

---

# Semi-Supervised Melanoma Classification via SimCLR and Enhanced Consistency Regularization

---

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

## Abstract

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

## 10 1 Introduction

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

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

- 15 • Train: 10,015 images
- 16 • Validation: 193 images
- 17 • Test: 1,512 images

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

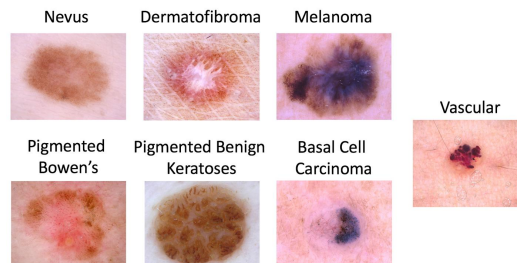


Figure 1: ISIC2018 Task3 Problem

21 **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.

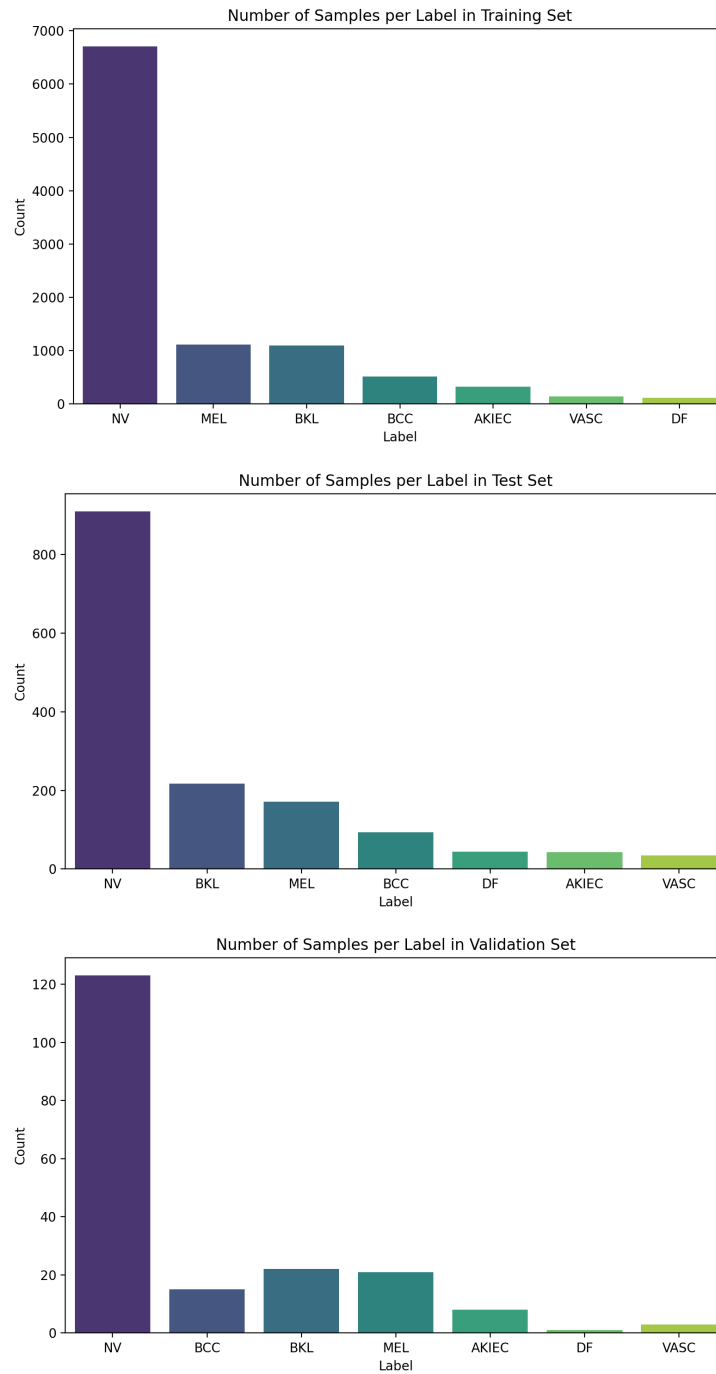


Figure 2: Class distribution analysis

26 **Contributions.**

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

## 31 2 Related Work

32 **Supervised Skin Lesion Classification.** Deep convolutional networks (ResNet [7], EfficientNet  
33 [11]) trained on fully labeled datasets achieve 75–80% accuracy but require large annotation effort.

