DermaNet: Flexible-View Fusion for Dermoscopic–Clinical Skin Lesion Classification

Anil Tan Aktan*, Mehmet Umut Sen, Erchan Aptoula, and Berrin Yanikoglu

Sabanci University, Istanbul, Türkiye
{tan.aktan,umut.sen,erchan.aptoula,berrin}@sabanciuniv.edu

Abstract. Skin lesion classification remains challenging because of class imbalance, substantial intra-class variation, and subtle differences between diagnostic categories. Paired dermoscopic and clinical images provide complementary visual evidence, but effectively leveraging widely available single-view data when learning a dual-view model remains challenging. We propose *DermaNet*, a unified flexible-view framework that learns a shared lesion representation from paired dermoscopic and clinical images while accommodating single-view inputs. Separate EfficientNetV2 backbones extract representations from multiple network depths, which are combined within each view through learned weighted depth aggregation. The resulting view-specific embeddings are fused before classification, while modality dropout exposes the model to paired- and single-view conditions during training. DermaNet also uses a clinically motivated hierarchical prediction strategy, in which a group-level classifier first separates melanocytic from non-melanocytic lesions and group-specific classifiers then perform fine-grained diagnosis within each category. Progressive experiments show that multi-level aggregation, hierarchical prediction, and unified feature-level fusion provide complementary improvements and support both paired and single-view inference. DermaNet achieves 49.7% Macro-F1 on MILK10k and 77.42% Micro-F1 on the official ISIC2019 test set while increasing the parameter count by only 0.4%. An implementation is available at `https://github.com/aniltanaktan/DermaNet`.

Keywords: Skin lesion classification · feature-level fusion · modality dropout · hierarchical classification · multi-level representation learning.

1 Introduction

Skin cancer is among the most common malignancies worldwide, and early detection can improve treatment outcomes [4]. Although deep learning has advanced computer-aided skin lesion diagnosis [7], classification remains challenging because of class imbalance, visual similarity between diagnostic categories, intra-class variation, and domain shift across devices and clinics [11].

The International Skin Imaging Collaboration (ISIC) provides benchmarks for both dermoscopy-only and paired-view lesion classification [6,10,17,23]. The

* Corresponding author.

ISIC2019 dataset contains dermoscopic images from eight lesion classes, whereas the MILK10k dataset provides paired dermoscopic and clinical images and includes unknown benign and malignant categories. The paired dermoscopic and clinical views provide complementary diagnostic information about the same lesion, forming a paired dual-view classification setting [9,12].

Prior work has combined dermoscopic images with clinical photographs or patient metadata [12]. Skin lesion studies have also used multi-scale inputs or intermediate backbone features [3], while broader image-analysis work has demonstrated the value of intermediate representations and deep supervision [1,28]. Recent dermatological AI has moved toward richer multimodal and clinically informed systems, including MIDAS [5], seven-point checklist knowledge integration [25], and agentic multimodal reasoning [15]. Missing-modality learning has also been studied through modality masking [18], including incomplete multimodal skin lesion diagnosis [24]. However, multi-level representation learning, dermoscopic–clinical fusion, flexible view availability, and clinically structured prediction are typically addressed separately. DermaNet brings these elements together within a unified lesion-classification framework.

Building on these directions, we propose *DermaNet*, a unified flexible-view framework that combines multi-level aggregation, dermoscopic–clinical feature fusion, modality dropout, and clinically structured prediction. Separate modality-specific backbones extract and aggregate representations from multiple depths before feature-level fusion. Modality dropout exposes the model to paired- and single-view configurations, while a melanocytic/non-melanocytic hierarchy supports fine-grained diagnosis. We evaluate these components through progressive and matched experiments on MILK10k and ISIC2019.

