mSpecFusion-Net: Intelligent Multimodal Smartphone Imaging for Mobile Differential Diagnosis of Psoriasis and Dermatitis and Deep Learning

Sewoong Kim[1⋆], Seong-Hyun Kim[2⋆], Thiago Coutinho Cavalcanti[1,3⋆], Jonghun Lee[4], Moon Hwan Lee[1,5], Donghun Lee[6⋆⋆], and Jae Youn Hwang[1,2,4,7⋆⋆]

[1] Dept. of Electrical Engineering and Computer Science, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea
[2] Dept. of Artificial Intelligence, DGIST, Daegu, Korea
[3] KAIST InnoCORE AI-CRED Institute, KAIST, Daejeon, Korea
[4] Dept. of Biomedical Science and Engineering, DGIST, Daegu, Korea
[5] Max Planck Institute for Medical Research, Heidelberg, Germany
[6] Dept. of Dermatology, Seoul National University Hospital, Seoul, Korea
[7] InnoCORE Research Group for Bio-Embodied Physical AI, DGIST, Daegu, Korea
rlatjdgus224@dgist.ac.kr; ivymed27@snu.ac.kr; jyhwang@dgist.ac.kr

Abstract. Psoriasis and seborrheic dermatitis (SD) often present with highly similar clinical features such as erythema and scaling, leading to frequent diagnostic ambiguity, particularly in primary care and teledermatology. Existing AI approaches based on mobile RGB imaging have shown limited discrimination performance because the two conditions can exhibit overlapping visual manifestations. To address this challenge, we propose a smartphone-based portable multimodal optical imaging and analysis framework that acquires co- and cross-polarized multispectral reflectance images and ultraviolet (UV)-excited autofluorescence, together with a modality-aware deep learning model, mSpecFusion-Net, designed for robust cross-modality integration. The proposed model encodes each modality in parallel and employs an iterative fusion Transformer that repeatedly aggregates complementary information across modalities to form a unified representation for differential diagnosis. Across patient-wise splits, mSpecFusion-Net (PCA + iterative fusion) delivers the strongest overall performance (F1 = 0.9057, ROC-AUC = 0.9461), surpassing RGB-only baselines and widely used backbones. This demonstrates that explicitly reducing inter-wavelength redundancy and progressively refining cross-modality evidence is critical for separating clinically similar conditions. The proposed smartphone multimodal pipeline is readily deployable for teledermatology and scalable to other multimodal mobile diagnostics.

Keywords: Mobile Diagnosis · Multimodal Imaging · Multispectral Imaging · Transformer Fusion.

⋆ Equal contribution.
⋆⋆ Corresponding authors.

1 Introduction

Psoriasis and seborrheic dermatitis (SD) are prevalent inflammatory skin diseases that often exhibit highly similar clinical features, including erythema and scaling. This overlap leads to frequent diagnostic ambiguity and can delay appropriate management or result in suboptimal treatment decisions, particularly in primary care and teledermatology where rapid and reliable triage is clinically valuable. In real-world clinical and remote-care settings, mobile skin images are difficult to standardize in terms of acquisition distance, viewing angle, and illumination. Moreover, the observed contrast is easily confounded by specular reflection from the skin surface and light scattering within tissue, which can obscure subtle morphological cues. These factors limit the robustness and reproducibility of automated classifiers that rely on RGB images or a single imaging contrast in point-of-care workflows.

Recent advances in dermatologic imaging and deep learning have explored complementary optical contrasts beyond RGB, including multispectral reflectance, polarization, and autofluorescence imaging [1–3]. Spectral imaging has also been investigated for dermatologic assessment using hyperspectral or multispectral data coupled with machine learning [4, 5]. Smartphone-based multispectral imaging has been directly studied for discriminating psoriasis and SD, motivating mobile deployment [6]. A recent Sensors review summarizes multispectral imaging systems for skin disease assessment and highlights practical constraints for portable implementations [7]. Unlike recent RGB-based fusion approaches augmented with auxiliary cues such as metadata [27], spatial–spectral features [28], or dermoscopic–clinical pairs [29], our framework fuses modalities captured by distinct physical principles without auxiliary metadata.

