
ENHANCING DIAGNOSIS THROUGH AI-DRIVEN ANALYSIS OF REFLECTANCE CONFOCAL MICROSCOPY *

Hong-Jun Yoon
Oak Ridge National Laboratory
Oak Ridge, TN USA
yoonh@ornl.gov

Chris Keum
Farragut High School
Knoxville, TN USA

Alexander Witkowski
Oregon Health and Science University
Portland, OR USA

Joanna Ludzik
Oregon Health and Science University
Portland, OR USA

Tracy Petrie
Oregon Health and Science University
Portland, OR USA

Heidi A. Hanson
Oak Ridge National Laboratory
Oak Ridge, TN USA

Sancy A. Leachman
Oregon Health and Science University
Portland, OR USA

ABSTRACT

Reflectance Confocal Microscopy (RCM) is a non-invasive imaging technique used in biomedical research and clinical dermatology. It provides virtual high-resolution images of the skin and superficial tissues, reducing the need for physical biopsies. RCM employs a laser light source to illuminate the tissue, capturing the reflected light to generate detailed images of microscopic structures at various depths. Recent studies explored AI and machine learning, particularly CNNs, for analyzing RCM images. Our study proposes a segmentation strategy based on textural features to identify clinically significant regions, empowering dermatologists in effective image interpretation and boosting diagnostic confidence. This approach promises to advance dermatological diagnosis and treatment.

Keywords reflectance confocal microscopy · skin cancer · dermatology · vision transformer · self supervised learning · artificial intelligence

1 Introduction

Reflectance Confocal Microscopy (RCM) marks a paradigm shift in biomedical imaging, offering a sophisticated, non-invasive technique to acquire high-resolution images of the skin and superficial tissues. Its development [1] represents a milestone in medical imaging, transitioning from early exploratory stages to becoming a cornerstone in clinical dermatology. RCM's capability for *in vivo* imaging, capturing live tissue images without the need for biopsies or tissue excision, has made it an indispensable tool in modern medical diagnostics. The inception of RCM can be traced back to its early conceptualization, where the need for less invasive, more accurate diagnostic methods in dermatology was recognized. Over the years, the technology has undergone significant advancements, evolving in its design and functionality. This evolution has been marked by improvements in laser source quality, detector sensitivity, and image processing algorithms, resulting in enhanced image clarity and depth of tissue analysis. RCM's operation relies on a focused laser light to illuminate the target tissue. The tissue interaction with this light, primarily through backscattering and reflection, forms the basis of image creation. A detector captures this reflected light, and sophisticated computer

*This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

algorithms transform these signals into detailed, cross-sectional images of the tissue. This allows for the visualization of cellular and sub-cellular structures with remarkable clarity and precision.

In dermatology, RCM has proven particularly beneficial. It provides a real-time, non-invasive alternative to traditional skin biopsies, allowing for immediate assessment of skin lesions. Conditions like melanoma, basal cell carcinoma, and squamous cell carcinoma can be diagnosed with greater accuracy, significantly improving patient outcomes [2]. Moreover, RCM facilitates the monitoring of treatment responses and the management of chronic conditions, offering a dynamic tool for personalized patient care. Comparatively, RCM holds distinct advantages over other imaging modalities such as ultrasound, magnetic resonance imaging (MRI), or traditional dermoscopy. Its higher resolution and ability to image at a cellular level give it an edge in early disease detection and diagnosis, particularly in cases where other modalities may fall short.

In recent years, the integration of AI and ML technologies, especially convolutional neural networks (CNNs), has further expanded the capabilities of RCM [3, 4, 5, 6]. These technologies enable the processing and interpretation of complex image data, facilitating automated detection and diagnosis of skin pathologies. This synergy of RCM and AI presents an exciting frontier in dermatological imaging, offering prospects for more accurate, efficient, and personalized patient care. Our study contributes to this evolving landscape by proposing a novel segmentation strategy for RCM images, focusing on textural features to identify key clinical regions. This approach aims to augment the diagnostic accuracy and efficiency, potentially transforming how dermatologists interpret RCM images and make clinical decisions.