Our main contributions are:

- **Unified flexible-view learning:** We introduce a dual-backbone fusion framework that supports paired dermoscopic–clinical inputs as well as single-view data through feature-level fusion and view-level modality dropout.
- **Multi-level representation aggregation:** DermaNet combines shallow, intermediate, and deep features using learned, modality-specific depth weights. This provides a compact representation that integrates information from multiple backbone depths with negligible computational overhead.
- **Clinically structured classification:** We organize lesion classes into melanocytic and non-melanocytic groups and compare hard hierarchical routing with soft probabilistic routing for fine-grained classification.

2 Proposed Approach

DermaNet supports dermoscopic-only, clinical-only, and paired dermoscopic–clinical inputs. Each available view is processed by a modality-specific EfficientNetV2 backbone, with multi-depth features aggregated by learned depth weights and fused into a shared lesion embedding (Sec. 2.1). During training, modality dropout randomly masks one view in paired samples, enabling the fusion module to learn from both complete and incomplete inputs (Sec. 2.2). The shared

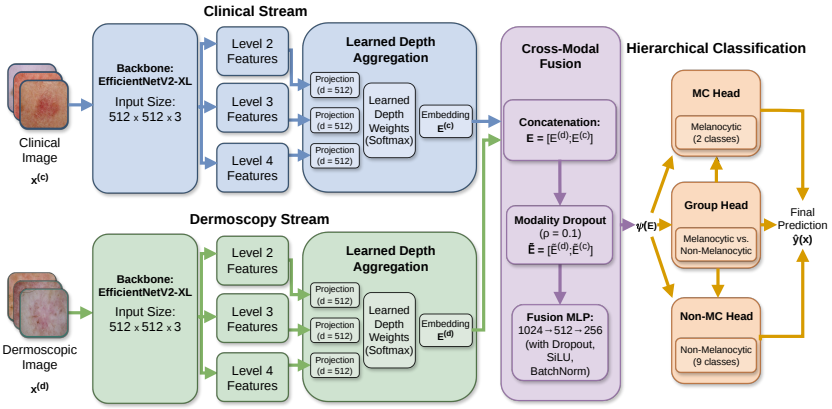


Fig. 1: Overview of DermaNet. Modality-specific backbones extract multi-level features, which are aggregated within each view and fused at the feature level. During training, modality dropout masks one available view. The fused representation is evaluated using melanocytic/non-melanocytic structured prediction.

lesion representation is classified with a melanocytic/non-melanocytic hierarchy where a group head predicts the parent group, and group-specific heads perform fine-grained diagnosis (Sec. 2.3).

2.1 Multi-Level Aggregation and Feature-Level Fusion

For each view, DermaNet extracts feature maps from three EfficientNetV2 stages, indexed as levels 2, 3, and 4, to capture shallow, intermediate, and deep cues. Each feature map is globally pooled and projected to a common embedding space before learned aggregation and cross-view fusion.

For each modality $m \in \{d, c\}$, where d and c denote dermoscopic and clinical inputs, respectively, we use a modality-specific EfficientNetV2 backbone $F^{(m)}(\cdot)$ and extract intermediate feature maps at three depths $\ell \in \{2, 3, 4\}$:

$$R_\ell^{(m)} = F_\ell^{(m)}(x^{(m)}) \in \mathbb{R}^{H_\ell \times W_\ell \times C_\ell}.$$

The three levels capture progressively deeper representations. We compress each block by global average pooling (GAP):

$$\boldsymbol{h}_\ell^{(m)} = \mathrm{GAP}\left(R_\ell^{(m)}\right) \in \mathbb{R}^{C_\ell}. \tag{1}$$

Since channel dimensions vary across depths, we project all levels into a shared 512-dimensional embedding space:

$$\boldsymbol{e}_\ell^{(m)} = \phi_\ell\left(\boldsymbol{h}_\ell^{(m)}\right), \quad \phi_\ell : \mathbb{R}^{C_\ell} \to \mathbb{R}^{512} \tag{2}$$