In parallel, multispectral/hyperspectral learning has progressed in other domains (e.g., remote sensing), including improved spectral-spatial modeling, few-shot/generalization methods, and multimodal fusion Transformers [11, 12, 9, 8, 13]. Parameter-efficient fine-tuning of multispectral foundation models for hyperspectral classification further suggests strategies for limited labeled data [10]. Despite these advances, two practical gaps remain for mobile dermatology deployment. For example, acquisition and calibration are often not standardized in unconstrained mobile settings, making performance sensitive to imaging variability [6, 7]. Also, when multiple modalities are available, effective integration of complementary information across modalities is still limited: naïve early fusion or single-stream learning can underutilize modality-specific cues or become sensitive to artifacts present in a single modality, reducing generalization in real-world use.

To address these challenges, we propose DermaSpecFusion, a smartphone-based multimodal optical imaging and analysis framework for differential diagnosis of psoriasis and seborrheic dermatitis. The system allows acquisition of co-polarized and cross-polarized multispectral reflectance and UV-excited autofluorescence in a unified mobile workflow, and applies PCA-based spectral compression to reduce inter-wavelength redundancy. We further develop a modality-aware deep learning model, *mSpecFusion-Net*, which encodes each modality in

parallel and uses a modality-aware iterative-fusion architecture to progressively integrate complementary cross-modal cues into a unified representation for diagnosis. Performance is evaluated using patient-wise 5-fold cross-validation to estimate generalization reliably[14, 15].

The key contributions of this work are as follows:

1. **Smartphone-based portable multimodal optical imaging workflow**: We integrate co-/cross-polarized multispectral reflectance and UV-excited autofluorescence into a mobile acquisition pipeline (a smartphone) to provide complementary optical contrasts for differentiating inflammatory skin diseases at the point of care.
2. **mSpecFusion-Net**: We propose an iterative-fusion Transformer architecture that repeatedly exchanges and refines complementary information across modalities to learn a robust unified representation for differential diagnosis.
3. **Clinical evaluation**: We show that the proposed multimodal acquisition and fusion framework consistently outperforms RGB-only and single-modality baselines under patient-wise evaluation, supporting translational potential for point-of-care screening and teledermatology.

2 Methodology

2.1 System overview

Figure 1 summarizes the proposed end-to-end framework. We first acquire complementary optical contrasts using a portable smartphone-based multimodal dermascope, and then perform automated differential diagnosis between psoriasis and seborrheic dermatitis using mSpecFusion-Net, an iterative-fusion Transformer that integrates co-polarized multispectral reflectance, cross-polarized multispectral reflectance, and UV-excited autofluorescence information.

2.2 Smartphone-based multimodal optical imaging and calibration

The multimodal dermascope and acquisition scheme are illustrated in Fig. 1. For multispectral reflectance imaging, eight narrowband LEDs (435, 470, 490, 505, 525, 570, 610, 640 nm) are sequentially activated via an LED multiplexer to form an 8-band reflectance stack, consistent with dermatologic MSI system designs [1, 7]. To acquire polarization information from the interaction of a target tissue with applied light at different wavelengths, we acquire both co-polarized and cross-polarized multispectral stacks by controlling the relative orientation between the illumination polarization and the analyzer on the detection path. Cross-polarized imaging suppresses specular reflection and enhances subsurface scattering-driven texture cues [2, 3]. We additionally acquire UV-excited autofluorescence (AF) images to provide biochemical-sensitive contrast complementary to reflectance imaging [1].

