

Improving Skin Lesion Diagnosis: A Hybrid Approach Using Orthogonal Combination of Local Binary Pattern Features and Ensemble Learning for Diagnostic Accuracy

Abstract

Background: As the world becomes wealthier and people expect higher standards of care, the demand for healthcare services is growing rapidly. This puts significant pressure on existing medical resources and systems, making it harder to meet everyone's needs. In dermatology, for instance, the rising demand calls for creative and efficient solutions, especially in diagnosing conditions like skin cancer. Early diagnosis of skin lesions is necessary not only for effective treatment but also for providing the best possible outcomes for patients. **Methods:** In this paper, we present a solution using machine learning (ML) to assist in automated skin diagnosis, with a particular focus on the early detection of skin lesions, which is a key factor for effective treatment and better patient outcomes. Our method utilizes a Gaussian mixture model (GMM) with geometric features to enhance image quality by removing artifacts. We then use a color descriptor based on hybrid orthogonal combination of local binary patterns to capture the unique characteristics of the lesions. To identify the most important features for accurate classification, we apply ReliefF feature selection, prioritizing those that contribute most significantly to the results. Finally, we used several ML models such as decision tree, random forest, k-nearest neighbors, multilayer perceptron, and ensemble extra tree (ET) to classify eight different types of skin lesions effectively. **Results:** Remarkably, ensemble ET achieves superior performance with an accuracy of 97.31%. **Conclusions:** This research advances early skin lesion diagnosis, enhancing patient care in dermatology.

Keywords: Classification, dermatology, Gaussian mixture model, health care, hybrid orthogonal combination of local binary patterns, skin lesion

Submitted: 21-Nov-2024

Revised: 23-May-2025

Accepted: 14-Jul-2025

Published: 02-Feb-2026

Introduction

Skin cancer, particularly melanoma (MEL), is among the most aggressive, rapidly spreading, and life-threatening forms of cancer, accounting for the majority of skin cancer-related deaths worldwide. The global incidence of skin cancer has been steadily rising over the past few decades due to factors such as increased exposure to ultraviolet radiation, lifestyle changes, and environmental influences. Among all types of skin cancers, MEL is especially dangerous due to its high potential for metastasis even in early stages. According to the World Health Organization, millions of new cases are diagnosed annually, and the mortality rate remains significant despite medical advances.

Early and accurate detection of suspicious skin lesions is, therefore, a critical step in effective clinical management. Early detection of MEL significantly increases the likelihood of successful treatment and favorable patient outcomes. In contrast, delayed diagnosis often leads to disease progression and reduces the effectiveness of therapeutic interventions, underscoring the critical importance of timely medical response. Conventional diagnostic approaches often begin with visual inspection using the naked eye or dermoscopy, followed by biopsy and histopathological analysis for confirmation. While these methods are essential and form the backbone of clinical practice, they are subjective, labor-intensive, and heavily reliant on the experience and judgment of dermatologists. Moreover, inter-observer variability, where different clinicians may

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License (CC BY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Rahmani N, Ebrahimnezhad H. Improving skin lesion diagnosis: A hybrid approach using orthogonal combination of local binary pattern features and ensemble learning for diagnostic accuracy. *J Med Sign Sens* 2026;16:5.

Nasrin Rahmani,
Hossein
Ebrahimnezhad

Department of Electrical and
Computer Engineering, Sahand
University of Technology,
Tabriz, Iran

Address for correspondence:
Prof. Hossein Ebrahimnezhad,
Department of Electrical and
Computer Engineering, Sahand
University of Technology,
Tabriz, Iran.
E-mail: ebrahimnezhad@sut.
ac.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.jmss_74_24

Quick Response Code:



arrive at different diagnoses for the same lesion, can compromise diagnostic consistency, especially in regions with limited access to dermatological expertise.

Given these limitations, the medical community has increasingly turned to computer-aided diagnosis systems to enhance diagnostic accuracy, reduce workload, and provide second opinions in both urban and remote healthcare settings. In particular, the integration of machine learning (ML) and deep learning (DL) techniques has shown promise in automatically analyzing dermoscopic images and classifying skin lesions with performance comparable to or even surpassing that of human experts in some cases. These systems are trained on large datasets of annotated images to detect subtle patterns and features that may not be visible to the human eye, thus offering a more objective and scalable diagnostic tool.

Among the datasets used to train and evaluate these models, the International Skin Imaging Collaboration (ISIC) 2019 dataset stands out as one of the most comprehensive and widely adopted benchmarks in this domain. It includes tens of thousands of dermoscopic images covering a broad spectrum of lesion types, making it a valuable resource for developing data-driven models. However, despite its widespread use, the ISIC 2019 dataset is not without its challenges, which can significantly impact model performance. These include visual inconsistencies, class imbalance, and the presence of image artifacts such as hair occlusion, rulers, markers, and uneven lighting, all of which complicate the learning process and reduce the generalizability of trained models in real-world applications.

These challenges include marker colors, variations in brightness and contrast,^[1] dermoscopic overlays, low image quality, and noisy images.^[2] Additional issues, such as physical rulers, cropped lesions, stickers, and hair interference, further complicate analysis.^[3] Figure 1 illustrates these challenges and issues related to the dataset. Furthermore, the imbalanced classes in the ISIC 2019 dataset limit model generalization, negatively impacting diagnostic accuracy.^[4]

To address these issues, this study introduces a hybrid framework designed to enhance both the preprocessing and feature extraction stages of the classification pipeline. The proposed method incorporates Gaussian mixture modeling (GMM) and color hybrid orthogonal combination of local binary patterns (OC-LBP) to improve image quality and capture discriminative texture and color features. In parallel, random oversampling is applied to mitigate class imbalance and improve the model's sensitivity to minority classes. Multiple traditional ML classifiers are then evaluated to identify the most robust model for this task. By integrating advanced preprocessing techniques and a carefully designed feature extraction strategy, this framework aims to achieve improved classification

accuracy and better generalization across diverse lesion types.

Related works

Recent studies have explored various techniques to enhance the diagnosis and classification of skin diseases. These efforts focus on developing noninvasive, efficient, and accurate diagnostic tools as alternatives to traditional biopsy methods.

Jena *et al.*^[5] introduced a segmentation method combining minimum generalized cross entropy with the Opposition African Vulture Optimization Algorithm, demonstrating high segmentation accuracy in brain magnetic resonance imaging and dermoscopic images. However, the study relies on manual processes, with future research aiming to automate these tasks using reinforcement learning. Similarly, Houssein *et al.*^[6] proposed the improved golden jackal optimizer (IGJO) for skin cancer image segmentation, an enhancement of the golden jackal optimization algorithm. By using Otsu's method for multilevel thresholding, IGJO outperformed other meta-heuristic algorithms in segmentation accuracy. This method improves accuracy in skin cancer detection and can be extended to other medical imaging applications.