Learned Depth Aggregation. Rather than concatenating all depth-specific embeddings directly, we learn their relative contribution within each modality. This produces a compact representation while allowing the network to emphasize the depth that is most useful for the classification objective. For each modality $m \in \{d, c\}$, we define three modality-specific trainable depth logits, shared across all samples, $\boldsymbol{w}^{(m)} \in \mathbb{R}^3$ and normalize them using softmax:

$$\boldsymbol{\alpha}^{(m)} = \mathrm{softmax}\Big(\boldsymbol{w}^{(m)}\Big), \qquad \boldsymbol{E}^{(m)} = \sum_{\ell \in \{2,3,4\}} \alpha_{\ell}^{(m)} \boldsymbol{e}_{\ell}^{(m)}. \tag{3}$$

The learned weights $\boldsymbol{\alpha}^{(m)}$ determine the contribution of each backbone depth to the modality representation $\boldsymbol{E}^{(m)}$. The clinical and dermoscopic embeddings are then concatenated and mapped to a fused representation:

$$\boldsymbol{E} = \Big[\boldsymbol{E}^{(c)}; \boldsymbol{E}^{(d)}\Big] \in \mathbb{R}^{1024}, \quad \boldsymbol{u} = \psi(\boldsymbol{E}) \in \mathbb{R}^{256}, \tag{4}$$

where $\psi(\cdot)$ is an MLP with layers $1024 \to 512 \to 256$, BatchNorm, SiLU, and dropout. The fused representation $\boldsymbol{u}$ is shared by all classification heads.

2.2 View-Level Modality Dropout

Paired datasets such as MILK10k provide both dermoscopic and clinical images, whereas ISIC2019 contains dermoscopic images only. This imbalance can bias joint training toward the consistently available dermoscopic stream and weaken the contribution of clinical features. Following established missing-modality training strategies [18,24], we adopt view-level *modality dropout*, which randomly masks one available view in paired training samples and deterministically masks unavailable views in single-view samples.

Let $a^{(m)} \in \{0, 1\}$ denote the availability of modality $m \in \{d, c\}$. For paired samples, the availability vector is sampled as

$$(a^{(d)}, a^{(c)}) \leftarrow \begin{cases} (0,1), & \text{with probability } \rho, \\ (1,0), & \text{with probability } \rho, \\ (1,1), & \text{with probability } 1-2\rho, \end{cases} \tag{5}$$

where $\rho = 0.1$ in the reported experiments. Thus, each view is masked in 10% of paired training samples, while both views remain available in the remaining 80%. For single-view inputs, the unavailable modality is deterministically masked; for example, dermoscopy-only samples use $(a^{(d)}, a^{(c)}) = (1, 0)$.

2.3 Clinically Structured Hierarchical Classification

To incorporate clinically meaningful structure into fine-grained lesion classification, we organize the diagnostic labels into melanocytic (MC) and non-melanocytic (NonMC) groups. DermaNet jointly predicts the parent group and

the diagnostic class using group-specific heads. This coarse-to-fine formulation explicitly represents relationships among diagnostic classes rather than treating all labels as unrelated alternatives [20]. We define:

$$\mathcal{Y}_{\mathrm{MC}} = \{\texttt{MEL}, \texttt{NV}\}, \quad \mathcal{Y}_{\mathrm{NonMC}} = \mathcal{Y} \setminus \mathcal{Y}_{\mathrm{MC}}.$$

Let $g(y)$ map each class $y \in \mathcal{Y}$ to its parent group. From the fused representation $\boldsymbol{u}$, the group head predicts the coarse group and two group-specific heads predict the final class within each group:

$$\boldsymbol{z}_g = W_g \boldsymbol{u} + \boldsymbol{b}_g, \tag{6}$$

$$\boldsymbol{z}_r = W_r \boldsymbol{u} + \boldsymbol{b}_r, \qquad r \in \{\mathrm{MC}, \mathrm{NonMC}\}. \tag{7}$$

Here, $\boldsymbol{z}_g \in \mathbb{R}^2$ and $\boldsymbol{z}_r \in \mathbb{R}^{|\mathcal{Y}_r|}$. The group posterior is $p_\theta(g \mid x) = [\mathrm{softmax}(\boldsymbol{z}_g)]_g$. We evaluate two inference rules based on the same trained heads: hard hierarchical routing, which restricts prediction to the most probable group, and soft hierarchical probability composition, which retains uncertainty over both groups.