To improve robustness under point-of-care acquisition variability, all modalities are calibrated using an embedded white reference target to correct shading

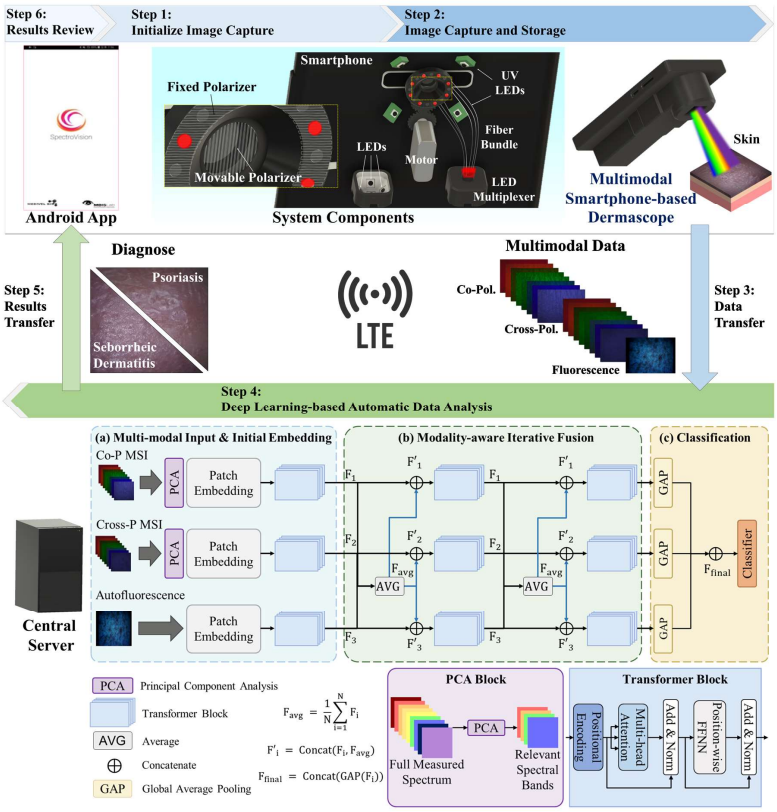

Fig. 1. Overview of the proposed framework. We introduce a standardized smartphone-based multimodal dermascope that jointly captures complementary optical contrasts: co-/cross-polarized multispectral reflectance and UV-excited autofluorescence (AF). To explicitly reduce inter-wavelength redundancy before fusion, the multispectral stream is compressed via PCA into a compact spectral representation. The resulting multimodal data are uploaded to a central server and analyzed by mSpecFusion-Net, a modality-aware iterative-fusion Transformer that progressively refines cross-modal evidence through repeated interactions, enabling robust differential diagnosis of psoriasis vs. seborrheic dermatitis.

and illumination non-uniformity [6, 7]. Figure 2 provides representative co- and cross-polarized multispectral stacks and reflectance profiles from psoriasis and seborrheic dermatitis, demonstrating wavelength-dependent differences and the benefit of polarization information.

2.3 Clinical dataset and patient-wise evaluation

We collected 2,032 multimodal images from 32 patients (14 SD, 18 psoriasis), diagnosed by board-certified dermatologists. The study was conducted with the

