Enhancing Anomaly Detection in Melanoma Diagnosis Through Self-Supervised Training and Lesion Comparison

Jules Collenne, Rabah Iguernaissi, Séverine Dubuisson, Djamal Merad

Enhancing Anomaly Detection in Melanoma Diagnosis through Self-Supervised Training and Lesion Comparison*

Jules Collenne[1 [0000−0002−7540−0610]], Rabah Iguernaissi[1 [0000−0002−1728−3532]], Séverine Dubuisson[1 [0000−0001−7306−4134]], and Djamal Merad[1 [0000−0001−9233−0020]]

Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

Abstract. Melanoma, a highly aggressive form of skin cancer notorious for its rapid metastasis, necessitates early detection to mitigate complex treatment requirements. While considerable research has addressed melanoma diagnosis using convolutional neural networks (CNNs) on individual dermatological images, a deeper exploration of lesion comparison within a patient is warranted for enhanced anomaly detection, which often signifies malignancy. In this study, we present a novel approach founded on an automated, self-supervised framework for comparing skin lesions, working entirely without access to ground truth labels. Our methodology involves encoding lesion images into feature vectors using a state-of-the-art representation learner, and subsequently leveraging an anomaly detection algorithm to identify atypical lesions. Remarkably, our model achieves robust anomaly detection performance on ISIC 2020 without needing annotations, highlighting the efficacy of the representation learner in discerning salient image features. These findings pave the way for future research endeavors aimed at developing better predictive models as well as interpretable tools that enhance dermatologists' efficacy in scrutinizing skin lesions.

Keywords: Melanoma, Skin Cancer, Lesion Comparison, Representation Learning, Anomaly Detection, Early Detection.

1 Introduction

The concept of comparing skin lesions within a single patient, known as the "ugly duckling sign", was introduced by Grob et al. [3] in 1998. According to this concept, visually distinct skin lesions within a patient are more likely to be indicative of cancers. Despite the extensive research on skin lesion classification, the exploration of lesion comparison within patients remains relatively limited, hampering the detection of anomalies that may signify malignancy. Concurrently, convolutional neural networks (CNNs) applied to dermoscopic images have exhibited

* Thanks to the French National Research Agency for the funding of the DIAMELEX project (ANR-20-CE45-0026).

remarkable success in diagnosing skin diseases [5, 13], yet the incorporation of lesion comparison remains underutilized. We posit that integrating the "ugly duckling sign" concept can significantly enhance these diagnostic outcomes.

Previous studies investigating the comparison of skin lesions for melanoma detection [2,4,12] have shown promising results, demonstrating increased detection rates when leveraging the "ugly duckling sign", whether analyzed by dermatologists or artificial intelligence algorithms. Automating this process commonly involves extracting features from a base model trained on the dataset, followed by comparing these features—an approach shared by our work. However, several crucial parameters must be considered in pursuit of this goal.

Soenksen et al. [10] utilize a CNN for classification, extracting features reduced to two dimensions using PCA and compared with Euclidean distance. Yu et al. [14] propose a hybrid model combining classification CNNs and transformers for lesion comparison. However, classification CNNs rely exclusively on labelled data and exhibit bias towards the primary task. Surprisingly, representation learners, purpose-built models for generating comparable feature vectors, have not been explored in prior work. These models employ a modified CNN architecture with a distinct training setup, comparing vector representations of two images. By bringing similar images closer and pushing dissimilar images apart, representation learners offer a promising approach for lesion comparison, distinct from the traditional use of classification CNNs in existing skin lesion studies.

In our approach, we take a significant stride by conducting our analysis without relying on any annotations. The underlying concept is rooted in the understanding that melanoma, like any cancer, entails the uncontrolled proliferation of malignant cells. Given a sufficient amount of data, the aberrant growth patterns exhibited by cancerous cells can be viewed as anomalies, especially when compared to other lesions of a patient. Thus, we posit that with a substantial volume of medical images, cancer detection could potentially be achieved without the need for annotations, as exemplified by our current methodology. To carry out our experiments, we employed the SimSiam [1] siamese neural network, leveraging the ResNet-50 [6] architecture as the backbone. This choice offers the advantage of reduced computational complexity compared to alternative models, while still yielding a robust representation of the skin lesion images.

