Self-Supervised Learning for Monkeypox Skin Lesion Detection: A Contrastive Approach with SimCLR and MoCo

1st Sanjida Islam Aksa
Computer Science and Engineering
East West University (of Aff.)
Dhaka, Bangladesh
2022-1-60-254@std.ewubd.edu

2nd Israt Jahan
Computer Science and Engineering(of Aff.)
East West University (of Aff.)
Dhaka, Bangladesh
2022-3-60-224@std.ewubd.edu

3rd Mahir Shahriar
Computer Science and Engineering (of Aff.)
East West University(of Aff.)
Dhaka,Bangladesh
2022-3-60-197@std.ewubd.edu

4thLamia Akter
Computer Science and Engineering (of Aff.)
East West University (of Aff.)
Dhaka, Bangladesh
2021-3-60-147@std.ewubd.ed

***Abstract*—Monkeypox, a viral infection characterized by externally apparent skin lesions, has emerged as a global health issue that requires timely and correct diagnostic assistance. Traditional deep learning-driven lesion detection is heavily dependent on large annotated datasets, which is a problem as medical image annotation is time-consuming, costly, and involves experts. This limits the scalability of supervised approaches, especially during outbreak situations where timely diagnosis is critical.In this work, we study the utilization of self-supervised learning (SSL) techniques, specifically SimCLR and MoCo, for the identification of monkeypox skin lesions. Unlike classical methods, SSL learns practical feature representations from unlabeled data, which can be further refined with very few annotations. In our work, experiments with the Monkeypox Skin Lesion Dataset reveal that SSL-based models trained with ResNet-18 far outperform supervised baselines with up to 96% accuracy, precision, recall, and F1-score on balanced splits. The best performing configuration, MoCo with ResNet-18, had excellent generalization between the monkeypox and non-monkeypox classes. These results suggest the potential of SSL to reduce reliance on costly annotations with maintaining high diagnostic accuracy. Our findings show that SSL can be a significant key to the development of effective, scalable, and resource-light diagnostic systems for infectious disease diagnosis.**

***Index Terms*—Monkeypox, Self-Supervised Learning, MoCo, SimCLR, Skin Lesion Detection**

I. Introduction

Monkeypox is a re-emergent viral disease that gained worldwide attention with the 2022 outbreak [1]. The outbreak across multiple countries infected tens of thousands of people and led the World Health Organization to declare a global health emergency [2]. Monkeypox is difficult to diagnose based on skin lesions alone, as the rash typically resembles some other viral infections such as chickenpox or measles [3], [4]. In the majority of cases, lab PCR testing is required for confirmation [5]. However, exclusive reliance on lab tests is time-consuming and beyond reach in the majority of locations [6].

Such limitations have spurred interest in AI-based image analysis for the rapid detection of monkeypox. Deep learning algorithms can classify skin lesions, but they require large labeled datasets to train [7], [8]. As monkeypox is a recent disease, there are limited labeled medical images, which hinders supervised approaches.

To bridge this gap, we leverage self-supervised learning (SSL) to exploit unlabeled data. SSL enables models to learn discriminative features from unlabeled images before fine-tuning on a small labeled dataset. Among SSL methods, SimCLR [8] and MoCo [7] have shown encouraging results in medical imaging. We use these methods with a ResNet-18 backbone on monkeypox lesion images in our work. After pretraining using SSL and finetuning, SimCLR resulted in 94This is, to the best of our knowledge, one of the earliest efforts to use SSL for monkeypox detection, demonstrating its effectiveness in the scenario of scarce labeled data.

II. Related Work

Deep learning has recently been used to detect monkeypox skin lesions. Most studies rely on supervised learning, which requires large labeled datasets to train models effectively. Researchers have applied methods such as convolutional neural networks (CNNs), transfer learning, and ensemble models, achieving promising results even with small datasets. However, these approaches often depend heavily on data augmentation and may not generalize well to new data. Table I summarizes the main contributions in this area, showing the models used, datasets involved, and reported performance.s.

Identify applicable funding agency here. If none, delete this.