Patil and Bellary^[7] developed a noninvasive MEL classification system using an improved convolutional neural network (CNN) with similarity measure for text processing (SMTP) loss, achieving high accuracy on the MEL dataset. Similarly, Rasheed *et al.*^[8] introduced hybrid-6, a DL model combined with feature optimization for eczema classification. Their model achieved 88.29% accuracy on the Eczema image resource (EIR) dataset, addressing challenges related to data complexity. Ren *et al.*^[9] proposed the modified differential evolution algorithm for pathological image segmentation, which demonstrated superior performance on breast and skin cancer images. A limitation of this study is the lack of parallel computing; future work will focus on integrating it to improve efficiency. Shekar and Hailu^[10] combined DenseNet-169 with transfer learning and feature extraction, achieving 89.09% accuracy on the ISIC dataset. The study is limited to binary classification, with future research aimed to investigate multiclass classification to broaden its applicability. Okur and Turkan^[11] presented an automated MEL detection approach using deep feature extraction combined with the bag of visual words (BoVW), achieving 96.2% accuracy on ISIC 2017. Further refinement of BoVW and clustering techniques is planned to enhance accuracy.

Akilandasowmya *et al.*^[12] proposed a skin cancer detection system using SCSO-ResNet50 and enhanced harmony search, achieving 92.04% accuracy on Kaggle and 94.24% on ISIC 2019. Their method outperformed existing techniques, showing promise for early diagnosis. Reis and Turk^[13] introduced multi-head attention block

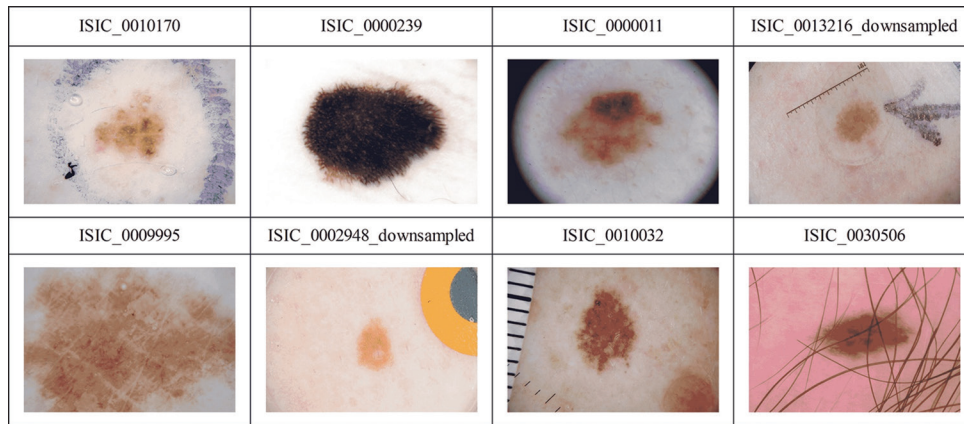


Figure 1: Examples of artifacts from International Skin Imaging Collaboration 2019 images

depthwise separable convolution network (MABSCNET), a hybrid depthwise separable convolution and vision transformers (ViTs) that integrates traditional ML algorithms. They incorporated contrast-limited adaptive histogram equalization and hypercolumn techniques for image enhancement, achieving an accuracy of 92.74% through ensemble learning. Their results emphasize the effectiveness of hybrid models in dermatological image classification.

Wang *et al.*^[14] proposed DSML-UNet for medical image segmentation, improving accuracy through multiscale large kernel convolutions and ResVgg modules. Zloto *et al.*^[15] designed an automated framework for diagnosing malignant eyelid tumors, achieving 93.8% sensitivity and 73.7% specificity. However, dataset limitations impact robustness, necessitating dataset expansion. They plan to expand the dataset and enhance model robustness.

Rasel *et al.*^[16] introduced a deep convolutional neural network (DCNN) to detect the blue-white veil in skin lesions, achieving 95.05% accuracy. The model incorporates Explainable artificial intelligence, showing potential for early MEL detection. Singh *et al.*^[17] developed an attention-learning DCNN integrating Zernike moments and transfer learning with DenseNet201 and Xception, addressing intraclass variation and data imbalance but facing computational and slower training times pose challenges.

Saha *et al.*^[18] proposed YoTransViT, a ViT-based framework using YOLOv8 for segmentation, achieving strong results on ISIC 2019. The model includes a web-based application, though additional datasets are needed for improved robustness. Khanna *et al.*^[19] designed a CNN with Swin Transformers, achieving 73% accuracy on ISIC 2019. Further enhancements are needed to improve performance on complex patterns through preprocessing and postprocessing techniques.

Mir *et al.*^[20] presented LesNet, a transfer learning-based architecture using DenseNet, VGG-16, and Inception with data augmentation, achieving 98% accuracy on HAM10000

and 94% on ISIC 2019. Future work will enhance interpretability through texture analysis.

Huang *et al.*^[21] utilized ADF-Net, a hybrid transformer-CNN framework for skin lesion segmentation, featuring adaptive feature fusion and focal attention decoding, with strong performance and efficient inference. Challenges include handling large-scale lesion variations and low-contrast regions. Thapar *et al.*^[22] combined K-means and GOA for segmentation with CNN-based classification, improving dermoscopy image analysis, and reducing overfitting. However, reliance on labeled medical datasets and limited exploration of DL models restricts its scalability.

Shankar *et al.*^[23] presented a DL framework using Xception CNN with transfer learning, achieving 97% accuracy and 0.93 area under the curve-receiver-operating characteristic (AUC-ROC) on 19,500 images. They addressed data imbalance using augmentation and class weights. Future work aims to expand the dataset and improve clinical validation. Natha *et al.*^[24] employed a max voting ensemble method using random forest (RF), support vector machine (SVM), and multi-layer perceptron network (MLPN), achieving 94.7% accuracy on HAM10000 and ISIC 2018. Future work includes integrating DL features and optimizing for clinical deployment. Noaman *et al.*^[25] introduced a skin lesion detection system using transfer learning with AlexNet, achieving 90.9% accuracy. However, it is not 100% accurate and should be used as a complementary tool to expert diagnosis. Future work will focus on fine-tuning for better performance across datasets. Pacal *et al.*^[26] improved the Swin Transformer by implementing hybrid shifted window-based multihead self-attention to capture finer details and better handle overlapping skin cancer regions. Their proposed method achieved 89.36% accuracy on the ISIC 2019 dataset. It outperformed CNN- and ViT-based methods.