By addressing the research gaps outlined above, we strive to develop more intelligent and interpretable tools that augment dermatologists' capabilities in analyzing skin lesions, facilitating early melanoma detection and improving patient outcomes.

2 Methods

Our study comprises two modules, as illustrated in Figure 1. Firstly, we train an encoder, employing a representation learner, to generate vector representations of images in the latent space. Subsequently, for each patient, we compare the generated vectors of their lesions using a k-nearest neighbors-inspired algorithm.

This anomaly detection process yields an "Ugly-Duckling Index" (UDI) which quantifies the degree of atypicality of a lesion compared to others.

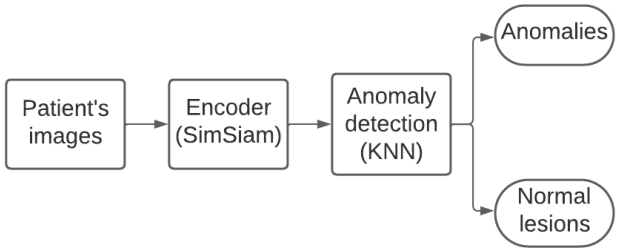

Fig. 1. Pipeline of our anomaly detection

2.1 Features extraction

For feature extraction, we utilized the SimSiam architecture [1] in conjunction with a ResNet-50 model [6]. This siamese architecture consists of two identical branches, both accepting the same input image that has undergone distinct augmentations. Notably, it does not necessitate the inclusion of negative pairs. During training, one branch undergoes a stop-gradient operation, preventing the model from producing identical vectors for all inputs. We maintained the original model and architecture, making only one adjustment after feature extraction: reducing the size of the vectors from 2048 to 100 by using the UMAP algorithm [7]. This modification was made to ensure that our feature vectors were represented in a dimension that strikes a balance between capturing pertinent information and avoiding excessive dimensionality, which could produce the curse of dimensionality. We also normalize the values per feature using the Min-Max scaler to facilitate anomaly detection. Our model is trained on 150 epochs, the loss function is the cosine similarity, the optimizer is SGD and the learning rate is 0.05. Before entering the model, the images are resized from their original size to 224x224, which is the expected size for ResNet-50. We kept the original data augmentation techniques of the original SimSiam paper which is a composition of random operations such as resize, crop, random grayscale, gaussian blur, horizontal flip and color jitter. The model seems to have converged after the 150 epochs (Figure 2).

Following the completion of training on the designated training set, we extracted the features from all images encompassing the dataset, which encompassed both the train and test sets. The acquired feature vectors served as the foundation for the subsequent stage of anomaly detection, facilitating the progression of our analysis.

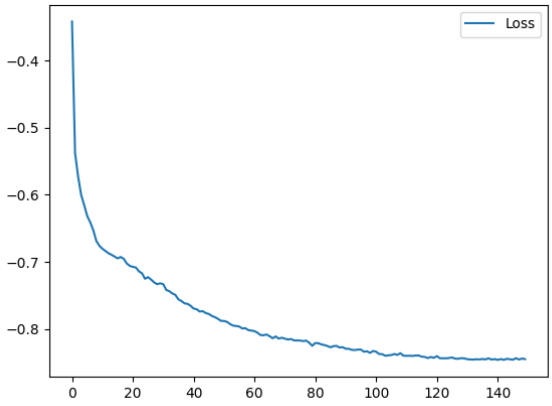

Fig. 2. Loss of ResNet-50 per epoch during SimSiam training.