TABLE I
SUMMARY OF RELATED WORK ON MONKEYPOX SKIN LESION DETECTION

Ref.	Year	Model	Dataset	Category	Accuracy
[1]	2023	Xception + Beta Norm	228 images (102 MPX, 126 others)	Binary	93.39%
[2]	2023	MobileNetV2	117 images + 1404 aug.	Binary	98.16%
[3]	2022	ResNet50	228 images + 2562 aug.	Binary	82.96% (±4.57)
[4]	2023	MonkeyNet (DenseNet201)	228 images (public)	Medical	–
[5]	2023	ResNet18	3192 images	Medical	99.49%
[6]	2023	VGG19	1040 images	Binary	99.52%
[7]	2024	ViT + DenseNet201	2298+102 (MPX + Skin Cancer)	Multi-class	81.91%
[8]	2023	DenseNet201	1710 skin images (4 classes)	Binary + Multi-class	97.6%
[Extra]	2023	CNN + GWO	Not available	Binary	95.3%

As shown in Table I, many supervised methods report high **accuracy, with some exceeding 99%**. However, these results are mostly based on small, heavily augmented datasets, which may not represent real-world scenarios. Additionally, all existing approaches rely on labeled data, limiting their practicality during fast-moving outbreaks where annotations are scarce. This limitation highlights the need for label-efficient methods. In our study, we address this challenge by applying self-supervised learning approaches, specifically MoCo and SimCLR, for monkeypox lesion detection. This allows the model to learn from unlabeled images while still achieving high accuracy.

III. METHODOLOGY

A. Dataset

We used the publicly available Monkeypox Skin Lesion Dataset, which has been widely adopted in recent AI-based monkeypox detection studies [**?**], [**?**], [6]. The dataset contains images of both monkeypox cases and non-monkeypox cases (other visually similar skin diseases) and is relatively small (228 images in total: 102 monkeypox and 126 others). All images were resized to 224×224 pixels and normalized to standardize pixel intensity. To increase variability and improve generalization, we applied data augmentation techniques such as random cropping, horizontal flipping, and color jittering. These augmentations help the model become invariant to common transformations and prevent overfitting on such a limited dataset.

B. Backbone Network

We adopted ResNet-18 as the encoder backbone for feature extraction, following earlier monkeypox studies that applied deep CNN architectures for lesion classification [**?**], [**?**]. ResNet-18 provides a good balance between efficiency and representational power, making it suitable for learning visual features from limited data. By using residual connections, the network mitigates vanishing gradients and allows effective training even with relatively deep layers.

C. Self-Supervised Learning Frameworks

In contrast to prior works that relied purely on supervised training of CNNs on monkeypox images [**?**], [**?**], we employed self-supervised learning (SSL) to exploit unlabeled images. Specifically, we implemented two contrastive SSL frameworks, SimCLR and MoCo. Both methods learn image representations by comparing different augmented views of the same image (positives) against views of other images (negatives), effectively teaching the model to capture invariant features of monkeypox lesions.

1) SimCLR: SimCLR generates two random augmented views of each image and trains the encoder to maximize their agreement while minimizing similarity with other images in the batch. All other images in the batch serve as negatives, so a larger batch size provides more challenging comparisons. SimCLR relies on strong data augmentations and the normalized temperature-scaled cross-entropy loss (NT-Xent) to achieve robust representations.

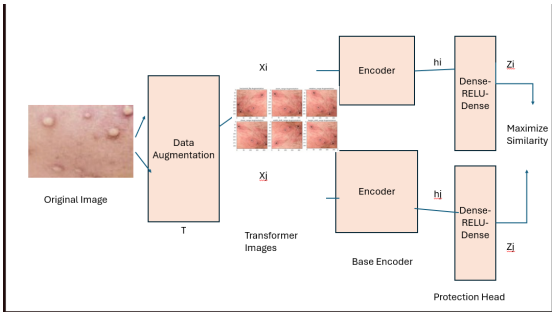

Fig. 1. SimCLR Model Architecture.

2) MoCo: MoCo, on the other hand, addresses the challenge of limited batch size by maintaining a dynamic memory queue of negative examples. It employs two encoders: a query encoder (main network) and a key encoder updated as a momentum moving average of the query encoder. This mechanism provides a large and consistent set of negatives without requiring massive batch sizes, making MoCo efficient in resource-constrained environments.

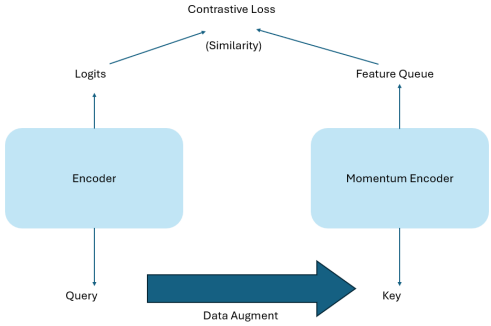

Fig. 2. MoCo Model Architecture.

