

Dual-Head Based Contrastive Self-Supervised Learning for Imbalanced Acne Severity Classification

Suhwan Yun*, Semin Kim**

*Student, Dept. of Software Convergence, Gyeongsang National University, Andong, Korea

**Professor, Dept. of Software Convergence, Gyeongsang National University, Andong, Korea

[Abstract]

In medical imaging, acquiring large-scale and balanced datasets is challenging due to privacy constraints and high annotation costs, and acne severity data are typically small-scale and highly imbalanced, making accurate classification of minority severe grades particularly important. This study proposes a dual-head self-supervised learning framework tailored for small-scale, imbalanced settings. To integrate the strengths of global alignment in SimCLR and ranking optimization in AUC-CL, two projection heads are constructed in parallel on a shared encoder, where InfoNCE loss and hard negative mining-based AUC ranking loss are respectively applied. Experimental results on ACNE04 demonstrate that the proposed method outperforms conventional supervised and existing self-supervised approaches overall, with notable improvements in F1-score and significant gains in classification performance for minority severe grades.

▶ **Key words:** ACNE04, Self-Supervised Learning, Dual-head Architecture, Imbalanced Classification, Contrastive Learning

[요약]

의료 영상 분야에서는 개인정보 보호와 높은 레이블 비용으로 인해 대규모·균형 데이터 확보가 어렵고, 여드름 중증도 데이터는 소규모이면서 클래스 불균형이 심해 소수 중증 등급의 정확한 분류가 중요하다. 이에 본 연구는 소규모 불균형 환경에 적합한 Dual-head 자가학습 프레임워크를 제안한다. SimCLR의 전역 정렬과 AUC-CL의 순위 최적화 장점을 통합하기 위해 공유 인코더 위에 두 개의 projection head를 병렬로 구성하고, 각각 InfoNCE loss와 hard negative mining 기반 AUC ranking loss를 적용하였다. ACNE04 데이터셋 실험 결과, 제안 방법은 기존 지도학습 및 자가학습 기법 대비 전반적으로 우수한 성능을 보였으며, 특히 F1-score와 소수 중증 등급 분류 성능이 유의미하게 향상되었다.

▶ **주제어:** ACNE04, 자가학습, 이중 헤드 구조, 불균형 분류, 대조 학습

• First Author: Suhwan Yun, Corresponding Author: Semin Kim
*Suhwan Yun (ghksl0414@student.anu.ac.kr), Dept. of Software Convergence, Gyeongsang National University
**Semin Kim (smkim@gknu.ac.kr), Dept. of Software Convergence, Gyeongsang National University
• Received: 2026. 03. 10, Revised: 2026. 04. 19, Accepted: 2026. 04. 20.

I. Introduction

여드름은 일반적으로 청소년뿐만 아니라 일반 성인들에게도 흔히 발생하는 피부 질환이다. 그러나 여드름을 치료하지 않고 방치하다보면 피부 통증이 생기거나 흉터가 생기기 때문에 검진 및 치료가 중요하다[1]. 여드름은 조기에 검진하고 치료하면 효과가 좋으나, 전문의의 진단을 받기 위해 병원을 방문해야 하기 때문에 접근성이 떨어진다. 이를 극복하기 위하여 최근에 얼굴 이미지로부터 여드름 중증도를 판단하는 다양한 연구가 진행되고 있다[2].

그러나 이러한 여드름 데이터는 의료정보와 관련된 개인정보이기 때문에 데이터의 확보가 어렵다. 또한 데이터를 확보하더라도 여드름 중증도에서 심각한 경우가 적고 대부분 낮은 경우가 많기 때문에 클래스 간 편차가 클 수밖에 없다[3]. 이러한 이유로 충분한 레이블 데이터셋을 확보하기 어려워 지도 학습 기반의 모델이 한계를 보이는 경우가 많다. 레이블링이 부족한 상황에서 유용한 특징을 학습할 수 있는 접근법으로 자가학습(Self-Supervised Learning, SSL)[4]이 다양한 비전 분야에서 대안으로 주목받고 있다.

자가학습에서 대표적인 방법으로는 SimCLR(A Simple Framework for Contrastive Learning of Visual Representations)[5] 방법이 있다. SimCLR은 대조 학습[6]을 통해 강력한 표현을 학습하지만 대규모 배치에 의존적인 단점이 있다. 그리고 또 다른 방법인 AUC-CL[7]은 AUC 최대화[8]를 통해 작은 배치에서도 안정적인 성능을 보였다. 하지만 이러한 방법들은 Positive-Negative 쌍 간 유사도 최적화에 초점을 두어 전반적인 표현 구조를 형성하지만, 클래스 간 경계 정보를 직접적으로 반영하지 못하였다. 특히 소규모 및 불균형 데이터 환경에서는 인접 클래스 간 표현 중첩이 심화되어 경계에 위치한 클래스의 분리도가 충분히 확보되지 못하였다[9]. AUC 기반 학습은 불균형 데이터 환경에서 순위 관계를 직접 최적화함으로써 이러한 경계 샘플에 대한 구분 능력을 향상시켰지만, Hard Negative 선택에 대한 민감성, 전역 표현 구조의 안정성 저하 가능성이라는 한계 또한 존재한다[8][10].