Despite advancements, they still face key challenges, such as class imbalance, visually similar lesion types, and artifacts such as hair, clinical markings, and lighting. While DL models, especially transformer-based architectures,

have improved classification performance, they require large datasets and high computational power, limiting their practical application. To address these limitations, our study introduces a hybrid approach that integrates an efficient feature extraction method (hybrid OC-LBP) with an ensemble learning model (ET). In addition, our model includes explicit artifact removal using GMM, ensuring cleaner input data, improving accuracy, and enhancing generalization while reducing computational overhead. This combination provides a more practical solution for real-world clinical applications.

Methods

To support automated skin lesion classification, we designed a structured framework combining several image analysis and ML components. The overall structure of our proposed method is summarized in Figure 2, which outlines the key stages of the pipeline, including preprocessing, feature extraction, feature selection, and classification. This block diagram serves as a visual overview to guide the reader through the technical details that follow.

Introduction to the database

The dataset used in this study is sourced from the publicly available ISIC 2019 Challenge dataset.^[27-29] It contains 25331 images of skin lesions across eight diagnostic categories, as shown in Table 1. All images are stored in JPEG format for consistency and ease of processing. Importantly, patient information remains confidential. Figure 3 provides

examples of these skin lesions. There is a noticeable class imbalance, with some categories underrepresented. This imbalance can hinder model generalization. To address this, we applied random oversampling techniques to improve the model’s ability to learn from the underrepresented classes. In this approach, samples from the minority class are randomly duplicated to match the number of instances in the majority class, creating a more balanced dataset

Image preprocessing

Image preprocessing enhances the quality of input data for better analysis. Here, we introduce a novel method based on the GMM combined with geometric features. This approach removes dermoscope measurement overlays, such as black regions, without requiring image cropping, thereby preserving more of the original image content for analysis. Figure 4, where multiple images with various artifacts such as dermoscope overlays, stickers, and clinical markings are shown before and after preprocessing, illustrating the removal of these artifacts using GMM.

GMM algorithm

The Gaussian mixture model (GMM) probabilistically segments image artifacts (hair, overlays, and clinical markings) by modeling pixel intensity distributions. Unlike fixed-threshold methods, GMM dynamically distinguishes artifacts while preserving lesion structures, enhancing input quality for feature extraction.

The GMM models the image data as a combination of several Gaussian distributions, each representing a cluster in the mixture distribution. The number of clusters can vary depending on the complexity and color distribution in the image. Each Gaussian cluster is characterized by parameters, including the mean, which represents the center of the cluster; the covariance matrix, which describes the shape and orientation of the cluster; and the weight, which indicates the proportion of data points in the cluster.

The expectation-maximization algorithm is employed to optimize these parameters iteratively. The process involves two key steps, including the Expectation step which assigns

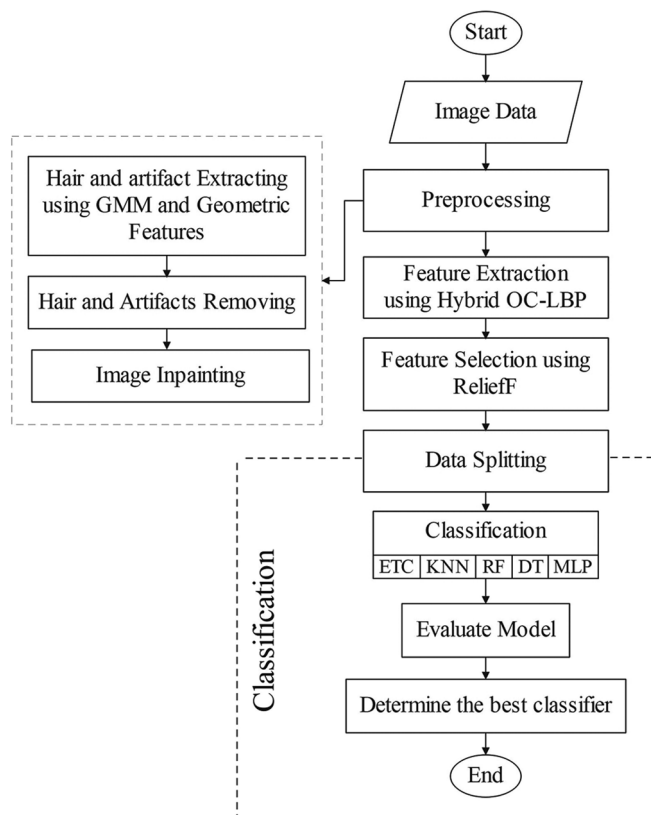


Figure 2: Block diagram of the proposed algorithm

Table 1: Class distribution of the International Skin Imaging Collaboration 2019 dataset

Abbreviation	Number of samples
AK	867
BCC	3323
BKL	2624
DF	239
MEL	4522
NV	12,875
SCC	628
VASC	253

AK – Actinic keratosis; BCC – Basal cell carcinoma; BKL – Benign keratosis; DF – Dermatofibroma; MEL – Melanoma; NV – Nevus; SCC – Squamous cell carcinoma; VASC – Vascular lesions

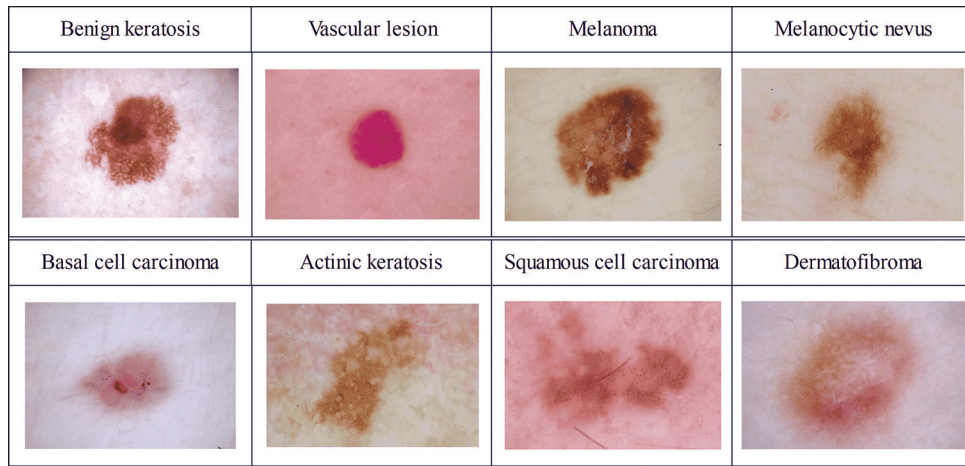


Figure 3: Examples of skin images from the eight categories in the International Skin Imaging Collaboration 2019 dataset