2.2 Deriving the ugly duckling index through *k*-nearest neighbors computation

The process of feature extraction yielded a collection of vector representations for each patient, providing the means to compare and analyze these vectors. For this, we devise a straightforward algorithm inspired by the k-nearest neighbors approach, wherein the relative distances between all lesions are compared. Lesions that exhibit greater distance from their respective k-nearest neighbors than other lesions are assigned a higher Ugly-Duckling Index (UDI). This index, computed per patient, quantifies the distinctiveness of a given lesion in relation to its k-neighbors. The higher the index, the more different the lesion is from its neighbors. First, we compute the distances between all vectors corresponding to the patient's lesions, and subsequently, for each vector, we compute the average distance to its k-nearest neighbors (Equation 1).

$$D_i = \frac{\sum_{j \in K_i} d(v_i, v_j)}{k} \tag{1}$$

with v_i the features' vector of image i, k the number of neighbours, K_i the ensemble of k neighbours of i and d the distance function (in our case the cosine distance). We divide the resulting values by the average of these values (denoted as $\overline{D}$) in order to obtain the UDI (Equation 2).

$$UDI_i = \frac{D_i}{\overline{D}} \tag{2}$$

In order to facilitate comparison, we also normalize the resulting values by removing the mean and scaling to unit variance, considering their range in $\mathbb{R}_+^*$. Among various distance metrics available, we opt for the cosine distance, aligning with the cosine similarity utilized by our siamese network for vector comparison.

By leveraging this choice, we effectively capitalize on the learned representations, as the cosine distance and cosine similarity are related by the simple equation $cosine\_distance = 1 - cosine\_similarity$. Figure 3 presents an example of a two-dimensional visualization of the results.

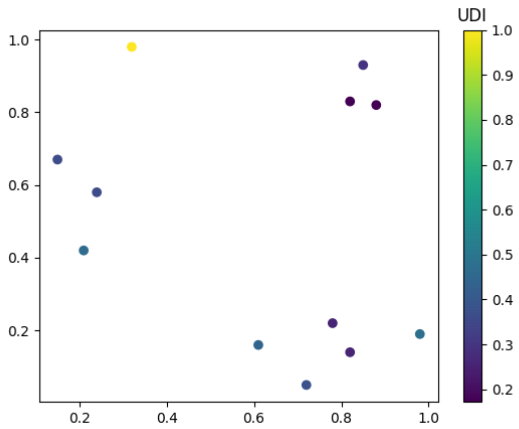

Fig. 3. Visualisation of our anomaly detection algorithm when using $k = 1$ on a 2 dimensional cloud of dots. Each point represents a lesion's features' vector while its colour represents its UDI.

Our algorithm was intentionally designed to mirror the approach employed by dermatologists when comparing lesions. Unlike a singular, ideal representation of normal lesions, typical lesions exhibit diversity and can be characterized by multiple clusters. Therefore, it is important to note that lesions that deviate significantly from the average appearance of a normal lesion may not necessarily be considered "ugly ducklings". By analyzing the k nearest vectors, our algorithm effectively identifies vectors that are distant from any other cluster, enabling the accurate prediction of true "ugly ducklings".

We manage cases where patients have a lower number of lesions than the k parameter by computing the distance to the centroid of the lesions (similar to prior approaches [10]). To further evaluate the effectiveness of this approach, we conducted a comparative analysis, contrasting the outcomes obtained using this method with those achieved through our current algorithm. The results of this comparative assessment will be presented in the subsequent section.

3 Results

The evaluation of our model on the test set involves comparing the computed Ugly-Duckling Index (UDI) with the ground truth annotations. In this section,

we present details about the dataset utilized in our experiments, analyze the performance of our model while varying the k parameter, and subsequently examine the results obtained using the optimal value of k.

