Southeast Asian Journal of Sciences, Vol. 12, No 1 (2026), pp.22-34
https://doi.org/10.36853/sajs0033

SPD-SKINNET: SPECTRAL RIEMANNIAN LEARNING WITH COVARIANCE REPRESENTATIONS FOR SKIN LESION CLASSIFICATION

Dinh Que Tran, Van Anh Thi Trinh and **Thao Van Doan**

Department of Information Technology
Posts and Telecommunications Institute of Technology (PTIT)
Hanoi, Vietnam
E-mail: quetd@ptit.edu.vn, anhttv@ptit.edu.vn, vandt@stu.ptit.edu.vn

Abstract

Automatic skin lesion diagnosis is a fundamental challenge in medical image analysis because of requiring accurate and interpretable representations of complex visual patterns. Most existing deep learning approaches rely on Euclidean feature embeddings, which primarily capture first-order statistics and often fail to model the intrinsic geometric and textural structure of skin lesions. In this paper, we propose **SPD-SkinNet**, a spectral–Riemannian deep learning framework that represents skin lesion features as symmetric positive definite (SPD) matrices, enabling the modeling of second-order statistics via feature covariance. By performing spectral decomposition on the SPD manifold, we extract eigenvalues that serve as compact and interpretable descriptors of lesion texture. The initial experimental results on the synthetic highlight the potential of integrating Riemannian geometry and spectral analysis into deep learning for medical imaging.

Key words: Symmetric positive definite matrices, Riemannian geometry, covariance representation, spectral learning, deep learning, skin lesion classification.
2020 Mathematics Classification: 68T07, 53C35, 62H25, 68U10.

1 Introduction

Automatic skin lesion classification is a critical problem in medical image analysis, with significant implications for early detection of melanoma and other dermatological conditions. Accurate and timely diagnosis can substantially improve patient outcomes, motivating the development of reliable computer-aided diagnostic systems. In recent years, the convolutional neural networks (CNNs) with architectures such as ResNet [3] and EfficientNet [13] learn hierarchical feature representations, progressing from low-level visual primitives to high-level semantic abstractions. These models have demonstrated strong performance across a wide range of medical imaging tasks, including skin lesion analysis [11, 3, 13]. However, standard CNNs operate in Euclidean space and predominantly encode first-order statistics, treating feature channels independently. This design limits their ability to explicitly capture inter-channel dependencies and higher-order statistical structures, which are essential for modeling complex visual textures.

Skin lesions exhibit heterogeneous patterns characterized by subtle color variations, irregular boundaries, and multi-scale textures. Critically, these patterns emerge not solely from individual feature responses but from structured interactions among features, such as co-occurrences between pigmentation, edges, and micro-patterns. Such interactions are fundamentally second-order and cannot be fully represented by conventional CNN embeddings. While some studies have explored higher-order or region-based descriptors, most approaches continue to rely on Euclidean embeddings, leaving the rich geometric structure of skin textures underutilized [15, 4, 5].

Covariance-based descriptors offer a principled way to capture these second-order statistics [15]. By modeling pairwise correlations between feature channels, covariance matrices encode complex texture interactions in a compact form. Notably, these matrices lie on the manifold of symmetric positive definite (SPD) matrices, which has well-defined Riemannian geometry [4, 5]. Recent geometry-aware learning approaches on SPD manifolds have demonstrated improved robustness and discriminative power in visual recognition tasks [6]. Additionally, spectral analysis of SPD matrices through eigen-decomposition yields interpretable representations of variation patterns, providing a frequency-like decomposition of feature interactions that is particularly useful in medical contexts. Despite the promise of SPD-based and spectral representations, their integration with deep CNN feature extraction in skin lesion analysis remains largely unexplored. Most existing CNN-based methods continue to rely on first-order Euclidean embeddings, which are insufficient to capture the structured, second-order texture patterns that are characteristic of skin lesions.

To address this gap, we propose **SPD-SkinNet**, a spectral–manifold deep learning framework that combines CNN feature extraction with covariance-based SPD representations and spectral decomposition. The key idea is to

transform CNN feature maps into covariance matrices, capturing second-order statistics of skin textures. The resulting SPD matrices are then spectrally decomposed to extract eigenvalues that serve as compact, interpretable, and discriminative descriptors of variation patterns within lesions.

