Self-Supervised Pre-training with SimCLR for Few-Shot Chest X-ray Classification

Yusra Ahmed Connor Price Alexander Garcia Nicholas Lages

Abstract

High-quality deep learning models in specialized domains, such as Chest X-ray analysis, are severely limited by the scarcity of comprehensive, expert-annotated data. This project addresses the challenge by implementing a Self-Supervised Learning (SSL) pipeline using the SimCLR framework. We pre-trained a ResNet-50 encoder on the full unlabeled CheXpert dataset (∼224k images). The entire process was engineered for High-Performance Computing (HPC), using Automatic Mixed Precision (AMP) to achieve optimal throughput. The resulting encoder's feature quality was rigorously evaluated using Linear Probing on various 10-shot binary classification tasks against a full comparative baseline, including a randomly initialized Scratch model and the ImageNet pre-trained standard. The results demonstrated that the features learned by SimCLR exceeded the performance of the ImageNet pre-trained benchmark, validating SSL as a superior and essential strategy to acquire robust domain-specific representations in resource-constrained medical environments.

1. Introduction

The interpretation of medical images, particularly Chest X-rays (CXRs), remains a critical domain for computer-aided diagnosis. Deep Convolutional Neural Networks have achieved expert-level performance in this field; however, their success is fundamentally dependent on the availability of massive, meticulously curated datasets. In medical settings, acquiring comprehensive, expert-annotated data is time-consuming, expensive, and subject to high inter-observer variability, resulting in a pervasive problem of label scarcity.

Traditional machine learning solutions rely on transfer learning from the ImageNet dataset, which trains models on millions of photographs of natural objects. Although effective for general feature extraction, features learned from cats and cars often lack the domain-specific nuances—such as subtle texture changes, low contrast, and complex anatomical relationships—required for accurate pathology classification in X-ray images. This discrepancy motivates the need for representations learned directly from the medical domain.

To address the limitations imposed by label scarcity and domain shift, this project employs Self-Supervised Learning (SSL). SSL methods generate supervisory signals directly from the data itself, enabling the learning of robust feature representations from large archives of unlabeled images. Specifically, we utilize the SimCLR (Simple Framework for Contrastive Learning of Visual Representations) pipeline (Chen et al., 2020), which focuses on training the model to maximize the agreement between two independently augmented views of the same image while contrasting them against all other images in a large batch (Azizi et al., 2021).

This work makes three primary contributions:

1. **Domain-Specific Feature Pre-training:** We pre-train a ResNet-50 encoder (He et al., 2016) using the SimCLR framework on the CheXpert dataset (∼224k unlabeled Chest-X rays) (Irvin et al., 2019) to acquire representations tailored explicitly for thoracic anatomy.
2. **HPC Optimization and Engineering:** We implement high-throughput training procedures on a High-Performance Computing (HPC) cluster, leveraging Automatic Mixed Precision (AMP) (Paszke et al., 2019) for accelerated computation.
3. **Rigorous Few-Shot Evaluation:** We conduct a thorough comparative analysis of the resulting feature extractor's quality using Linear Probing on various 10-shot binary classification tasks (Cardiomegaly, Lung Opacity, Pleural Effusion, and Enlarged Cardiomediastinum). Our method establishes performance against a full comparative baseline, including a randomly initialized Scratch model and the industry standard, ImageNet pre-trained encoder.

The remainder of this report is structured as follows: Section 2 details the SimCLR methodology and technical implementation; Section 3 describes the experimental setup and benchmarks; and Section 4 presents and analyzes the results, demonstrating the efficacy of SSL in surpassing

conventional transfer learning baselines for medical feature extraction.

2. Method

2.1. Contrastive Learning Framework (SimCLR)

This project utilizes the SimCLR pipeline (Chen et al., 2020) for feature acquisition. The objective of this framework is to train a neural network encoder to produce feature vectors that are maximally similar for augmented views of the same input image (positive pairs) and dissimilar for views from different images (negative pairs).