따라서 본 연구는 단순한 손실 함수 결합이 아닌, 불균형 데이터 환경에서 특징 표현 학습과 클래스 간 분리 최적화를 동시에 수행하기 위한 Dual-head 자가학습 프레임워크를 제안한다. 제안 방법은 공유 인코더 위에 두 개의 projection head를 병렬로 구성하여, 하나는 InfoNCE loss를 통해 전체적인 특징 표현을 학습하고, 다른 하나는 AUC ranking loss를 통해 소수 클래스에 민감한 분류 경

계를 학습하도록 설계하였다. 이를 통해 기존 자가학습 기반 표현 학습의 안정성과 AUC 기반 학습의 불균형 대응 능력을 동시에 확보할 수 있도록 하였다. 또한 두 학습 방식의 특성을 이론적, 실험적으로 분석하고 불균형 데이터 환경에서 발생하는 구조적 한계를 체계적으로 규명하였다. 이를 바탕으로 표현 정렬과 순위 최적화를 통합한 새로운 결합 손실 함수를 설계하였으며, ACNE04 데이터셋을 통해 제안한 손실 함수가 실제 소규모 불균형 의료 이미지 환경에서 효과적으로 작동함을 입증하였다.

본 논문의 주요 기여는 다음과 같다.

- 클래스 불균형 의료 영상 데이터 환경에서 기존 자가학습(SSL) 모델의 한계를 분석하였다. SimCLR의 배치 크기 의존성과 AUC 기반 학습의 표현 학습 한계를 체계적으로 고찰하여 문제 해결의 근거를 제시하였다.
- 본 연구는 기존 Dual-head 기반 자가학습 구조를 단순 적용하는 것이 아니라, 불균형 데이터 환경에 최적화하기 위하여 InfoNCE와 AUC ranking loss를 각각 분리된 head에 적용한 목적 지향적 Dual-head 최적화 구조를 제안하였다.
- 소수 클래스(minority class) 판별 성능의 실질적 향상을 입증하였다. ACNE04 데이터셋을 활용한 실험에서 제안 방법이 기존 단일 SSL 방식 대비 F1-score와 Recall을 유의미하게 개선하여, 데이터 불균형 도메인에서의 실용적 활용 가능성을 확인하였다.

각 장의 구성은 다음과 같다. II장에서는 자가학습 및 AUC 최적화와 관련된 기존 연구를 고찰한다. III장에서는 제안하는 Dual-head 프레임워크의 구조와 손실 함수 설계를 상세히 기술한다. IV장에서는 실험 환경과 데이터셋을 설명하고, 기존 방법과의 성능 비교 및 분석 결과를 제시한다. 마지막으로 V장에서는 결론과 한계, 향후 연구 방향을 논의한다.

II. Related works

2.1 Self-Supervised Contrastive Learning

자가학습(Self-Supervised Learning, SSL)은 레이블 없이 데이터 자체의 구조를 활용하여 표현을 학습하는 방법으로, 최근 다양한 비전 과제에서 효과를 보이고 있다. 대표적으로 대조학습(Contrastive Learning) 기반 방법과 비대조학습(Non-Contrastive Learning) 기반 방법으로 구분할 수 있다.

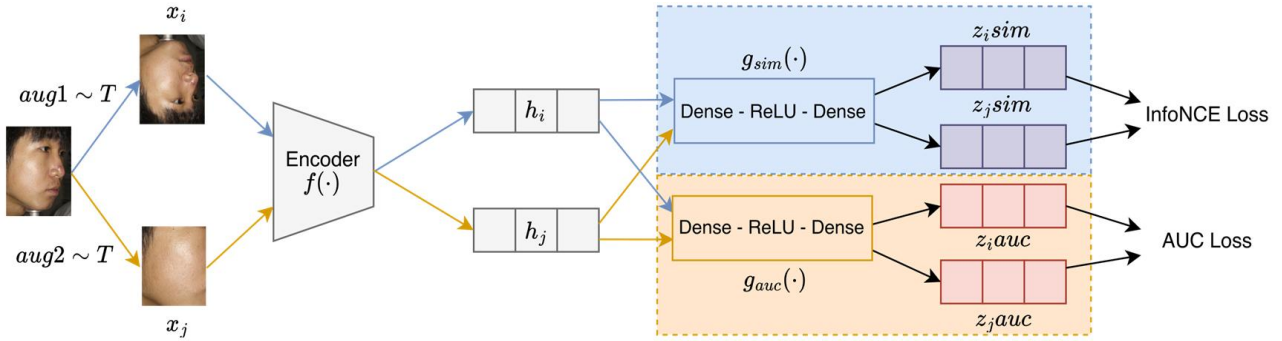


Fig. 1. Dual-head contrastive Framework

Contrastive Learning 기반 방법으로는 SimCLR과 MoCo(Momentum Contrast)[11]가 대표적이다. SimCLR은 동일 이미지의 두 augmentation을 positive pair로 정의하고, 나머지 샘플을 negative로 간주하여 InfoNCE loss[12]를 통해 표현을 학습한다. InfoNCE는 positive 쌍의 유사도를 높이고 negative와의 유사도를 낮추는 방식으로 동작하며, 대규모 배치 환경과 긴 학습에서 강력한 표현을 학습하는 것으로 보고되었다. 또한 MoCo는 Momentum Encoder와 Memory Queue를 활용하여 대규모 Negative Sample을 효율적으로 구성함으로써 안정적인 학습을 가능하게 하였으며, 배치 크기에 대한 의존성을 완화하였다. 그러나 이러한 방법들은 여전히 Negative Sample 구성에 의존하며, 학습 안정성과 계산 비용 측면에서 한계를 가진다.

