2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635546

A TIERED QUADRUPLET NETWORK WITH PATIENT-SPECIFIC MINING AND DYNAMIC MARGIN FOR IMPROVED UGLY DUCKLING LESION CLASSIFICATION

Nathasha Naranpanawa[⋆] *H. Peter Soyer*[†] *Adam Mothershaw*[†]
Gayan K. Kulatilleke[⋆] *Zongyuan Ge*[‡§] *Brigid Betz-Stablein*[†] *Shekhar S. Chandra*[⋆]

[⋆] School of Electrical Engineering and Computer Science, The University of Queensland, Australia
[†] Frazer Institute, The University of Queensland, Dermatology Research Centre, Australia
[‡] Monash Medical AI, Monash University, Australia
[§] Department of Data Science and AI, Faculty of IT, Monash University, Australia

ABSTRACT

The concept of the 'ugly duckling sign' is a criterion used in diagnosing cutaneous melanoma by identifying visually distinct skin lesions. However, variations in pigmented lesions among patients make it challenging to visually distinguish ugly ducklings. Hence, we propose DMT-Quadruplet - a deep metric learning network to learn lesion features at two tiers - patient-level and lesion-level. We introduce a patient-specific quadruplet mining approach within a tiered quadruplet network, to capture contextual information both globally and locally. We further incorporate a dynamic margin within the patient-specific mining to allow more useful quadruplets to be mined within individuals. Comprehensive experiments show that our proposed method outperforms traditional classifiers, achieving significantly higher sensitivity than baseline models, successfully classifying ugly duckling lesions both in patient-specific and patient-agnostic contexts.

Index Terms— Deep Learning, Metric Learning, Ugly Ducklings, Skin Cancer

1. INTRODUCTION

The concept of an 'Ugly Duckling' lesion (UD) was introduced in 1998 [1], as an additional criterion for when a melanoma could not be identified with common criteria such as the ABCDE method [2]. The ugly duckling sign was developed based on the fact that naevi on an individual tend to resemble one another, whereas malignant melanoma often deviates from this nevus pattern and stands out from its peers on a common body region. Hence, an 'Ugly Duckling' lesion is defined as a visually different outlier from its surrounding naevi on an individual, and is considered suspicious for malignancy.

In a clinical setting, practitioners usually observe all lesions on a particular body-site and make a judgement on the visual similarity of the lesions to identify outlier ugly ducklings. However, disparities among individuals and their respective set of lesions can exist in terms of colour, shape, size and distribution (Figure 1). The visual lesion characteristics of one individual might be completely different to that of another [3]. This inconsistency is further indicated by the considerable inter-observer variability in selecting ugly duckling lesion, as expert physicians also differ in clinical experience and visual perception [4, 5].

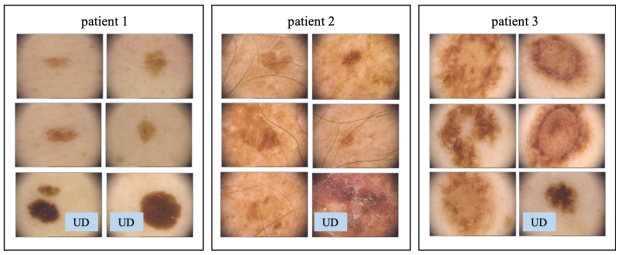


Fig. 1: Visual lesion characteristics vary among patients. Figure shows sets of lesions from 3 different patients. Each lesion set is from the same body-site of the particular patient, and ugly ducklings are marked with a 'UD' label.

When it comes to implementing AI-based ugly duckling recognition methods, disregarding this pattern variability between individuals and training a model only at lesion-level might be disadvantageous, leading to incorrect predictions. Additionally, the ontology of a skin lesion dataset is such that the sub-classes (normal and ugly duckling) are not mutually exclusive among individuals even though intra-patient features might be different. Thus, the classification of ugly duckling lesions pose a unique and challenging task of taking into account both inter-patient (patient-level) and intra-patient (lesion-level) feature representations. Additionally, while some studies consider melanoma as ugly ducklings [6, 7], it's incorrect to interchange annotations for malignant melanoma with ugly ducklings because not all melanomas exhibit the distinctive characteristics associated with ugly duckling lesions [8].

To tackle the above challenges, we aim to use metric learning to learn a similarity metric within data. While

Authorized licensed use limited to: Northwestern University. Downloaded on February 03,2025 at 20:39:28 UTC from IEEE Xplore. Restrictions apply.