Looking forward, the potential of RCM extends beyond dermatology. Its application could revolutionize areas such as ophthalmology, neurology, and oncology, where non-invasive imaging plays a critical role. Furthermore, the ongoing integration of RCM with emerging technologies such as augmented reality, 3D imaging, and telemedicine could further enhance its utility, making it a versatile tool in various medical disciplines. The future of RCM is poised to be marked by continuous innovation, with the promise of significantly impacting patient care and medical research.

2 Methods

2.1 Dataset

The dataset utilized in this study comprised a total of 519 de-identified Reflectance Confocal Microscopy (RCM) images provided by the esteemed Department of Dermatology at Oregon Health and Science University. The dataset was balanced, consisting of 233 images that were recommended for biopsy and 286 images that were not recommended for biopsy. This collection is composed of various skin conditions of varying cancer likelihoods. Note that the RCM images included in the dataset exhibit two pixel resolutions. Specifically, a subset of the images possesses a resolution of 1,000 by 1,000 pixels, whereas the remaining images have a resolution of 1,024 by 1,024 pixels. All the images are rendered in single-channel 8-bit grayscale. All images were de-identified to ensure patient confidentiality, with any personal identifiers removed.

2.2 Image patch generation

We encountered a challenge with the image sizes of the RCM images in the dataset, as they exceeded the commonly used de-facto standard resolution of 256 by 256 pixels in AI/ML models for image processing and classification. In order to align with the current standards in AI/ML, we made the decision to divide each RCM image into multiple patches, each measuring 256x256 pixels. This division allowed us to maintain compatibility with AI/ML models while exploring the intricate characteristics of the images. Furthermore, this approach aligns with the primary objective of our study, which is to investigate image characteristics using computer algorithms and enable automatic segmentation with minimal expert intervention.

Recognizing the significance of boundary information within these patches, we implemented a step size of 64 pixels between adjacent patches. This step size was chosen to preserve the boundary characteristics and fine details within the images. In order to ensure comprehensive coverage of the image space, we employed a mirroring technique to extend the boundaries of the RCM images by 128 pixels. This allowed for a thorough exploration of image boundaries. Consequently, a single RCM image with a resolution of 1,000 by 1,000 pixels resulted in the generation of 256 distinct image patches. Similarly, an image with a slightly higher resolution of 1,024 by 1,024 pixels produced 289 patches. Figure 1 illustrates the procedure of mirroring and image patch creation from the RCM images.

2.3 Feature extraction

In our study, we employed a cutting-edge self-supervised learning algorithm known as DINO (self-distillation with no labels) [7], to effectively train a Vision Transformer (ViT) model [8] from scratch for the purpose of extracting