이후 Negative Sample 없이 학습하는 Non-Contrastive 방법으로 BYOL(Bootstrap Your Own Latent)[13]과 DINO(Distillation with No Labels)[14]가 제안되었다. BYOL은 Teacher-Student 구조를 통해 Representation Collapse를 방지하며 안정적인 표현 학습을 수행하고, DINO는 Self-distillation 기반으로 Teacher Network의 출력을 활용하여 의미론적 특징을 효과적으로 학습한다.

이러한 방법들은 강력한 표현 학습 성능을 보이지만, 주로 전역적인 Representation Alignment에 초점을 두고 있어 클래스 간 경계 정보나 불균형 데이터 환경에서의 minority class 분리 문제를 직접적으로 고려하지 않는 한계를 가진다. 이는 소수 클래스 샘플이 결정 경계 근처에 집중되는 여드름 중증도 분류와 같은 불균형 환경에서 성능 저하로 이어질 수 있다.

2.2 AUC Contrastive Learning

AUC[12](Area Under the ROC Curve)는 임의의 Positive 샘플이 Negative 샘플보다 높은 점수를 가질 확률로 해석되며, 분류기의 순위 능력을 평가하는 지표이다. AUC Maximization은 점수의 절대값이 아닌 Positive와

Negative 간의 순위 관계를 직접 최적화하는 접근으로, 일반적으로 Pairwise Ranking Loss 형태로 구현된다[15].

AUC-CL[7]은 이러한 AUC 기반 순위 최적화를 대조 학습에 결합한 방법으로 소규모 배치 및 불균형 데이터 환경에서 안정적인 성능을 보인다는 점에서 주목받았다. 순위 기반 학습은 경계에 위치한 샘플에 대해 더 강한 구분 신호를 제공할 수 있으며, 다수 클래스에 의해 지배되는 현상을 완화할 가능성이 있다.

Ranking 기반 최적화는 Positive와 Negative 간의 순위 관계를 직접적으로 학습하는 방식으로, 표현간 상대적 차이를 강조한다. 그러나 Softmax 기반 대조 학습과 달리 전체 배치에 대한 확률적 정규화를 수행하지 않으므로 [16], 표현 공간의 전역 분포를 명시적으로 정렬하지 못해 전역 표현 구조의 안정성이 상대적으로 약화될 가능성이 존재한다[17][18].

III. The Proposed Method

3.1 Method Framework

두 개의 서로 다른 데이터 증강 $aug1 \sim T$, $aug2 \sim T$ 를 동일한 증강 분포로 샘플링하여 하나의 입력 이미지에 적용함으로써 서로 상관된 두 개의 뷰를 생성한다(Figure 1). T는 Data Augmentation 종류를 의미하며, 본 연구에서는 RandomResizedCrop, HorizontalFlip, ColorJitter, ToGray, GaussianBlur, Solarize, Normalize를 포함하는 조합으로 구성하였다. $aug1$, $aug2$ 는 해당 분포로부터 독립적으로 Random Sampling된 데이터 증강 기법을 의미한다. 공유 인코더 $f(\cdot)$ 는 각 증강 이미지를 표현 공간으로 매핑하여, 생성된 특징은 두 개의 projection head $g_{sim}(\cdot)$ 와 $g_{auc}(\cdot)$ 로 전달된다. sim head는 대조 손실을 통해 표현 간의 일치도를 최대화하고, AUC head는 순위 기반 최적화를 통해 양성-음성 쌍 간의 분리도를 향상시킨

다. 두 목적 함수를 결합하여 인코더를 자기학습 방식으로 사전 학습한다. 사전 학습이 완료된 후에는 projection head를 제거하고, 학습된 인코더 $f(\cdot)$ 와 representation h 를 Downstream Task를 위한 초기화 가중치로 사용한다. 클래스 불균형 환경에서의 표현 학습을 개선하기 위해 Self-Supervised Contrastive Learning 기반의 전역 표현 정렬 능력과 AUC 기반 Ranking 최적화의 경계 분리 특성을 결합한 Dual-Head Contrastive Framework를 제안한다. pretraining 단계에서는 동일 이미지로부터 생성된 두 개의 Augmentation을 입력으로 사용하며, 하나의 Shared Encoder를 통해 Feature Representation을 추출한다. 이후 두 개의 projection head를 통해 각각 SimCLR 기반의 InfoNCE loss와 AUC 기반의 Ranking Loss를 동시에 최적화한다. 이후 두 개의 projection head를 통해 각각 다른 Embedding Space로 사상한다.

3.2 Agreement Maximization

sim head에서는 InfoNCE Loss를 사용하여 Positive Pair 간의 표현 일치도를 최대화한다. 각 이미지 x 에 대해 두 개의 Augmentation x_i, x_j 를 생성하고, 공유 Encoder $f(\cdot)$ 와 projection head $g_{sim}(\cdot)$ 를 통과시켜 Representation z_i^{sim}, z_j^{sim} 를 얻는다. InfoNCE Loss는 다음과 수식 (1)과 같이 정의된다.

$$L_{infoNCE} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(z_i^{(1)\top} z_i^{(2)}/\tau)}{\sum_{j=1}^N \exp(z_i^{(1)\top} z_j^{(2)}/\tau)} \quad (1)$$

여기서 $z_i^{(1)}, z_i^{(2)}$ 는 동일한 원본 이미지에 서로 다른 Augmentation을 적용하여 얻은 두 표현이며, τ 는 Temperature 하이퍼파라미터이다. 분자 항은 Positive Pair 간의 유사도를 최대화하며, 분모 항은 배치 내 모든 샘플과의 유사도를 정규화하여 Negative 샘플과의 대비를 형성한다. 이를 통해 Representation 공간에서 Positive Pair의 정렬을 강화하고, Negative Pair와의 분리를 유도한다.