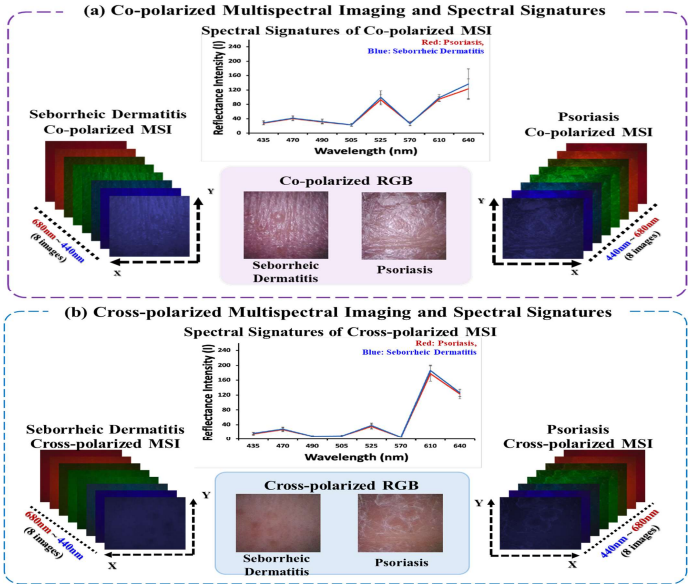

Fig. 2. Representative co- and cross-polarized multispectral imaging (MSI) examples and corresponding spectral signatures for psoriasis and seborrheic dermatitis. (a) Co-polarized MSI: 8-band reflectance stacks and representative RGB lesion images are shown together with mean reflectance spectra extracted from lesion regions. (b) Cross-polarized MSI: corresponding 8-band stacks, RGB appearances, and spectral profiles under cross-polarized illumination. Cross-polarization suppresses surface specular reflection and enhances subsurface texture cues. The mean reflectance spectra (435–640 nm) for psoriasis (red) and seborrheic dermatitis (blue) exhibit subtle, partially overlapping wavelength-dependent trends, highlighting the challenge of direct spectral discrimination. Error bars indicate variability across samples.

Department of Dermatology at Seoul National University Hospital and approved by its IRB (No. 1908-161-1059); all participants provided informed consent. We adopt patient-wise 5-fold cross-validation [14, 15]: 32 patients are split into 10 disjoint groups; per round, 2 groups (∼20%) form the test set, 1 the validation, and 7 the training. Each patient thus appears in exactly one test set with no identity leakage. Tables 2–5 report patch-level metrics (micro-averaged for AUROC/AUPRC) [16, 17]; for deployment, patch predictions can be aggregated to lesion/patient level by averaging or voting.

2.4 mSpecFusion-Net

All images are calibrated as described above and then spatially aligned across modalities. We crop around the lesion region to reduce background and resize inputs to a fixed spatial resolution. To increase training samples and to capture local texture patterns, each image is partitioned into a fixed grid of non-

Table 1. Dataset obtained from clinical trials.

Class	Cases (Patients)	Images
Psoriasis	18	1,248
Seborrheic Dermatitis	14	784
Total	32	2,032

Table 2. Principal component analysis (PCA) effects on mSpecFusion-Net performance. Co-P MSI: co-polarized multispectral imaging; Cross-P MSI: cross-polarized multispectral imaging; AF: autofluorescence.

Configuration				F1	Acc.	ROC AUC	PRC AUC
Co-P MSI	Cross-P MSI	AF	PCA				
✓	✓			0.8279	0.8406	0.9134	0.9024
✓	✓		✓	0.8899	0.8926	0.9453	0.9395
✓	✓	✓		0.8485	0.8577	0.9118	0.9032
✓	✓	✓	✓	**0.9057**	**0.9068**	**0.9461**	**0.9358**

overlapping patches. Patches inherit the patient label (psoriasis vs SD) and are used as training instances, while patient-wise splitting is strictly maintained to avoid leakage [14, 15]. To reduce redundancy across wavelengths, we apply principal component analysis (PCA) as a discrete band selection step rather than a learned projection [18]. The eight calibrated bands are ranked by their cumulative absolute PCA loading, and the top $K = 4$ bands are retained per polarization stream. Since this is discrete index selection rather than a data-dependent transformation, it is performed once on the calibrated dataset and fixed across all five folds; parameter learning itself remains strictly fold-independent, so the leakage pattern typically associated with global preprocessing does not apply.

The architecture of mSpecFusion-Net is shown in Fig. 1. The model processes each modality with a parallel Transformer encoder branch after the pretrained PCA block [26]. Each branch begins with patch embedding and positional encoding, followed by stacked Transformer blocks. We then perform *iterative fusion*: after each fusion stage, a shared representation is computed by aggregating modality features and is re-injected into each modality branch for further refinement. This repeated cross-modality interaction is motivated by iterative fusion designs shown effective in multispectral learning [8, 9]. Finally, modality-level pooled features are concatenated and fed to an MLP classifier for binary prediction. We compare iterative fusion against early, middle, and late fusion baselines, following common multimodal fusion patterns [9]. We train using binary focal cross-entropy loss and optimize with Adam [19]. To improve robustness to real-world capture variability, we apply on-the-fly data augmentation such as brightness/contrast jitter and mild geometric transforms [20]. We also compare against strong backbone baselines adapted to our inputs, including ResNet [21], EfficientNet [22, 23], ConvNeXt [24], Swin Transformer [25], and ViT [26]. For context on robust multispectral learning under limited data, the representative works from other domains were cited [11–13, 10].