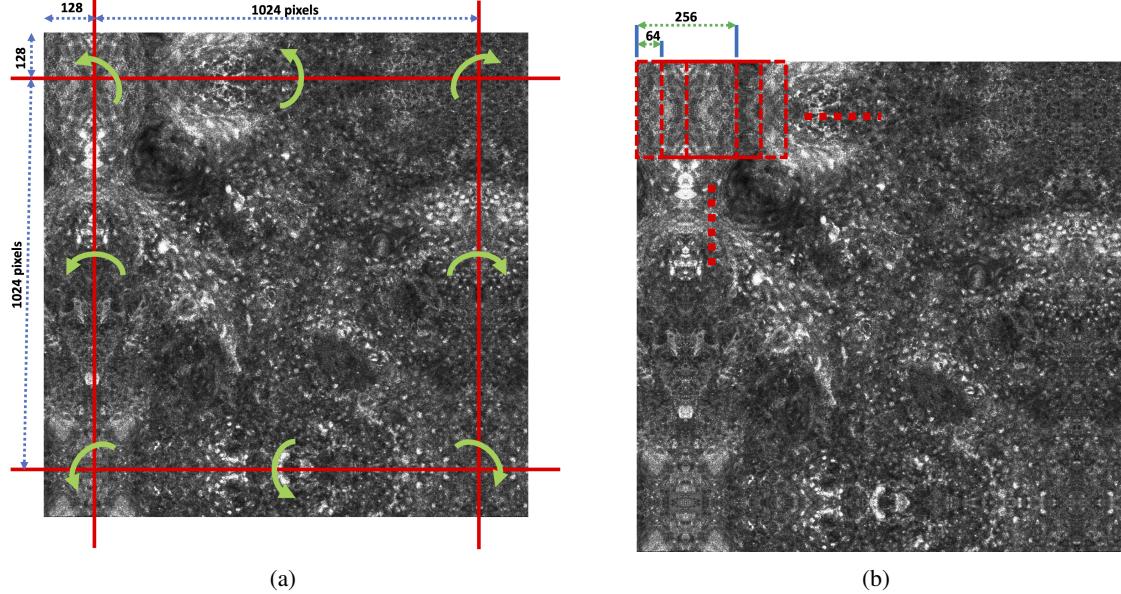


Figure 1: Patch generation from RCM images. (a) applying mirroring to the edges of image and (b) making image patches of 256x256 pixels.

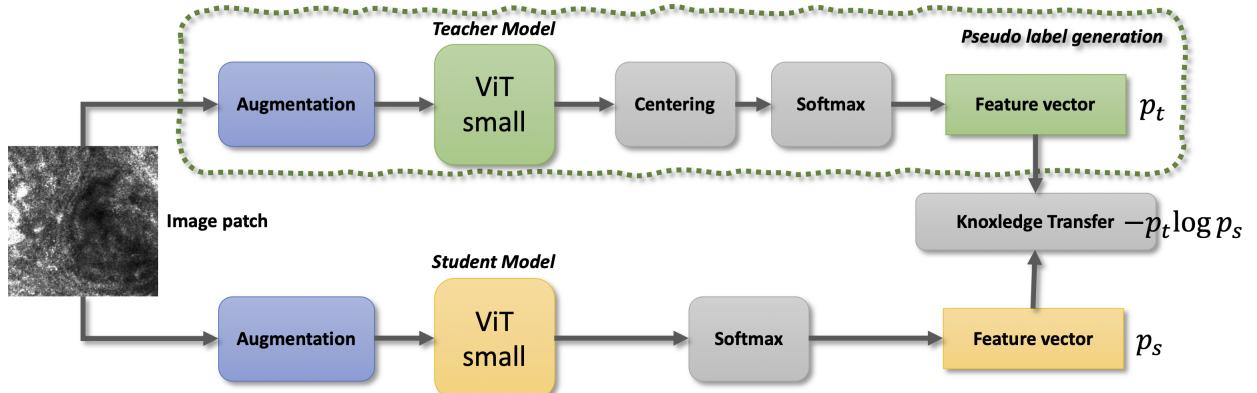


Figure 2: ViT small feature extraction model trained using the image patches from RCM images with the DINO algorithm.

features from image patches from RCM images within our dataset. The DINO algorithm is a self-supervised learning technique that uses knowledge distillation, where a student network learns to mimic the output of a teacher network, to train vision models without labeled data, enabling them to capture rich, high-level representations of images. Its unique approach leverages knowledge distillation techniques in an unsupervised setting, allowing the model to learn robust and meaningful representations of the data without the need for labeled training samples. In our study, we employed the ViT small model to extract a feature embedding consisting of 768 numbers. The DINO algorithm is illustrated in Figure 2.

2.4 Clustering

Following the feature extraction process, we employed the well-regarded k -means clustering algorithm [9] in statistical data analysis to categorize the image patches. The essence of k -means clustering is to partition the data into distinct groups (clusters) by assigning each data point to the cluster with the nearest mean, serving as a prototype. This method effectively grouped the image patches based on their extracted features, revealing patterns and similarities that may not be immediately apparent.

To determine the most effective number of clusters, we utilized the silhouette score [10] as a guiding metric. The silhouette score measures the similarity of an object to its own cluster (cohesion) compared to other clusters (separation).