3.3 AUC Maximization

AUC head에서는 Positive Similarity와 Hard Negative Similarity[19]를 기반으로 Pairwise Ranking Loss를 적용한다. 각 Anchor에 대해, Positive Similarity는 $s_i^+ = sim(z_i^{auc}, z_i^{+auc})$ 이며, i^+ 는 동일 이미지의 대응 Augmentation Index를 의미한다.

Hard Negatives는 다음 수식 (2)와 같이 정의된다.

$$N_i = \text{top-k} \{ \text{sim}(z_i^{auc}, z_j^{auc}) \mid j \neq i, j \neq j^+ \} \quad (2)$$

AUC loss는 다음 수식 (3)과 같이 정의된다.

$$L_{AUC} = \frac{1}{B} \sum_{i=1}^B [(s_i^+ - c)^2 - 2as_i^+ + \sum_{k=1}^K ((s_{i,k}^- - b)^2 + 2as_{i,k}^-)] - B \quad (3)$$

여기서 s_i^+ 는 Anchor i 의 Positive Similarity, $s_{i,k}^-$ 는 Anchor i 의 Hard Negative k Similarity이다. B 는 batch size, K 는 Hard Negative 개수이며 상수 c 와 학습 가능한 파라미터 b 를 도입하여 양성 및 음성 샘플의 유사도 분포를 특정 영역으로 군집화한다. 동시에 α 를 통해 양성 샘플이 음성 샘플보다 높은 유사도 순위를 갖도록 강제한다.

따라서 자기 자신과 Positive Pair를 제외한 Negative Similarity 중 상위 k 개를 선택한다. 이 때 각 Anchor에 대해 Ranking Loss는 $s_i^+ > s_{i,k}^-$ 가 되도록 최적화된다. 결과적으로 AUC 기반 최적화는 Positive-Negative 간 순위 관계를 직접적으로 조정하며, 어려운 Negative에 대해 더 큰 구분력을 학습하게 되고, 클래스간 분리도를 효과적으로 강화한다.

3.4 Joint Loss

전역적 표현 정렬(Global Alignment)과 순위 최적화를 동시에 활용하기 위해, InfoNCE Loss와 AUC 기반 Ranking Loss를 수식 (4)와 같이 하나의 통합 목적 함수로 결합한다.

$$L_{total} = L_{AUC} + \lambda L_{sim} \quad (4)$$

여기서 $\lambda \in [0,1]$ 는 표현 간 일치도(Agreement)와 순위 기반 판별력(Discrimination) 사이의 균형을 조절하는 하이퍼파라미터이다. 공유 인코더 $f(\cdot)$ 는 총 손실 L_{total} 을 기준으로 업데이트되며, 각 projection head는 자신에게 대응되는 목적 함수에 의해 최적화된다. L_{total} 에서 각 헤드의 기여도를 조절하는 λ 는 0.2를 사용하였다. 불균형한 여드름 데이터셋에서 클래스 간 경계 분리도를 높이는데 있음을 반영한 것이며, 낮은 λ 설정을 통해 L_{auc} 의 순위 최적화 기여도를 높여 소수 클래스 판별력을 극대화 하였다.

IV. Experiments and Results

4.1 Experimental Setup

ACNE04 데이터셋(총 1,457장)을 활용하여 선행 연구[3]를 참조하여 랜덤하게 8:2 비율로 분할하여 1,165장을 학습용, 292장을 테스트용으로 구성하였다. 데이터 분할은 실험

의 재현성을 위해 고정된 랜덤 시드 3407을 사용하여 데이터를 분할하였으며, 초기화에 따른 성능 변동을 고려하기 위해 서로 다른 난수 초기값을 사용한 3회 반복 실험을 수행하였다. 실험은 PyTorch 프레임워크를 기반으로 NVIDIA RTX 4090 GPU 환경에서 수행되었다. ResNet-50[20]을 백본으로 사용하여 초기 가중치는 ImageNet[21]에서 사전 학습된 pretrained weight로 설정하였으며 Early stopping은 적용하지 않았다. 총 200 epochs 동안 128의 배치 사이즈에서 실험을 진행하였다. AUC 최적화의 안정성을 위해 초기 20 epochs 동안 L_{sim} Loss만을 사용하여 Warm-up 단계를 거쳤으며, 이후 Hard Negative Mining(상위 10% 또는 최소 5개)을 포함한 AUC 기반 학습을 병행하였다. 또한 두 손실 함수의 결합 가중치 λ 는 0.2로 설정하였다. projection head는 2-layer MLP 구조로 구성되며, Linear(2048 → 512) - Batch Normalization - ReLU - Linear(512 → 128)로 이루어진다. 제안 방법에서는 InfoNCE Loss와 AUC Ranking Loss를 각각 최적화하기 위해 두 개의 독립적인 projection head를 사용하며, 두 Head는 동일한 구조를 가진다. 최종 임베딩 차원은 128로 설정하였다. Learning rate는 1e-3, Weight decay는 5e-4, InfoNCE Loss 계산 시 Temperature Parameter τ 는 0.5로 설정하였다. 데이터 증강은 Albumentations 라이브러리를 사용하여 RandomResizedCrop(224×224), HorizontalFlip(p=0.5), ColorJitter(p=0.8), Grayscale(p=0.2), GaussianBlur(p=0.5)를 적용하였다.

이후 해당 Pre-trained Weight를 초기값으로 사용한 Fine-tuning의 손실 함수는 Cross-entropy Loss를 적용하여 Epoch은 50, Optimizer는 AdamW[22]를 사용하였고, Cosine Annealing Learning Rate Scheduler[23]를 적용하였다.