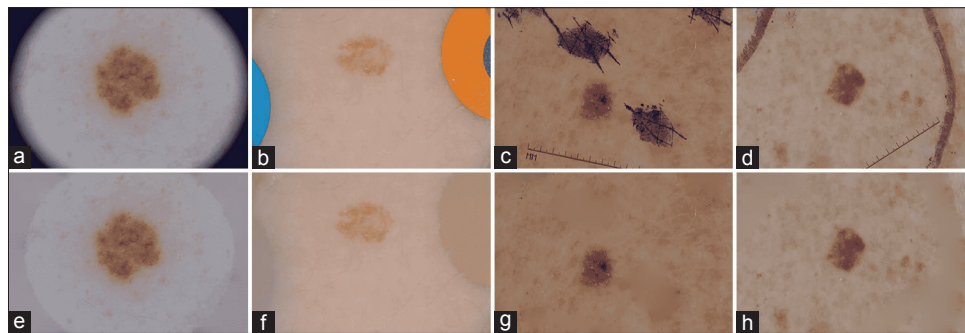


Figure 4: Before and after preprocessing: Dermoscope overlay (a and e), stickers (b and f), and clinical markings (c-h)

probabilities to each data point, representing the likelihood of belonging to each cluster and the Maximization step which updates the parameters (mean, covariance, and weight) based on the probabilities assigned in the previous step. This iterative process continues until a termination condition is met, such as a fixed number of iterations or predefined convergence criteria. The trained GMM then uses the calculated probabilities for each cluster to identify and separate artifacts from relevant lesion data. Finally, the artifact regions are filled with the average color of the surrounding lesion area, ensuring a smooth transition and preserving the visual integrity of the lesion for further processing. We selected the number of Gaussian components dynamically using the Bayesian information criterion (BIC) over a range of $k = 2-6$. The model automatically chooses the optimal k -value based on the lowest BIC score. Regularization ($1e-5$) was applied to maintain numerical stability.

Mathematically, the GMM^[30] is expressed as a weighted sum of normal distributions. The probability density function of a GMM with components is given by:

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \Sigma_k) \quad (1)$$

$p(x)$ represents the mixed density function at the point x , x denotes a d dimensional observation vector, and π_k

indicates the mixing weight, in which $\sum_{k=1}^K \pi_k = 1$. The Gaussian distribution for the k -th mixture component can be defined as:

$$N(x | \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp\left\{-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right\} \quad (2)$$

where μ_k is the mean and Σ_k is the covariance matrix for the k -th Gaussian component.

Σ_k^{-1} is the inverse of covariance matrix

The GMM was chosen for this study due to its ability to probabilistically model the complex distributions of pixel intensities in dermoscopic images, enabling precise segmentation of artifacts while preserving lesion information critical for classification tasks.

Feature extraction employing color hybrid orthogonal combination of local binary patterns

Feature extraction is the process of identifying important characteristics or features from an image that are critical for classification or analysis. In this study, we employed the hybrid OC-LBP method to extract features from abnormal skin regions. This method enhances the functionality of

the standard local binary patterns (LBP) by incorporating both color information and multiscale radii, enabling more robust and discriminative feature extraction. Standard LBP detects grayscale intensity differences to capture local texture patterns. However, this method may overlook subtle color variations and complex textures, which are crucial for medical image analysis. The OC-LBP^[31] method addresses this limitation by integrating combinations of LBP with color channels, allowing it to extract more comprehensive features, especially for lesions with distinct color patterns. The hybrid OC-LBP extends the OC-LBP by incorporating multiple radii around the central pixel, thereby capturing features at different spatial scales. This multiscale approach enables the detection of both fine-grained and large-scale structures, improving robustness to variations in lesion size and lighting conditions. Mathematically, LBP^[32] is expressed as:

$$LBP_{p,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \times 2^p \quad ;$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

where g_p is the intensity of the p -th neighboring pixel, g_c is the center pixel Intensity. The dimension of LBP is 2^p . For Hybrid OC-LBP, the method applies this operator across multiple radii R_1, R_2, \dots, R_n . For instance, using 4 neighbors at two different radii R_1 and R_2 , the method generates two histograms, each with 16 dimensions (one for each radius). These histograms are then concatenated, resulting in a total histogram dimension of:

$$2^4 \times \left[\frac{P}{4}\right] \times 2 = 4 \times P \times 2 \quad (4)$$

Where is the number of neighboring pixels forming a circle around c . Unlike dimension-reduction techniques that

may compromise feature discrimination, Hybrid OC-LBP retains the discriminative ability of OC-LBP while adapting to multiscale textures. The inclusion of multiple radii enhances its capacity to handle variations in lesion size, lighting conditions, and color patterns. By leveraging orthogonal combinations and multiscale histograms, Hybrid OC-LBP provides a compact yet comprehensive feature representation. The proposed method is shown in Figure 5, which visually depicts the process of feature extraction using hybrid OC-LBP.

Feature selection

We used the ReliefF algorithm^[33] for feature selection. This algorithm assigns weights to features based on their ability to distinguish between classes, ensuring that only the most relevant features are used for classification. In our study, we selected eight salient features from a total of 864 based on preliminary testing. These top 8 features were ranked using ReliefF. Selecting fewer than 8 features resulted in reduced classification accuracy, while using more than 8 increased the dimensionality without any measurable improvement in performance. Therefore, 8 features were found to strike an effective balance.

Machine learning approach

We applied several supervised learning algorithms to classify the skin lesions based on extracted features: Multilayer perceptron (MLP), k-nearest neighbors (KNN), decision tree (DT), extra tree (ET), and RF. MLP is a neural network that excels at learning complex patterns, while KNN classifies based on the majority of neighbors. DT makes hierarchical decisions based on feature importance, and ET improves upon DT for better generalization. RF aggregates multiple DTs to predict the most frequent class. We optimized hyperparameters for the classifiers used in

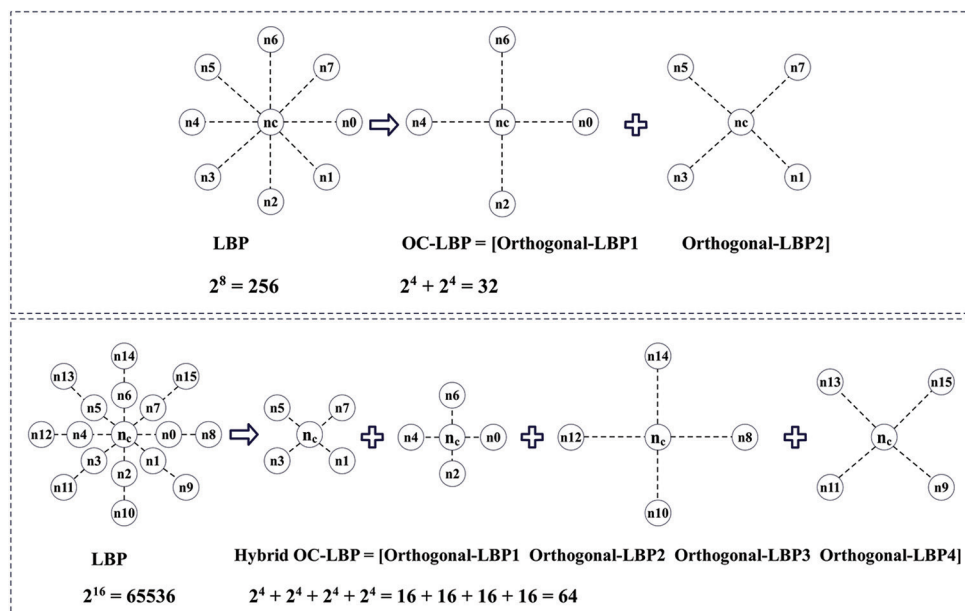


Figure 5: Block diagram comparing local binary patterns (LBP), orthogonal combination of LBP (OC-LBP), and hybrid OC-LBP