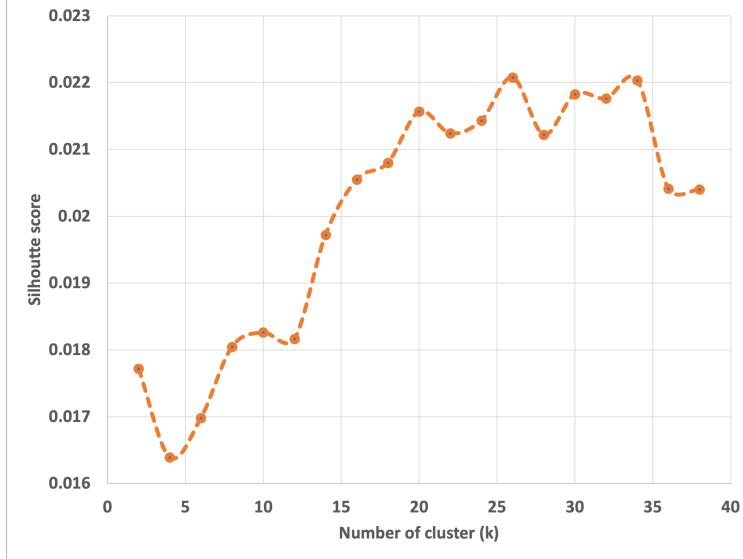


Figure 3: The silhouette scores of the clusters obtained from the image patch features using the k -means clustering algorithm are displayed. The x-axis represents the number of clusters k , while the y-axis denotes the silhouette score corresponding to each k .

By analyzing the silhouette scores for different values of k , we were able to identify the optimal number of clusters that best represented the inherent structure of our data, ensuring a robust and meaningful clustering outcome.

In order to add a layer of clinical relevance and validity to our computational findings, expert dermatologists (AW, JL) were actively involved in the analysis process. These experts meticulously examined the characteristics and clinical implications of each identified cluster. This comprehensive approach, integrating advanced AI algorithms with expert medical knowledge, underscores the multidisciplinary nature of our study.

3 Results

The ViT small model was trained using the DINO algorithm in an unsupervised manner on the Summit supercomputer at the Oak Ridge Leadership Computing Facility (OLCF). The image patches were then used to extract features, which were clustered using the k -means clustering algorithm. Figure 3 illustrates the silhouette scores of the clusters determined by the k -means clustering algorithm for varying values of the number of clusters, k . Despite obtaining the highest silhouette score at $k = 26$, we note a remarkable similarity in the silhouette scores from $k = 18$ to 34 . Consequently, we determine the optimal number of clusters to be $k = 18$, as it marks the inflection point where the trend shifts.

Two expert dermatologists (AW, JL), who have undergone extensive training, conducted a rigorous analysis of the patch images associated with each cluster. They meticulously identified the distinct characteristics and clinical implications of each cluster, as documented in Table 1. Furthermore, the severity of each cluster is visually represented using color codes: green represents regular epidermis, yellow and orange represent irregular epidermis, red represents atypical cells, and blue represents artifacts. In order to provide a visual reference, Figure 4 showcases sample images for each cluster.

Figure 5 displays a set of sample RCM images overlaid with cluster maps. The cluster maps indicate the class numbers and their associated risk levels using color codes. It is important to note that the images showcase the segmentation of cluster map regions based on their clinical characteristics, thereby aiding dermatologists in reading the cases more effectively and confidently diagnosing skin conditions.

4 Discussion

We conducted a proof-of-concept study that combined the latest advancements in AI and machine learning techniques with the clinical expertise of expert clinicians. This interdisciplinary collaboration yielded highly promising outcomes that can significantly advance the field of cancer diagnosis analytics and visualizations. Specifically, our study focused on automatically segmenting regions and visualizing severity in RCM images, and the results were remarkably effective.

Table 1: Clinical implications hold from the image patches in clusters determined by an expert dermatologist with the color codes: green (regular epidermis), yellow and orange (irregular epidermis), red (atypical cells), and blue (artifacts).