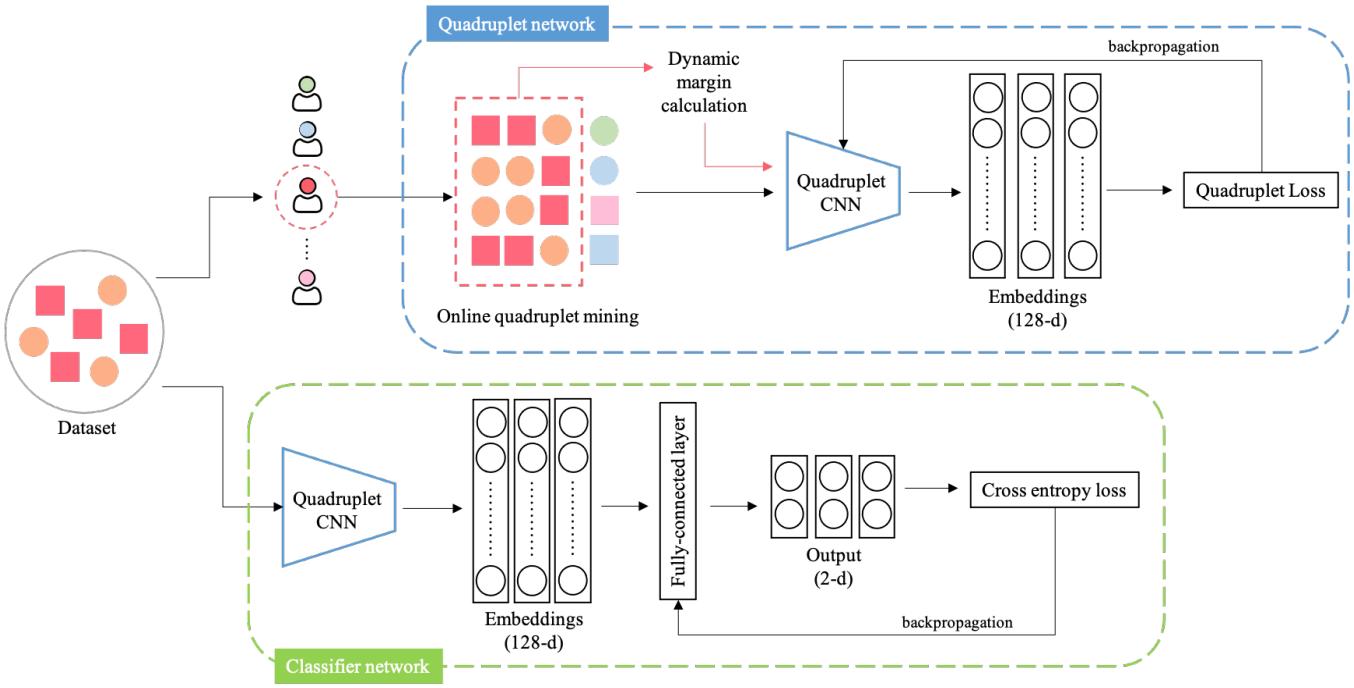


Fig. 2: The proposed DMT-Quadruplet architecture. In the first stage, a Quadruplet network is trained with online patient-specific quadruplet mining and a dynamic margin. The trained quadruplet network is then used as a backbone feature extractor to train a simple CNN classifier in the second stage. During the testing phase, only the trained classifier network from the second stage is used for generating a lower dimensional embedding for each test image and performing binary classification.

a few previous studies have employed wide-field images for ugly duckling classification with deep learning methods [9, 10, 11, 6], no metric learning approaches have been employed for ugly duckling classification. Specifically, our proposed novel framework employs a quadruplet network with patient-specific sample mining and dynamic separating margins, that is capable of remaining patient-agnostic while learning patient-specific representations. Our main contributions are:

- We propose a novel patient-specific metric-learning method for improved classification of ugly duckling lesions in largely imbalanced skin lesion datasets.
- We propose a new patient-specific tiered quadruplet mining strategy that is capable of capturing both local and global feature representations accurately.
- We introduce a novel patient-specific dynamic margin to incorporate better patient-level information.

To the best of our knowledge, this is the first work employing a tiered and patient-specific quadruplet network for ugly duckling lesion classification.

2. METHODS

The overview of the proposed two-stage framework is presented in Figure 2. The first stage involves training a quadruplet network, and the second stage utilizes the trained quadruplet network to generate a latent representation of samples in metric space and to train a classifier on them.

