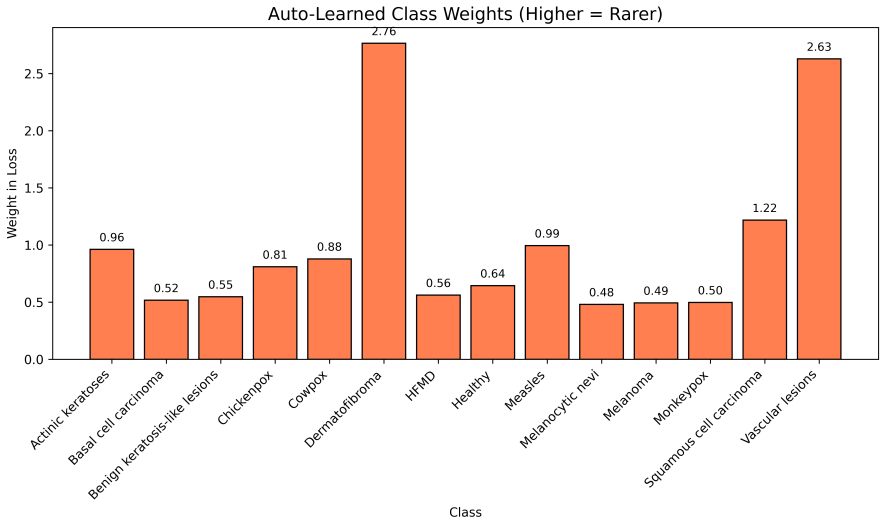

Figure 2: ENS loss weights learned for each class. Higher values correspond to rarer classes.

7.4 Confusion Matrix Analysis

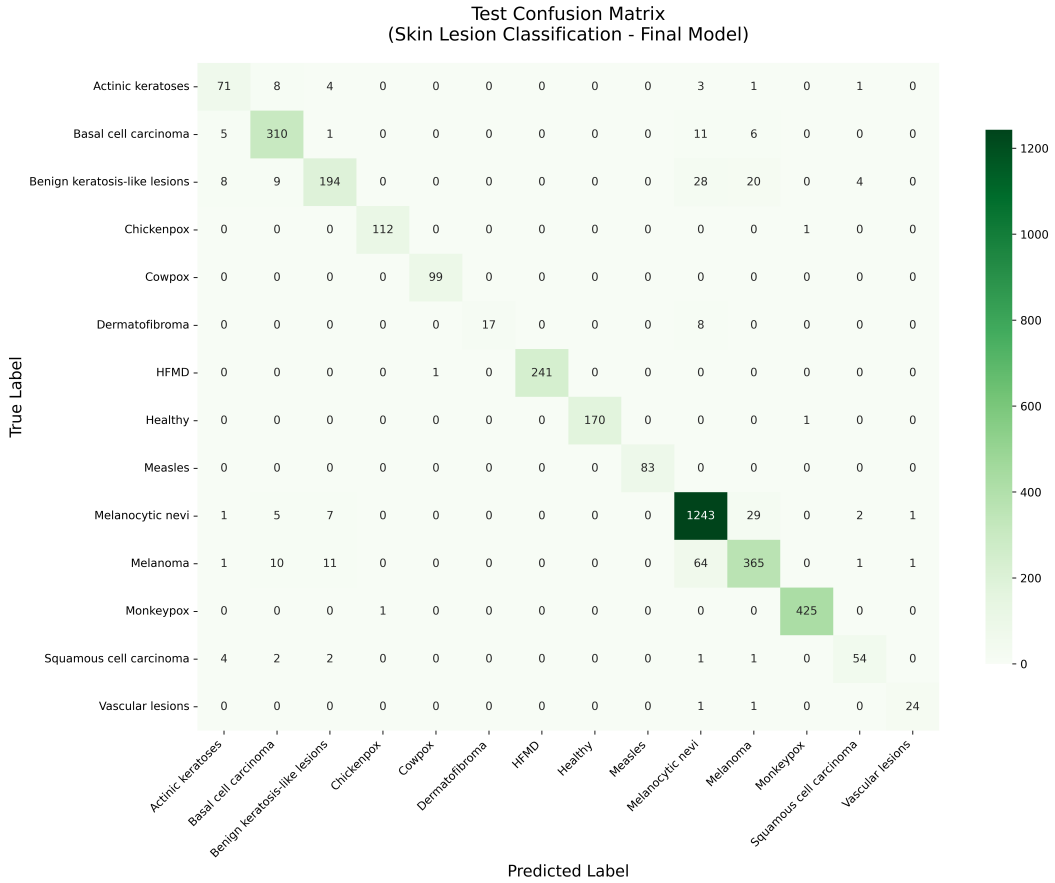

Figure 3: Confusion matrix on the test set.