3.1 Dataset

We conducted our experiments using the ISIC 2020 dataset [8], which is the largest available dataset for skin lesions (examples of images are presented Figure 4). It consists of over 32,000 dermoscopy images, which we randomly divided into three subsets for training, validation and test. The training set comprises 70% of patients, while the validation and test sets contain 15% each. Although the ISIC 2020 dataset is extensive, it contains images with various artifacts, including excessive body hair, bubbles, ink, and colored marks. In a traditional classification task, a well-trained model learns to disregard these artifacts as they are irrelevant to the classification process. However, in the context of detecting atypical images in a self-supervised manner, the model may perceive these artifacts as significant factors when determining the atypicality of a lesion. By utilizing the ISIC 2020 dataset, we aimed to leverage its rich diversity and substantial size to train and evaluate our model effectively. In the subsequent sections, we present the results of our experiments, highlighting the performance achieved and providing insights into the effectiveness of our proposed methodology.

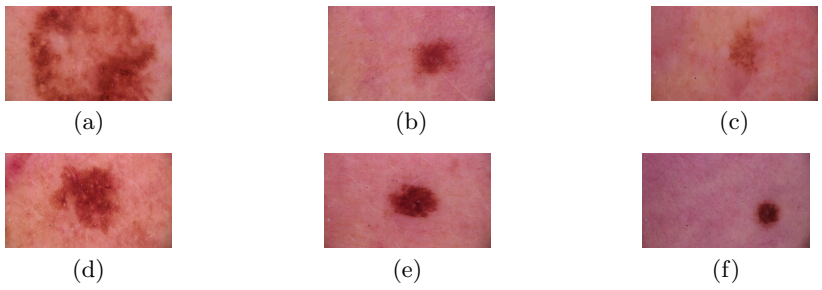

Fig. 4. Examples of images from the ISIC 2020 dataset from a single patient are shown above. All images except for (a) depict nevus.

3.2 Anomaly detection

We conducted a thorough analysis of our anomaly detection results, comparing them to the ground truth annotations. The evaluation focused on determining the optimal value for the parameter k, and the summarized findings are presented in Figure 5. This analysis was exclusively performed on the validation data. Our study revealed a clear trend where increasing the value of k resulted in improved

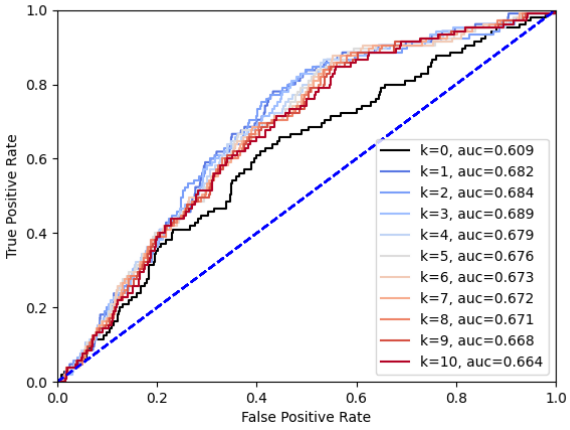

Fig. 5. AUC comparison on the validation data with varying values of the parameter k

performances until reaching a peak at $k = 3$, after which the performance gradually declined. Notably, when utilizing the distance to the centroid of all patient's lesions, the Area Under the ROC Curve (AUC) reached its lowest value, indicating the limited relevance of this approach to our problem. We hypothesize that "ugly duckling" lesions, being typically isolated, can be adequately assessed by comparing them to their three nearest neighbors to determine their deviation from other lesions. As lesions are represented by multiple clusters, increasing the number of neighbors may lead to the averaging out of distances, resulting in similar values for all lesions. Using a high value of k leads to the inclusion of the distance to the centroid, which may inadvertently diminish the potential information from patients with a low number of lesions. Using the average of coordinates is comparable to using a high k value, which implies the assumption of a unique normality for lesions, located at a single point in space, which may not accurately reflect reality. Skin lesions present a diverse spectrum of patterns that can still be benign, suggesting that representing them as multiple distinct clusters in the latent space would be more appropriate. Additionally, images that do not belong to any cluster (being too far apart from any cluster) are classified as 'ugly duckling' and are more likely to be malignant.