2.1. Tiered Quadruplet Network with a Patient-specific Mining Approach

In the first stage of our system, we implement a two-tier quadruplet network to capture both inter-patient and intra-patient features in the dataset. Tier 1 (lesion-level/inter-patient) considers different individuals in the dataset, while Tier 2 (patient-level/intra-patient) looks at the collection of lesions of each individual (Figure 3). With this, our aim is to identify the similarities between the lesions of individuals, as well as dissimilarities between individuals in the data.

The sample mining procedure during the quadruplet network training is illustrated in Figure 3. We mine a quadruplet of samples across the two tiers of the network in a patient-specific manner. First, we randomly select an X number of individuals at Tier 1, and randomly sample k images from each individual between both classes of lesions. This ensures each mini-batch would see sufficient information of each individual's lesions. Next, at Tier 2, we consider only the samples belonging to a particular individual (x) to mine all triplets of anchor (a^x), positive (p^x), and negative (n^x) samples, where positive sample is of the same class as the anchor and negative sample of the opposite class. Next, we mine a fourth sample

Authorized licensed use limited to: Northwestern University. Downloaded on February 03,2025 at 20:39:28 UTC from IEEE Xplore. Restrictions apply.

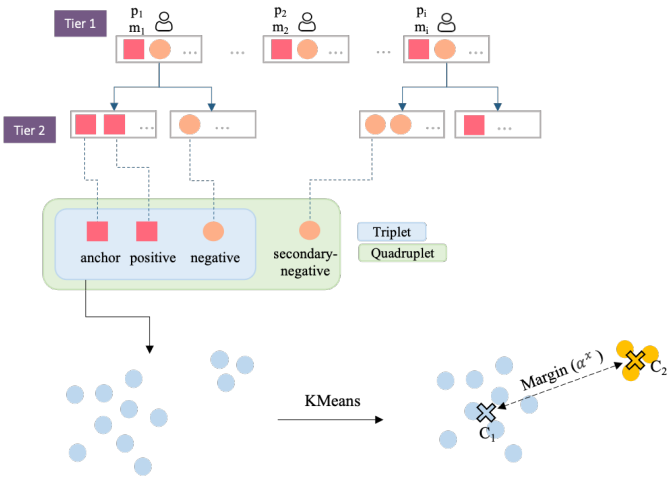


Fig. 3: Patient-specific mining approach. The blue boxes indicate the mining of a triplet within a patient, and the fourth sample is mined as indicated by the green box. The patient-specific dynamic margin is calculated between two distinct clusters of embeddings of an individual using KMeans.

for each (a^x, p^x) pair of individual x. We refer to this additional sample as a secondary-negative (sn) which is mined from a different patient (y) from Tier 1. This secondary-negative can be of the same class or opposite class as the anchor (a^x). We perform online random-hard mining to select the furthest negative and furthest secondary-negative samples. Thus, the Tiered Quadruplet (T-Quad) loss takes in quadruplets of samples in the form (a^x, p^x, n^x, sn^y) for each batch, and the loss is computed as in Equation 1 (patient-specific updates to the traditional triplet loss are shown in blue). Margin α is the fixed Tier 2 margin within patients and β ($> \alpha$) is the Tier 1 margin between patients. Thus, with the tiered loss, the network learns a more global representation of lesions as Tier 2 classes (normal and UD lesions) are mutual across Tier 1 classes (individuals).

$$
\begin{aligned}
L_{patient\ level} &= d(a^x, p^x) - d(a^x, n^x) + \alpha \\
L_{lesion\ level} &= d(a^x, p^x) - d(a^x, sn^y) + \beta \\
L_{T-Quad} &= \frac{1}{N}\sum_{i=1}^{N}[L_{lesion\ level^i} + L_{patient\ level^i}]
\end{aligned}
\tag{1}
$$

2.2. Dynamic margin involved tiered quadruplet loss

We argue that the fixed margin α would be different between individuals based on the previously mentioned observation of their phenotypic features being different. Therefore, based on our assumption that metric learning benefits from patient-specific information to learn a better separation of features, we further extend our loss to include a patient-specific dynamic margin instead of a global fixed margin. As shown in Figure 3, we cluster the embeddings of an individual's samples into two groups using the K-Means algorithm, and the distance between the two cluster centroids is used as the dynamic margin α^x for mining (a^x, p^x, n^x) triplets from that particular individual. Equation 2 shows the updated loss for our final quadruplet network which we refer to as the Dynamic Margin Tiered Quadruplet (DMT-Quad) network.