From a geometric perspective, the eigenvalues can be interpreted as intrinsic *energy modes* of texture, providing a frequency-like decomposition of feature interactions. This spectral characterization enhances both discriminative capability and interpretability—critical for clinical applications requiring model transparency.

We validate the proposed framework on a synthetic dataset, demonstrating improved performance over standard CNN baselines while also providing meaningful insights into texture structure.

The main contributions of this paper are summarized as follows:

- We introduce a covariance-based SPD representation for modeling skin lesion textures, explicitly capturing second-order feature interactions beyond standard CNN embeddings.
- We propose a spectral learning mechanism that leverages eigenvalues of SPD matrices as compact, discriminative, and interpretable descriptors.
- We develop an end-to-end framework, *SPD-SkinNet*, integrating CNN feature extraction with Riemannian geometry and spectral analysis for skin lesion classification.
- We demonstrate the effectiveness of the proposed approach on a synthetic dataset and the initial results indicate its potential for the accurate and robust skin lesion classification.

2 Methodology

2.1 Motivation and Problem Statement

Automated skin lesion analysis remains a challenging task due to the intrinsic complexity and heterogeneity of dermatological patterns. Unlike natural images, skin lesions exhibit subtle chromatic variations, irregular and fuzzy boundaries, multi-scale textures, and intricate interactions between structural and color-based cues. These characteristics are inherently second-order, arising from correlations among local patterns rather than isolated feature activations.

Deep convolutional neural networks (CNNs) have achieved remarkable success in medical image analysis, including dermatology [1, 2, 3, 11]. Their hierarchical structure enables progressive abstraction from low-level visual primitives to high-level semantic representations. However, conventional CNN architectures operate in Euclidean space and predominantly encode first-order

statistics, treating feature channels independently. Consequently, they are not explicitly designed to capture inter-channel dependencies or higher-order statistical relationships, which are critical for modeling complex skin textures.

Recent advances in dermatological imaging further emphasize the importance of capturing structural variability and texture correlations for accurate diagnosis [12, 14]. This motivates the need for extending representations of first-order features to explicitly encode second-order interactions in a geometry-aware manner.

Problem Statement. Given an input skin image $I \in \mathbb{R}^{3\times H\times W}$, the objective is to learn a representation that:

- jointly captures local feature responses and their interactions,
- encodes intrinsic texture structure in a geometry-aware space,
- supports accurate, robust, and interpretable classification of skin lesions.

To address these challenges, we propose a unified framework that integrates CNN-based feature extraction with covariance-based symmetric positive definite (SPD) representations and spectral analysis.

2.2 Feature Extraction

Let the input image be $I \in \mathbb{R}^{3\times H\times W}$. A deep convolutional backbone extracts hierarchical feature representations:

$$F = \mathrm{CNN}(I), \quad F \in \mathbb{R}^{C\times H'\times W'}.$$

Modern CNN architectures such as VGG, ResNet, and EfficientNet have demonstrated strong capability in extracting discriminative visual features [8, 3, 13]. In this work, the backbone serves as a learnable encoder transforming raw pixels into semantically meaningful representations.

Interpretation of Feature Maps. The tensor F can be interpreted as a set of feature maps encoding patterns at multiple abstraction levels. Early layers capture edges, gradients, and color transitions, while deeper layers encode high-level structures such as lesion boundaries, asymmetry, and texture irregularities.

Spatial-to-Sample Transformation. For statistical modeling, the feature tensor is reshaped into:

$$X \in \mathbb{R}^{C\times N}, \quad N = H' \times W',$$

where each column corresponds to a spatial sample and each row corresponds to a feature channel.

Limitation of First-Order Features. Each vector $f_i \in \mathbb{R}^C$ represents local responses but does not explicitly capture dependencies between channels. Dermatological textures are governed by co-occurrence patterns (e.g., alignment between pigmentation and structural edges), motivating the introduction of covariance-based second-order representations.

2.3 SPD Representation Layer

To capture second-order statistical structure, we compute the empirical covariance matrix:

$$C = \frac{1}{N}\sum_{i=1}^{N}(f_i - \mu)(f_i - \mu)^T, \quad \mu = \frac{1}{N}\sum_{i=1}^{N} f_i.$$

The resulting matrix $C \in \mathbb{R}^{C\times C}$ encodes both marginal variances and pairwise correlations:

- diagonal entries: channel-wise variance,
- off-diagonal entries: inter-channel dependencies.

