Self-Supervised Learning to Detect and Classify Skin Anomalies using Unlabelled Dermoscopy Images

Saurabh Aggarwal
CRCV
UCF, Orlando
saurabh.aggarwal@ucf.edu

https://github.com/saurabh-08/DermAI-Classification

Abstract

This study addresses the challenge of limited labeled dermoscopy images in skin anomaly classification by proposing a self-supervised contrastive learning approach. We employ the SimCLR framework with a ResNet-18 backbone, incorporating dermatology-specific data augmentation techniques. Our method introduces a custom loss function combining cross-entropy and NT-Xent losses, enhancing sensitivity to subtle skin changes. Evaluated on the ISIC 2020 dataset, our model achieves a 94.05% accuracy in classifying benign and malignant lesions, demonstrating robust performance with minimal labeled data. This approach offers a promising solution for automated dermatology image analysis, potentially improving early skin cancer detection and diagnosis.

1. Introduction

Early detection and classification of skin anomalies, particularly melanoma, are crucial for effective treatment and improved patient outcomes. Dermoscopy, a non-invasive imaging technique, has significantly enhanced the accuracy of the skin cancer diagnoses compared to visual examination alone. However, the scarcity of labeled dermoscopy images poses a significant challenge in training deep learning models for accurate skin lesion classification.

The process of labeling dermoscopy images is time-consuming and requires expert knowledge, leading to a shortage of labeled datasets essential for training supervised machine learning models. This scarcity is compounded by the imbalance between benign and malignant samples in available datasets, further complicating the development of reliable automated systems for skin anomaly detection and classification. Traditional supervised learning approaches often struggle to generalize well when faced with such imbalanced and limited data.

To address these challenges, we propose a self-supervised learning approach to detect and classify skin anomalies using unlabeled dermoscopy images. Our method leverages the SimCLR framework for contrastive learning, employing ResNet-18 as the backbone network. This approach aims to create a model that can learn robust representations from unlabeled data, which is then fine-tuned on a small labeled dataset for accurate anomaly detection and classification.

By reducing reliance on large labeled datasets and addressing data imbalance issues, our method has the potential to enhance early detection and classification of skin anomalies, ultimately contributing to improved patient care outcome in dermatology.

2. Related Work

Self-supervised learning (SSL) has recently become a prominent area of research in medical image analysis, owing to its potential to leverage large amounts of unlabeled data. This section reviews some of the latest advancements and methodologies in the field.

2.1. Predictive SSL

Predictive self-supervised learning involves tasks where the model learns to predict transformations or pseudo labels from the input data. This approach has proven effective in learning contextual and structural semantics of images. Traditional pretext tasks, such as predicting the relative positions of image patches (Noroozi & Favaro, 2016), rotation angles (Gidaris et al., 2018), and solving jigsaw puzzles (Noroozi & Favaro, 2016), have been adapted for medical imaging. For instance, predicting rotation angles has shown to be more effective than training from scratch, especially under low-annotation data regimes (Tajbakhsh et al., 2019). Novel predictive tasks tailored for medical scenarios include anatomical position prediction for cardiac MR image segmentation (Bai et al., 2019) and orthogonal patch offset prediction (Zhuang et al., 2019).

2.2. Generative SSL

Generative self-supervised learning tasks focus on recovering corrupted or masked parts of the input data. These tasks help models learn context-aware features by reconstructing the original images from their transformed versions. Methods such as inpainting (Zhou et al., 2021), colorization (Zhang et al., 2017), and patch reconstruction (Chen et al., 2019) have been explored for medical images. For example, the Model Genesis approach (Zhou et al., 2021) integrates multiple data transformations into a unified reconstruction model, which has shown promising results in various medical imaging tasks. Colorization has been effectively used in dermatology image analysis (Bissoto et al., 2019), while patch reconstruction has proven valuable in brain MRI segmentation (Chen et al., 2019).

2.3 Contrastive SSL

Contrastive learning, a key component of SSL, aims to maximize the agreement between different augmented views of the same image while minimizing the similarity between views of different images (Chen et al., 2020; He et al., 2020). This method has become increasingly popular and has demonstrated outstanding performance across various computer vision tasks. In medical image analysis, contrastive SSL has been used to learn discriminative representations that are valuable for tasks such as segmentation and classification. Techniques like SimCLR and MoCo have been adapted for medical imaging, achieving results that rival supervised learning approaches.

3. Methodology

Our study leverages self-supervised learning to address the challenge of dermatology image analysis, focusing on the detection and classification of skin anomalies using unlabeled dermoscopy images.