$$
\begin{aligned}
L_{patient\ level} &= d(a^x, p^x) - d(a^x, n^x) + \alpha^{\mathbf{x}} \\
L_{lesion\ level} &= d(a^x, p^x) - d(a^x, sn^y) + \beta \\
L_{DMT-Quad} &= \frac{1}{N}\sum_{i=1}^{N}[L_{lesion\ level^i} + L_{patient\ level^i}]
\end{aligned}
\tag{2}
$$

2.3. Classification Network

In the second stage of our proposed method, we use the trained quadruplet network as the feature extractor to implement a binary CNN classifier (Figure 2). All layers of the quadruplet network are frozen, and only the parameters of the last layer are updated in the classifier.

3. EXPERIMENT SETUP

3.1. Dataset

The de-identified dataset used in this study is derived from the Brisbane Naevus Morphology Study (BNMS) [12] dataset. All lesions were annotated as normal or ugly duckling by a board-certified dermatologist (HPS) with 30+ years of experience using an in-house annotation software. The final dataset (SkinUD) had a total of 10,493 images from 37 participants, with approximately 1:32 ratio between UD to normal lesions. The same dataset split was employed for the training of both Quadruplet and classifier network. For the training data, images are center-cropped, resized and randomly flipped both vertically and horizontally, and UD samples are duplicated by 10-fold.

3.2. Implementation Details

The quadruplet network that serves as the main backbone of our proposed method is a ResNet18 model where the last layer is substituted with a fully-connected layer, generating a 128-dimensional embedding as the output. For this, we use Adam optimizer with a learning rate of 0.0001, and margin values α and β are 1.0 and 1.5 respectively. During classification, the original last layer of the backbone CNN is added back in for binary classification with Adam optimizer, a Cross Entropy loss and a learning rate of 0.0001. We compare our classification results with ResNet18 as a baseline traditional classifier. All experiments were carried out with PyTorch on a NVIDIA Tesla V100 32GB GPU.

4. RESULTS AND DISCUSSION

Table 1 shows that the Patient-specific Tiered Quadruplet with Dynamic Margin (DMT-Quadruplet) has the highest

Authorized licensed use limited to: Northwestern University. Downloaded on February 03,2025 at 20:39:28 UTC from IEEE Xplore. Restrictions apply.

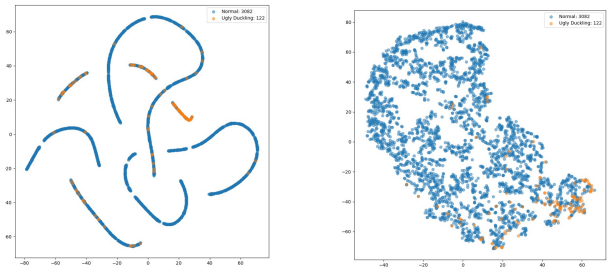
Table 1: Classification performance of metric learning-based classifiers on the test set

	Method	Specificity	Sensitivity	Recall	Precision	F1-score	AUC	Accuracy
Baseline CNN classifiers	ResNet18	96.4±1.3	46.2±6.2	71.3±2.8	66.3±3.0	68.1±2.6	84.0±0.0	94.0±1.4
	ResNet34	96.3±0.6	47.4±4.4	71.9±2.3	66.1±2.7	68.4±2.4	88.7±3.1	94.0±0.8
	VGG16	94.9±0.9	54.6±10.0	74.7±4.5	63.9±0.2	67.4±1.0	**90.7±1.4**	92.7±0.5
	EfficientNeB0	**97.6±0.2**	40.4±0.5	69.0±0.2	**68.9±0.8**	**68.9±0.3**	89.7±1.8	**95.0±0.0**
Traditional metric losses	Naïve Siamese	93.4±0.3	60.1±2.5	76.8±1.1	62.5±0.3	66.3±0.4	86.8±1.4	91.7±0.5
	Naïve Triplet	93.9±1.6	51.9±2.6	76.4±3.5	63.3±1.0	66.9±0.6	88.5±0.0	92.0±1.4
Patient-specific metric losses	Patient-specific Triplet (PS-Triplet)	94.5±3.1	58.8±6.1	73.9±4.2	64.5±3.6	66.8±2.4	87.1±0.0	92.3±2.5
	Patient-specific Tiered Quadruplet (T-Quad)	93.5±1.5	63.0±7.2	78.3±2.8	63.3±0.9	67.2±0.9	88.3±0.0	91.7±1.2
	Patient-specific Tiered Quadruplet + Dynamic Margin (DMT-Quad)	91.7±1.4	**71.2±5.9**	**81.4±2.2**	62.2±0.8	66.3±0.9	90.2±0.0	90.3±1.3