Cluster	Description	Color code	Cluster	Description	Color code
0	Regular epidermis	Green	9	Regular epidermis	Green
1	Inflammatory cells	Green	10	Epidermal invaginations	Green
2	Bulbous projections	Green	11	Epidermis streaming with dentritic cells	Orange
3	Bulbous projections, epidermal invaginations	Green	12	Pagetoid spread	Red
4	Artifacts	Blue	13	Irregular epidermis type 4	Yellow
5	Irregular epidermis type 1	Orange	14	Meshwork pattern (moderate atypia)	Red
6	Bundles of atypical cells	Red	15	Meshwork pattern (severe atypia)	Red
7	Irregular epidermis type 2	Yellow	16	Irregular epidermis type 5	Yellow
8	Irregular epidermis type 3	Yellow	17	Regular epidermis	Green

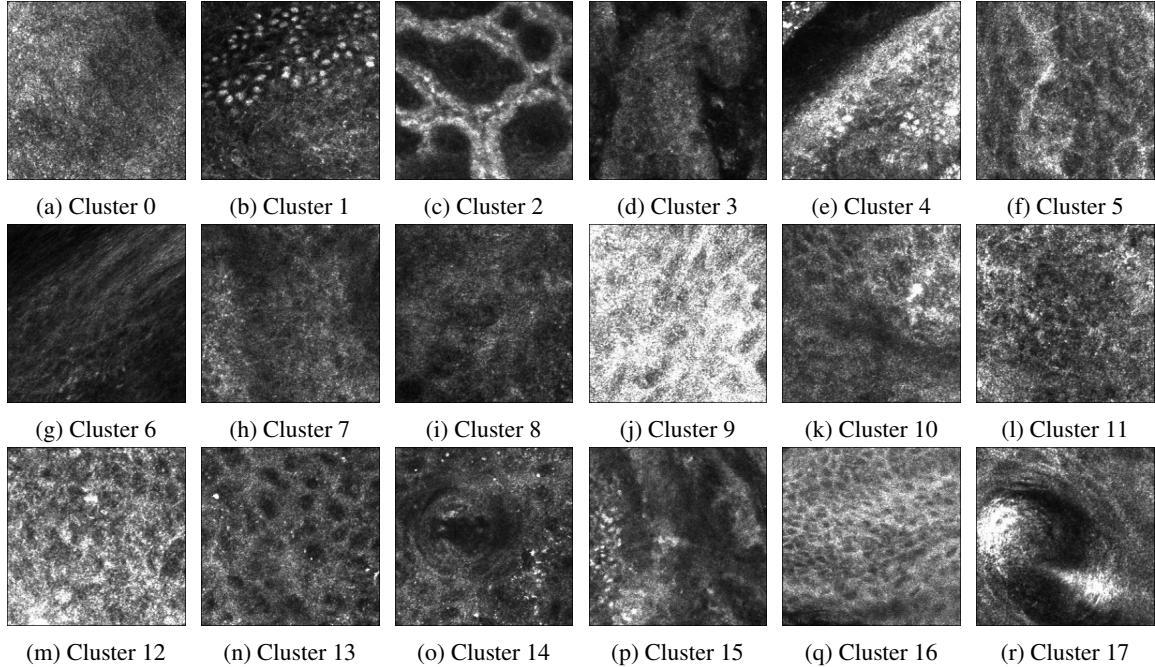


Figure 4: Example of patch images clustered by their image features determined by the ViT model trained by the DINO algorithm, associated with (a) cluster 0, (b) cluster 1, (c) cluster 2, and up to (r) cluster 17.