Following the identification of the optimal value for k on the validation data, we proceeded to assess the performance of our model on the test set using the selected value of $k = 3$. Surprisingly, the corresponding AUC is 0.719, indicating even higher performance compared to the validation data. The peak balanced accuracy achieved by threshold selection is 0.70, aligning with a sensitivity of 0.77 and a specificity of 0.62, thereby illustrating the optimal discriminative performance. Although these results remain below the performance of models that leverage ground-truth annotations, they are notably promising, particularly

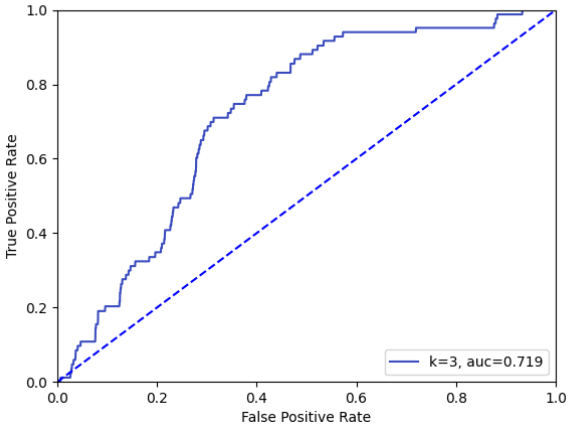

Fig. 6. ROC curve of the model on the test set for $k = 3$.

when taking into account that annotations are entirely omitted from the entire pipeline. It also demonstrates capability of the representation learner to learn visual features in a way that can be leveraged for subsequent tasks. The threshold also affords the flexibility to fine-tune the trade-off between detecting a greater number of melanomas or upholding a higher level of specificity in our model. For a visual depiction of the model's performance, see the ROC curve in Figure 6.

4 Conclusions and future work

Our study showcases compelling outcomes by leveraging representation learners in a fully self-supervised manner on dermoscopic images. Despite the inherent difficulty of this task, our developed architecture demonstrates proficiency in melanoma detection, avoiding the reliance on labeled data and relying solely on inter-lesion comparisons within patients. Subsequent investigations can investigate performance gains achieved by integrating this information into state-of-the-art classification models, as well as exploring novel representation learning techniques focused on the interpretability of the learnt vectors. Recent advancements in one-class classification models [9] and specialized loss functions for siamese networks [11] have diligently addressed the challenge of vector representation, thereby facilitating more seamless image comparisons. Although our approach presently hinges on the k-nearest neighbors algorithm, necessitating diligent parameter tuning, we acknowledge its potential for refinement, making it an enticing avenue for future research endeavors aimed at automating this anomaly detection paradigm. It is noteworthy that this approach could have a substantial impact on broader applications, particularly in the realm of diagnosing various types of cancers and illnesses.