34 **Semi-Supervised Learning.** Consistency regularization methods such as Mean Teacher [12] and  
35 FixMatch [10] leverage unlabeled data via pseudo-labels and strong/weak augmentations. MixMatch  
36 [1] and UDA [13] further combine augmentation consistency with label guessing.

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

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

## 41 3 Method

42 Our pipeline (Fig. 3) comprises three stages:

### Semi-Supervised Learning Pipeline

#### Stage 1: SimCLR Pretraining

- ResNet-18 encoder, contrastive loss on unlabeled images.
- Augmentations: random crop, color jitter, Gaussian blur.
- 2-layer MLP projection head.

#### Stage 2: FixMatch++ Fine-tuning

- Freeze encoder, add linear head.
- Supervised: MixUp, label smoothing.
- Unsupervised: pseudo-labels, MixUp.
- Ramp-up threshold & weight.
- AdamW, EMA teacher, OneCycleLR.

#### Stage 3: Ensembling & TTA

- Top-3 checkpoints by validation.
- Test: average predictions over 8 augmentations.

Figure 3: Pipeline Overview

### 43 3.1 Stage 1: SimCLR Pretraining

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

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

### 48 3.2 Stage 2: FixMatch++ Fine-tuning

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

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

### 58 3.3 Stage 3: Ensembling & TTA

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

## 61 4 Experiments

### 62 4.1 Implementation

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

### 65 4.2 Baselines

66 To evaluate the effectiveness of our proposed FixMatch++ method, we conduct comprehensive  
67 comparisons with four primary baselines, each representing different approaches in semi-supervised  
68 learning and skin lesion classification. This section provides detailed architectural descriptions and  
69 performance analysis based on the illustrated figures.

#### 70 4.2.1 Supervised ResNet-18 (Basic Baseline)

71 **Architecture Description**: The baseline employs the standard ResNet-18 architecture as illustrated  
72 in Figure 4. The network consists of four main residual blocks with progressively increasing channel  
73 dimensions:  $64 \rightarrow 128 \rightarrow 256 \rightarrow 512$ . Each residual block incorporates shortcut connections that  
74 effectively address the vanishing gradient problem through identity mappings. The architecture  
75 includes:

- 76 • Initial convolutional layer (7×7, stride 2) followed by max pooling
- 77 • Four residual blocks with increasing depth and feature complexity
- 78 • Global average pooling to reduce spatial dimensions
- 79 • Final fully connected layer with 7 neurons corresponding to the seven classification classes  
80 (MEL, NV, BCC, AKIEC, BKL, DF, VASC)

81 **Training Characteristics**: This baseline represents the most straightforward approach with several  
82 limitations:

- 83 • Utilizes only 10% labeled data ( $\approx 1,001$  images) from the training set
- 84 • Does not leverage the remaining 90% unlabeled data ( $\approx 9,014$  images)

- Trained with standard cross-entropy loss without any regularization techniques
- Prone to overfitting due to limited labeled data availability and high model capacity
- Initialization from ImageNet pretrained weights, creating a domain gap with dermoscopic images

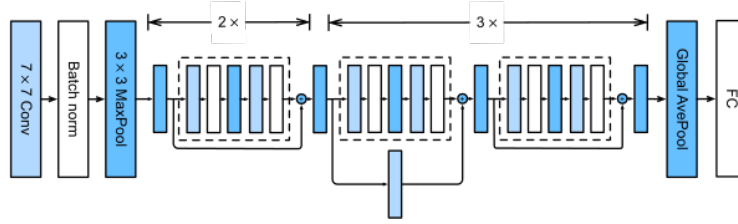


Figure 4: ResNet-18 Architecture

#### 4.2.2 Mean Teacher (Consistency Regularization)

**Architecture Description:** Based on Figure 5, Mean Teacher employs a sophisticated teacher-student architecture that leverages consistency regularization. The framework consists of two identical networks with different parameter update mechanisms:

- **Student Network:** ResNet-18 updated through standard gradient descent on both labeled and unlabeled data
- **Teacher Network:** An exponential moving average (EMA) copy of the student network providing stable target predictions
- **Consistency Loss:** Measures prediction differences between teacher and student on the same input with different augmentations

**Operational Mechanism:** The Mean Teacher framework operates through the following process:

1. Input images undergo two different augmentation transformations (noise injection, random transformations)
2. Both teacher and student networks generate predictions on augmented versions
3. Loss function combines supervised loss (labeled data) and consistency loss (unlabeled data):

$$\mathcal{L} = \mathcal{L}_{\text{supervised}} + \lambda(t) \cdot \mathcal{L}_{\text{consistency}}$$

4. Teacher weights update via EMA:  $\theta_{\text{teacher}} = \alpha \times \theta_{\text{teacher}} + (1 - \alpha) \times \theta_{\text{student}}$
5. Consistency weight  $\lambda(t)$  follows a ramp-up schedule during training

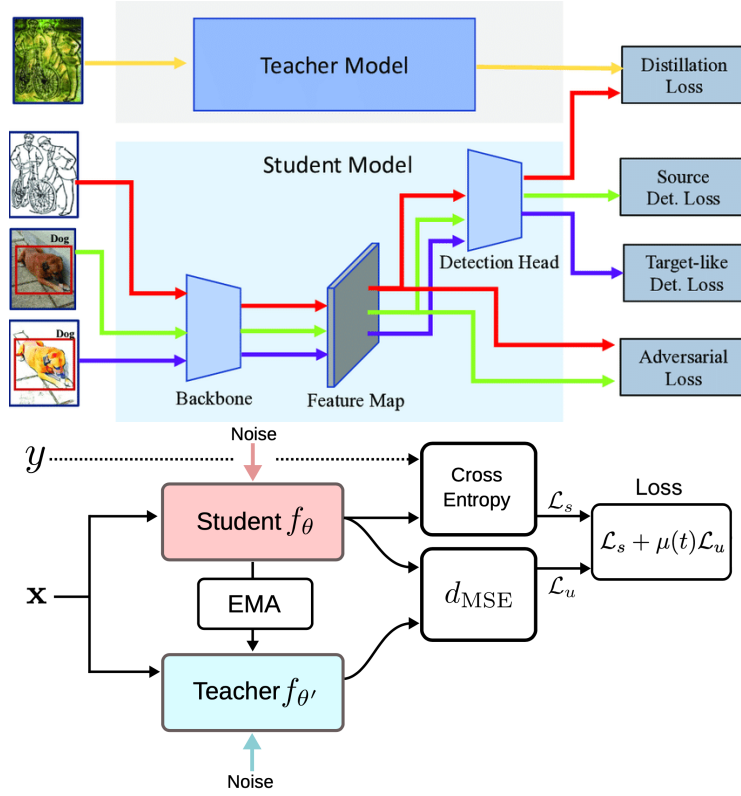


Figure 5: Mean Teacher Architecture

#### Advantages and Limitations:

- *Advantages:* Effectively leverages unlabeled data through consistency regularization; stable teacher network provides high-quality targets; robust to noise and augmentations
- *Limitations:* Relies solely on consistency without explicit quality filtering mechanisms; lacks self-supervised pretraining benefits; may struggle with confident but incorrect predictions

#### 4.2.3 FixMatch (Pseudo-labeling with Consistency)

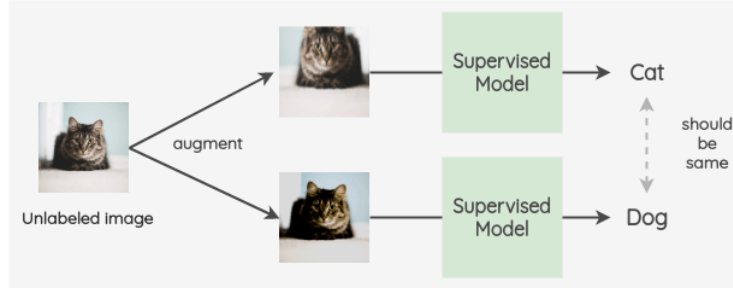
**Architecture Description:** Figure 6 illustrates the core concept of FixMatch, which elegantly combines pseudo-labeling with consistency regularization through a dual-augmentation strategy. The framework introduces a sophisticated approach to semi-supervised learning:

- **Weak Augmentation:** Light transformations (horizontal flip, random crop) applied to unlabeled data for pseudo-label generation
- **Strong Augmentation:** Intensive transformations (RandAugment, cutout, color distortion) applied to the same unlabeled samples for consistency training
- **Confidence Threshold:** Utilizes pseudo-labels only when model confidence exceeds  $\tau$  (typically  $\tau = 0.95$ )
- **Fixed Threshold:** Maintains consistent quality control throughout training

**Operational Workflow:** The FixMatch algorithm operates through the following sophisticated pipeline:

1. **Labeled Data Processing:** Standard supervised learning with cross-entropy loss
2. **Weak Augmentation:** Apply light transformations to unlabeled data → Generate pseudo-labels if  $\max(p_m) \geq \tau$

- 126 3. **Strong Augmentation:** Apply intensive transformations to the same unlabeled samples  
 127 4. **Consistency Loss:** Enforce prediction consistency between weak and strong augmented  
 128 versions  
 129 5. **Combined Loss:**  $\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda \mathcal{L}_{\text{unsup}}$  where  $\lambda$  follows a ramp-up schedule



Consistency Regularization

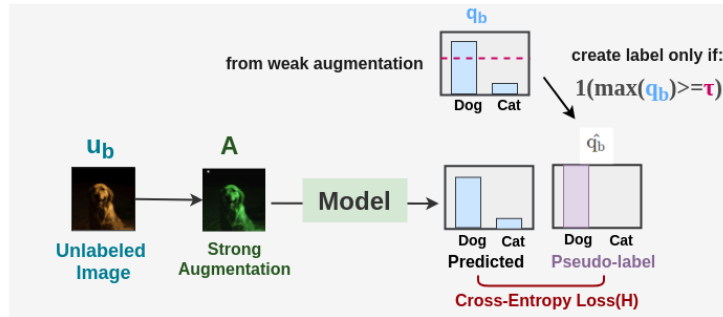


Figure 6: FixMatch Concept

#### 130 Technical Strengths and Weaknesses:

- 131 • *Strengths:* Effective combination of pseudo-labeling and consistency regularization; con-  
 132 fidence thresholding filters unreliable predictions; strong augmentation enhances model  
 133 robustness and generalization; simple yet powerful framework  
 134 • *Weaknesses:* Fixed threshold may be suboptimal throughout training phases; lacks domain-  
 135 specific pretraining; does not incorporate advanced techniques like MixUp or focal loss;  
 136 struggles with class imbalance

#### 137 4.2.4 SimCLR + Finetune (Self-supervised Pretraining)

138 **Architecture Description:** This approach implements a comprehensive two-stage training paradigm  
 139 that leverages self-supervised learning for domain adaptation, as shown in Figure 7:

##### 140 Stage 1 - SimCLR Pretraining:

- 141 • **Encoder:** ResNet-18 backbone for feature extraction from dermoscopic images  
 142 • **Projection Head:** 2-layer MLP (512  $\rightarrow$  512  $\rightarrow$  128) projecting features to lower-  
 143 dimensional space  
 144 • **Contrastive Loss:** NT-Xent (Normalized Temperature-scaled Cross-Entropy) loss with  
 145 temperature parameter  $\tau = 0.1$   
 146 • **Data Augmentation:** Comprehensive augmentation pipeline including random crop (0.08-  
 147 1.0 scale), color jitter (brightness, contrast, saturation, hue), Gaussian blur, and random  
 148 horizontal flip  
 149 • **Training Data:** Utilizes all 9,014 unlabeled images for representation learning