4.2 Confusion Matrix by Acne Severity Grade

본 논문에서는 여드름 중증도를 4단계로 구분하여 분류를 수행하였다. Figure 2는 ACNE04 데이터셋의 예시 이미지와 각 여드름 중증도 등급(Grade)을 함께 보여준다.

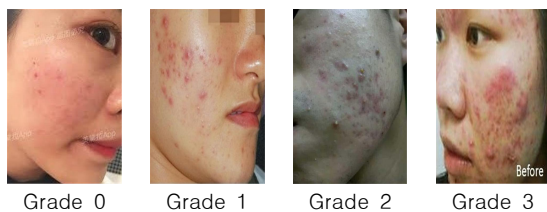


Fig. 2. Examples of ACNE04 Grade (0-3)

Grade 0 ~ 3으로 갈 수록 중증 단계의 피부 상태를 의미하며 Grade 0과 1이 다수를 차지하고, Grade 2와 3은 상대적으로 적은 수를 차지하는 불균형 구조를 가진다.

Table 1-3은 각각 SimCLR, AUC-CL, 제안 방법의 혼동행렬을 나타낸다. 각 표에서 행은 실제 등급(True Label)을 의미하고, 열은 예측한 등급(Predicted Label)을 의미한다.

Table 1의 SimCLR 결과를 보면, Grade 0에서는 95장을 정확히 분류하여 안정적인 성능을 보였다. 그러나 Grade 2에서는 17장이 Grade 1로 오분류되는 경향이 나타났다. 이는 중증도 단계의 구분 경계가 명확히 형성되지 않았음을 의미한다. Table 2의 AUC-CL은 Grade 2에서 27장을 정확히 분류하여 SimCLR 대비 개선된 성능을 보였다. 이는 순위 기반 학습이 소수 클래스 판별에 긍정적인 영향을 미쳤음을 시사한다. 그러나 Grade 3에서는 18장만 정확히 분류되어 중증 단계에서의 성능 저하가 확인되었다. Table 3의 제안 방법은 Grade 2에서 26장을 정확히 분류하여 SimCLR 대비 크게 향상된 결과를 나타내는 동시에 Grade 3에서 22장을 정확히 분류하여 SimCLR과 동일한 성능을 유지하였다. 종합적으로, 제안한 Dual-head 구조는 다수 클래스의 정확도를 유지하면서도 소수 클래스의 판별력을 안정적으로 향상시켰으며, 불균형 데이터 환경에서 전반적인 분류 경계를 개선하는데 효과적임을 확인하였다.

Table 1. Confusion Matrix for SimCLR

		Predicted Label			
		Grade 0	Grade 1	Grade 2	Grade 3
True Label	Grade 0	95	16	0	0
	Grade 1	24	84	9	1
	Grade 2	0	17	17	4
	Grade 3	0	1	2	22

Table 2. Confusion Matrix for AUC-CL

		Predicted Label			
		Grade 0	Grade 1	Grade 2	Grade 3
True Label	Grade 0	92	18	1	0
	Grade 1	17	89	12	0
	Grade 2	0	6	27	5
	Grade 3	0	1	6	18

Table 3. Confusion Matrix for Method

		Predicted Label			
		Grade 0	Grade 1	Grade 2	Grade 3
True Label	Grade 0	92	19	0	0
	Grade 1	24	92	9	1
	Grade 2	0	10	26	2
	Grade 3	0	1	2	22

Table 4. Classification Performance Comparison

Method	Class	Precision	Recall	F1-score
SimCLR	Grade 0	0.8	0.86	0.83
	Grade 1	0.71	0.71	0.71
	Grade 2	0.61	0.45	0.52
	Grade 3	0.81	0.88	0.85
	Avg (Macro)	0.73	0.72	0.72
	Accuracy	-	-	0.75
AUC-CL	Grade 0	0.84	0.83	0.84
	Grade 1	0.78	0.75	0.77
	Grade 2	0.59	0.71	0.64
	Grade 3	0.78	0.72	0.75
	Avg (Macro)	0.75	0.75	0.75
	Accuracy	-	-	0.77
Proposed	Grade 0	0.85	0.83	0.84
	Grade 1	0.75	0.78	0.77
	Grade 2	0.7	0.68	0.69
	Grade 3	0.88	0.88	0.88
	Avg (Macro)	0.8	0.79	0.8
	Accuracy	-	-	0.79

실험 결과 제안 구조는 데이터 불균형 환경에서 가장 우수한 성능을 보임을 확인하였다. Table 4는 각 방법에 대한 클래스별 분류 성능을 Precision, Recall, F1-score 기준으로 비교한 결과를 나타낸다. 전체적으로 제안 방법 (Proposed)은 정확도(Accuracy) 0.79와 Macro F1-score 0.80을 기록하며 모든 비교 방법 중 가장 우수한 성능을 보였다. 기존 방법인 SimCLR은 Grade 0과 Grade 1과 같은 다수 클래스에서는 비교적 안정적인 성능 (F1-score 0.83, 0.71)을 보였으나, 소수 클래스인 Grade 2에서는 F1-score가 0.52로 크게 감소하는 한계를 보였다. AUC-CL은 Grade 2에서 Recall이 0.71로 향상되며 일부 개선된 모습을 보였으나, Grade 3에서 F1-score가 0.75로 감소하는 등 클래스 간 성능 편차가 존재하였다. 반면, 제안 방법은 모든 클래스에서 균형 잡힌 성능을 보이며, 특히 Grade 2에서 F1-score를 0.69까지 향상시켜 기존 방법 대비 가장 큰 개선을 달성하였다. 또한 Grade 3에서도 F1-score 0.88로 가장 높은 성능을 기록하여, 중증 클래스에 대한 판별 능력이 크게 향상되었음을 확인할 수 있다.

