Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD38378

Lightweight Deep Learning for Skin Lesion Image Classification

Qiyang Shen

School of Data Science, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China
123090485@link.cuhk.edu.cn

Abstract. Skin lesion image classification is important for early skin-cancer screening, but dermoscopic datasets are severely class-imbalanced, which can impair the recognition of minority classes such as melanoma. Using MobileNetV2 as the backbone, this paper studies, through ablation experiments, how channel attention (squeeze-and-excitation, SE) and class-imbalance handling strategies affect seven-class skin lesion classification. On the HAM10000 dataset split by lesion identifier, class-balanced weighted cross-entropy raises the macro-F1 from 0.702 to 0.776, and class-balanced sampling further raises it to 0.797, minority-class recall improves (melanoma 0.514 to 0.577, vascular lesions 0.467 to 0.867). SE alone improves class balance, but its gain does not add up with the imbalance strategies. The results provide empirical guidance for comparing structural enhancements and imbalance-handling strategies in lightweight skin lesion classification.

Keywords: skin lesion classification, lightweight network, class imbalance, attention mechanism, deep learning

1. Introduction

Skin cancer is among the malignancies with an increasing incidence. Melanoma is aggressive, but when detected early, it is highly treatable. Dermoscopy is a non-invasive examination widely used in dermatology, yet manual diagnosis depends on clinical expertise and is time-consuming, which motivates the development of automated assistive diagnosis. The release of the HAM10000 dataset [1] and the International Skin Imaging Collaboration (ISIC) challenge [2] have advanced the field, and deep convolutional neural networks (CNNs) have reached dermatologist-level performance on this task [3].

Two problems remain. First, dermoscopic datasets are extremely imbalanced: in HAM10000, melanocytic nevi (nv) account for about 6705 images, while dermatofibroma (df) has only 115, which can cause minority classes, especially melanoma, to be underdiagnosed [4, 5]. Second, deployment on edge and mobile devices demands lightweight models, for which MobileNet [6-8] and EfficientNet [9] offer feasible solutions.

This paper takes MobileNetV2 as the backbone and studies how channel attention [10] and two mild class-balancing strategies (weighted cross-entropy and class-balanced sampling) affect seven-class classification, using ablation experiments to examine whether the gains are additive. The contributions are twofold. First, we verify on MobileNetV2 that two mild balancing strategies

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD38378

improve minority-class recall. Second, we find that combining attention with imbalance handling does not simply accumulate the gains, which may inform method selection in similar lightweight models.

2. Related work

2.1. Skin lesion classification

Esteva et al. [3] showed that a deep CNN could classify skin lesions at dermatologist level. Tschandl et al. [1] released HAM10000, a 10015-image, seven-class dermoscopic benchmark. The ISIC 2018 challenge [2] extended this with segmentation, attribute detection and classification tasks involving more than 12500 images. Recently, lightweight Transformers and hybrid architectures have also been introduced: Abbas et al. [11] proposed a separable vision Transformer for lightweight multi-class classification, and Xu et al. [12] designed a CNN-Transformer hybrid with inverse class-frequency weighting to ease imbalance.

Two difficulties dominate the task. One is the minority-class underdiagnosis caused by imbalance, the other is confusion between visually similar classes, such as melanoma and nevi. Raising minority recall without sacrificing overall accuracy is therefore the key challenge.

2.2. Lightweight networks and attention

Edge deployment has driven lightweight network design. Howard et al. [6] proposed MobileNetV1 with depthwise separable convolutions, Sandler et al. [7] proposed MobileNetV2 with inverted residuals and linear bottlenecks, Howard et al. [8] proposed MobileNetV3 combining NAS and attention. EfficientNet [9] balances accuracy and efficiency via compound scaling. Knowledge distillation is another route to lightweight deployment and is left for future work.

Attention recalibrates features adaptively. Hu et al. [10] proposed SE, which models channel dependencies via global average pooling and gating with a small parameter cost. Woo et al. [13] added spatial attention (Convolutional Block Attention Module, CBAM), and Hou et al. [14] embedded position information into channel attention (coordinate attention), both of which can be plugged into lightweight backbones.

2.3. Class imbalance

Class imbalance occurs when the numbers of training samples differ substantially across classes, biasing the classifier toward majority classes and lowering minority recall [5]. Buda et al. [4] studied this systematically in CNNs. Existing methods fall into loss-level and data-level groups. Loss-level methods reweight gradients via class weights or hard-example focus (e.g., weighted cross-entropy, Focal Loss [15]), Cui et al. [16] proposed class-balanced loss based on the effective number of samples. Data-level methods adjust the class distribution via over/under-sampling or synthetic samples (e.g., Synthetic Minority Over-sampling Technique, SMOTE [17]). In medical imaging, minority classes are often the clinically important ones, so choosing the right balancing strategy matters.