sensitivity measure for the SkinUD dataset at 71.2% accuracy. This is 54% better than the baseline ResNet18 classifier, while being 37% better than the naive triplet classifier. All metric-based classifiers consecutively improve upon the sensitivity measure, and the bias caused by the majority class is reduced. Therefore, the specificity and overall accuracy of the DMT-Quadruplet is lower than the baseline ResNet18 only by 4.8% and 3.9% respectively. The DMT-Quadruplet reports the highest recall and ROC AUC as well. The results indicate that the tiered quadruplet architecture is capable of incorporating more lesion-level information with the help of the secondary-negative, which provides the network with an additional global view of lesion features.

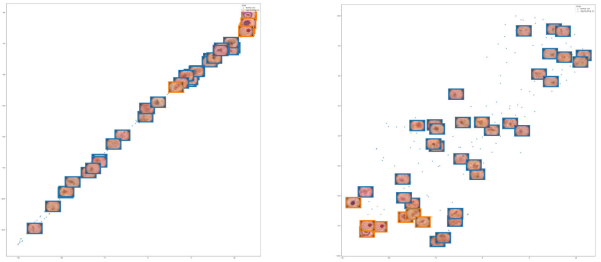
Overall, coupled with the patient-specific mining approach, computing an online dynamic margin of separation boosts the performance of the DMT-Quadruplet network as a fixed margin would have discarded useful triplets inaccurately during the online mining process. Further, by setting a large value for the patient-level separating margin β of the DMT-Quadruplet, the network prioritizes learning global features over local features during training, allowing the network to remain patient-agnostic at lesion-level while separating lesions at patient-level accurately.

We applied t-SNE algorithm to obtain 2d embeddings for samples in the test set to illustrate the cluster separations as shown in Figure 4. The manifold of the Naive triplet does not show a good separation for UD lesions (Figure 4(a)) with UDs spread over the manifold, but the DMT-Quadruplet shows a better separation by clustering UD lesions closer together (Figure 4(b)). Similarly, the DMT-Quadruplet shows a better separation for all lesions of the same individual (Figure 5). Thus, the DMT-Quadruplet, despite being patient-specific in triplet mining, can effectively separate lesions within an individual in a patient-agnostic manner.



(a) For the Naive triplet (b) For the DMT-Quadruplet

Fig. 4: The 2D separation manifold in metric space for all samples in the test set using (a) the Naive triplet and (b) the DMT-Quadruplet to generate embeddings.



(a) For the Naive triplet (b) For the DMT-Quadruplet

Fig. 5: The 2D separation manifold with image tiles in metric space for the same individual in the test set using (a) the Naive triplet and (b) the DMT-Quadruplet to generate embeddings.

5. CONCLUSION

In this work, we present a a novel patient-specific metric learning method for improved classification of ugly duckling lesions. A patient-specific quadruplet sampling strategy between two tiers of information is explored to incorporate more global context of features, and an inter-patient dynamic margin of separation is included to mine more useful local features. Experimental results demonstrate that the proposed model effectively surpasses traditional classification methods for identifying minority class lesions, while also successfully handling datasets with major class imbalances.

In conclusion, with our two-stage pipeline combining a quadruplet-based feature extractor and a classifier, we can effectively separate lesions of an individual and accurately classify ugly duckling lesions. In clinical application, our method will be particularly useful for patients who have many naevi (>100) as assessing each lesion individually is time consuming and unrealistic for a clinician. Applying this alongside an ABCDE algorithm might provide an image triage so the dermatologist/clinician only needs to assess the most suspicious lesions, saving time and effort. As such, the proposed method can successfully assist clinicians in early melanoma detection

Authorized licensed use limited to: Northwestern University. Downloaded on February 03,2025 at 20:39:28 UTC from IEEE Xplore. Restrictions apply.

as ugly duckling lesions are an indicator of a potential malignant melanoma developing.

6. COMPLIANCE WITH ETHICAL STANDARDS

The images used in this study are derived from the BNMS – Brisbane Naevus Morphology Study which was approved by the Metro South Human Research Ethics Committee (#HREC/09/QPAH/ 162, 26/08/2009) and The University of Queensland (#2009001590, 14/10/ 2009) and conducted in accordance with the Declaration of Helsinki.