this study to achieve the best possible performance. For instance, ET classifier was tested with various values for the number of estimators and maximum tree depth, with optimal settings identified as and . Similarly, KNN was evaluated with different neighbor counts ($n_neighbors = 1, 3, 5$), while the MLP classifier was tuned by varying hidden layer sizes (1000, 1500, and 2000) and maximum iterations ($max_iter = 30$). These optimized parameters, as listed in Table 2, were used in the final evaluation.

Performance metrics

We evaluated model performance using several metrics derived from the confusion matrix: Accuracy, precision, recall, F1-score, TPR, FPR, and log loss. These metrics provide a comprehensive view of model effectiveness. Formulas for these metrics are provided in Table 3 for precise model evaluation.

Table 2: Optimized hyperparameters of machine learning algorithms

Algorithm	Optimized hyperparameters
ETC	$n_estimators=200, criterion="entropy", max_depth=300$
KNN	$n_neighbors=1, metric="euclidean"$
MLP	$hidden_layer_sizes=(1500, 1500), max_iter=30$
DT	$max_depth=100, criterion="entropy", splitter="best"$
RF	$n_estimators=20, criterion="entropy", max_depth=100$

$n_neighbors$ – Number of nearest neighbors for the KNN algorithm;
 $n_estimators$ – Number of DTs in the RF. KNN – K-nearest neighbors;
 MLP – Multilayer perceptron; DT – Decision tree; RF – Random forest; ETC – Estimate to complete

Table 3: Equations for evaluating performance

Performance measure	Definition	Definition	Equation
Accuracy	Overall correctness	$\frac{TP + TN}{TP + FP + FN + TN} \times 100$	(5)
Precision	Correct positive predictions	$\frac{TP}{TP + FP} \times 100$	(6)
Recall	Correctly identified positive cases	$\frac{TP}{TP + FN} \times 100$	(7)
F1-score	Harmonic mean of precision and recall	$2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100$	(8)
TPR	TPR	$\frac{TP}{TP + FN}$	(9)
FPR	FPR	$\frac{FP}{FP + FN}$	(10)
Log loss	Cross-entropy loss	$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(P_{ij}) \times 100$	(11)

FPR – False-positive rate; TPR – True-positive rate

Results

MATLAB 21b (MathWorks, Natick, MA, USA) was used for preprocessing, feature extraction, and selection, while Python 3.10.9 (Python Software Foundation, Wilmington, DE, USA) and Google Colab (Google LLC, Mountain View, CA, USA) were utilized for the classification process. This study evaluates the performance of our classification model using the ISIC 2019 dataset. We compare several classifiers, including Ensemble ET, KNN, DT, RF, and MLP, focusing on their strengths and limitations. The classification results, with accuracy scores for each model on the imbalanced dataset, are presented in Table 4. Our proposed hybrid OC-LBP method outperforms other approaches, including standard OC-LBP. The selection of radii values in the hybrid OC-LBP method is based on the principles of multiscale texture analysis and spatial feature representation, ensuring an optimal balance between local and global feature extraction. The choice of radii directly impacts the model’s ability to capture fine-grained textural details while preserving broader structural information critical for lesion classification. To determine the most effective radius combination, we theoretically considered the impact of different values on feature discrimination. Smaller radii emphasize local micropatterns, which are useful for capturing fine textures but may introduce sensitivity to noise. In contrast, larger radii extract broader lesion structures, which improve generalization but may incorporate irrelevant background information. Based on these principles, multiple radius combinations – (1, 4),

(2, 5), and (3, 6) – were assessed: $r = 1$ captured high-frequency textures but was prone to noise, leading to unstable feature extraction. $r = 3$ improved differentiation but lacked robustness for capturing large-scale structures. $r = 4$ and $r = 5$ provided a more comprehensive multiscale representation by integrating both fine and coarse texture variations. $r = 6$ included excessive background information, reducing feature distinctiveness. Ultimately, $r = 2$ and $r = 5$ were chosen as an optimal trade-off, balancing local and global information: $r = 2$ captures fine-grained textures, ensuring sensitivity to subtle lesion patterns. $r = 5$ captures broader lesion structures, improving overall feature representation. This selection aligns with multiscale texture analysis principles, optimizing classification robustness while mitigating noise interference. Additionally, to reinforce this theoretical foundation, an empirical validation was conducted, confirming that the (2, 5) combination achieved the highest classification accuracy. This ensures that the choice of radii is well-supported and effectively enhances feature extraction performance.

We conducted comprehensive analyses using two validation techniques on a balanced dataset, including 80/20 cross-validation and 10-fold cross-validation. These methods were employed to evaluate the robustness and generalization of the proposed model. Key performance metrics, including accuracy, precision, recall, F1-score, and log loss, were calculated for five classifiers. The results are summarized in Table 5.

The ET classifier outperformed other models, reaching 97.31% accuracy (10-fold CV) with superior precision

and the lowest log loss (0.06). In comparison, KNN performed competitively (96.84% accuracy) but had higher log loss (1.13–1.24), indicating greater misclassification likelihood. The RF classifier exhibited reliable performance, achieving accuracy scores of 96.38% in 10-fold cross-validation and 96.22% on the 80/20 split. While RF’s ensemble-based approach ensured strong generalization, its performance marginally lagged behind ET in terms of accuracy and log loss. The DT classifier, as a simpler model, showed lower accuracy and F1-scores than ensemble methods such as ET and RF. These results indicate that DT struggled to generalize effectively, particularly for minority classes, reaffirming the advantages of ensemble techniques in improving classification performance. The MLP recorded the lowest accuracy among the classifiers, with scores of 92.68% in 10-fold cross-validation and 92.05% on the 80/20 split. Its relatively high log loss (0.29–0.30) and lower precision and recall metrics suggest that MLP may require further optimization or a larger dataset to achieve comparable performance with other models. Ultimately, the ET classifier consistently outperformed the other models across all evaluation metrics, making it the most robust and reliable choice for this study. While KNN and RF delivered competitive accuracy, their higher log loss values indicate a slight compromise in prediction confidence. On the other hand, DT and MLP, despite their potential, underperformed in comparison, further emphasizing the effectiveness of ensemble-based methods like ET in achieving superior classification results.