The model achieves near-perfect performance on infectious diseases and the Healthy class. Most errors occur among visually similar pigmented lesions, particularly between Melanoma and Melanocytic nevi, which reflects known clinical difficulty in dermatological diagnosis.

7.5 Class-wise Performance

To further analyze the model's reliability across diverse conditions, we report the detailed precision, recall, and F1-score for each class in Table 3.

Table 3: Detailed classification report on the test set.

Class	Precision	Recall	F1-score	Support
Actinic keratoses	0.7889	0.8068	0.7978	88
Basal cell carcinoma	0.9012	0.9309	0.9158	333
Benign keratosis-like lesions	0.8858	0.7376	0.8050	263
Chickenpox	0.9912	0.9912	0.9912	113
Cowpox	0.9900	1.0000	0.9950	99
Dermatofibroma	1.0000	0.6800	0.8095	25
HFMD	1.0000	0.9959	0.9979	242
Healthy	1.0000	0.9942	0.9971	171
Measles	1.0000	1.0000	1.0000	83
Melanocytic nevi	0.9146	0.9651	0.9392	1288
Melanoma	0.8629	0.8057	0.8333	453
Monkeypox	0.9953	0.9977	0.9965	426
Squamous cell carcinoma	0.8710	0.8438	0.8571	64
Vascular lesions	0.9231	0.9231	0.9231	26

The results highlight the effectiveness of the dual-weighted learning strategy on rare classes. For instance, *Vascular lesions*, which has only 26 test samples, achieves a high F1-score of 0.9231. Similarly, distinct infectious diseases like *Chickenpox*, *Measles*, and *Monkeypox* are classified with near-perfect accuracy (F1 > 0.99). However, visually similar pigmented lesions remain challenging; specifically, *Melanoma* (recall 0.8057) shows some confusion with *Melanocytic nevi*, a known difficulty in dermoscopic diagnosis due to high inter-class similarity.

8 Comparative Study

Table 4: Comparison with recent benchmark studies.

Method	Dataset	Classes	Params	Key Result
This work (MobileViTv2)	HAM10000+MSLDv2.0	14	∼4M	wF1=0.9263
Aruk et al. Aruk et al. [2025]	HAM10000	7	38M	F1=91.11%
Aruk et al. Aruk et al. [2025]	ISIC 2019	8	38M	F1=90.38%
Elhadidy et al. Elhadidy et al. [2025]	MSLDv2.0	6	–	Val Acc=99.92%

Despite using significantly fewer parameters, the proposed approach achieves competitive performance on a more diverse and imbalanced task.

9 Conclusion

We present a compact MobileViTv2-based model for 14-class skin lesion classification. By combining inverse square-root sampling and ENS-weighted loss, we effectively handle severe class imbalance. The model achieves strong weighted F1-score while remaining efficient and deployable in mobile devices with < 4M parameters. Future work includes stronger regularization, additional backbone comparisons, and modbile deployment.

References

Shams Nafisa Ali, Md. Tazuddin Ahmed, Tasnim Jahan, Joydip Paul, S. M. Sakeef Sani, Nawshaba Noor, Anzirun Nahar Asma, and Taufiq Hasan. A web-based mpox skin lesion detection system using state-of-the-art deep learning models considering racial diversity. *arXiv preprint*, arXiv:2306.14169, 2023. URL `https://arxiv.org/abs/2306.14169`.

Ibrahim Aruk, Ishak Pacal, and Ahmet Nusret Toprak. A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification. *Expert Systems with Applications*, 283:127721, 2025. doi: 10.1016/j.eswa.2025.127721. URL `https://doi.org/10.1016/j.eswa.2025.127721`.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9268–9277. IEEE, 2019. doi: 10.1109/CVPR.2019.00949.

Mohamed S. Elhadidy, Abdelrahman T. Elgohr, Ahmed Mousa, Ahmed Safwat, Roayat Ismail Abdelfatah, and Hossam M. Kasem. Benchmarking pre-trained CNNs and vision transformers for mpox-related dermatological image classification on MSLD v2.0. *Results in Engineering*, 28:108071, 2025. doi: 10.1016/j.rineng.2025.108071. URL `https://doi.org/10.1016/j.rineng.2025.108071`.

Amirreza Mahbod, Gerald Schaefer, Chunliang Wang, Rupert Ecker, and Isabella Ellinge. Skin lesion classification using hybrid deep neural networks. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1229–1233. IEEE, 2019.

Sachin Mehta and Mohammad Rastegari. Separable self-attention for mobile vision transformers. *arXiv preprint*, arXiv:2206.02680, 2022. URL `https://arxiv.org/abs/2206.02680`.

Su Myat Thwin and Hyun-Seok Park. Skin lesion classification using a deep ensemble model. *Applied Sciences*, 14(13):5599, 2024.

Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 Dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1): 180161, 2018. doi: 10.1038/sdata.2018.161. URL `https://doi.org/10.1038/sdata.2018.161`.