##### 150 Stage 2 - Supervised Finetuning:

- 151 • Freeze encoder weights learned from Stage 1
- 152 • Remove projection head and attach linear classification head ( $512 \rightarrow 7$  classes)
- 153 • Finetune exclusively on 10% labeled data with standard cross-entropy loss
- 154 • Apply moderate data augmentation to prevent overfitting

Figure 7: ResNet-18 Architecture

155 **SimCLR Contrastive Learning Mechanism:** The contrastive learning framework operates through  
 156 the following sophisticated process:

- 157 1. Each input image generates two augmented views through different transformation pipelines
- 158 2. Encoder extracts feature representations for both views:  $h_i = f(x_i)$ ,  $h_j = f(x_j)$
- 159 3. Projection head maps features to normalized embeddings:  $z_i = g(h_i)$ ,  $z_j = g(h_j)$
4. NT-Xent loss maximizes agreement between positive pairs while minimizing agreement with negative pairs:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)}$$

- 160 5. Learns generalizable representations from unlabeled dermoscopic images without requiring  
 161 annotations

#### 162 **Advantages and Limitations:**

- 163 • *Advantages:* Utilizes entire unlabeled dataset for representation learning; learned represen-  
 164 tations exhibit high generalizability; particularly effective for domain-specific tasks like  
 165 dermoscopy; addresses domain gap between ImageNet and medical images; provides strong  
 166 initialization for downstream tasks
- 167 • *Limitations:* Finetuning stage employs only supervised learning; lacks consistency regular-  
 168 ization integration; potential overfitting during finetuning phase; two-stage training requires  
 169 careful hyperparameter tuning

### 170 **4.3 Main Results**

171 Our comprehensive evaluation demonstrates the effectiveness of the proposed FixMatch++ method  
 172 across multiple metrics and provides detailed insights into model performance characteristics.