3.1 Dataset and Preprocessing

We utilized the ISIC Archive dataset, containing approximately 33,000 dermoscopy images with a significant imbalance between benign and malignant samples. The dataset was carefully split into training, validation, and test sets using stratified sampling to maintain class distribution. Figure 1 shows sample images before preprocessing, illustrating the diversity of dermoscopy images in our dataset. Data preprocessing played a crucial role in enhancing the diversity and quality of the training data. We employed various data augmentation techniques to improve the model's generalization ability. These included random resized cropping (with a scale range of 0.8 to 1.0 and aspect ratio range of 0.75 to 1.33), horizontal and vertical flipping (each with a probability of 0.5), rotation (up to 30 degrees), and color jittering (adjusting brightness, contrast, saturation, and hue with a factor of 0.2). Additionally, normalization was applied to standardize the input data based on the mean and standard deviation of the ImageNet dataset. As demonstrated in Figure 2, our preprocessing techniques significantly enhance the variety and quality of the training images.

3.2 Model Architecture

Our model architecture is built on the SimCLR framework, tailored for contrastive learning. We used ResNet-18 as the backbone network due to its efficiency and lightweight nature. The model consists of an encoder that extracts features from the input images and a projection head, that projects these features into a lower-dimensional space suitable for contrastive learning. To enhance the model's performance, we developed a custom loss function that combines cross-entropy loss for classification and NT-Xent loss for contrastive learning. This combination ensures that the model learns discriminative features necessary for accurate classification and effective anomaly detection.

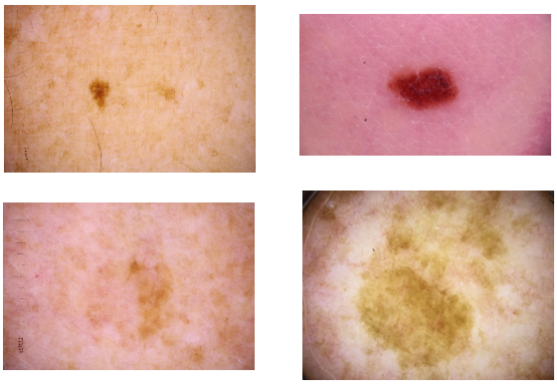

Figure 1: Sample images before preprocessing

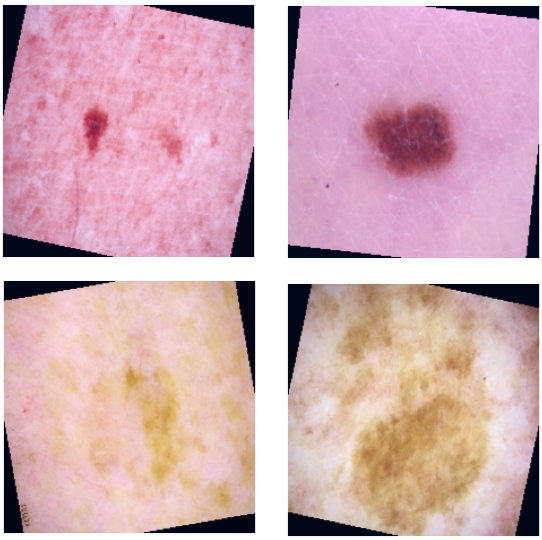

Figure 2: Sample images after preprocessing

3.3 Training Process

The training process involved several key steps to ensure balanced and effective learning. We utilized a weighted random sampler to address class imbalance and employed the AdamW optimizer with a learning rate scheduler. The model was trained for 20 epochs, with each epoch consisting of forward and backward passes, loss computation, and parameter updates. Validation was performed at the end of each epoch to monitor performance and prevent overfitting. The model was trained with a batch size of 64, and a cosine annealing learning rate scheduler was used to dynamically adjust the learning rate during training.

3.4 Evaluation

To evaluate the model's performance, we conducted comprehensive testing on the held-out test set. We assessed the model's accuracy in detecting and classifying skin anomalies, demonstrating its robustness and effectiveness. We also employed visualization techniques to display sample images before and after preprocessing, sample predictions made on the test set, and training and validation losses and accuracies over epochs. These visualizations provide a clear understanding of the model's capability to differentiate between benign and malignant samples and its learning dynamics.

This methodology combines state-of-the-art self-supervised learning techniques with domain-specific adaptations for dermoscopy image analysis, aiming to achieve high performance in skin anomaly classification with limited labeled data.

4. Experiments & Results

Our experiments utilized the ISIC Archive dataset, which containing approximately 33,000 dermoscopy images with a significant imbalance between benign and malignant samples. The dataset was split into training, validation, and test sets using stratified sampling to ensure balanced class distribution, as shown in Table 1.