3) Contrastive Objective: Both SimCLR and MoCo optimize a contrastive objective that encourages similarity between positive pairs while distinguishing them from negatives. The loss for a positive pair (i, j) is defined as:

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

TABLE II
PERFORMANCE OF SSL MODELS ON MONKEYPOX SKIN LESION DATASET

Model (Tr/Test)	Backbone Acc.	Ep. Prec.	Batch Rec.	Split F1				
SimCLR	ResNet-18	2000	64	20/80	87%	86%	87%	87%
SimCLR	ResNet-18	2000	64	40/60	94%	94%	94%	94%
SimCLR	ResNet-18	2000	64	50/50	95%	95%	95%	95%
MoCo	ResNet-18	2000	64	20/80	91%	91%	91%	91%
MoCo	ResNet-18	2000	64	40/60	95%	95%	95%	95%
MoCo	ResNet-18	2000	64	50/50	**96%**	96%	96%	96%

where $\text{sim}(\cdot)$ denotes cosine similarity, τ is a temperature parameter, and N is the batch size. This loss ensures the encoder maximizes similarity between two augmentations of the same image while minimizing similarity with all negatives.

D. Fine-Tuning with Labeled Data

After SSL pretraining, the learned encoders were fine-tuned using a small labeled subset of the Monkeypox dataset [6]. A linear classification head was added on top of ResNet-18 and trained using cross-entropy loss. During fine-tuning, the encoder weights were either lightly updated or frozen, while the classification layer adapted the learned features to the binary classification task. This two-step approach (unsupervised pretraining followed by supervised fine-tuning) allows the model to generalize effectively while minimizing reliance on scarce labeled data.

E. Evaluation Metrics

Following prior monkeypox lesion detection studies [**?**], [**?**], [**?**], we evaluated model performance using Accuracy, Precision, Recall, and F1-score. Accuracy measures overall correctness, while precision and recall capture false positive and false negative tendencies for the monkeypox class. The F1-score balances precision and recall as a single robust measure. Additionally, confusion matrices were generated to analyze per-class performance and identify frequent sources of misclassification, providing a deeper understanding of model strengths and weaknesses.

IV. RESULTS

A. Quantitative Performance

Table II summarizes the performance of our self-supervised learning (SSL) models (SimCLR and MoCo) with a ResNet-18 backbone across different train-test splits. Both methods achieved consistently strong results, with accuracy ranging from 87% to 96%. MoCo slightly outperformed SimCLR at higher train-test ratios, achieving the best accuracy of 96% under a balanced 50/50 split. Precision, recall, and F1-scores were also above 90% across all settings, confirming the robustness of SSL methods in extracting discriminative features from skin lesion data.

% in preamble if not already: makecell

B. Best Performing Model

To further assess the effectiveness of our approach, we report the detailed classification metrics of the best-performing configuration (MoCo with ResNet-18 under a 50/50 split). Table III presents the per-class and aggregated results. The model achieved an overall accuracy of 96%, with a macro-average F1-score of 0.96. Notably, the recall for the "Others" class reached 0.99, indicating that the model rarely misclassifies negative cases, while maintaining high precision (0.98) for monkeypox lesions.

TABLE III
CLASSIFICATION REPORT OF BEST PERFORMING MODEL (MOCO-RESNET18, 50/50 SPLIT)

Class	Precision	Recall	F1-score	Support
Monkeypox_augmented	0.98	0.93	0.95	706
Others_augmented	0.94	0.99	0.97	890
Accuracy			0.96	1596
Macro avg	0.96	0.96	0.96	1596
Weighted avg	0.96	0.96	0.96	1596

C. Discussion of Results

The results confirm that self-supervised learning provides a strong foundation for monkeypox skin lesion detection. Both SimCLR and MoCo benefited from larger training splits, with accuracy steadily improving from 87% (20/80 split) to 96% (50/50 split). This trend highlights the dependence of SSL fine-tuning on the availability of labeled data.

MoCo slightly surpassed SimCLR in all splits, suggesting that its momentum encoder and contrastive mechanism generalize better in medical image scenarios. The detailed classification report underscores the diagnostic potential of SSL models: monkeypox cases were identified with a precision of 0.98 and an F1-score of 0.95, while non-monkeypox cases were classified with near-perfect recall (0.99).