Hard Hierarchical Routing. For hard routing, the group with the highest posterior probability is selected:

$$\hat{g}(x) = \underset{g \in \{\mathrm{MC}, \mathrm{NonMC}\}}{\arg\max} \; p_\theta(g \mid x). \tag{8}$$

The group head and both group-specific heads are evaluated in a single forward pass. Each sample contributes to the group loss and to the loss of its corresponding subgroup head. For sample (x_i, y_i), let $g_i = g(y_i)$ denote its group and $y_i^{(g_i)}$ its within-group class index. Using cross-entropy loss $\mathrm{CE}(\cdot, \cdot)$, the joint objective for sample (x_i, y_i) is

$$\mathcal{L}(i) = \lambda_g \, \mathrm{CE}(g_i, \boldsymbol{z}_g(x_i)) + \lambda_s \, \mathrm{CE}\Big(y_i^{(g_i)}, \boldsymbol{z}_{g_i}(x_i)\Big), \tag{9}$$

where each sample updates the group head and its corresponding subgroup head. The weights λ_g and λ_s control the relative contribution of group-level and within-group supervision.

Soft Probability Composition. The same trained heads can instead be combined without committing to a single parent group. For each class y, we compose its parent-group probability with its conditional within-group probability:

$$p_\theta(y \mid x) = p_\theta\big(g(y) \mid x\big) \, p_\theta\big(y \mid x, g(y)\big), \qquad y \in \mathcal{Y}. \tag{10}$$

This produces a normalized posterior over the complete class set and preserves uncertainty when the group prediction is ambiguous. On the official ISIC2019 test set, soft routing achieved 77.40% Micro-F1, close to the 77.42% obtained with hard routing. We therefore use hard routing in the main progression experiments.

3 Experiments and Results

Datasets and Evaluation Protocols. We evaluate DermaNet on MILK10k [17] and ISIC2019 [6,10,23]. MILK10k is the paired dermoscopic–clinical benchmark with 11 classes and official-test Macro-F1 as the primary metric. ISIC2019 is the dermoscopy-only benchmark with eight diagnostic classes. We evaluate the official test split after excluding unknown (UNK) samples, consistent with the prior literature considered here, and additionally report a 70:10:20 patient-level train–validation–test split.

Implementation Details. For the main progression experiments, MILK10k and ISIC2019 models were trained separately. ISIC2019 models used dermoscopic images only, with the clinical branch masked in DermaNet configurations. Model-0 is the EfficientNetV2 baseline, Model-1 adds learned multi-level aggregation, and Model-2 adds hierarchical prediction. Since clinical images are unavailable for ISIC2019, Model-3 is not applicable.

Motivated by prior ISIC2019 benchmarking that reported strong performance of EfficientNet-style backbones [2], we use EfficientNetV2-XL [22]. All input images (RGB) are resized to 512×512 pixels. Models are trained with AdamW [16] (lr 9.5×10^{-5}, wd 7.25×10^{-5}) using batch size 16 and gradient accumulation 4. Dropout and drop-path are set to 0.4 and 0.1, respectively.

Table 1: Progressive performance on the official MILK10k and ISIC2019 test sets. Models 0–2 define the single-view component progression. For MILK10k, Model 3 replaces score-level fusion with paired feature-level fusion. For ISIC2019, paired fusion is not applicable because clinical images are unavailable.