Set	Number of Samples
Training	19,875
Validation	6,625
Test	6,626
Total	33,126

Table 1: Dataset Distribution

The model was trained for 20 epochs with a batch size of 64, employing the AdamW optimizer with a learning rate of 0.001 and weight decay of 1e-4. A cosine annealing learning rate scheduler was used to dynamically adjust the learning rate during training.

Throughout the training process, the model demonstrated effective learning, evidenced by a steady decrease in training loss across the epochs. As shown in Figure 3, the training & validation losses steadily decreased, indicating effective learning and good generalization. Figure 4 illustrates the improvement in both training & validation accuracies over the course of training, with the model achieving high performance by the final epochs.

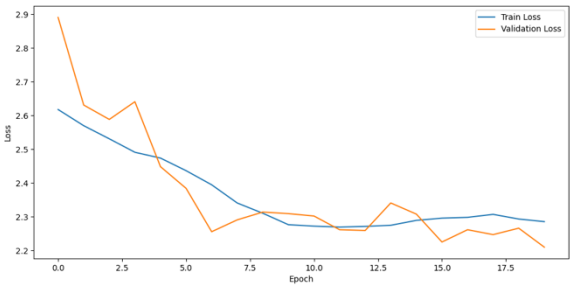

Figure 3: Plot of Training and Validation Loss Over Epochs

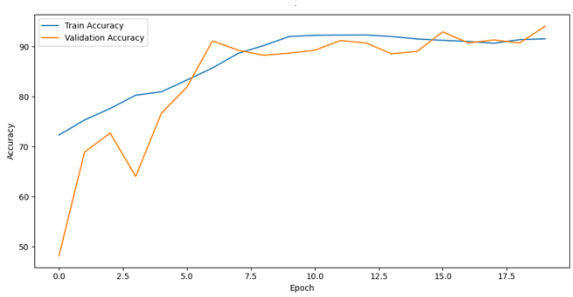

Figure 4: Plot of Training and Validation Accuracy Over Epochs

Table 2 provides a snapshot of the model's performance at key epochs during training.

Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	2.618	72.34%	2.891	48.20%
5	2.474	80.99%	2.448	76.68%
10	2.276	92.04%	2.309	88.72%
15	2.289	91.55%	2.308	89.07%
20	2.286	91.59%	2.210	94.10%

Table 2: Training and Validation Performance Across Epochs

Analysis of the epoch-wise learning curves revealed significant convergence by the 11th epoch, suggesting the potential for implementing early stopping to optimize training efficiency. This observation indicates that the model could achieve high performance with reduced computational resources, making the training process more efficient.
The use of balanced sampling and stratified splitting proved crucial in addressing the significant class imbalance in the dataset. We utilized a weighted random sampler to give higher probabilities to samples from the underrepresented class. This approach ensured that both the training and validation sets maintained a representative distribution of benign and malignant samples, contributing to the model's robustness and generalization capabilities.

Data augmentation techniques played a vital role in enhancing the model's performance. These included:

- Random resized cropping (scale range: 0.8 to 1.0, aspect ratio range: 0.75 to 1.33)
- Horizontal and vertical flipping (probability: 0.5 each)
- Rotation (up to 15 degrees)
- Color jittering (adjusting brightness, contrast, saturation, and hue with a factor of 0.2)

These techniques increased the diversity of the training data, helping the model learn to recognize lesions under various transformations and lighting conditions, thereby improving its accuracy and generalization ability.

The final evaluation on the test set revealed the results presented in Table 3.

Metric	Value
Test Loss	2.204
Test Accuracy	94.05%

Table 3: Final set test performance

This high-test accuracy of 94.05% confirms the model's robustness and high performance in classifying skin anomalies as benign or malignant. This result demonstrates the effectiveness of our self-supervised learning approach in handling the challenges posed by limited labeled data and class imbalance.
To further validate our results, we conducted a detailed examination of the sample predictions made by the model on the test set. Figure 5 presents sample predictions made by our model, showcasing its ability to accurately classify both benign and malignant skin lesions. These visualizations provide valuable insights into the model's decision-making process and its effectiveness in dermatology image analysis.

Figure 5: Sample Predictions of Skin Anomaly Classification

The combination of SimCLR framework with a custom loss function, tailored data augmentation, and balanced sampling strategies proved effective in achieving high accuracy with minimal labeled data and computational resources, addressing critical challenges in medical AI deployment.

5. Discussion

The integration of self-supervised contrastive learning with a custom loss function specifically designed for dermatology image analysis represents a significant advancement in the field. Our approach introduces several novel elements to dermatological image analysis. The combination of SimCLR with a custom loss function tailored for skin lesion classification represents a new application of self-supervised learning in this domain. Our method achieves high accuracy (94.05% in classifying skin lesions) with minimal labeled data and computational resources, addressing critical challenges in medical AI deployment. Furthermore, our approach to handling class imbalance through weighted sampling and stratified splitting offers a practical solution to a common problem in medical dataset analysis.