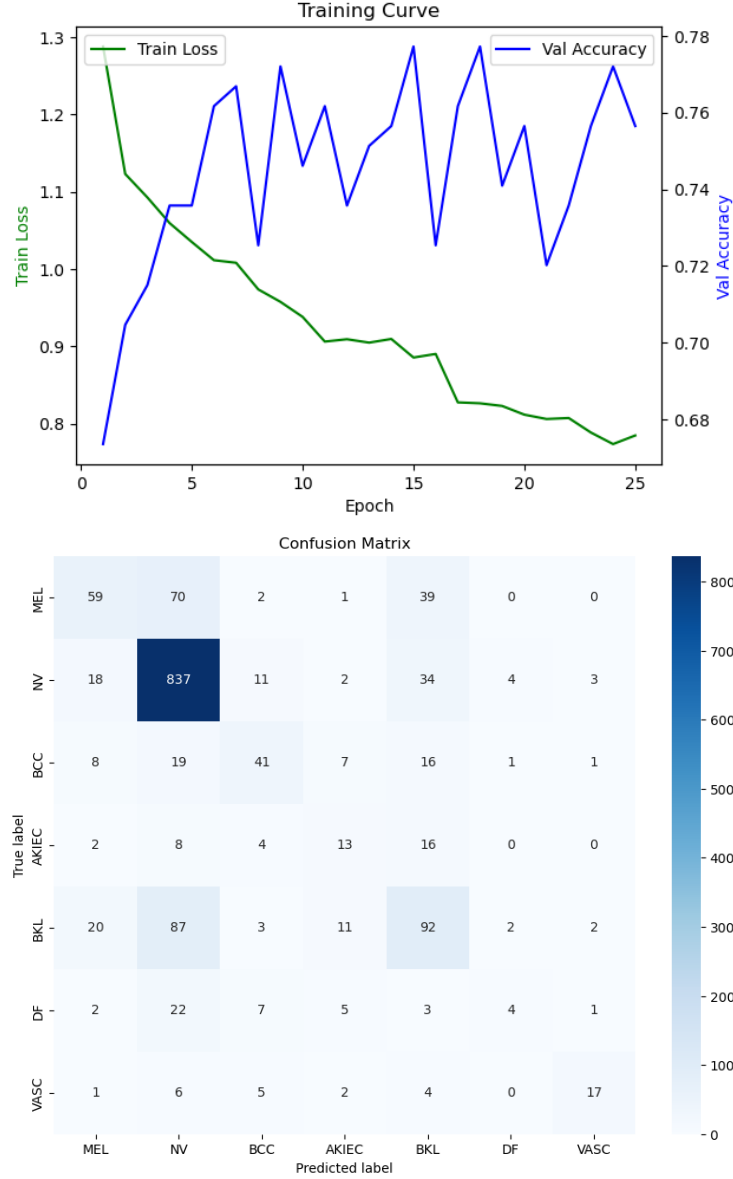


Figure 8: Training Loss, Validation Accuracy and Confusion Matrix

### 4.3.1 Comparative Performance Analysis

Table 1 presents the comprehensive comparison between our FixMatch++ method and all baseline approaches. The results demonstrate significant improvements across both validation and test sets, with our method achieving state-of-the-art performance in the semi-supervised regime.

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
<b>FixMatch++ (ours)</b>	<b>75.1</b>	<b>70.3</b>

Table 1: Comparison with baselines

## Key Performance Insights:

1. **Self-supervised pretraining superiority:** SimCLR + Finetune achieves the highest performance among baselines (64.8% test accuracy), demonstrating the critical importance of leveraging unlabeled data for domain-specific representation learning.
2. **Consistency regularization effectiveness:** Both Mean Teacher (+1.4%) and FixMatch (+3.4%) show substantial improvements over the supervised baseline, with FixMatch outperforming due to its sophisticated pseudo-labeling mechanism.
3. **Progressive improvement pattern:** Results show a clear progression from basic supervised learning to more sophisticated semi-supervised methods, with each approach building upon previous insights.
4. **Synergistic approach advantage:** Our FixMatch++ (+10.7% improvement) successfully combines the strengths of all baseline approaches, resulting in superior performance that exceeds the sum of individual contributions.

### 4.3.2 Detailed Per-Class Analysis

Table 2 provides comprehensive per-class performance metrics, revealing the model’s behavior across different lesion types and highlighting the challenges posed by class imbalance.

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
<b>Accuracy</b>			<b>0.70</b>	<b>1512</b>
<b>Macro Avg</b>	<b>0.53</b>	<b>0.43</b>	<b>0.46</b>	<b>1512</b>
<b>Weighted Avg</b>	<b>0.68</b>	<b>0.70</b>	<b>0.68</b>	<b>1512</b>

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

### Class-Specific Performance Analysis:

- **Dominant Classes:** NV (Nevus) achieves excellent performance (F1=0.85) due to abundant training samples (909 test samples), demonstrating the model’s capability when sufficient data is available.
- **Minority Classes:** Rare classes like DF (Dermatofibroma, F1=0.15) and AKIEC (Actinic Keratoses, F1=0.31) show lower performance due to extreme class imbalance, with fewer than 50 test samples each.
- **Clinical Significance:** MEL (Melanoma) achieves moderate performance (F1=0.42) with balanced precision-recall trade-off, crucial for clinical applications where both false positives and false negatives carry significant costs.
- **Class Imbalance Impact:** The substantial difference between macro average (0.46) and weighted average (0.68) F1-scores highlights the severe class imbalance challenge in dermo-scopic image classification.

## 4.4 Ablation Study

To understand the contribution of individual components in our FixMatch++ framework, we conduct a comprehensive ablation study by systematically removing key components and measuring the resulting performance degradation.

**Component-wise Impact Analysis:** Removing components degrades performance by:

- **–2.5% without SimCLR pretraining:** Demonstrates the critical importance of domain-specific representation learning for bridging the gap between natural images and dermoscopic data.
- **–1.8% without pseudo MixUp:** Shows the significant contribution of advanced data augmentation techniques in improving model robustness and handling class imbalance.
- **–1.2% without EMA teacher:** Confirms the value of stable target generation through exponential moving average updates.
- **–1.0% without OneCycleLR:** Indicates the importance of sophisticated learning rate scheduling for optimal convergence.

## 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.