The confusion matrix provided valuable insights into the classifier’s performance across different lesion types,

Table 4: Performance of orthogonal combination of local binary patterns and hybrid orthogonal combination of local binary patterns on an imbalanced dataset

Methodology	K-P-R	ET (%)	RF (%)	DT (%)	KNN (%)	MLP (%)
Hybrid OC-LBP	1-4-2,5	57.81	54.78	51.59	50.91	50.83
OC-LBP	1-4-2	50.55	53.83	50.28	50.67	44.83
OC-LBP	1-4-5	50.95	51.93	51.02	49.84	43.33

K – Number of key points in the image used for feature extraction; P – Number of neighbors surrounding the key points; R – Radius; OC-LBP – Orthogonal combination of local binary patterns; RF – Random forest; KNN – K-nearest neighbors; MLP – Multilayer perceptron; DT – Decision tree; ET – Extra tree

Table 5: Comparison of classification algorithms on balanced datasets using cross-validation

Cross-validation	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Log loss (%)
80/20	ET	96.82	96.06	96.82	96.71	0.06
	KNN	96.54	96.68	96.54	96.37	1.24
	DT	96.44	96.61	96.44	96.26	1.28
	RF	96.22	96.54	96.22	96.01	0.17
	MLP	92.05	92.17	92.05	91.89	0.30
10-fold	ET	97.31	97.54	97.31	97.23	0.06
	KNN	96.84	96.12	96.84	96.67	1.13
	DT	96.79	96.91	96.79	96.62	1.15
	RF	96.38	96.65	96.38	96.18	0.13
	MLP	92.68	93.23	92.68	92.53	0.29

RF – Random forest; KNN – K-nearest neighbors; MLP – Multilayer perceptron; DT – Decision tree; ET – Extra tree

Table 6: Comparative analysis of results from existing literature

Author	Dataset	Algorithm	Methodology (features)	Classes	Accuracy (%)
This study	ISIC-2019	ETC, KNN, DT, RF, MLP	Hybrid OC-LBP	8	97.31, 96.84, 96.79, 96.38, 92.68
Natha <i>et al.</i> ^[24]	ISIC-2018	Max voting ensemble method	GLCM + color histograms + color moments	7	94.70
Ahmed <i>et al.</i> ^[34]	ISIC-2019	KNN, DT	GLCM + statistical	8	94, 93
Yang <i>et al.</i> ^[35]	ISIC-2018	SLP-Net	-	7	93.87
Ghahfarrokhi <i>et al.</i> ^[36]	PH 2	KNN	Nonlinear + GLCM + texture	2	92.46
Gupta <i>et al.</i> ^[37]	ISIC-2019	CNN	-	8	90
Ozkan and Koklu ^[38]	PH 2	DT, KNN	ABCD	3	90.00, 82.00
Albawi <i>et al.</i> ^[39]	ISIC	KNN	GLCM + 2D-DWT	3	87.46

KNN – K-nearest neighbors; RF – Random forest; KNN – K-nearest neighbors; MLP – Multilayer perceptron; DT – Decision tree; RF – Random forest; ISIC – International skin imaging collaboration; OC-LBP – Orthogonal combination of local binary patterns; ETC – Estimate to complete; DWT – Discrete wavelet transform; GLCM – Gray-level co-occurrence matrix; SLP – Single-layer perceptron; CNN – Convolutional neural network

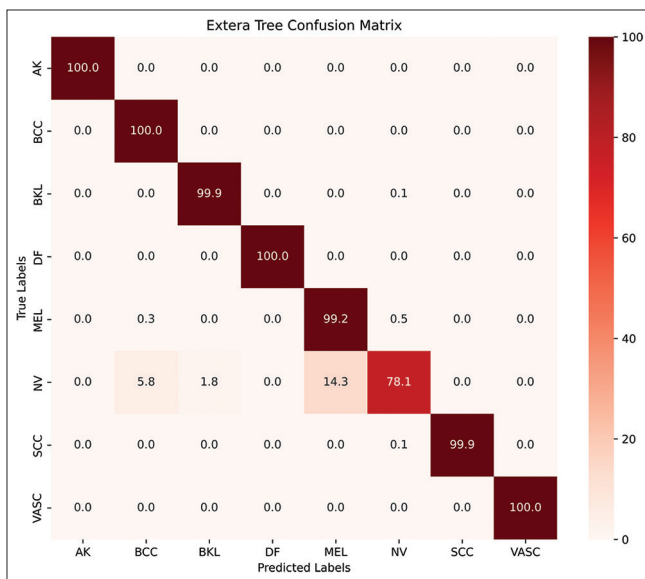


Figure 6: Confusion matrix of the International Skin Imaging Collaboration 2019 dataset with ET classifier

helping us identify misclassifications. As shown in Figure 6, the model was highly accurate in classifying lesions such as AK, basal cell carcinoma (BCC), DF, squamous cell carcinoma, MEL, and vascular lesions. While our method demonstrated high overall accuracy, the classification of NV lesions presented a challenge, with an accuracy of 78.1%. NV lesions were classified with 78.1% accuracy, mainly due to textural and color similarities with benign keratosis, BCC, and MEL. Early-stage MEL lesions closely resemble NV, complicating boundary distinction in dermoscopic images.

The ROC curve, which assesses the trade-off between sensitivity TPR and FPR, was used to evaluate the ET classifier. As shown in Figure 7, the ET model achieved perfect AUC values of 1.00 for all classes, indicating its strong performance in distinguishing each skin lesion type. The macroaverage AUC was also 1.00, reflecting the model’s consistent accuracy across all classes.