Covariance representations have been shown to capture region-level statistics effectively in visual recognition tasks [15].

Geometric Perspective. Under mild conditions, C is symmetric positive definite, i.e., $C \in \mathcal{S}^C_{++}$. The space of SPD matrices forms a Riemannian manifold with non-Euclidean geometry [4, 5], providing a principled framework for modeling structured variability beyond standard Euclidean representations.

2.4 Spectral Representation Layer

To extract compact and informative descriptors, we perform eigen-decomposition:

$$C = U\Lambda U^T,$$

where U is orthogonal and

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_C)$$

contains the eigenvalues.

Interpretation. Eigenvalues quantify variation along principal directions. In skin lesion analysis:

- large eigenvalues correspond to dominant structural or textural patterns,
- small eigenvalues correspond to smoother or less informative variations.

The spectrum provides a compact characterization of texture complexity.

Dimensionality Reduction. To obtain a low-dimensional representation, we retain the top-k eigenvalues:

$$\Lambda_{\text{top-}k} = (\lambda_1, \ldots, \lambda_k),$$

forming a discriminative feature vector. This is supported by geometry-aware dimensionality reduction on SPD manifolds [6].

2.5 Classification Head

The spectral representation is passed through a multi-layer perceptron:

$$\hat{y} = f_\theta(\Lambda_{\text{top-}k}),$$

where $\hat{y}$ denotes predicted class probabilities. This module maps geometry-aware features into semantic lesion predictions.

2.6 Loss Function

We employ the standard cross-entropy loss:

$$\mathcal{L}_{\text{cls}} = -\sum_i y_i \log(\hat{y}_i).$$

Spectral Regularization (Optional). To enhance interpretability and encourage clinically meaningful representations:

$$\mathcal{L}_{\text{spec}} = \|\Lambda - \Lambda_{\text{ref}}\|_2^2,$$

where Λ_{ref} is a reference spectrum (e.g., from healthy skin).

Total Objective.

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{spec}}.$$

This encourages the model to learn deviations from normal texture patterns while maintaining discriminative performance.

3 Theoretical Properties

Motivation

CNNs effectively capture local spatial patterns but do not explicitly model inter-channel dependencies. Discriminative cues in dermatology often arise from structured feature interactions. Covariance-based representations address this by encoding second-order statistics in a principled manner.

Covariance matrices naturally lie on the SPD manifold $\mathcal{S}_{++}^C$, enabling integration of statistical modeling with Riemannian geometry [4, 5].

Notation. Let $\mathcal{S}_{++}^C$ denote the set of SPD matrices and $\mathcal{S}_{+}^C$ the set of positive semidefinite matrices, with $\mathcal{S}_{++}^C \subset \mathcal{S}_{+}^C$.

Proposition 1 (Positive Definiteness of Covariance [4, 5])**.** *Let* $X = [f_1, \ldots, f_N] \in \mathbb{R}^{C \times N}$ *and* $X_c = X - \mu$*. Then*

$$C = \frac{1}{N} X_c X_c^T \in \mathcal{S}_{+}^C.$$

If $\operatorname{rank}(X_c) = C$*, then* $C \in \mathcal{S}_{++}^C$.

Regularization. To address rank deficiency:

$$C_\epsilon = C + \epsilon I, \quad \epsilon > 0,$$

ensuring numerical stability and preserving the SPD property.

Proposition 2 (Second-Order Statistical Encoding [15])**.** *Covariance matrices satisfy:*

1. $C_{ii} = \operatorname{Var}(X_i)$,
2. $C_{ij} = \operatorname{Cov}(X_i, X_j)$.

Proposition 3 (Spectral Properties and Optimal Approximation [6])**.** *Let* $C = U \Lambda U^T$*. Then:*

1. *Eigenvalues are invariant under orthogonal transformations.*
2. *Truncated decomposition provides the optimal rank-k approximation in Frobenius norm.*

Summary

The framework is underpinned by three principles:

- SPD representations define a stable, geometrically meaningful feature space.
- Second-order statistics capture essential inter-channel interactions.
- Spectral decomposition yields compact, invariant, and discriminative descriptors.

These properties justify the effectiveness of the proposed SPD-SkinNet in modeling complex skin lesion textures and improving classification performance in high-dimensional, low-sample settings.