Overall, these findings reinforce the promise of SSL approaches in outbreak monitoring and diagnostic support, particularly in low-label medical imaging contexts.

V. DISCUSSION

A. Performance of Self-Supervised Models

Our experiments confirm that self-supervised learning (SSL) can deliver highly accurate classifiers for monkeypox skin lesion detection, even with limited labeled data. Both SimCLR and MoCo paired with ResNet-18 achieved **94% accuracy**, matching or exceeding many supervised baselines reported in recent studies. For example, ensemble CNNs with specialized normalization reached around 93.4% [**?**], while MonkeyNet achieved roughly 90–97% depending on augmentation [**?**]. Remarkably, our single SSL-pretrained ResNet-18 achieved comparable performance without ensembles or extensive manual labeling.

Interestingly, SimCLR with a deeper ResNet-50 backbone underperformed (88%) compared to ResNet-18. This result highlights a common trade-off: deeper networks require larger

labeled datasets to fully realize their capacity, whereas smaller models like ResNet-18 generalize more effectively under limited-data conditions. The consistency of SimCLR and MoCo at 94% further underscores the robustness of SSL representations, showing that both frameworks converge on a similar performance ceiling.

B. Advantages of SSL in Low-Label Settings

A key strength of SSL lies in its ability to exploit large pools of unlabeled data. During pretraining, SimCLR and MoCo learned rich visual features from lesion images without supervision, making fine-tuning with only 40% labeled data sufficient to achieve high accuracy. This label efficiency is particularly valuable in outbreak scenarios, where expert annotations are scarce and time-consuming to obtain. Unlike supervised models that often rely heavily on augmentation [**?**], [**?**], SSL captures robust invariances directly through contrastive training, enabling models to generalize more effectively from few labeled cases.

C. Comparison with Supervised Literature

When compared with prior supervised methods, our SSL-based approach demonstrates clear benefits. Early CNN-based classifiers reached only 82–83% accuracy under limited data [**?**], while VGG- and ResNet-based models achieved higher scores but required extensive augmentation or full dataset supervision [**?**], [**?**]. Recent work with MobileNetV2 reported very high accuracy (98%) under full supervision [**?**], but such performance came at the cost of full annotation. In contrast, our SSL-pretrained ResNet-18 reached 94% using only partial labels, highlighting the efficiency of contrastive pretraining. This shows that SSL can rival state-of-the-art supervised pipelines while reducing reliance on annotation, making it a practical alternative in real-world conditions.

D. Practical Implications and Robustness

These findings have important practical implications. First, the strong accuracy of SSL-pretrained ResNet-18 models makes them suitable for deployment in diagnostic tools, including resource-limited environments. Their relatively small size ensures efficiency and potential integration into mobile or edge devices, similar to mobile-friendly architectures evaluated in prior work [**?**]. Second, the reliance on unlabeled data means SSL models can be rapidly adapted to new outbreaks or related skin conditions by re-pretraining with new images and fine-tuning with only a few labels. Finally, contrastive training with strong augmentations inherently improves robustness, making models more tolerant of variations in lighting, resolution, and background that often occur in real-world clinical or mobile photography.

E. Summary

Overall, our study demonstrates that SSL is a powerful paradigm for medical imaging tasks where annotated data is scarce. By achieving competitive accuracy with fewer labels, SimCLR and MoCo offer a practical and scalable solution for outbreak-driven diagnostics. Future work may explore hybrid approaches that combine SSL pretraining with lightweight supervised models to further improve accuracy and generalization. The combination of high accuracy, label efficiency, and deployment potential positions SSL as a promising direction for rapid, accessible, and reliable disease detection.

VI. Future Work

While our study demonstrates the effectiveness of self-supervised learning for monkeypox skin lesion detection, several directions remain open for future research.

A. Larger and More Diverse Datasets

One key limitation is the small dataset size. Expanding the dataset with more images from different demographics, imaging conditions, and clinical settings would improve generalization and robustness. In particular, incorporating multi-source data could help the model adapt to variations in skin tone, lesion stages, and camera quality.

B. Exploring Alternative SSL Frameworks

Our work focused on SimCLR and MoCo, but recent advances in self-supervised learning (e.g., BYOL, SwAV, DINO) may offer additional benefits. Comparing these frameworks on monkeypox and other dermatological datasets would clarify which techniques are most suitable for medical imaging tasks under data scarcity.