Table 3. Comparison of multimodal fusion strategies on mSpecFusion-Net performance.

Fusion Strategy	F1	Acc.	ROC AUC	PRC AUC
Early Fusion	0.8570	0.8603	0.9148	0.9129
Middle Fusion	0.8445	0.8688	0.9393	0.9365
Late Fusion	0.8657	0.8667	0.9207	0.9100
mSpecFusion-Net (Iterative Fusion)	**0.9057**	**0.9068**	**0.9461**	**0.9358**

Table 4. Multimodal datasets vs. performance of mSpecFusion-Net.

Dataset			F1	Acc.	ROC AUC	PRC AUC
Co-P MSI	Cross-P MSI	AF				
(RGB Image)			0.6484	0.6774	0.7823	0.8023
✓			0.8747	0.8760	0.9133	0.8950
	✓		0.6851	0.7478	0.7978	0.8307
✓		✓	0.8232	0.8400	0.8933	0.8762
	✓	✓	0.6487	0.6713	0.7805	0.7964
✓	✓		0.8899	0.8926	0.9453	**0.9395**
✓	✓	✓	**0.9057**	**0.9068**	**0.9461**	0.9358

3 Experimental Results

Table 2 evaluates the effect of PCA-based spectral compression. Removing inter-wavelength redundancy consistently improves performance when using multispectral inputs, with the best results achieved when PCA is combined with multispectral modalities (Co-P MSI, Cross-P MSI). This suggests that compact spectral representations allow the model to focus on diagnostically informative spectral variation rather than redundant band-wise noise [18]. Notably, adding AF without PCA yields a smaller gain, indicating that the benefit of additional modalities is maximized when the spectral stream is appropriately regularized.

Table 3 compares multimodal fusion strategies. While early, middle, and late fusion provide competitive baselines, iterative fusion achieves the highest F1 and the most balanced AUROC/AUPRC, supporting the hypothesis that repeated cross-modality interaction better consolidates complementary cues than single-shot fusion [8, 9]. This trend is consistent with our clinical setting, where modality-specific cues may be weak or partially overlapping and benefit from progressive refinement.

Table 4 reports modality ablations. Single-modality configurations underperform multimodal combinations, confirming that no single optical contrast is sufficient for robust differentiation between psoriasis and seborrheic dermatitis (SD). In particular, combining co- and cross-polarized MSI yields a strong improvement, consistent with polarization enabling to suppress specular reflection and exposing subsurface texture cues [2, 6]. Adding AF further improves the final model, suggesting complementary biochemical-sensitive information beyond reflectance [1]. Overall, the full multimodal configuration with PCA and iterative fusion provides the best performance.

Table 5. mSpecFusion-Net vs. other deep learning models.

Model	Batch	F1	Acc.	ROC AUC	PRC AUC
mSpecFusion-Net	16	**0.9057**	**0.9068**	**0.9461**	**0.9358**
EfficientNet-B0 [22]	64	0.6054	0.6991	0.6865	0.6502
EfficientNet-V2L [23]	32	0.6570	0.7136	0.7185	0.6845
ConvNeXt-Base [24]	16	0.8330	0.8658	0.8951	0.8675
ConvNeXt-XLarge [24]	8	0.8293	0.8609	0.9047	0.8889
Swin-Tiny [25]	64	0.6188	0.7264	0.7237	0.6889
Swin-Small [25]	64	0.6673	0.7264	0.4084	0.6889
Swin-Base [25]	16	0.6888	0.6888	0.6120	0.6153
Swin-Large [25]	64	0.6453	0.7264	0.3968	0.6489
ViT-Base [26]	16	0.7081	0.7183	0.7312	0.6999
ViT-Large [26]	32	0.6889	0.6889	0.6611	0.6274
ViT-Huge [26]	8	0.6819	0.6839	0.6944	0.6643
ResNet-18 [21]	32	0.7162	0.7608	0.7511	0.7718
ResNet-50 [21]	16	0.6609	0.7105	0.6951	0.7281