Description	**MILK10k** (Macro-F1)	**ISIC2019** (Micro-F1)
Model-0: Baseline		
Two separately trained single-modality EfficientNetV2 models, score-level combined at inference.	40.3	73.63
Model-1: Baseline + Multi-Level Features		
Baseline model with multi-level features (levels 2/3/4) with learned depth aggregation.	41.2	75.18
Model-2: Model-1 + Hierarchical Decision		
A 2-group hierarchy (melanocytic vs. non-melanocytic) added to each single-modality model.	45.4	**77.42**
Model-3: Model-2 + Paired Feature-Level Fusion		
Feature-level dermoscopic–clinical fusion for MILK10k; dermoscopy-only operation for ISIC2019.	**49.7**	–

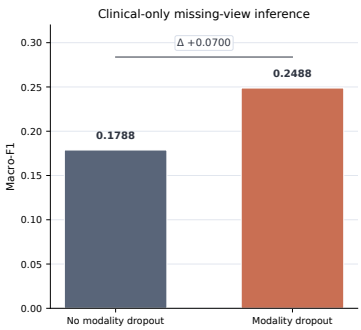
MILK10k: Contribution of Paired Views. On the MILK10k validation set, independently trained EfficientNetV2-XL baselines achieved 42.19% F1 with

clinical images and 36.95% F1 with dermoscopic images. Their score-level combination improved F1 to 43.67%, indicating complementary information across views. Since score-level fusion combines predictions only after independent classification, this motivates DermaNet's feature-level fusion before the final decision.

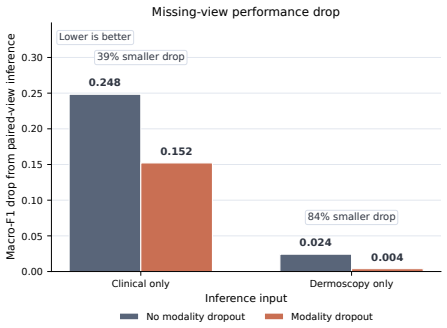
Progressive Analysis of Proposed Architecture. Table 1 summarizes the contribution of each component on MILK10k and ISIC2019. Across both datasets, learned depth aggregation provides consistent gains, indicating that useful diagnostic cues are distributed across backbone depths. For the final paired MILK10k model, the learned L2/L3/L4 weights were 0.078/0.203/0.719 for the clinical stream and 0.086/0.222/0.692 for the dermoscopic stream, with both assigning the largest weight to L4. The melanocytic vs non-melanocytic hierarchy yields a larger improvement, supporting clinically structured supervision for fine-grained classification.

Despite these gains, DermaNet adds only 1.67M parameters over the score-level baseline (413.72M vs. 412.05M; +0.4%) and keeps estimated per-lesion compute essentially unchanged (187.275 vs. 187.272 GMACs).

Modality Dropout and Missing-View Robustness. We assessed modality dropout with a matched mixed-view ablation on combined MILK10k+ISIC data: 4,716 paired samples (7.3%) and 60,057 dermoscopy-only samples (92.7%). Two identical dual-stream hierarchical models differed only in view-level masking: none for the reference model and random masking of either stream with probability 0.1 for the modality-dropout model. Held-out paired samples were evaluated with paired, clinical-only, and dermoscopy-only inputs. Fig. 2 reports clinical-only Macro-F1 and the Macro-F1 drop from paired-view inference. Modality dropout improves clinical-only performance and reduces the missing-view drop, while dermoscopy-only inference remains stable.



(a) Clinical-only Macro-F1.



(b) Missing-view performance drop.

Fig. 2: Effect of modality dropout on missing-view robustness. (a) Clinical-only Macro-F1. (b) Macro-F1 drop from paired-view inference; lower is better.

ISIC2019: Comparison to Prior Work. Table 2 compares DermaNet with selected ISIC2019 leaderboard entries and published methods under the official test protocol and a 70:10:20 split. DermaNet achieves competitive results in both settings, supporting the effectiveness of multi-level aggregation and hierarchical prediction for dermoscopy-only classification.

Table 2: ISIC2019 results on the official test set and a 70:10:20 train–validation–test split. Best results within each protocol are shown in bold.