이러한 결과는 제안 방법이 다수 클래스의 성능을 유지하면서도, 데이터 불균형 환경에서 소수 클래스의 표현 학습과 분류 경계 형성을 효과적으로 개선했음을 시사한다.

Table 5. Comparison with Baseline Methods for ACNE04

Method	Accuracy	Precision	Recall	F1-score
Supervised	74.54 ± 0.40	0.7241 ± 0.0077	0.7427 ± 0.0220	0.7306 ± 0.0083
SimCLR	75.68 ± 1.03	0.7326 ± 0.0035	0.7457 ± 0.0191	0.7332 ± 0.0073
AUC-CL	77.05 ± 0.59	0.7527 ± 0.0121	0.7652 ± 0.0114	0.7553 ± 0.0073
Proposed	79.68 ± 0.71	0.7841 ± 0.0144	0.7851 ± 0.0236	0.7920 ± 0.0022

4.3 Comparison with Baseline Methods

제안한 모델의 분류 성능을 평가하기 위해 기존의 대조 학습 기반 모델인 SimCLR 및 AUC-CL과의 비교 실험을 수행하였다. 분류 성능을 평가하기 위해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score 총 4개의 지표를 사용하였다.

Accuracy는 정확도를 나타내며 전체 샘플 중 올바르게 분류된 비율을 의미하며, Precision은 모델이 해당 클래스로 예측한 샘플 중 실제로 해당 클래스인 비율을 나타낸다. Recall은 실제 해당 클래스 샘플 중 모델이 정확히 탐지한 비율을 의미하며 F1-score는 Precision과 Recall의 조화 평균으로, 두 지표 간의 균형을 평가하는 데 사용된다.

Table 5는 ACNE04 데이터셋에 대한 분류 성능을 비교한 결과를 나타낸다. ACNE04 데이터셋의 클래스 불균형 및 높은 클래스 내 변동성으로 인해 전반적인 성능은 70% 대에 분포하며, 이는 선행 연구[3]의 재현 실험에서도 동일하게 관찰된다. 본 연구는 절대 수치보다 통제된 동일 환경에서의 상대적 성능 향상폭을 기준으로 제안 방법론의 유효성 검증을 진행하였다. 이러한 조건에서 제안 방법은 모든 지표에서 기존 방법 대비 가장 우수한 성능을 기록하였다. 특히 F1-score의 향상은 불균형 데이터 환경에서 소수 클래스에 대한 분류 안정성이 개선되었음을 의미한다. 이는 제안한 학습 전략이 클래스 불균형 상황에서 정밀도와 재현율 간의 균형을 효과적으로 개선했음을 보여준다.

Table 6. Effect of Hyperparameters λ and k

λ	k (Top-k /Ratio)	(Macro) F1-score
0.1	5 (10%)	0.77
0.2	5 (10%)	0.8
0.5	5 (10%)	0.76
0.2	10 (20%)	0.6

4.4 Ablation Study

λ 와 k 값에 따른 성능 변화를 분석하기 위해 ablation study를 수행하였다. Table 6은 λ 와 k 값 변화에 따른 macro-F1 성능 변화를 보여준다. $\lambda=0.2$, $k=5$ (10%) 설정

에서 가장 높은 macro-F1 (0.80)을 기록하였다. λ 가 0.1에서 0.2로 증가할 때 성능이 향상되었으나, 0.5로 증가할 경우 오히려 성능이 감소하는 경향을 보였다. 이는 적절한 가중치 설정이 모델의 표현 학습에 중요함을 시사한다.

또한, k 값을 5에서 10으로 증가시킬 경우 Macro-F1이 0.80에서 0.60으로 크게 감소하는 것을 확인할 수 있었다. 이는 과도한 샘플 선택이 오히려 노이즈를 증가시켜 성능 저하를 유발할 수 있음을 의미한다. 따라서, 적절한 비율의 샘플을 선택하는 것이 중요하며, 본 연구에서는 $k=5$ (10%)를 최적 설정으로 사용하였다.

4.5 Applicability

본 연구에서 제안하는 프레임워크의 일반화 가능성을 검증하고, 데이터 불균형 환경에서의 성능을 평가하기 위해 의료 데이터셋인 CheXpert를 이용한 추가 실험을 수행하였다.

이를 위해 CheXpert 데이터셋[24]에서 총 1,500장의 이미지를 추출하여 'No Finding' 라벨을 기준으로 정상(0)과 비정상(1)으로 재구성하여 이진 분류 태스크를 설정하였다. 모델의 테스트를 위해 정상 데이터의 비율을 각각 5%와 10%로 설정하여 극심한 불균형 환경을 조성하였다.

Table 7은 전체 데이터 중 비정상(Abnormal) 샘플의 비율을 각각 5%와 10%로 설정한 극심한 불균형 환경에서 각 모델의 성능을 비교한 결과이다. 실험 결과 제안 방법이 5%, 10%의 모든 비율에서 기존 베이스라인 모델들을 일관되게 상회하는 성능을 보였다. 불균형 데이터에서 Recall과 F1-score의 유의미한 상승은 제안하는 구조가 다수 클래스에 편향되지 않고 정밀한 분류 경계를 형성하고 있음을 시사한다.