Finally, Table 5 compares mSpecFusion-Net against widely used deep learning backbones [21–26]; mSpecFusion-Net consistently outperforms them. Across the five patient-wise folds, the proposed model achieves F1 = 0.9057 ± 0.0612, accuracy = 0.9068 ± 0.0576, ROC-AUC = 0.9461 ± 0.0540, and PRC-AUC = 0.9358 ± 0.0666 (mean ± SD, micro-averaged), with consistent relative method ordering across folds. With 4.18 M parameters and 44.7 GFLOPs, the 17.3 ms GPU inference latency is negligible compared to the ∼20 s acquisition.

4 Conclusion

In this work, we present DermaSpecFusion, a novel smartphone-based multimodal dermascope and *mSpecFusion-Net*, a modality-aware iterative-fusion deep learning model for mobile differential diagnosis of psoriasis and seborrheic dermatitis. The key novelty of our framework lies in three aspects: (1) standardized acquisition of complementary optical contrasts (co-/cross-polarized multispectral reflectance and UV-excited autofluorescence), (2) PCA-based spectral compression to reduce inter-wavelength redundancy before fusion, and (3) iterative cross-modality refinement that progressively consolidates complementary evidence, rather than relying on one-shot early/middle/late fusion. Systematic ablation studies support these design choices. PCA regularization consistently improves multispectral performance by suppressing redundant spectral noise, iterative fusion outperforms conventional fusion strategies by better integrating partially overlapping modality-specific cues, and multimodal combinations substantially outperform single-modality settings, confirming that no single optical contrast is sufficient for robust differentiation. Under patient-wise evaluation, the full model achieves F1 = 0.9057, Accuracy = 0.9068, ROC-AUC = 0.9461, and PRC-AUC = 0.9358, and surpasses widely used generic backbones, demonstrating that explicit modeling of cross-modality complementarity is more effec-

tive than simply scaling backbone capacity. Overall, these results show clinically grounded multimodal acquisition and progressive cross-modal reasoning can substantially improve robustness for visually similar inflammatory skin diseases in mobile settings. The proposed framework provides a practical foundation for point-of-care screening and teledermatology, offers a scalable template for multimodal AI diagnosis in other dermatologic and biomedical imaging applications.

Acknowledgements. This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (No. RS-2024-00354020 and No. RS-2024-00414968), and by the InnoCORE program of the Ministry of Science and ICT (N10260004 for AI-CRED and 26-InnoCORE-01 for Bio-Embodied). Code will be released at `https://github.com/Seong-Hyun-0224/mSpecFusion-Net`.

Disclosure of Interests. The authors are pursuing intellectual property protection and have no other competing interests.