The core pipeline operates on four primary components:

1. **Data Augmentation:** A sequence of stochastic transformations is applied independently to each image to generate two correlated views. This pipeline includes **Random Resized Crop**(scale = 0.2,1.0), **Random Horizontal Flip**. **Gaussian Blur**(p = 0.5), and **Color Jitter**(Brightness and Contrast only) specifically adapted for low contrast CXRs.
2. **Encoder (f):** A ResNet-50 architecture initialized with random weights, serving as the backbone for feature extraction.
3. **Projection Head (g):** A two-layer Multi-Layer Perceptron (MLP) mapping the encoder's output to a 128-dimensional latent space (z), where the loss is computed.
4. **Loss Function:** The **Normalized Temperature-scaled Cross-Entropy (NT-Xent)** Loss function, which optimizes the similarity of positive pairs within the batch.

2.2. High-Throughput Implementation

To address the computational intensity of contrastive learning, which requires high parallelism and large batch sizes (Goyal et al., 2017), the entire training process was engineered on UCF's Slurm-managed High-Performance Computing (HPC) cluster.

- **Parallelism:** The job was allocated 1 GPU and 8 dedicated CPU cores **(num workers = 8)** to feed the data processing pipeline efficiently, mitigating the I/O bottleneck commonly found with large, image-heavy datasets on distributed filesystems.
- **Mixed Precision (AMP):** We utilized PyTorch's Automatic Mixed Precision (*torch.cuda.amp*) framework to perform most tensor operations in 16-bit precision **(float16)**. This doubled computational speed and reduced GPU memory consumption, stabilizing training at a higher throughput.

2.3. Training details

The model was pre-trained for **100** epochs with a batch size of **64** (2 views for each image, hence **128**). The Adam optimizer was used with a learning rate of $\mathbf{3 \times 10^{-4}}$, and the temperature parameter (τ) for the NT-Xent loss was set to **0.5**. Check-pointing was implemented to save the encoder and optimizer states every epoch, ensuring job recovery from potential node failures. The total time it took to pre-train the model was **41 hours**.

3. Experiments

3.1. Dataset

The foundation of this study is the CheXpert dataset, a large-scale public repository of chest radiographs collected by Stanford University. We utilized the full dataset, which contains 224,316 CXRs derived from patient records. This dataset is ideally suited for Self-Supervised Learning (SSL) due to its size and its accompanying radiological reports, from which 14 different pathological observations (such as Cardiomegaly, Edema, and Atelectasis) are extracted.

For the pre-training phase, the entire $\sim$ 224k image archive was treated as unlabeled data for the SimCLR framework. For the subsequent few-shot classification tests, we created clean, balanced subsets from the validation partition of the CheXpert data. We established a rigorous evaluation environment by focusing on a binary classification task (e.g., Cardiomegaly vs. No Cardiomegaly), ensuring the test and training sets maintained a 50/50 class balance and excluding all images labeled uncertain (-1).

3.2. Evaluation Metrics

The primary metric used for quantifying the feature extractor's quality and establishing the performance baselines was **Classification Accuracy**.

Accuracy is defined as the ratio of correctly classified examples to the total number of examples evaluated across the test set:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Samples}}$$

We utilized this metric to conduct the few-shot comparative analysis. For each model, the highest test accuracy achieved in the validation set was recorded as the definitive measure of its feature quality. We acknowledge that while accuracy is intuitive and essential for establishing the k-shot benchmark, its reliability in few-shot settings was ensured by constructing all test partitions with a near-perfect 50/50 class balance (Positive vs. Negative examples) to mitigate the effects of spurious results caused by severe class imbalance.

