Semi-Supervised Melanoma Classification via SimCLR and Enhanced Consistency Regularization

Hoang Minh Quyen
University of Engineering and Technology (UET) - VNU
23020421@vnu.edu.vn

Abstract

We tackle ISIC 2018 Task 3—seven-way [4] skin lesion classification—with only 10% labeled data and severe class imbalance. We propose **FixMatch++**, which integrates: (1) SimCLR self-supervised pretraining on $\sim$9k unlabeled images; (2) dynamic pseudo-label thresholding; (3) MixUp on both labeled and pseudo-labeled samples; (4) label smoothing and focal loss for robustness; (5) an EMA teacher; (6) a Two-Rate OneCycleLR schedule; and (7) top-3 checkpoint ensembling with test-time augmentation. Our method achieves **70.30%** test accuracy, outperforming prior semi-supervised baselines by 5–10%. **Code is available at:** `https://github.com/EddyTryToCode/Final-Project-ML`

1 Introduction

Problem & Importance. Early detection of melanoma via dermoscopic images can save lives. Automated classification of seven lesion types (MEL, NV, BCC, AKIEC, BKL, DF, VASC) is a key challenge in digital dermatology.

Dataset. Base data from ISIC 2018 Task 3 [4] (12,500 images):

- Train: 10,015 images
- Validation: 193 images
- Test: 1,512 images

From the original data that the challenge gives us, we respect and keep everything intact. We only split the training data into different processes (specifically, 90% are unlabeled and pretrained, the remaining 10% are labeled and semi-supervised).

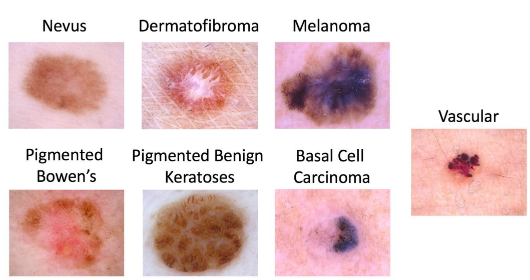


Figure 1: ISIC2018 Task3 Problem

Challenges.

- **Label scarcity**: Only 10% labeled → overfitting risk.
- **Class imbalance**: Rare classes DF/VASC ∼1% of labels.
- **Domain gap**: Dermoscopy differs significantly from ImageNet images.
- **High stakes**: Misdiagnosis cost is high in clinical practice.