4 Illustrated Numerical Example

To provide an intuitive understanding of the SPD-SkinNet pipeline, we present a simplified numerical example illustrating the transformation from CNN feature maps to spectral features via the SPD representation.

4.1 Step 1: Feature Extraction

Consider a CNN feature map with $C = 2$ channels and $N = 3$ spatial locations:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix} \in \mathbb{R}^{2\times 3},$$

where each column f_i corresponds to a spatial feature vector.

4.2 Step 2: Mean Centering

Compute the mean vector:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} f_i = \begin{bmatrix} 2 \\ 7/3 \end{bmatrix},$$

and center the features:

$$X_c = X - \mu = \begin{bmatrix} -1 & 0 & 1 \\ -1/3 & -4/3 & 5/3 \end{bmatrix}.$$

4.3 Step 3: Covariance (SPD) Matrix

The empirical covariance is:

$$C = \frac{1}{N} X_c X_c^T = \begin{bmatrix} 2/3 & 2/3 \\ 2/3 & 14/9 \end{bmatrix}.$$

The positive determinant and positive trace confirm $C \in \mathcal{S}_{++}^{C}$. Diagonal entries represent channel variances, and off-diagonal entries encode inter-channel correlations (texture interactions).

4.4 Step 4: Spectral Decomposition

Eigen-decomposition yields:

$$C = U\Lambda U^T, \quad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2), \quad \lambda_1 \approx 1.92,\ \lambda_2 \approx 0.30.$$

Interpretation: λ_1 captures dominant texture variation, while λ_2 reflects weaker or smoother patterns.

4.5 Step 5: Spectral Feature for Classification

Retaining the top-$k = 1$ eigenvalue:

$$\Lambda_{\text{top-1}} = [1.92],$$

provides a compact, discriminative feature for downstream classification.

4.6 Discussion

This example highlights key advantages:

- Covariance encodes second-order interactions ignored by first-order CNN features.
- Spectral decomposition extracts dominant structure in a compact and interpretable form.
- Eigenvalues provide intrinsic coordinates on the SPD manifold, invariant to rotation in feature space.

5 Initial Experiments on Synthetic Dataset

5.1 Experimental Setup

We evaluate the proposed SPD-SkinNet on a synthetic dataset designed to simulate structured skin texture patterns with controlled variability. The dataset includes multiple classes exhibiting distinct geometric and stochastic characteristics, enabling a controlled assessment of the model's ability to capture second-order feature interactions.

To ensure a fair evaluation, we additionally implement a CNN baseline using the same backbone architecture (ResNet18) and identical training protocol. Both models are trained for 20 epochs using cross-entropy loss with label smoothing. Performance is evaluated using accuracy, F1-score, and Area Under the Curve (AUC) on both training and validation splits.

5.2 Training Dynamics

Figure 1 illustrates the evolution of loss, accuracy, F1-score, and AUC during training for both SPD-SkinNet and the CNN baseline.

Both models exhibit stable convergence, with loss decreasing monotonically and validation loss closely tracking training loss, indicating well-behaved optimization without significant overfitting. Performance improvements saturate after approximately 5–8 epochs, suggesting that the dominant structural patterns in the dataset are learned rapidly.

Table 1: Comparison between SPD-SkinNet and CNN baseline on the synthetic dataset.

Model	Accuracy	F1-score	AUC
CNN Baseline	0.879	0.857	0.982
SPD-SkinNet	**0.879**	**0.873**	**0.985**

5.3 Performance Evaluation

Table 1 summarizes the final performance of both models. Notably, SPD-SkinNet achieves comparable accuracy to the CNN baseline (0.879), indicating that both first-order and second-order representations are sufficient to capture the primary class distinctions in this synthetic setting.

However, SPD-SkinNet consistently outperforms the CNN baseline in terms of F1-score and AUC. The F1-score improves from 0.857 to 0.873, while AUC increases from 0.982 to 0.985. These improvements suggest enhanced robustness to class imbalance and improved separability in the learned feature space.

5.4 Discussion

The experimental results demonstrate that incorporating second-order statistics via SPD representations provides measurable benefits over standard CNN feature aggregation:

- **Second-order feature modeling:** Covariance-based SPD matrices capture inter-channel dependencies that are not explicitly modeled by first-order pooling methods.
- **Improved class balance:** Higher F1-scores indicate better handling of class-level variations and minority patterns.
- **Enhanced separability:** The increase in AUC reflects improved ranking performance and more discriminative feature representations.