Method	Micro-F1	Macro-F1	AUC	Bal. Acc.
Official ISIC2019 Test Set				
DaisyLAB (2019) [8]	68.5%	–	91.3%	63.6%
DysionAI (2019) [27]	70.3%	–	78.0%	60.7%
AImageLab–PRHLT (2019) [19]	70.9%	–	88.6%	59.3%
DermaNet (official test)	**77.4%**	**69.5%**	**93.2%**	**67.0%**
Literature 70:10:20 Train–Validation–Test Split				
Sun *et al.* (2021) [21]	–	–	91.5%	66.2%
Yao *et al.* (2022) [26]	75.0%	–	–	–
Li *et al.* (2022) [13]	75.1%	–	–	–
Alsahafi *et al.* (2023) [3]	71.3%	–	–	–
Liu *et al.* (2026) [14]	79.9%	68.9%	–	–
DermaNet (70:10:20 split)	**82.6%**	**72.7%**	**95.4%**	**70.8%**

4 Conclusion

We presented *DermaNet*, a flexible-view framework for dermoscopic–clinical lesion classification. DermaNet combines learned multi-level aggregation, feature-level fusion, modality dropout, and melanocytic/non-melanocytic hierarchical prediction within a unified dual-backbone architecture. Progressive experiments show that clinically structured prediction and paired feature-level fusion provide the largest gains, while multi-level aggregation improves performance with negligible computational overhead. The MILK10k results support learning a shared representation from paired dermoscopic and clinical views, and the ISIC2019 results show that the multi-level and hierarchical components remain effective in dermoscopy-only classification. The missing-view ablation further suggests that modality dropout can improve robustness when one view is unavailable. Future work will systematically study modality dropout across masking probabilities, training-set compositions, and missing-view evaluation protocols, and explore alternative clinically motivated hierarchy definitions beyond melanocytic/non-melanocytic grouping.

Acknowledgments. This work was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under Project No. 124N417.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ahmed, S.A.A., Yanikoglu, B.: Within-network ensemble for face attributes classification. In: Image Analysis and Processing – ICIAP 2019. Lecture Notes in Computer Science, vol. 11751, pp. 466–476. Springer (2019). https://doi.org/10.1007/978-3-030-30642-7_42
2. Aktan, A.T., Almouradi, A., Aptoula, E., Yanikoglu, B.: Comparison of leading deep learning architectures on skin lesion classification using the isic 2019 dataset. In: 2025 Medical Technologies Congress (TIPTEKNO). pp. 1–4 (2025). https://doi.org/10.1109/TIPTEKNO68206.2025.11270128
3. Alsahafi, Y.S., Kassem, M.A., Hosny, K.M.: Skin-net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier. Journal of Big Data **10**(1), 105 (2023)
4. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **74**(3), 229–263 (2024)
5. Chiou, A.S., Omiye, J.A., Gui, H., Swetter, S.M., Ko, J.M., Gastman, B., Arbesman, J., Cai, Z.R., Gevaert, O., Sadée, C., Rotemberg, V.M., Han, S.S., Tschandl, P., Dickman, M., Bailey, E.E., Bae, G.H., Bailin, P., Boldrick, J., Yekrang, K., Caroline, P., Hanna, J., Kurtansky, N.R., Weber, J., See, N.A., Phung, M., Gallegos, M., Daneshjou, R., Novoa, R.A.: Multimodal image dataset for ai-based skin cancer (MIDAS) benchmarking. NEJM AI **2**(6) (2025). https://doi.org/10.1056/AIdbp2400732
6. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 168–172. IEEE (2018)
7. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. nature **542**(7639), 115–118 (2017)
8. Gessert, N., Nielsen, M., Shaikh, M., Werner, R., Schlaefer, A.: Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. MethodsX **7**, 100864 (2020)
9. Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I., Chang, S.E.: Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. Journal of Investigative Dermatology **138**(7), 1529–1538 (2018)
10. Hernández-Pérez, C., Combalia, M., Podlipnik, S., Codella, N.C., Rotemberg, V., Halpern, A.C., Reiter, O., Carrera, C., Barreiro, A., Helba, B., et al.: Bcn20000: Dermoscopic lesions in the wild. Scientific data **11**(1), 641 (2024)