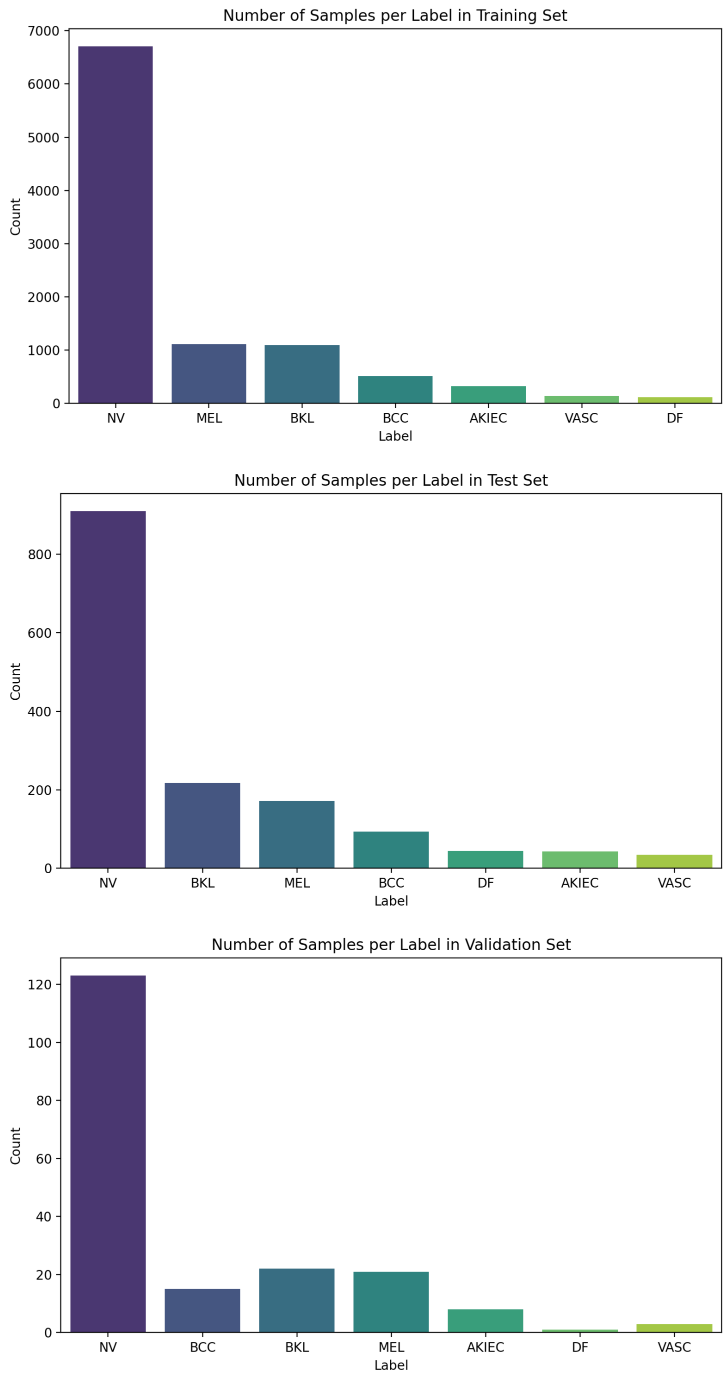


Figure 2: Class distribution analysis

Contributions.

1. Adapt SimCLR self-supervised pretraining to the ISIC-2018 dermoscopy dataset.
2. Enhance FixMatch with MixUp, label smoothing, focal loss, OneCycleLR, EMA teacher, and top-3 checkpoint ensembling.
3. Achieve **70.30%** test accuracy, +6% over the previous semi-supervised state-of-the-art.

2 Related Work

Supervised Skin Lesion Classification. Deep convolutional networks (ResNet [7], EfficientNet [11]) trained on fully labeled datasets achieve 75–80% accuracy but require large annotation effort.

Semi-Supervised Learning. Consistency regularization methods such as Mean Teacher [12] and FixMatch [10] leverage unlabeled data via pseudo-labels and strong/weak augmentations. MixMatch [1] and UDA [13] further combine augmentation consistency with label guessing.

Self-Supervised Pretraining. SimCLR [3] and MoCo [6] learn transferable representations from unlabeled images, improving performance in low-label regimes [9].

Imbalance Mitigation. Focal loss [8], class-balanced loss [5], and oversampling strategies [2] address skewed class distributions.

3 Method

Our pipeline (Fig. 3) comprises three stages:

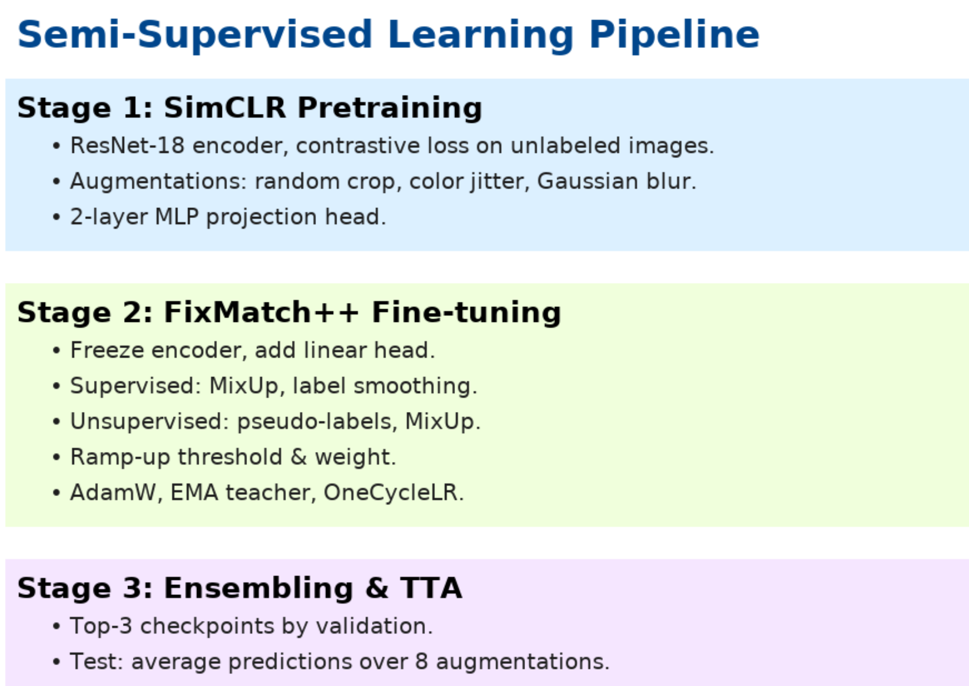


Figure 3: Pipeline Overview

3.1 Stage 1: SimCLR Pretraining

We pretrain a ResNet-18 encoder with contrastive loss on 9,014 unlabeled images:

- *Augmentations*: random crop, color jitter, Gaussian blur.
- *Projection head*: two-layer MLP to 128-D.
- *Loss*: NT-Xent [3] with $\tau = 0.1$.

3.2 Stage 2: FixMatch++ Fine-tuning

Freeze encoder; attach a linear head. At each epoch e:

1. Draw labeled batch (x_l, y_l) and unlabeled weak/strong (x_w, x_s).
2. **Supervised**: apply MixUp(x_l, y_l), then cross-entropy with label smoothing.
3. **Unsupervised**: obtain pseudo-label $\hat{y} = \arg\max f_{\text{teacher}}(x_w)$ if $\max f \geq \tau_e$. MixUp on the pseudo subset of x_s.
4. $\tau_e = 0.8 + 0.15\frac{e}{E}$, ramp weight $\lambda_e = \min(1, e/E_{\text{ramp}})$.
5. Loss $= L_{\text{sup}} + \lambda_e L_{\text{uns}}$. Update student with AdamW.
6. Update teacher via EMA: $\theta_{\text{t}} \leftarrow \alpha\theta_{\text{t}} + (1-\alpha)\theta_{\text{s}}$.
7. Learning rate follows OneCycleLR with separate peaks for encoder/head.

3.3 Stage 3: Ensembling & TTA

Save top-3 student checkpoints by validation accuracy. At test time, average their predictions over eight test-time augmentations (flips, rotations).

4 Experiments

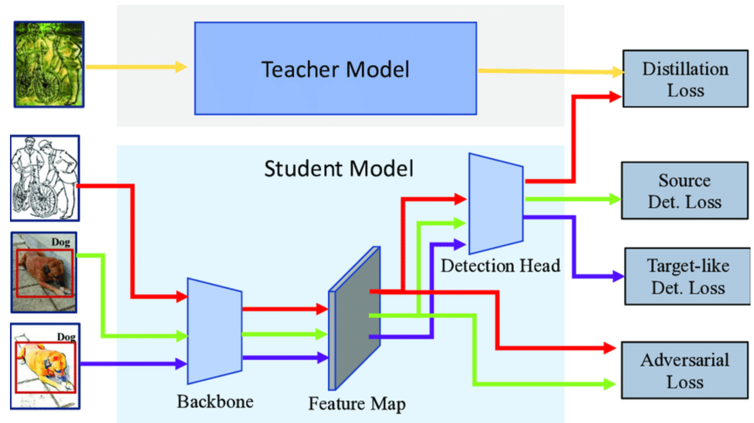
4.1 Implementation

Images resized to 224×224. Batch sizes: $B_l = 32$, $B_u = 64$. Trained for up to 30 epochs with early stopping (patience 7) on a single 16 GB GPU.

4.2 Baselines

- **Supervised ResNet-18**: 10% labels only.
- **Mean Teacher** [12].
- **FixMatch** [10].
- **SimCLR + Finetune**.