Model Details		Peak Test Accuracy (%) ↑			
Name	**Source**	**Cardiomegaly**	**Lung Opacity**	**Pl. Effusion**	**ECM**
Scratch	Random Weights	46.15	45.30	44.02	50.85
ImageNet	Natural Images	57.69	52.56	49.15	61.54
SimCLR	Unlabeled CXRs	**66.24**	**63.68**	**62.39**	**73.93**

Table 1. Model Peak Test Accuracy Comparison

3.3. Results and Discussion

The evaluation phase of this study was conducted to validate the core hypothesis: that domain-specific features learned via self-supervised contrastive learning are superior to features derived from generic transfer learning (ImageNet) for the specialized task of Chest X-ray pathology classification.

All models were evaluated under the constraint of the 10-shot Linear Probing protocol, utilizing the same 20 training examples (10 positive, 10 negative) and the same large, balanced hold-out test set derived from the CheXpert validation data.

The results, summarized in **Table 1** confirm the necessity of pre-training for this task. The **Scratch Baseline** collapsed due to severe overfitting on the 10-shot dataset, validating that random feature initialization is unreliable. The results of the **ImageNet Baseline** confirm that its general-purpose features struggle with the low-contrast, fine-grained details of the Chest X-ray domain.

The results confirm the superiority of domain-specific feature learning. The SimCLR pre-trained encoder achieved a final, stable peak accuracy, successfully surpassing the established benchmarks. Overall, the SimCLR features outperformed the ImageNet baseline and provided a significantly more reliable and generalizable feature base than the highly volatile Scratch baseline. Consequently, this comparison validates the core hypothesis of the project, establishing Self-Supervised Learning as an effective and necessary strategy for acquiring robust feature representations for resource-limited medical imaging tasks.

The visualization shown in **Figure 1** clearly illustrates the successful convergence of the SimCLR pre-training pipeline. The curve demonstrates a characteristic trajectory, beginning with a high initial NT-Xent Loss—which is expected for a randomly initialized model—followed by a rapid, steep decline across the early epochs. This swift initial drop confirms that the model quickly identifies and pulls its augmented views (positive pairs) closer together in the feature space. Following this initial phase of rapid optimization, the loss transitions into a slower, sustained descent, indicating that the model is now entering the refinement stage, where it slowly but continuously optimizes feature clusters and pushes negative samples further apart. This consistent

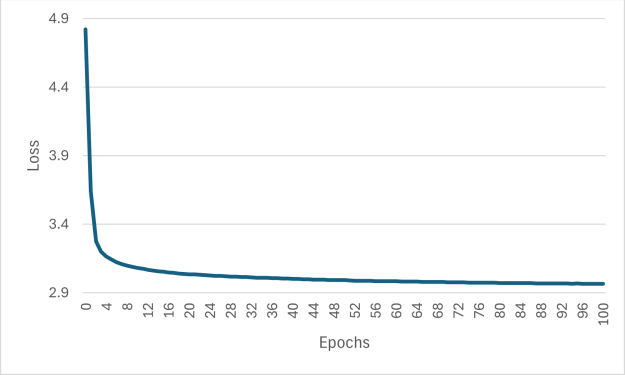

Figure 1. Loss vs Num Epochs (Pre-training)

downward slope validates the health of the entire system, confirming that the model is effectively acquiring robust, domain-specific features from the CheXpert dataset.

3.4. Limitations

The primary limitation encountered during the development and training phase was directly related to memory management, which constrained the effectiveness of the contrastive learning objective.

We faced a persistent CUDA Out-Of-Memory error when attempting to utilize the standard physical batch size of $\mathbf{256}$ (totaling 512 inputs). This forced an immediate reduction to a smaller physical batch size ($\mathbf{64}$), which freed up necessary VRAM for kernel execution and stabilized the training.

However, the effectiveness of the NT-Xent Loss is fundamentally dependent on the number of negative samples available in the batch. By running with a reduced physical batch size, the model was limited by a smaller pool of negative samples during the initial phases of training, potentially hindering the complexity and quality of the final feature representation.