DOI: 10.54254/2755-2721/2026.LD38378

3. Method

3.1. Backbone

We use MobileNetV2 [7] as the backbone. Inputs are resized to 224×224, after the backbone, global average pooling and a fully connected head produce seven-class probabilities. With inverted residuals and linear bottlenecks, MobileNetV2 has about 2.23M parameters, making it suitable for resource-constrained deployment. We initialize with ImageNet weights and use transfer learning.

3.2. SE module

We adopt SE [10] as the attention module. Global average pooling squeezes the feature map into a channel descriptor, two FC layers with a Sigmoid generate channel weights, and the input is recalibrated channel-wise:

$$z_c = (1/HW)\ sum_\{i,j\}\ x_c(i,j) \tag{1}$$

$$s = sigmoid(\ W2 * delta(W1 * z)\ ) \tag{2}$$

$$x'_c = s_c * x_c \tag{3}$$

We insert SE after each inverted residual block, it adds only about 0.04M parameters (2.23M to 2.27M).

3.3. Class-balancing strategies

We compare two mild strategies. Weighted cross-entropy weights class c by w_c = (1/sqrt(n_c)) / sum_k(1/sqrt(n_k)), giving rarer classes larger gradient contributions without the overfitting risk of full balancing. The loss is:

$$L\ =\ -\ sum\ _\ \{c = 1..C\}\ w\ _\ c\ *\ y\ _\ c\ *\ log(p\ _\ c) \tag{4}$$

Class-balanced sampling draws samples with a probability p_c proportional to 1/sqrt(n_c) per epoch, balancing class occurrence without changing the loss. In pilot experiments, more aggressive strategies (Focal Loss [15] gamma=2, full-balanced sampling 1/n, and strong augmentation on minority classes) all led to training-set memorization (train accuracy > 0.99 while minority recall collapsed below 0.1), with large rotations and affine transformations applied to minority classes appearing to be the main cause. We therefore adopt the mild strategies above with light augmentation.

4. Experiments

4.1. Dataset and setup

We use HAM10000 [1], comprising 10015 images across seven classes: nv 6705, mel 1113, bkl 1099, bcc 514, akiec 327, vasc 142, df 115. We split the dataset by lesion identifier so that images of the same lesion never cross sets, obtaining 8013 training images, 1010 validation images, and 992 test images. All models use 224×224 input, a batch size of 64, 50 epochs, Adam (lr 1e-4) with cosine annealing, mixed precision, and a fixed seed. Training uses light augmentation (horizontal flip,

DOI: 10.54254/2755-2721/2026.LD38378

random crop, small rotation), validation/test use only resize and normalization. We select the epoch with the highest validation macro-F1 and report accuracy, macro precision, macro recall, macro-F1, macro-AUC and per-class recall:

$$\mathrm{Recall_c} = \mathrm{TP_c} / (\mathrm{TP_c} + \mathrm{FN_c}) \tag{5}$$

$$\mathrm{F1_c} = 2 * \mathrm{P_c} * \mathrm{R_c} / (\mathrm{P_c} + \mathrm{R_c}) \tag{6}$$

$$\mathrm{F1_macro} = (1/\mathrm{C}) \, \mathrm{sum_}\{\mathrm{c} = 1..\mathrm{C}\} \, \mathrm{F1_c} \tag{7}$$

4.2. Ablation results

Table 1. Overall test-set metrics of the six groups

Group	Acc	P_macro	R_macro	F1_macro	AUC	dF1
Baseline	0.8407	0.7775	0.6681	0.7017	0.9706	—
+A (SE)	0.8296	0.7626	0.6888	0.7180	0.9676	+0.016
+B1 (weighted CE)	0.8518	0.8074	0.7540	0.7763	0.9720	+0.075
+A+B1 (SE+wCE)	0.8145	0.6758	0.7676	0.7115	0.9645	+0.010
+B2 (sampling)	0.8589	—	—	0.7974	0.9727	+0.096
+A+B2 (SE+samp)	0.8306	—	—	0.7431	0.9685	+0.041

Table 1 shows that imbalance handling is the main driver: +B1 and +B2 raise macro-F1 by 0.075 and 0.096 respectively, with small gains in accuracy and AUC. SE alone raises F1 by 0.016 but lowers accuracy by 0.011, a precision-recall trade-off. The +A+B1/+A+B2 combinations still outperform the Baseline but underperform the corresponding +B alone, so the gains are not additive. Specifically, +A+B1 drops macro precision from 0.8074 to 0.6758 while recall rises only marginally to 0.7676: after SE recalibrates the channels, loss weighting makes the model predict minority classes more aggressively and precision falls. Channel attention and imbalance handling therefore overlap on this backbone, whether to combine them should be decided by validation results.