11. Kassem, M.A., Hosny, K.M., Damaševičius, R., Eltoukhy, M.M.: Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. Diagnostics **11**(8), 1390 (2021)
12. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE journal of biomedical and health informatics **23**(2), 538–546 (2019). https://doi.org/10.1109/JBHI.2018.2824327
13. Li, T., Cao, P., Yuan, Y., Fan, L., Yang, Y., Feris, R.S., Indyk, P., Katabi, D.: Targeted supervised contrastive learning for long-tailed recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6918–6928 (2022)
14. Liu, H., Dou, Y., Cao, M., Wang, K., Zou, Y., Ma, G., Li, C.: Mfsf2-net: Towards improved skin lesion diagnosis via multi-scale and multi-frequency-domain fusion. Pattern Recognition **174**, 112918 (2026). https://doi.org/10.1016/j.patcog.2025.112918
15. Liu, Y., Yan, S., Hu, M., Ju, L., Li, X., Tang, F., Feng, W., Ge, Z.: Dermagent: A self-reflective agentic system for dermatological image analysis with multi-tool reasoning and traceable decision-making (2026), https://arxiv.org/abs/2605.14403
16. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
17. MILK study team: MILK10k (2025). https://doi.org/10.34970/648456
18. Neverova, N., Wolf, C., Taylor, G., Nebout, F.: Moddrop: adaptive multi-modal gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(8), 1692–1706 (2016). https://doi.org/10.1109/TPAMI.2015.2461544
19. Pollastri, F., Maronas, J., Parreno, M., Bolelli, F., Paredes, R., Grana, C., Albiol, A.: Aimagelab-prhlt at isic challenge 2019. Tech. rep., Università degli Studi di Modena e Reggio Emilia and Universitat Politècnica de València (2019)
20. Silla Jr, C.N., Freitas, A.A.: A survey of hierarchical classification across different application domains. Data mining and knowledge discovery **22**(1), 31–72 (2011)
21. Sun, Q., Huang, C., Chen, M., Xu, H., Yang, Y.: Skin lesion classification using additional patient information. BioMed research international **2021**(1), 6673852 (2021)
22. Tan, M., Le, Q.: Efficientnetv2: Smaller models and faster training. In: International conference on machine learning. pp. 10096–10106. PMLR (2021)
23. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 180161 (2018)
24. Wang, K., Qiu, L., Zhang, Y., Xie, F.: I2m2net: Inter/intra-modal feature masking self-distillation for incomplete multimodal skin lesion diagnosis. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 Workshops. Lecture Notes in Computer Science, vol. 15274, pp. 3–13. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-031-77610-6_1
25. Wang, Y., Yu, T., Cai, J., Kalia, S., Lui, H., Wang, Z.J., Lee, T.K.: Integrating clinical knowledge graphs and gradient-based neural systems for enhanced melanoma diagnosis via the seven-point checklist. IEEE Transactions on Neural Networks and Learning Systems **37**(1), 37–51 (2026). https://doi.org/10.1109/TNNLS.2025.3600443
26. Yao, P., Shen, S., Xu, M., Liu, P., Zhang, F., Xing, J., Shao, P., Kaffenberger, B., Xu, R.X.: Single model deep learning on imbalanced small datasets for skin le-

sion classification. IEEE transactions on medical imaging **41**(5), 1242–1254 (2022). https://doi.org/10.1109/TMI.2021.3136682
27. Zhou, S., Zhuang, Y., Meng, R.: Multi-category skin lesion diagnosis using dermoscopy images and deep cnn ensembles. ISIC Challenge 2019 (2019)
28. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: International workshop on deep learning in medical image analysis. pp. 3–11. Springer (2018)