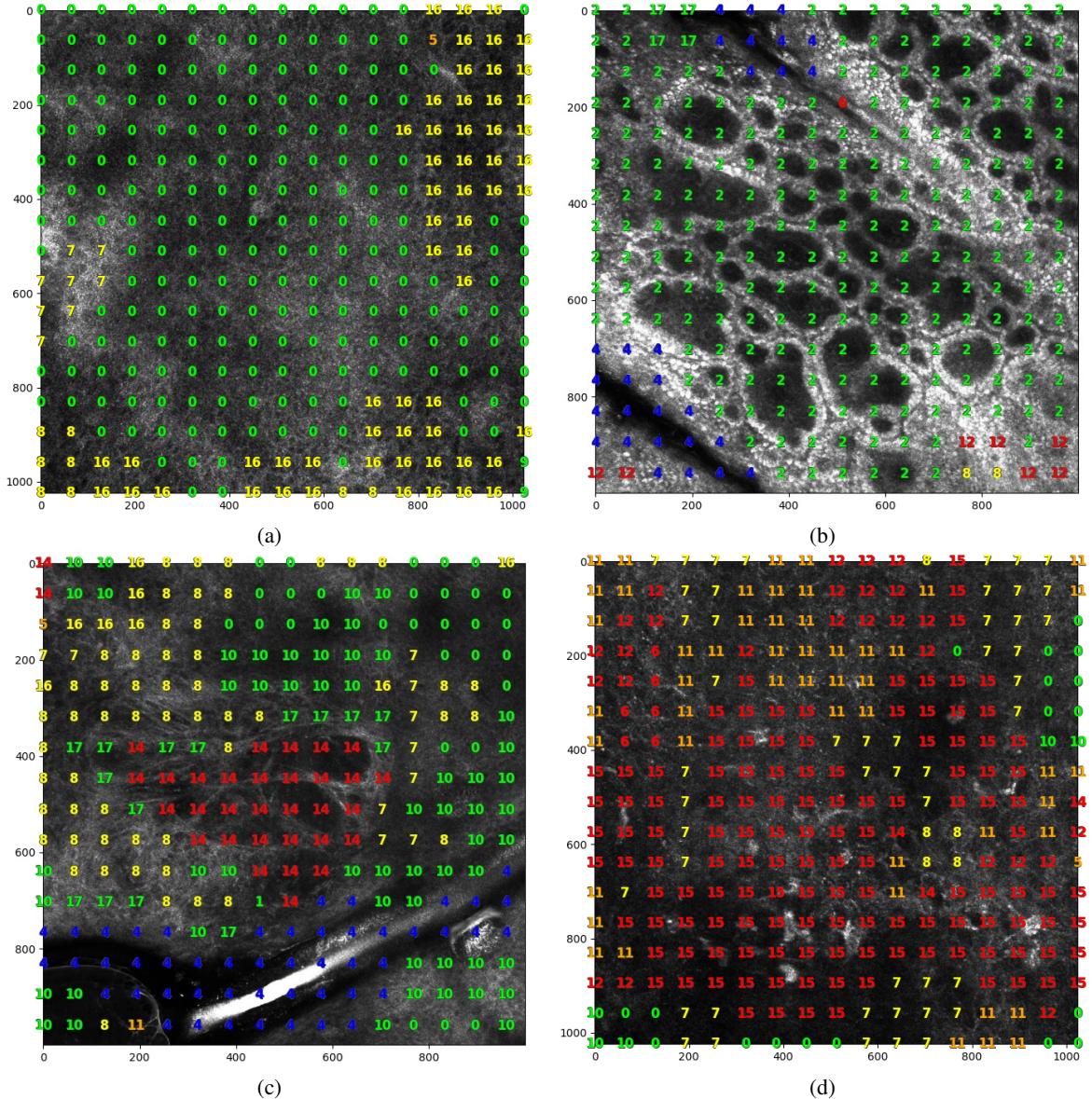


Figure 5: Sample RCM images are superimposed with the cluster map. (a) and (b) biopsy not recommended and (c) and (d) biopsy recommended cases.

For feature extraction, we chose to use the ViT small model, which proved to be a compelling approach. The combination of the ViT small model and the DINO algorithm for training demonstrated its effectiveness in our study. However, further comparative studies are needed to achieve a more comprehensive representation of image features. For example, the ViT models offer three variations (small, base, and large) based on the size of their trainable parameters. Additionally, exploring popular convolutional neural network (CNN) models [11, 12, 13, 14], known for their ability to infer image features rapidly and with lower computational power, could be a valuable avenue for future research. Furthermore, considering other self-supervised learning and unsupervised learning techniques, such as bootstrap your own latent (BYOL)[15], presents intriguing possibilities for follow-up investigations.

While the choice of using k -means clustering was commendable, further studies comparing different clustering algorithms[16, 17, 18, 19] are necessary to gain deeper insights into their performance. Evaluating these algorithms, along with selecting optimal feature extraction models, can be done using the silhouette score as a performance metric. A higher score indicates superior performance. It is important to note that the order of clusters and the association of samples may vary during the clustering algorithm, requiring repetitive evaluation and interpretation of the clinical significance associated with each cluster by expert clinicians if we repeat the clustering process. Since this step is time-consuming and demands significant human effort, minimizing this effort is crucial for achieving study efficiency.