4.3. Per-class analysis

Fig. 1 shows the confusion matrices of selected groups and the validation macro-F1 curves, Table 2 lists per-class recall. Panels (a)–(c) are the confusion matrices of Baseline, +A(SE) and +B1(weighted CE), and (d) the six validation macro-F1 curves.

Table 2 shows that balancing improves minority recall: weighted CE raises melanoma recall from 0.514 to 0.577, and sampling raises vascular recall from 0.467 to 0.867 and bkl from 0.604 to 0.772. This is accompanied by a slight drop in majority-class nv recall (0.962 to ~0.945), as expected. SE alone raises recall for vasc, bkl and df but lowers it for mel and bcc, trading some fine-grained discrimination (mel vs nv) for better class balance. In the +A+B1/+A+B2 combinations, nv recall falls further to 0.897–0.917 and most minorities gain nothing extra, consistent with the non-additive finding in Table 1. Note that vasc and df have only about a dozen test samples each, so per-class recall is noisy, e.g., vasc recall of 1.0 in +A+B1 may reflect this small-sample fluctuation.

DOI: 10.54254/2755-2721/2026.LD38378

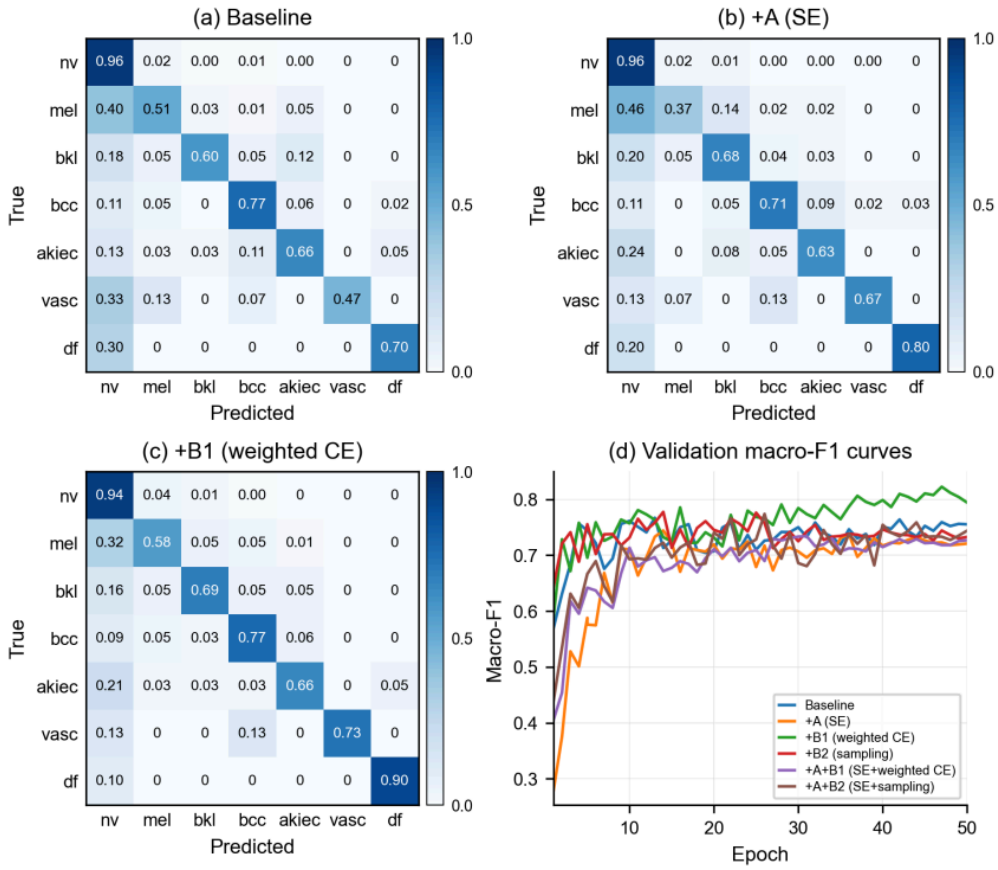

Figure 1. Ablation results: (a) Baseline, (b) +A(SE), (c) +B1(weighted CE) confusion matrices, (d) validation macro-F1 curves