이러한 결과는 기존 의료 데이터의 불균형 문제를 해결하는데 있어 제안하는 구조가 실용적인 대안이 될 수 있음을 보여준다.

V. Discussion

본 연구에서 사용한 ACNE04 데이터셋은 총 1,457장의 제한된 샘플로 구성되어 있으며, Grade 2 및 Grade 3의 비율이 상대적으로 낮은 불균형 구조를 가진다. 또한 여드름 병변은 클래스 간 시각적 차이가 미세하고 동일 클래스 내 병변 형태의 다양성이 커 분류 난도가 높은 특성을 가진다. 이러한 데이터셋 특성으로 인해 전체 분류 성능이 70% 후반 수준으로 형성되며, 이는 선행 연구[3]의 재현 실험에서도 유사하게 관찰된다. 특히 원 논문에서는 무작위 데이터 분할(Random Split)을 사용하여 실험을 수행하였기 때문에 동일한 학습 및 평가 환경을 완전히 재현하는데 한계가 존재하며, 데이터 분할 방식에 따라 성능 편차가 발생할 수 있다. 따라서 본 연구에서는 절대적인 성능 수치보다 동일 통제 환경 내에서의 상대적 성능 향상을 중심으로 제안 방법의 효과를 평가하였다.

또한 본 연구에서는 CheXpert 데이터셋 기반 추가 실험을 통해 제안 방법의 일반화 가능성을 보완적으로 검증하였으나, 다양한 도메인 및 대규모 데이터셋에 대한 확장 검증은 아직 수행되지 못하였다. 따라서 향후에는 보다 다양한 의료 및 일반 영상 데이터셋에 대한 추가 실험을 통해 제안 방법의 일반화 성능을 폭넓게 검증할 필요가 있다.

VI. Conclusions

본 연구에서는 대조 학습 기반 자가학습 방법인 SimCLR과 AUC 최적화 기반 학습 기법의 특성을 분석하고, 소규모 불균형 데이터 환경에서 발생하는 한계를 고찰하였다. 특히 여드름 데이터셋인 ACNE04는 클래스 간 불균형이 심하고 병변 간 시각적 차이가 미세하여, 기존의 Cross-entropy 기반 학습이나 단일 목적 함수 기반 표현 학습으로는 소수 클래스에 대한 구분 성능이 저하되는 문제가 발생한다. 따라서 기존 Contrastive Learning 목적

Table 7. Performance Comparison Under 5% and 10% Imbalance Ratios

Ratio	Method	Accuracy	Precision	Recall	F1-score
5%	Supervised	67.24 ± 3.32	0.6450 ± 0.0119	0.7604 ± 0.0209	0.6122 ± 0.0306
	SimCLR	74.64 ± 1.50	0.6599 ± 0.0010	0.7888 ± 0.0018	0.6478 ± 0.0041
	AUC-CL	75.07 ± 0.89	0.6394 ± 0.0036	0.7513 ± 0.0095	0.6383 ± 0.0181
	Proposed	76.35 ± 0.65	0.6830 ± 0.0030	0.8093 ± 0.0123	0.6932 ± 0.0010
10%	Supervised	80.34 ± 1.32	0.6531 ± 0.0091	0.7594 ± 0.0154	0.6692 ± 0.0098
	SimCLR	84.47 ± 0.25	0.7103 ± 0.0041	0.8156 ± 0.0029	0.7232 ± 0.0067
	AUC-CL	85.15 ± 0.41	0.7293 ± 0.0057	0.7822 ± 0.0311	0.7336 ± 0.0097
	Proposed	87.04 ± 0.25	0.7170 ± 0.0068	0.8241 ± 0.0096	0.7426 ± 0.0077

함수가 대규모 배치 크기에 의존한다는 점과, AUC 기반 학습이 전역적 표현 정렬 측면에서 제한적일 수 있다는 점에 주목하였다. 이를 바탕으로 공유 encoder 위에 두 개의 projection head를 구성한 Dual-head 구조를 설계하고, InfoNCE Loss와 AUC 기반 Ranking Loss를 동시에 최적화하는 통합 학습 목적 함수를 제안하였다. 실험 결과, 제안 방법은 기존 방법 대비 소수 클래스에 대한 분별력 향상 및 전체 분류 성능 개선을 보였으며, 특히 불균형 환경에서의 안정적인 성능을 확인하였다. 다만 실험이 ACNE04 데이터셋에 국한되어 수행되어 데이터셋의 불균형 정도나 특성에 따라 파라미터를 최적화하는 기법에 대한 연구가 부족하다. 향후 데이터셋의 불균형 정도에 따라 λ 를 스스로 조절하는 파라미터 최적화 기법에 대한 연구가 추가로 필요하다.

ACKNOWLEDGEMENT

Following are results of a study on the "Glocal University 30" Project, supported by the Ministry of Education and Gyeongsangbuk-do Province