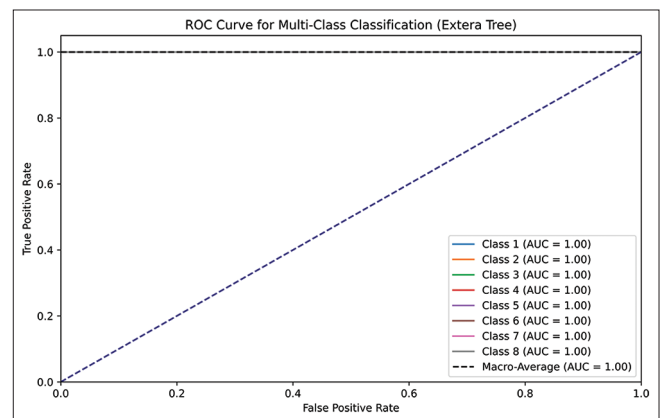


Figure 7: Receiver-operating characteristic curves for International Skin Imaging Collaboration 2019: Evaluating the ET Classifier. AUC: Receiver-operating characteristic

Table 6 provides a comparative analysis of recent methods for skin lesion classification. Our proposed hybrid OC-LBP + ET achieves the highest accuracy (97.31%), surpassing various traditional and DL-based approaches. Methods such as Natha *et al.* (94.70%) and Ahmed *et al.* (94%) utilize GLCM and statistical techniques, which, while effective, do not fully capture multiscale texture and color variations within skin lesions. Our approach overcomes this limitation by integrating hybrid texture descriptors and ensemble learning, leading to improved feature representation. In addition, CNN-based models, such as Gupta *et al.*(90%), achieve competitive performance but require high computational resources and large datasets, making them less practical for real-time clinical applications. Our method, by contrast, maintains high accuracy while being computationally efficient, making it more suitable for real-world dermatology use. Furthermore, traditional ML methods (e.g., Ghahfarrokhi *et al.* – 92.46%) lack the advanced feature selection mechanisms that our approach employs, limiting their classification robustness. These comparisons emphasize the advantages of our method in achieving state-of-the-art

accuracy while ensuring computational feasibility and clinical applicability, reinforcing its contribution to the field of dermatological image analysis.

Discussion

This study presents a novel and computationally efficient hybrid framework tailored for the classification of skin lesions, combining the strengths of a Gaussian mixture model (GMM)-based preprocessing stage with a multiscale hybrid OC-LBP feature extraction mechanism. This integration addresses several inherent challenges in dermoscopic image analysis, such as heterogeneous lighting conditions, speckle noise, and dermoscopic artifacts, all of which can obscure diagnostically relevant patterns and reduce the performance of automated classifiers. By leveraging the probabilistic modeling capabilities of GMM, our preprocessing pipeline is capable of suppressing background noise and nonlesion regions, thereby enhancing the contrast and saliency of lesion-specific features before extraction.

The hybrid OC-LBP component operates over multiple radii, enabling the capture of both micro- and macrotextural features. This multiscale design is particularly advantageous in dermatological imaging, where lesions exhibit a wide spectrum of spatial granularity. Moreover, by orthogonally combining complementary local descriptors, our feature extraction method ensures a richer and more discriminative representation of lesion textures compared to traditional LBP variants.

In terms of classification, the use of an ensemble-based strategy built upon the Extremely Randomized Trees (ET) algorithm facilitates robustness to outliers and improves generalizability across lesion subtypes. The achieved classification accuracy of 97.31% not only exceeds that of conventional ML baselines but also outperforms state-of-the-art DL models, including Transformer-based architectures such as Swin Transformer, ViT, and hybrid CNN-Transformer models. While Transformers are theoretically well-suited for global feature modeling, their real-world applicability in medical imaging remains constrained by their sensitivity to dataset scale, class imbalance, and the lack of inductive biases essential for texture-centric domains like dermatology. In addition, their performance often deteriorates when confronted with subtle interclass variations and visually similar lesion types in small datasets. Contrastingly, our framework capitalizes on handcrafted features that are intrinsically less reliant on the volume and diversity of training data. This renders the system more resilient in resource-limited settings, such as low-resource clinics. Furthermore, the explicit artifact suppression and targeted feature selection embedded in our pipeline result in a compact and noise-robust representation, enhancing overall model performance, which is a critical factor for real-time or embedded deployment.

Nonetheless, our method is not without limitations. A notable challenge remains the accurate classification of clinically confounding classes, particularly melanocytic nevus and early-stage MEL, which share overlapping color distributions and textural features. While preliminary investigations suggested that the inclusion of shape-based descriptors might enhance class separability, practical constraints imposed by the ISIC-2019 dataset limited this direction. Many of the dataset's images are precropped or framed in such a way that lesion borders are truncated or missing entirely, thereby precluding the reliable extraction of global shape attributes. Furthermore, the development of a dedicated and accurate lesion segmentation module, which would be a prerequisite for extracting meaningful shape features, falls outside the scope of the present framework and would require a significantly different architectural pipeline.

In conclusion, the proposed hybrid framework demonstrates that by thoughtfully combining classical ML principles with domain-specific insights and preprocessing, it is possible to attain high classification performance without the computational overhead and data demands of contemporary DL models. Its ability to balance accuracy, interpretability, and computational efficiency positions it as a viable candidate for integration into dermatological diagnostic workflows, particularly in environments constrained by hardware or data limitations.

Conclusions

This study presents a novel hybrid framework for classifying skin lesions, combining Hybrid OC-LBP for feature extraction with GMM for preprocessing. The method effectively tackles key challenges in medical image analysis, such as noise, lighting variations, and artifacts. These qualities make it a practical alternative to complex, resource-intensive approaches. Our results showcase the framework's high accuracy and robustness, underscoring its potential to enhance diagnostic workflows in dermatology. However, the model does face some limitations, particularly in differentiating visually similar lesion types. Addressing these challenges in future research through advanced feature engineering, hyperparameter optimization can further strengthen its performance. With continued refinement and validation, this hybrid approach has the potential to seamlessly integrate into clinical diagnostic systems, improving patient outcomes and pushing the boundaries of medical image analysis.

Ethical approval

The data used in this study were obtained from the publicly available International Skin Imaging Collaboration (ISIC 2019) dataset. Ethical approval and participant consent were obtained by the ADNI investigators, and all procedures were conducted in accordance with relevant guidelines and regulations.

Funding

This research did not receive any specific grant funding.

Availability of data and materials

The datasets utilized in this study are publicly available from the ISIC 2019 Challenge repository (<https://challenge.isic-archive.com/data/#2019>).