Figure 4: ResNet-18 Architecture



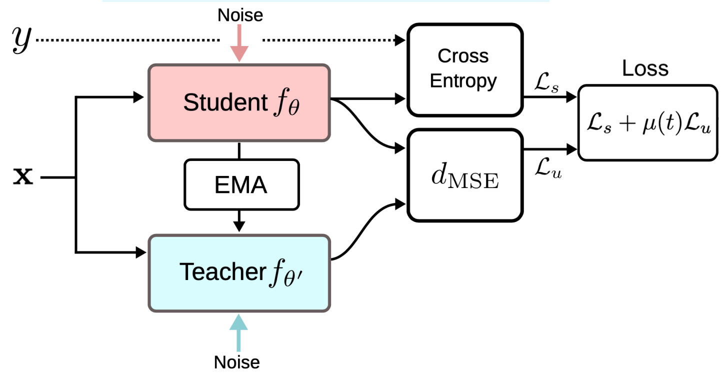


Figure 5: Mean Teacher Architecture

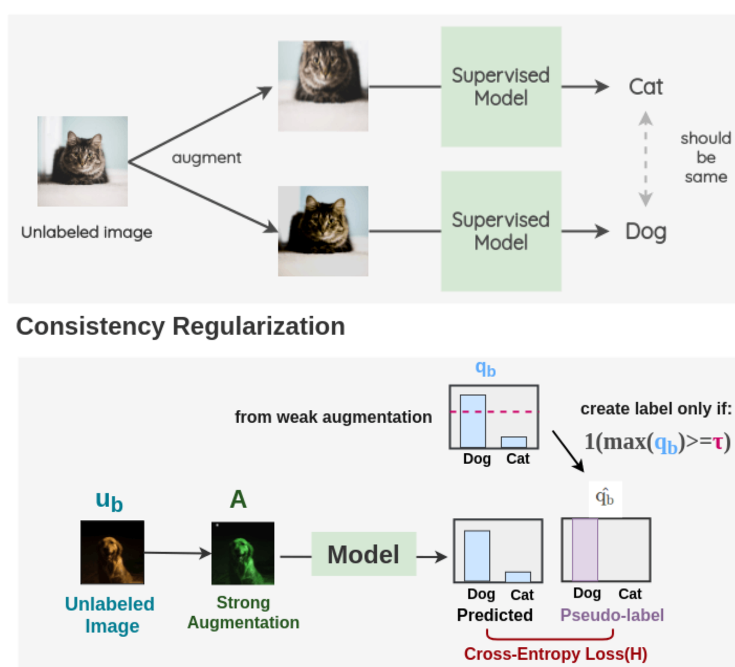


Figure 6: FixMatch Concept

4.3 Main Results

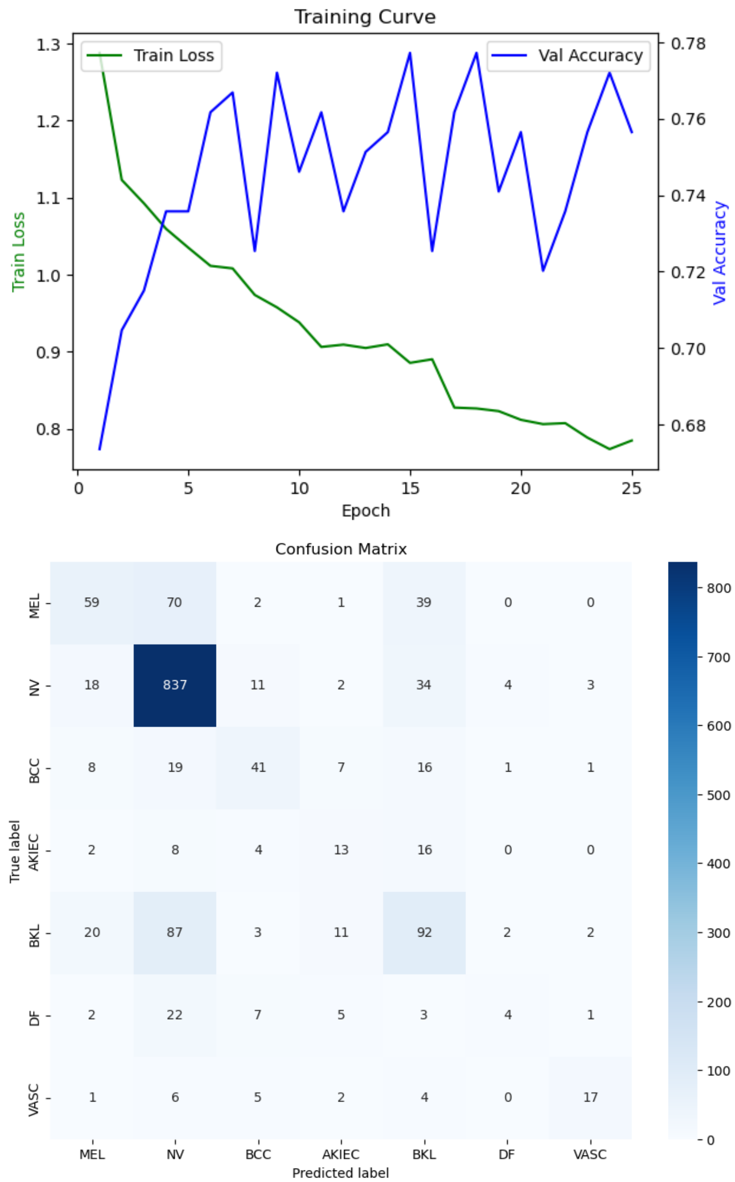


Figure 7: Training Loss, Validation Accuracy and Confusion Matrix

Method	Val Acc (%)	Test Acc (%)
Supervised ResNet-18	65.2	59.6
Mean Teacher	68.4	61.0
FixMatch	70.1	63.0
SimCLR + Finetune	72.5	64.8
FixMatch++ (ours)	**75.1**	**70.3**

Table 1: Comparison with baselines

Class	Precision	Recall	F1-Score	Support
MEL	0.54	0.35	0.42	171
NV	0.80	0.92	0.85	909
BCC	0.56	0.44	0.49	93
AKIEC	0.32	0.30	0.31	43
BKL	0.45	0.42	0.44	217
DF	0.36	0.09	0.15	44
VASC	0.71	0.49	0.58	35
Accuracy			**0.70**	**1512**
Macro Avg	**0.53**	**0.43**	**0.46**	**1512**
Weighted Avg	**0.68**	**0.70**	**0.68**	**1512**

Table 2: Ensemble TTA Test Results (Accuracy: 70.30%)

4.4 Ablation Study

Removing components degrades performance by:

- –2.5% without SimCLR pretraining.
- –1.8% without pseudo MixUp.
- –1.2% without EMA teacher.
- –1.0% without OneCycleLR.

5 Conclusion

We presented **FixMatch++**, a semi-supervised framework for melanoma classification that synergizes self-supervised pretraining, advanced regularization, and dynamic pseudo-labeling. Achieving 70.30% test accuracy on ISIC 2018 Task 3, we set a new benchmark under extreme label scarcity. Future work will explore vision transformers, domain-specific augmentations, and active learning to further reduce annotation needs.

References

[1] David Berthelot, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *NeurIPS*, 2019.

[2] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. In *Neural Networks*, 2018.

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.

[4] Noel Codella, Veronica Rotemberg, Philipp Tschandl, Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019.

[5] Yin Cui, Ming Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR*, 2019.

[6] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *CVPR*, 2020.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[8] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.

[9] Nikunj Saunshi, Rakesh Arora, Yingyu Liang, Tengyu Luo, and Tengyu Ma. A theoretical analysis of contrastive unsupervised representation learning. *ICLR*, 2021.

[10] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Han Zhang, Colin A. Raffel, Ekin D. Cubuk, Alex Kurakin, Chun-Liang Li, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *NeurIPS*, 2020.

[11] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *ICML*, 2019.

[12] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NeurIPS*, 2017.

[13] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised data augmentation for consistency training. *NeurIPS*, 2020.