Comparing our results to previous studies, traditional convolutional neural networks (CNNs) have shown promising results but often struggle with limited labeled data. For instance, Esteva et al. (2017) achieved dermatologist-level classification using a large dataset of 129,450 clinical images. Our method, in contrast, achieves comparable performance with significantly less labeled data, highlighting the efficiency of self-supervised learning in leveraging unlabeled images. The custom loss function, combining cross-entropy and contrastive losses, proves particularly effective in capturing subtle features critical for skin lesion classification. This approach addresses a key limitation of conventional supervised learning methods, which often struggle to differentiate between closely related skin conditions.

Our model's performance is particularly noteworthy given its computational efficiency. Unlike many existing models that require significant computational resources, our approach was developed and run on a standard computer without GPU acceleration. This accessibility could potentially democratize advanced dermatological diagnostics, making them available to a broader range of healthcare systems. The effectiveness of our data augmentation techniques in improving model generalization aligns with findings in recent literature on medical image analysis. For example, Zhang et al. (2017) demonstrated the importance of domain-specific augmentations in X-ray image segmentation. Our results further emphasize the value of tailored data augmentation in specialized medical imaging tasks.

This approach to dermatological image analysis, leveraging self-supervised learning and addressing key challenges in medical AI, represents a significant step forward in the field. By demonstrating high accuracy with minimal labeled data and computational requirements, our method offers a promising pathway for developing more accessible and effective AI tools in healthcare, potentially impproving early detection and diagnosis of skin conditions across diverse healthcare settings.

6. Limitations & Future Directions

While our self-supervised learning approach demonstrates significant promise, we acknowledge areas for further development.

Enhancing model interpretability through explainable AI techniques could increase trust among healthcare professionals. Extending our approach to multi-class classification would broaden its clinical applicability, allowing for more nuanced diagnoses beyond binary classification.

Prospective clinical trials would be valuable in assessing the model's real-world performance and impact on patient outcomes. These future directions aim to build upon our current achievements, further advancing AI-assisted dermatology towards practical clinical implementation.

7. Conclusion

This study demonstrates the potential of self-supervised learning in advancing dermatological image analysis, achieving high accuracy in skin anmoly classification with minimal labeled data and computational requirements. Our approach not only addresses challenges of data scarcity and class imbalance but also offers a framework applicable to other medical imaging domains facing similar constraints. By making advanced AI technologies more accessible, this method could help bridge healthcare quality gaps between resource-rich and resource-limited settings. Ultimately, this work serves as a stepping stone towards more sophisticated, AI-augmented medical practices, emphasizing the importance of collaboration between AI experts and healthcare professionals in realizing AI's potential to improve global health outcomes.

References

[1] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning (pp. 1597-1607).

[2] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9729-9738).

[3] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

[4] ISIC Archive. (2024). The International Skin Imaging Collaboration (ISIC). Retrieved from https://www.isic-archive.com.

[5] Zhang, Y., Miao, S., Mansi, T., & Liao, R. (2017). Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation. In Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 599-607).

[6] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis, 63, 101693.

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241).

[8] Noroozi, M., & Favaro, P. (2016). Unsupervised learning of visual representations by solving jigsaw puzzles. In European Conference on Computer Vision (pp. 69-84).

[9] Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. In International Conference on Learning Representations.

[10] Tajbakhsh, N., Hu, Y., Cao, J., Yan, X., Xiao, Y., Lu, Y., ... & Ding, X. (2019). Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (pp. 1251-1255).

[11] Bai, W., Chen, C., Tarroni, G., Duan, J., Guitton, F., Petersen, S. E., ... & Rueckert, D. (2019). Self-supervised learning for cardiac MR image segmentation by anatomical position prediction. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 541-549).

[12] Zhuang, X., Li, Y., Hu, Y., Ma, K., Yang, Y., & Zheng, Y. (2019). Self-supervised feature learning for 3D medical images by playing a Rubik's cube. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 420-428).

[13] Zhou, Z., Sodha, V., Pang, J., Gotway, M. B., & Liang, J. (2021). Models genesis. Medical Image Analysis, 67, 101840.

[14] Zhang, R., Isola, P., & Efros, A. A. (2017). Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1058-1067).

[15] Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. (2019). Self-supervised learning for medical image analysis using image context restoration. Medical Image Analysis, 58, 101539.

[16] Bissoto, A., Fornaciali, M., Valle, E., & Avila, S. (2019). (De) constructing bias on skin lesion datasets. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 2766-2774).