Table 2. Per-class recall

Class	Baseline	+A	+B1	+B2	+A+B1	+A+B2
akiec	0.6579	0.6316	0.6579	0.7105	0.6579	0.6842
bcc	0.7727	0.7121	0.7727	0.6970	0.6970	0.7273
bkl	0.6040	0.6832	0.6931	0.7723	0.6535	0.6931
df	0.7000	0.8000	0.9000	0.9000	0.9000	0.9000
mel	0.5135	0.3694	0.5766	0.5405	0.5676	0.5586
nv	0.9616	0.9585	0.9447	0.9508	0.8971	0.9171
vasc	0.4667	0.6667	0.7333	0.8667	1.0000	0.8000

5. Conclusion

This paper studied lightweight deep learning for seven-class skin lesion classification on HAM10000, using MobileNetV2 as the backbone and comparing channel attention (SE) with two class-imbalance strategies: class-balanced weighted cross-entropy and class-balanced sampling.

Imbalance handling was the dominant source of improvement: it raised the macro-F1 from 0.702 to 0.776 (weighted cross-entropy) and 0.797 (sampling), with minority-class recall increasing from

DOI: 10.54254/2755-2721/2026.LD38378

0.514 to 0.577 for melanoma and from 0.467 to 0.867 for vascular lesions, at only a modest cost to the majority class. This suggests that, for severely imbalanced medical datasets, how the training distribution is handled can matter as much as the network structure, and that a mild weighting scheme suffices: more aggressive choices, including Focal Loss, full rebalancing and strong minority-class augmentation, led to training-set memorization and collapsed test recall.

SE attention behaved differently. It improved per-class balance with almost no parameter overhead (about 0.04M), confirming that channel recalibration offers a cheap way to re-emphasize discriminative features. However, combining SE with either strategy did not further improve performance, the combinations even underperformed the balancing strategies alone. A plausible explanation is that the two mechanisms act along the same axis: loss weighting already amplifies minority-class gradients, and SE recalibration shifts feature emphasis in a similar direction, so their effects overlap rather than add. This non-additivity cautions against blindly stacking attention on top of reweighting in imbalanced classification.

The main limitations are the use of a single dataset, the absence of on-device latency and quantization evaluations, and the small test sets for the rarest classes, which make per-class estimates noisy. Future work will validate the approach across multiple datasets, evaluate quantized and distilled models on edge hardware, and add explainability to make predictions more trustworthy in clinical use.

References

[1] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data, 5*, 180161.

[2] Codella, N., Rotemberg, V., Tschandl, P., et al. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC). arXiv:1902.03368.

[3] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

[4] Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks, 106*, 249–259.

[5] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE TKDE, 21*(9), 1263–1284.

[6] Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.

[7] Sandler, M., Howard, A., Zhu, M., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *CVPR* (pp. 4510–4520).

[8] Howard, A., Sandler, M., Chu, G., et al. (2019). Searching for MobileNetV3. In *ICCV* (pp. 1314–1324).

[9] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *ICML* (pp. 6105–6114).

[10] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *CVPR* (pp. 7132–7141).

[11] Abbas, Q., Daada, A., Rashid, U., et al. (2023). Assist-Dermo: A lightweight separable vision transformer model for multiclass skin lesion classification. *Diagnostics, 13*(15), 2531.

[12] Xu, J., Zhao, X., & Li, X. J. (2026). A lightweight classification network for skin lesion based on the hybridization of CNN and Transformer. *Journal of Shanxi University (Nat. Sci. Ed.), 49*(2), 284–294.

[13] Woo, S., Park, J., Lee, J. Y., et al. (2018). CBAM: Convolutional block attention module. In *ECCV* (pp. 3–19).

[14] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *CVPR* (pp. 13708–13717).

[15] Lin, T. Y., Goyal, P., Girshick, R., et al. (2017). Focal loss for dense object detection. In *ICCV* (pp. 2999–3007).

[16] Cui, Y., Jia, M., Lin, T. Y., et al. (2019). Class-balanced loss based on effective number of samples. In *CVPR* (pp. 9268–9277).

[17] Chawla, N. V., Bowyer, K. W., Hall, L. O., et al. (2002). SMOTE: Synthetic minority over-sampling technique. *JAIR, 16*, 321–357.