7. CONFLICTS OF INTEREST

HPS is a shareholder of MoleMap NZ Limited and e-derm consult GmbH and undertakes regular teledermatological reporting for both companies. HPS is a Medical Consultant for Canfield Scientific Inc, Blaze Bioscience Inc, and a Medical Advisor for First Derm.

8. REFERENCES

[1] J. Grob and J. Bonerandi, "The 'ugly duckling' sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening," *Archives of Dermatology*, vol. 134, no. 1, pp. 103–a–104, jan 1998.

[2] D. S. Rigel, R. J. Friedman, A. W. Kopf, and D. Polsky, "ABCDE - An evolving concept in the early detection of melanoma," pp. 1032–1034, aug 2005.

[3] Y. Wazaefi, C. Gaudy-Marqueste, M. F. Avril, J. Malvehy, G. Pellacani, L. Thomas, R. Triller, Y. Bruneu, S. Monestier, M. A. Richard, B. Fertil, and J. J. Grob, "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," *The Journal of investigative dermatology*, vol. 133, no. 10, pp. 2355–2361, 2013.

[4] C. Gaudy-Marqueste, Y. Wazaefi, Y. Bruneu, R. Triller, L. Thomas, G. Pellacani, J. Malvehy, M. F. Avril, S. Monestier, M. A. Richard, B. Fertil, and J. J. Grob, "Ugly duckling sign as a major factor of efficiency in melanoma detection," *JAMA Dermatology*, vol. 153, no. 4, pp. 279–284, apr 2017.

[5] A. Scope, S. W. Dusza, A. C. Halpern, H. Rabinovitz, R. P. Braun, I. Zalaudek, G. Argenziano, and A. A. Marghoob, "The "ugly duckling" sign: Agreement between observers," *Archives of Dermatology*, vol. 144, no. 1, pp. 58–64, jan 2008.

[6] J. B. Garcia, S. Tanadini-Lang, N. Andratschke, M. Gassner, and R. Braun, "Suspicious Skin Lesion Detection in Wide-Field Body Images using Deep Learning Outlier Detection," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2022-July, pp. 2928–2932, 2022.

[7] Z. Yu, V. Mar, A. Eriksson, S. Chandra, P. Bonnington, L. Zhang, and Z. Ge, "End-to-End Ugly Duckling Sign Detection for Melanoma Identification with Transformers," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12907 LNCS. Springer Science and Business Media Deutschland GmbH, 2021, pp. 176–184.

[8] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, M. Combalia, S. Dusza, P. Guitera, D. Gutman, A. Halpern, B. Helba, H. Kittler, K. Kose, S. Langer, K. Lioprys, J. Malvehy, S. Musthaq, J. Nanda, O. Reiter, G. Shih, A. Stratigos, P. Tschandl, J. Weber, and H. P. Soyer, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific Data*, vol. 8, no. 1, pp. 1–8, jan 2021.

[9] J. S. Birkenfeld, J. M. Tucker-Schwartz, L. R. Soenksen, J. A. Avilés-Izquierdo, and B. Marti-Fuster, "Computer-aided classification of suspicious pigmented lesions using wide-field images," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105631, 2020.

[10] M. Mohseni, J. Yap, W. Yolland, A. Koochek, and M. S. Atkins, "Can self-training identify suspicious ugly duckling lesions?" in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2021, pp. 1829–1836.

[11] L. R. Soenksen, T. Kassis, S. T. Conover, B. Marti-Fuster, J. S. Birkenfeld, J. Tucker-Schwartz, A. Naseem, R. R. Stavert, C. C. Kim, M. M. Senna, J. Avilés-Izquierdo, J. J. Collins, R. Barzilay, and M. L. Gray, "Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images," *Science Translational Medicine*, vol. 13, no. 581, 2021.

[12] R. A. Sturm, C. Fox, P. McClenahan, K. Jagirdar, M. Ibarrola-Villava, P. Banan, N. C. Abbott, G. Ribas, B. Gabrielli, D. L. Duffy, and H. P. Soyer, "Phenotypic characterization of nevus and tumor patterns in MITF E318K mutation carrier melanoma patients," *Journal of Investigative Dermatology*, vol. 134, no. 1, pp. 141–149, 2014.

Authorized licensed use limited to: Northwestern University. Downloaded on February 03,2025 at 20:39:28 UTC from IEEE Xplore. Restrictions apply.