REFERENCES

- [1] Liu, Y. T., Wang, Y. W., Tu, C., Ren, J. W., Huo, J., Nan, X. J., Dou, J. H., Peng, Z. H., & Zeng, W. H., "Recommendations for managing adult acne and adolescent acne based on an epidemiological study conducted in China," *Scientific Reports*, vol. 14, Art. no. 16327, 2024. DOI: 10.1038/s41598-024-67215-2
- [2] Kim, S., Lee, C., Jung, G., Yoon, H., Lee, J., & Yoo, S., "Facial Acne Segmentation based on Deep Learning with Center Point Loss," 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), pp. 678-683, L'Aquila, Italy, 2023. DOI: 10.1109/CBMS58004.2023.00299
- [3] Wu, X., Wen, N., Liang, J., Lai, Y.-K., She, D., Cheng, M.-M., & Yang, J., "Joint Acne Image Grading and Counting via Label Distribution Learning," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10642-10651, 2019.
- [4] Dosovitskiy, A., Springenberg, J. T., Riedmiller, M., & Brox, T., "Discriminative unsupervised feature learning with convolutional neural networks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014. DOI: 10.48550/arXiv.1406.6909
- [5] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G., "A Simple Framework for Contrastive Learning of Visual Representations," *Proceedings of the 37th International Conference on Machine Learning (ICML)*, in *Proceedings of Machine Learning Research*, vol. 119, pp. 1597-1607, 2020. DOI: 10.48550/arXiv.2002.05709
- [6] Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F., "A Survey on Contrastive Self-Supervised Learning," *Technologies*, vol. 9, no. 1, Art. no. 2, 2021. DOI: 10.3390/technologies9010002
- [7] Sharma, R., Ji, K., Xu, Z., & Chen, C., "AUC-CL: A Batchsize-Robust Framework for Self-Supervised Contrastive Representation Learning," *The Twelfth International Conference on Learning Representations (ICLR)*, 2024. URL: <https://openreview.net/forum?id=YgMdQB09U>
- [8] Yang, T., & Ying, Y., "AUC maximization in the era of big data and AI: A survey," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1-37, 2022. DOI: 10.1145/3554729
- [9] Wang, T., & Isola, P., "Understanding contrastive representation learning through alignment and uniformity on the hypersphere," *Proceedings of the 37th International Conference on Machine Learning (ICML)*, in *Proceedings of Machine Learning Research*, vol. 119, pp. 9929-9939, 2020. DOI: 10.48550/arXiv.2005.10242
- [10] Lu, X., Liu, X., Xiao, Z., Zhang, S., Huang, J., Yang, C., & Liu, S., "Self-supervised dual-head attentional bootstrap learning network for prostate cancer screening in transrectal ultrasound images," *Computers in Biology and Medicine*, vol. 165, Art. no. 107337, 2023. DOI: 10.1016/j.combiomed.2023.107337
- [11] Van den Oord, A., Li, Y., & Vinyals, O., "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018. DOI: 10.48550/arXiv.1807.03748
- [12] Fawcett, T., "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006. DOI: 10.1016/j.patrec.2005.10.010
- [13] Grill, J. B., Strub, F., Althé, F., Tallec, C., Richemond, P., Buchatskaya, E., ... & Valko, M., "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 21271-21284, 2020. DOI: 10.48550/arXiv.2006.07733
- [14] Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., et al., "DINO: DETR with improved denoising anchor boxes for end-to-end object detection," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2023. DOI:10.48550/arXiv.2203.03605
- [15] Gao, W., & Zhou, Z.-H., "On the consistency of AUC pairwise optimization," *arXiv preprint arXiv:1208.0645*, 2012. DOI: 10.48550/arXiv.1208.0645
- [16] Li, D., Jin, R., & Ren, B., "Revisiting recommendation loss functions through contrastive learning (technical report)," *arXiv preprint arXiv:2312.08520*, 2023. DOI: 10.48550/arXiv.2312.08520
- [17] Jiang, Z., et al., "Improving contrastive learning on imbalanced data via open-world sampling," *Advances in Neural Information*

- Processing Systems (NeurIPS), vol. 34, pp. 5997–6009, 2021. DOI: 10.48550/arXiv.2111.01004
- [18] Roh, S., et al., "Equilibrium contrastive learning for imbalanced image classification," arXiv preprint arXiv:2602.09506, 2026. DOI: 10.48550/arXiv.2602.09506
- [19] Robinson, J., Chuang, C. Y., Sra, S., & Jegelka, S., "Contrastive learning with hard negative samples," arXiv preprint arXiv:2010.04592, 2020. DOI: 10.48550/arXiv.2010.04592
- [20] He, K., Zhang, X., Ren, S., & Sun, J., "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 770–778, 2016. DOI: 10.48550/arXiv.1512.03385
- [21] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L., "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255, Miami, FL, USA, June 2009. DOI: 10.1109/CVPR.2009.5206848
- [22] Loshchilov, I., & Hutter, F., "Decoupled Weight Decay Regularization," International Conference on Learning Representations (ICLR), 2019. URL: <https://openreview.net/forum?id=Bkg6RiCqY7>
- [23] Loshchilov, I., & Hutter, F., "SGDR: Stochastic Gradient Descent with Warm Restarts," International Conference on Learning Representations (ICLR), 2017. URL: <https://openreview.net/forum?id=Skq89Scxx>
- [24] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 1, pp. 590–597, 2019. DOI: 10.1609/aaai.v33i01.3301590

Authors



Suhwan Yun is currently pursuing the B.S. degree in Software Convergence at Gyeongbuk National University, Andong, Korea, since 2020. His research interests include computer vision, self-supervised learning.



Semin Kim received his B.S. degree in Multimedia Engineering from Andong National University in 2006 and his Ph.D. degree in Information and Communications Engineering from the Korea Advanced

Institute of Science and Technology (KAIST) in 2014. He has been serving as an Assistant Professor in the Software Convergence major at Gyeongbuk National University since 2025. He worked at the Korea Institute for Advancement of Technology (KIAT) in 2008 and at Samsung Electronics in 2014. He also worked at Hyundai Mobis in 2019 and at Lululab in 2021. His research interests include vision-based image recognition and large language models (LLMs).