Importantly, while classification accuracy remains unchanged, the improvements in F1-score and AUC highlight the advantage of second-order representations in capturing finer structural variations.

5.5 Limitations and Future Work

Despite promising results, several limitations should be acknowledged. First, the evaluation is conducted on a synthetic dataset, which may not fully reflect the complexity of real-world dermatological images. Second, the observed performance gains, while consistent, are moderate in magnitude.

Future work will focus on:

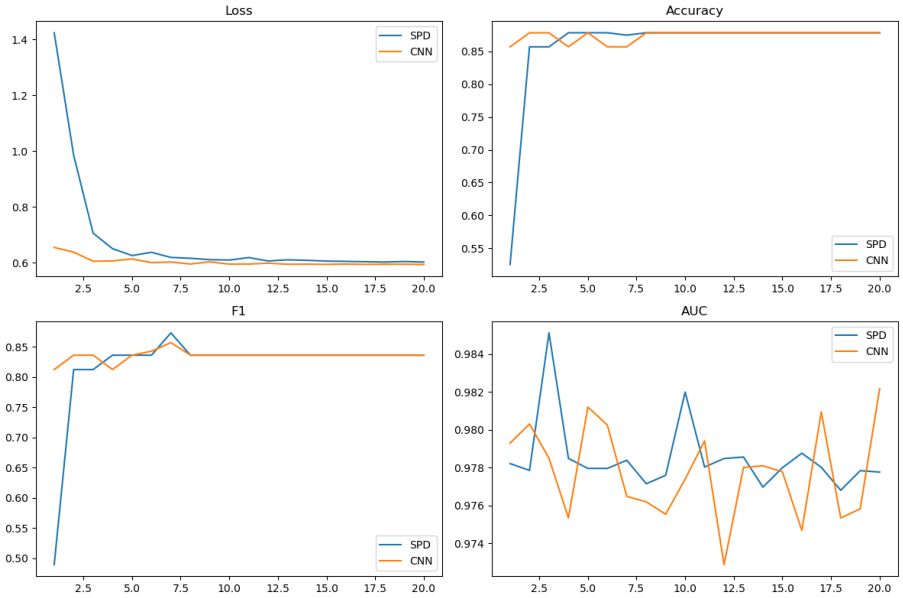

Figure 1: Training and validation performance comparison between SPD-SkinNet and CNN baseline. Top-left: loss; top-right: accuracy; bottom-left: F1-score; bottom-right: AUC.

- Extending evaluation to real dermatology datasets.
- Incorporating more complex textures and domain variability.
- Exploring geometry-aware extensions, including log-Euclidean mappings and Lie algebra representations.

6 Discussion

The results suggest that SPD matrices provide a principled mechanism for encoding second-order statistics of convolutional features, capturing interactions between feature channels that are not accessible through conventional first-order pooling. Spectral decomposition further transforms these representations into compact and interpretable descriptors.

Although the accuracy gains are not observed, the consistent improvements in F1-score and AUC indicate that SPD-based representations enhance the quality of the learned feature space, particularly in terms of robustness and discriminative power. This makes the approach a promising foundation for future extensions incorporating geometric learning on SPD manifolds.

7 Conclusion

We introduced SPD-SkinNet, a spectral–manifold deep learning framework for skin lesion classification. CNN features are transformed into SPD covariance matrices, and spectral analysis yields interpretable, compact representations. Initial experiments demonstrate robust performance on synthetic data, validating the approach. In the future, we will consider applying Graph Neural Networks for patient similarity analysis, utilise multi-scale covariance for cross-layer texture modeling and apply spectral regularization to improve generalization and robustness.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012, pp. 1097–1105.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[4] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, Jan. 2006.

[5] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 411–421, Aug. 2006.

[6] M. T. Harandi, M. Salzmann, and R. Hartley, "From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 17–32.

[7] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv preprint arXiv:1602.07360*, 2016.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[10] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[11] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[12] N. C. Codella, D. Gutman, M. Celebi, B. Helba, M. A. Marchetti, S. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the ISIC 2017 workshop," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2018, pp. 168–172.

[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114.

[14] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, Article no. 180161, 2018.

[15] F. Porikli, O. Tuzel, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *European Conference on Computer Vision (ECCV)*, 2006, pp. 589–600.