References

1. Jiang, Z., Gu, X., Chen, D., Zhang, M., Xu, C.: Deep learning-assisted multispectral imaging for early screening of skin diseases. *Photodiagn. Photodyn. Ther.* **48**, 104292 (2024)
2. Zhang, T., Zeng, H., Ahad, T., Lui, H., Lee, T. K., Wu, Z., Zhao, J., Kalia, S.: Enhancing in vivo morphological assessment of skin conditions using a four-state polarized multispectral imaging system. *Opt. Lett.* **50**(15), 4774–4777 (2025)
3. Jung, G., Lee, J., Kim, S.: Integrated deep learning approach for generating cross-polarized images and analyzing skin melanin and hemoglobin distributions. *Biomed. Eng. Lett.* **14**(6), 1355–1364 (2024)
4. Lin, T.-L., Mukundan, A., Karmakar, R., Avala, P., Chang, W.-Y., Wang, H.-C.: Hyperspectral imaging for enhanced skin cancer classification using machine learning. *Bioengineering* **12**(7), 755 (2025)
5. Parasca, S. V., Calin, M. A., Manea, D., Radvan, R.: Hyperspectral imaging with machine learning for in vivo skin carcinoma margin assessment: a preliminary study. *Phys. Eng. Sci. Med.* **47**(3), 1141–1152 (2024)
6. Kim, S., Kim, J., Hwang, M., Kim, M., Jo, S. J., Je, M., Jang, J. E., Lee, D. H., Hwang, J. Y.: Smartphone-based multispectral imaging and machine-learning based analysis for discrimination between seborrheic dermatitis and psoriasis on the scalp. *Biomed. Opt. Express* **10**(2), 879–891 (2019)
7. Ilişanu, M.-A., Moldoveanu, F., Moldoveanu, A.: Multispectral imaging for skin diseases assessment—State of the art and perspectives. *Sensors* **23**(8), 3888 (2023)
8. Shen, J., Chen, Y., Liu, Y., Zuo, X., Fan, H., Yang, W.: ICAFusion: Iterative Cross-Attention Guided Feature Fusion for Multispectral Object Detection. arXiv:2308.07504 (2023)
9. Roy, S. K., Deria, A., Hong, D., Rasti, B., Plaza, A., Chanussot, J.: Multimodal Fusion Transformer for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **61**, 1–20 (2023)

10. Ligan, B., Jbilou, K., Kalloubi, F., Ratnani, A.: Parameter-Efficient Fine-Tuning of Multispectral Foundation Models for Hyperspectral Image Classification. arXiv:2505.15334 (2025)
11. Li, W., Wang, J., Gao, Y., Zhang, M., Tao, R., Zhang, B.: Graph-Feature-Enhanced Selective Assignment Network for Hyperspectral and Multispectral Data Classification. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–14 (2022)
12. Yu, Q., Wang, Y.: Hyperspectral image few-shot classification based on spatial–spectral information complementation and multilatent domain generalization. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **18**, 13212–13224 (2025)
13. Nkegoum, M., Pham, M.-T., Fromont, É., Avignon, B., Lefèvre, S.: FSMODNet: A closer look at few-shot detection in multispectral data. arXiv:2509.20905 (2025)
14. Arlot, S., Celisse, A.: A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010)
15. Wong, T.-T., Yeh, P.-Y.: Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **32**(8), 1586–1594 (2020)
16. Hanley, J. A., McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**(1), 29–36 (1982)
17. Davis, J., Goadrich, M.: The relationship between Precision–Recall and ROC curves. In: *ICML* (2006)
18. Jolliffe, I.: *Principal Component Analysis*. Springer (2002)
19. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. In: *ICLR* (2015)
20. Shorten, C., Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR* (2016)
22. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: *ICML* (2019)
23. Tan, M., Le, Q.: EfficientNetV2: Smaller models and faster training. In: *ICML* (2021)
24. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: *CVPR* (2022)
25. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using shifted windows. In: *ICCV* (2021)
26. Dosovitskiy, A. *et al.*: An image is worth 16×16 words: Transformers for image recognition at scale. In: *ICLR* (2021)
27. Yang, Y.-C., Chung, W.-C., Wu, C.-Y., Hung, C.-L., Chen, Y.-J.: Skin disease classification using transfer learning model and fusion strategy. *Intell.-Based Med.* **12**, 100271 (2025)
28. Siddiqui, M. S. B., Bhuiyan, M. I. H.: S^3F-Net: A multi-modal approach to medical image classification via spatial–spectral summarizer fusion network. *IEEE J. Biomed. Health Inform.* (2026)
29. Wang, J., Zhang, Y., Xie, F., Liu, J.: Enhancing diagnosis of psoriasis and inflammatory skin diseases: A spatially aligned multimodal model integrating clinical and dermoscopic images. *J. Invest. Dermatol.* **145**(11), 2736–2744 (2025)