References

1. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15750–15758 (June 2021)
2. Gaudy-Marqueste, C., Wazaefi, Y., Bruneu, Y., Triller, R., Thomas, L., Pellacani, G., Malvehy, J., Avril, M.F., Monestier, S., Richard, M.A., Fertil, B., Grob, J.J.: Ugly Duckling Sign as a Major Factor of Efficiency in Melanoma Detection. JAMA Dermatology **153**(4), 279–284 (04 2017). https://doi.org/10.1001/jamadermatol.2016.5500, https://doi.org/10.1001/jamadermatol.2016.5500
3. Grob, J.J., Bonerandi, J.J.: The 'Ugly Duckling' Sign: Identification of the Common Characteristics of Nevi in an Individual as a Basis for Melanoma Screening. Archives of Dermatology **134**(1), 103–104 (01 1998)
4. Grob, J.J., Wazaefi, Y., Bruneu, Y., Gaudy-Marqueste, C., Monestier, S., Thomas, L., Avril, M.F., Triller, R., Pellacani, G., Malvehy, J., Fertil, B.: Diagnosis of melanoma: Importance of comparative analysis and "ugly duckling" sign. Journal of Clinical Oncology **30**(15_suppl), 8578–8578 (2012). https://doi.org/10.1200/jco.2012.30.15_suppl.8578, https://doi.org/10.1200/jco.2012.30.15_suppl.8578
5. Ha, Q., Liu, B., Liu, F.: Identifying melanoma images using efficientnet ensemble: Winning solution to the siim-isic melanoma classification challenge. arXiv:2010.05351 (2020). https://doi.org/10.48550/ARXIV.2010.05351, https://arxiv.org/abs/2010.05351
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015). https://doi.org/10.48550/ARXIV.1512.03385, https://arxiv.org/abs/1512.03385
7. McInnes, L., Healy, J., Saul, N., Großberger, L.: Umap: Uniform manifold approximation and projection. Journal of Open Source Software **3**(29), 861 (2018). https://doi.org/10.21105/joss.00861, https://doi.org/10.21105/joss.00861
8. Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Lioprys, K., Malvehy, J., Musthaq, S., Nanda, J., Reiter, O., Shih, G., Stratigos, A., Tschandl, P., Weber, J., Soyer, P.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data **8**, 34 (2021). https://doi.org/10.1038/s41597-021-00815-z
9. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: Dy, J., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 4393–4402. PMLR (10–15 Jul 2018), https://proceedings.mlr.press/v80/ruff18a.html
10. Soenksen, L.R., Kassis, T., Conover, S.T., Marti-Fuster, B., Birkenfeld, J.S., Tucker-Schwartz, J., Naseem, A., Stavert, R.R., Kim, C.C., Senna, M.M., Avilés-Izquierdo, J., Collins, J.J., Barzilay, R., Gray, M.L.: Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. Science Translational Medicine **13**(581), eabb3652 (2021). https://doi.org/10.1126/scitranslmed.abb3652, https://www.science.org/doi/abs/10.1126/scitranslmed.abb3652
11. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: III, H.D., Singh, A. (eds.) Proceedings of the 37th International Conference on Machine Learning. Proceedings

of Machine Learning Research, vol. 119, pp. 9929–9939. PMLR (13–18 Jul 2020), https://proceedings.mlr.press/v119/wang20k.html
12. Wazaefi, Y., Gaudy-Marqueste, C., Avril, M.F., Malvehy, J., Pellacani, G., Thomas, L., Triller, R., Bruneu, Y., Monestier, S., Richard, M.A., Fertil, B., Grob, J.J.: Evidence of a limited intra-individual diversity of nevi: Intuitive perception of dominant clusters is a crucial step in the analysis of nevi by dermatologists. Journal of Investigative Dermatology **133**(10), 2355–2361 (2013). https://doi.org/https://doi.org/10.1038/jid.2013.183, https://www.sciencedirect.com/science/article/pii/S0022202X15359911
13. Winkler, J.K., Sies, K., Fink, C., Toberer, F., Enk, A., Deinlein, T., Hofmann-Wellenhof, R., Thomas, L., Lallas, A., Blum, A., Stolz, W., Abassi, M.S., Fuchs, T., Rosenberger, A., Haenssle, H.A.: Melanoma recognition by a deep learning convolutional neural network—performance in different melanoma subtypes and localisations. European Journal of Cancer **127**, 21–29 (2020). https://doi.org/https://doi.org/10.1016/j.ejca.2019.11.020, https://www.sciencedirect.com/science/article/pii/S0959804919308640
14. Yu, Z., Mar, V., Eriksson, A., Chandra, S., Bonnington, P., Zhang, L., Ge, Z.: End-to-end ugly duckling sign detection for melanoma identification with transformers. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 176–184. Springer International Publishing, Cham (2021)

5 Acknowledgement

This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-45676-3_16. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted- manuscript-terms