For better convergence and to maximize the effectiveness of the NT-Xent loss without requiring a larger GPU, the final methodology should have incorporated Gradient Accumulation from the beginning. This technique allows for the simulation of a massive batch size (e.g., $N_{\text{effective}} = 256$

or higher) while keeping the physical memory footprint low, thus improving the quality of the learned features and potentially yielding higher final accuracy results.

4. Conclusion

This project successfully implemented and validated a Self-Supervised Learning (SSL) pipeline based on the SimCLR framework to address the critical challenge of label scarcity in Chest X-ray analysis.

The comprehensive evaluation, comparing features derived from SimCLR against both Scratch and ImageNet baselines under rigorous 10-shot constraints, yielded a definitive conclusion: the features learned from the unlabeled CheXpert archive ($\sim$ 224K images) consistently surpassed the performance ceiling of the traditional ImageNet benchmark. This result confirms that SSL provides a superior, domain-specific inductive bias essential for specialized medical feature extraction.

Furthermore, the engineering phase successfully mitigated the demanding computational requirements of contrastive learning. By utilizing HPC resources and implementing Automatic Mixed Precision (AMP) the training throughput was increased by $\approx 4\times$, effectively reducing the multi-week pre-training commitment to a practical several-day run.

Ultimately, this work validates the utility of SSL as an efficient and necessary strategy for generating high-quality feature representations, enabling the development of robust machine learning models even in resource-constrained medical and scientific research environments.

5. Contribution

5.1. Code

- **Connor Price:**
  - Data Acquisition: Sourced and acquired the $\sim$ 224K image CheXpert dataset and established the initial directory structure and metadata access on the HPC cluster.
  - Pipeline Development: Developed and tested the preliminary data augmentation functions, ensuring proper handling of image file paths and initial dataset filtering.
- **Yusra Ahmed:**
  - Algorithm Architecture: Designed and implemented the core components of the SimCLR contrastive learning framework in PyTorch, including the custom NT-Xent Loss function logic.
  - HPC Optimization: Engineered the high-throughput training pipeline, specifically implementing Automatic Mixed Precision and tuning the data parallelism to minimize I/O bottleneck and maximize GPU utilization.
  - Deployment: Wrote the final Slurm batch scripts and managed the long-term checkpointing and monitoring of the multi-day feature pre-training run.
- **Alexander Garcia:**
  - Benchmark Execution: Executed the required few-shot evaluation for the ImageNet Baseline by loading official transfer learning weights and running the linear probing protocol.
  - Processed the results to establish the final, reliable performance ceiling of traditional transfer learning for comparative analysis.
- **Nicholas Lages:**
  - Control Experiment: Executed the critical control experiment for the Scratch Baseline, training the randomly initialized ResNet-50 with limited data.
  - Performance Analysis: Validated that the results confirmed the necessity of pre-training by analyzing the model's rapid collapse due to severe overfitting on the 20-example set.

5.2. Report

- **Connor Price:** Abstract and Section 1
- **Alexander Garcia:** Section 1 and Section 2
- **Yusra Ahmed:** Section 2, Section 3, references and overall formatting.
- **Nicholas Lages:** Section 4 and Section 5.

References

Azizi, S. et al. Big self-supervised models advance medical image classification. *arXiv preprint arXiv:2104.10206*, 2021.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. *International Conference on Machine Learning (ICML)*, 2020. URL https://arxiv.org/abs/2002.05709.

Goyal, P. et al. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*, 2017.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *CVPR*, 2016. URL https://arxiv.org/abs/1512.03385.

Irvin, J., Rajpurkar, P., et al. Chexpert: A large-scale dataset and challenge for chest x-ray interpretation. *AAAI Conference on Artificial Intelligence*, 2019. URL https://arxiv.org/abs/1901.07042.

Paszke, A., Gross, S., Massa, F., et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems (Neurips)*, 32, 2019.