Acknowledgments

No financial or advisory support was received for this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Malik S, Islam SM, Akram T, Naqvi SR, Alghamdi NS, Baryannis G. A novel hybrid meta-heuristic contrast stretching technique for improved skin lesion segmentation. *Comput Biol Med* 2022;151:106222.
- Khan MA, Akram T, Zhang YD, Sharif M. Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recognit Lett* 2021;143:58-66.
- Cassidy B, Kendrick C, Brodzicki A, Jaworek-Korjakowska J, Yap MH. Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Med Image Anal* 2022;75:102305.
- Yao P, Shen S, Xu M, Liu P, Zhang F, Xing J, *et al.* Single model deep learning on imbalanced small datasets for skin lesion classification. *IEEE Trans Med Imaging* 2022;41:1242-54.
- Jena B, Naik MK, Panda R, Abraham A. A novel minimum generalized cross entropy-based multilevel segmentation technique for the brain MRI/dermoscopic images. *Comput Biol Med* 2022;151:106214.
- Houssein EH, Abdelkareem DA, Emam MM, Hameed MA, Younan M. An efficient image segmentation method for skin cancer imaging using improved golden jackal optimization algorithm. *Comput Biol Med* 2022;149:106075.
- Patil R, Bellary S. Machine learning approach in melanoma cancer stage detection. *J King Saud Univ Inf Sci* 2022;34:3285-93.
- Rasheed A, Umar AI, Shirazi SH, Khan Z, Nawaz S, Shahzad M. Automatic eczema classification in clinical images based on hybrid deep neural network. *Comput Biol Med* 2022;147:105807.
- Ren L, Zhao D, Zhao X, Chen W, Li L, Wu T, *et al.* Multi-level thresholding segmentation for pathological images: Optimal performance design of a new modified differential evolution. *Comput Biol Med* 2022;148:105910.
- Shekar BH, Hailu H. An efficient stacked ensemble model for the detection of COVID-19 and skin cancer using fused feature of transfer learning and handcrafted methods. *Comput Methods Biomech Biomed Eng Imaging Vis* 2023;11:878-94.
- Okur E, Turkan M. Weighted bag of visual words with enhanced deep features for melanoma detection. *Expert Syst Appl* 2024;237:121531.
- Akilandasowmya G, Nirmaladevi G, Suganthi SU, Aishwariya A. Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for early detection and classification. *Biomed Signal Process Control* 2024;88:105306.
- Reis HC, Turk V. Fusion of transformer attention and CNN features for skin cancer detection. *Appl Soft Comput* 2024;164:112013.
- Wang B, Qin J, Lv L, Cheng M, Li L, He J, *et al.* DSML-UNet: Depthwise separable convolution network with multiscale large kernel for medical image segmentation. *Biomed Signal Process Control* 2024;97:106731.
- Zloto O, Fogel O, Ben Simon G, Rosner M, Vishnevskia-Dai V, Hostovsky A, *et al.* Computer-aided diagnosis of eyelid skin tumors using machine learning. *Can J Ophthalmol* 2025;60:e275-80.
- Rasel MA, Abdul Kareem S, Kwan Z, Yong SS, Obaidallah U. Bluish veil detection and lesion classification using custom deep learnable layers with explainable artificial intelligence (XAI). *Comput Biol Med* 2024;178:108758.
- Singh C, Ranade SK, Singh SP. Attention learning models using local Zernike moments-based normalized images and convolutional neural networks for skin lesion classification. *Biomed Signal Process Control*. 2024;96:106512.
- Saha DK, Joy AM, Majumder A. YoTransViT: A transformer and CNN method for predicting and classifying skin diseases using segmentation techniques. *Inform Med Unlocked* 2024;47:101495.
- Khanna S, Tyagi A, Sharma S, Bharti AK. Pixels to prognosis: Unveiling skin lesion patterns through swin transformer. *Procedia Comput Sci* 2024;245:193-201.
- Mir AN, Nissar I, Rizvi DR, Kumar A. LesNet: An automated skin lesion deep convolutional neural network classifier through augmentation and transfer learning. *Procedia Comput Sci* 2024;235:112-21.
- Huang Z, Deng H, Yin S, Zhang T, Tang W, Wang Q. ADF-net: A novel adaptive dual-stream encoding and focal attention decoding network for skin lesion segmentation. *Biomed Signal Process Control* 2024;91:105895.
- Thapar P, Rakhra M, Alsaadi M, Quraishi A, Deka A, Ramesh JV. A hybrid grasshopper optimization algorithm for skin lesion segmentation and melanoma classification using deep learning. *Healthc Anal* 2024;5:100326.
- Shankar M, Rehman S, Naveed S, Khan PA, Mushraq S. Skin disease detection using CNN. In: *Challenges in Information, Communication and Computing Technology*. London: CRC Press; 2025. p. 437-42.
- Natha P, Tera SP, Chinthaginjala R, Rab SO, Narasimhulu CV, Kim TH. Boosting skin cancer diagnosis accuracy with ensemble approach. *Sci Rep* 2025;15:1290.
- Noaman A, Ahmad R, Khan MF, Mohammed AS, Farooq M, Adnan KM. Beyond binary: Multi-class skin lesion classification with AlexNet transfer learning-towards enhanced dermatological diagnosis. *Discov Appl Sci* 2025;7:1-18.
- Pacal I, Alaftekin M, Zengul FD. Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP. *J Imaging Inform Med* 2024;37:3174-92.
- Hernández-Pérez C, Combalia M, Podlipnik S, Codella NC, Rotemberg V, Halpern AC, *et al.* BCN20000: Dermoscopic lesions in the wild. *Sci Data* 2024;11:641.
- Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, *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)*. IEEE; 2018.

- p. 168-72.
29. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 2018;5:180161.
 30. Qiao J, Cai X, Xiao Q, Chen Z, Kulkarni P, Ferris C, *et al.* Data on MRI brain lesion segmentation using K-means and Gaussian mixture model-expectation maximization. *Data Brief* 2019;27:104628.
 31. Zhu C, Bichot CE, Chen L. Image region description using orthogonal combination of local binary patterns enhanced with color information. *Pattern Recognit* 2013;46:1949-63.
 32. Guo Z, Zhang L, Zhang D. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans Image Process* 2010;19:1657-63.
 33. Robnik-Šikonja M, Kononenko I. An Adaptation of Relief for Attribute Estimation in Regression. In: *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*. Citeseer; 1997. p. 296-304.
 34. Ahammed M, Al Mamun M, Uddin MS. A machine learning approach for skin disease detection and classification using image segmentation. *Healthc Anal* 2022;2:100122.
 35. Yang B, Zhang R, Peng H, Guo C, Luo X, Wang J, *et al.* SLP-Net: An efficient lightweight network for segmentation of skin lesions. *Biomed Signal Process Control* 2025;101:107242.
 36. Ghahfarrokhi SS, Khodadadi H, Ghadiri H, Fattahi F. Malignant melanoma diagnosis applying a machine learning method based on the combination of nonlinear and texture features. *Biomed Signal Process Control* 2023;80:104300.
 37. Gupta S, Jayanthi R, Verma AK, Saxena AK, Moharana AK, Goswami S. Ensemble optimization algorithm for the prediction of melanoma skin cancer. *Meas Sens* 2023;29:100887.
 38. Ozkan IA, Koklu M. Skin lesion classification using machine learning algorithms. *Int J Intell Syst Appl Eng* 2017;5:285-9.
 39. Albawi S, Abbas YA, Almadany Y. Robust skin diseases detection and classification using deep neural networks. *Int J Eng Technol* 2018;7:6473-80.