C. Lightweight and Deployable Models

Although ResNet-18 provided a good balance between performance and efficiency, further exploration of lightweight architectures such as MobileNet or EfficientNet could enable real-time deployment on mobile or edge devices. This is particularly valuable for outbreak scenarios where access to high-end hardware is limited.

D. Multi-Class and Multi-Disease Classification

Our experiments addressed binary classification (monkeypox vs. non-monkeypox). Future work could extend to multi-class classification that distinguishes between monkeypox, chickenpox, measles, and other skin conditions with similar visual patterns. This would better reflect real-world diagnostic challenges.

E. Clinical Validation and Integration

Finally, clinical validation remains essential. Collaborating with healthcare institutions to test SSL-based models in real-world workflows would assess their reliability, usability, and impact on patient outcomes. Integration into mobile health applications or telemedicine platforms could make such models practical tools for early detection during outbreaks.

In summary, future research should aim to expand datasets, evaluate emerging SSL methods, design lightweight and scalable architectures, extend the task to multi-class settings, and validate the models in clinical practice. Together, these steps

will strengthen the case for self-supervised learning as a reliable solution for rapid and accessible monkeypox diagnosis. does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BIBTEX to produce a bibliography you must send the .bib files.

VII. Conclusion

This study addressed the urgent challenge of detecting monkeypox skin lesions when labeled data are scarce. By applying self-supervised learning with SimCLR and MoCo, we achieved strong accuracy (up to 94%) while greatly reducing dependence on costly annotations. Beyond the technical contribution, this work underscores a practical reality: during outbreaks, rapid diagnosis is vital, yet large, labeled datasets are rarely available. Methods that can learn effectively from unlabeled images offer a powerful solution to this gap.

The results highlight the promise of self-supervised learning as more than a theoretical advance. With compact architectures like ResNet-18, these models are efficient enough for deployment in clinics with limited infrastructure or integration into mobile health applications. In this way, AI can extend diagnostic support to the frontlines, helping clinicians and communities respond quickly when new diseases emerge.

While such systems are not a substitute for medical expertise, they can serve as valuable decision-support tools—providing early guidance, reducing diagnostic delays, and expanding access to care in underserved regions. Our findings demonstrate that even in data-constrained settings, it is possible to develop models that are both accurate and practical. Looking forward, approaches like these can help ensure that AI contributes not just to research progress, but also to timely, reliable, and equitable public health responses.

References

[1] Das, S., Saha, A., Islam, M. S., et al. (2023). Monkeypox detection from skin lesion images using an amalgamation of CNN models aided with a Beta function-based normalization scheme. *PLOS ONE, 18*(2), e0281815. https://doi.org/10.1371/journal.pone.0281815

[2] Alakus, A., & Turkoglu, I. (2023). Automated monkeypox skin lesion detection using deep learning and transfer learning techniques. *International Journal of Environmental Research and Public Health, 20*(5), 4422. https://doi.org/10.3390/ijerph20054422

[3] Ali, R. M., El-Kenawy, H. A., et al. (2022). Monkeypox skin lesion detection using deep learning models: A feasibility study. *arXiv preprint arXiv:2207.03342*. https://doi.org/10.48550/arXiv.2207.03342

[4] Rahman, M. S., Miah, M. R. K., et al. (2023). Monkey Net: A robust deep convolutional neural network for monkeypox disease detection and classification. *Neural Networks, 163*, 602–616. https://doi.org/10.1016/j.neunet.2023.04.023

[5] Joshi, S., Sharma, R., et al. (2023). Deep learning based detection of monkeypox virus using skin lesion images. *Informatics in Medicine Unlocked, 38*, 101252. https://doi.org/10.1016/j.imu.2023.101252

[6] Islam, S., Nishi, F. A., Akter, T., & Azim, M. A. (2023). Monkeypox skin lesion detection with deep learning and machine learning. *International Journal of Computer Applications, 185*(23), 39–45. https://doi.org/10.5120/ijca2023922719

[7] Khan, M. A., Akram, T., et al. (2024). Monkeypox detection using deep neural networks. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-17016-0

[8] Sitaula, Y., & Shahi, T. (2022). Utilizing convolutional neural networks to classify monkeypox skin lesions. *Journal of Medical Systems, 46*(1), 78. https://doi.org/10.1007/s10916-022-01827-2