The findings depicted in Figure 5 underscore the efficacy of our algorithm and system in automatically segmenting and determining the severity of abnormalities in RCM images. It is important to highlight that the presence of atypical regions identified by the algorithm does not automatically imply the necessity of a biopsy, as exemplified in Figure 5b. This could be attributed to either misidentification of certain areas or erroneous readings. Furthermore, it is noteworthy that these abnormal regions may not encompass the entirety of the image but can still occupy a considerable portion, as demonstrated in Figure 5c.

Looking ahead, a promising avenue for further scientific investigation and development lies in the utilization of cluster maps in clinical decision-making. For instance, quantifying the extent of connected atypical regions and establishing a threshold could potentially serve as a criterion for clinical judgments. Nonetheless, it is crucial to acknowledge that such research would necessitate a more extensive dataset that accurately represents the entire disease spectrum in order to draw robust and dependable conclusions.

5 Conclusion

In this study, we utilized a self-supervised training algorithm, DINO, to develop an ML model that can extract local textual features from image patches of RCM images. We then grouped the images based on their features using the k -means clustering algorithm. Our results suggest that unsupervised clustering of the features extracted from the images can be used for downstream clinical classification tasks.

Our research indicates that the feature map has clinical relevance and that the cluster map can be utilized as a segmentation tool for the RCM images. This can assist dermatologists in reading cases more effectively and confidently diagnosing skin conditions. Furthermore, our approach could be utilized as an educational tool for dermatologic trainees.

Acknowledgments

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the DOE Office of Science under Contract No. DE-AC05-00OR22725.

The study was supported by the Seed Money Fund program of Oak Ridge National Laboratory, under the Laboratory Directed Research and Development (LDRD) No. 11013.

References

- [1] Piergiacomo Calzavara-Pinton, Caterina Longo, Marina Venturini, Raffaella Sala, and Giovanni Pellacani. Reflectance confocal microscopy for in vivo skin imaging. *Photochemistry and photobiology*, 84(6):1421–1430, 2008.
- [2] Amanda Levine and Orit Markowitz. Introduction to reflectance confocal microscopy and its use in clinical practice. *JAAD case reports*, 4(10):1014–1023, 2018.
- [3] Marek Wodzinski, Andrzej Skalski, Alexander Witkowski, Giovanni Pellacani, and Joanna Ludzik. Convolutional neural network approach to classify skin lesions using reflectance confocal microscopy. In *2019 41st Annual*

- International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4754–4757. IEEE, 2019.
- [4] Miroslawa Sikorska, Andrzej Skalski, Marek Wodzinski, Alexander Witkowski, Giovanni Pellacani, and Joanna Ludzik. Learning-based local quality assessment of reflectance confocal microscopy images for dermatology applications. *Biocybernetics and Biomedical Engineering*, 41(3):880–890, 2021.
 - [5] Gabriele Campanella, Cristian Navarrete-Decent, Konstantinos Liopyris, Jilliana Monnier, Saud Aleissa, Brahmteg Minhas, Alon Scope, Caterina Longo, Pascale Guitera, Giovanni Pellacani, et al. Deep learning for basal cell carcinoma detection for reflectance confocal microscopy. *Journal of Investigative Dermatology*, 142(1):97–103, 2022.
 - [6] Raj H Patel, Emilie A Foltz, Alexander Witkowski, and Joanna Ludzik. Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: A systematic review. *Cancers*, 15(19):4694, 2023.
 - [7] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.
 - [8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
 - [9] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.
 - [10] Irit Gat-Viks, Roded Sharan, and Ron Shamir. Scoring clustering solutions by their biological relevance. *Bioinformatics*, 19(18):2381–2389, 2003.
 - [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
 - [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
 - [13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
 - [14] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
 - [15] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.
 - [16] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.
 - [17] Douglas A Reynolds et al. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.
 - [18] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. Birch: an efficient data clustering method for very large databases. *ACM sigmod record*, 25(2):103–114, 1996.
 - [19] Daniel Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.