Mathematics and Computer Science
2025; 10(6): 92-101
http://www.sciencepublishinggroup.com/j/mcs
doi: 10.11648/j.mcs.20251006.11

Clinically Reliable Melanoma Classification from Darmoscopy Image Using Linear and Nonlinear MIL Approaches

Mahbub Hasan, Md Shohel Babu*

Department of Computer Science and Engineering, Southeast University, Dhaka, Bangladesh

Email address:
mahbub.hasan@seu.edu.bd (Mahbub Hasan), shohel.babu@seu.edu.bd (Md Shohel Babu)
*Corresponding author

To cite this article:
Mahbub Hasan, Md Shohel Babu. (2025). Clinically Reliable Melanoma Classification From Darmoscopy Image Using Linear and Nonlinear MIL Approaches. *Mathematics and Computer Science, 10*(6), 92-101. https://doi.org/10.11648/j.mcs.20251006.11

Received: 15 September 2025 ; **Accepted:** 25 September 2025 ; **Published:** 10 November 2025

Abstract: Melanoma is one of the most aggressive forms of skin cancer, accounting for an excessive number of skin cancer-related deaths despite its relatively low incidence. Early and accurate detection is crucial to improving survival rates, but manual diagnosis through dermoscopy remains challenging due to the subtle lesion characteristics. In this study, we propose a lightweight and resource-efficient framework for melanoma detection that integrates image segmentation with multiple instance learning (MIL)-based classification. Dermoscopic images from the ISIC 2018 dataset were preprocessed using resizing, noise removal, and hair artifact reduction to improve visual consistency. Lesion segmentation was performed using a statistical measure of pixel intensity standard deviation (SD), which adaptively determined the most suitable enhancement and thresholding techniques across three criteria. Segmentation outputs were then used to extract region-specific features, which were classified using three MIL-based approaches: mi-KSPSVM, mi-SPSVM, and MIL with Lagrangian relaxation, along with standard SVM baselines. Model performance was evaluated under 5-fold, 10-fold, and Leave-One-Out (LOO) cross-validation schemes using accuracy, sensitivity, specificity, F1-score, and computational efficiency as metrics. Experimental results demonstrate that segmentation significantly enhances classification performance, particularly in terms of sensitivity and F1-score. Notably, mi-KSPSVM achieved the highest sensitivity of 94.74% under 10-fold CV, while mi-SPSVM exhibited superior generalization under LOO validation. Additionally, preprocessing and segmentation consistently reduced misclassification of melanoma cases while preserving computational efficiency, with the SVM (RBF) baseline remaining the fastest model. Future work will explore the integration of deep learning-based segmentation to improve accuracy further and extend applicability to other skin cancer subtypes.

Keywords: Melanoma, Segmentation, MIL, SVM, Machine Learning, Classification, Multiple Instance Learning

1. Introduction

Melanoma is among the most lethal forms of skin cancer, despite representing a small fraction of total skin cancer cases. Its aggressive ability to spread to other parts of the body leads to disproportionately high death rates, and the number of cases around the world has been gradually rising over the past few decades [1]. Early identification is crucial because survival rates significantly decrease once the disease progresses beyond the localized stage [2].

Traditional diagnostics rely on dermoscopic visual inspection followed by histopathological analysis of biopsied tissue. However, these manual approaches are limited by inter-observer variability, subtle lesion presentations, and the time-intensive nature of expert interpretation [3]. To address these challenges, Computer Aided Diagnosis (CAD) systems have appeared as promising tools to support dermatologists in clinical decision making.

Using histopathological and dermoscopic images, CAD systems aim to improve diagnostic accuracy, consistency, and scalability. Advances in image processing, machine learning,

and deep learning have transformed melanoma analysis, with algorithms now achieving dermatologist-level performance in specific tasks [4]. Furthermore, frameworks such as Multiple Instance Learning (MIL) have gained attention for handling weakly labeled data, especially in Whole Slide Imaging (WSI), where pixel-level annotations are impractical [5].

The availability of benchmark datasets has played a crucial role in accelerating progress. The International Skin Imaging Collaboration (ISIC) has hosted annual challenges, providing thousands of dermoscopic images with segmentation masks and diagnostic labels, fostering standardization and reproducibility [6]. Datasets such as HAM10000 [7] and PH^2 [8] have further enriched the field with diverse, high-quality annotations.

In histopathology, the release of WSI datasets has enabled the development of MIL-based pipelines and weakly supervised learning strategies. Nevertheless, challenges such as domain shifts, label noise, class imbalance, and limited generalization across populations persist [9]. Given the rapid expansion of literature, synthesizing recent advances is essential to understand the current landscape and identify future directions. Prior surveys have often focused narrowly on segmentation or classification, or on specific methodological paradigms such as convolutional neural networks. However, the field has matured to encompass a broad spectrum of approaches from unsupervised clustering and saliency-based methods to supervised deep learning architectures and advanced classification pipelines integrating transformers, ensembles, and MIL frameworks.

The rest of the paper is organized as follows : (1) unsupervised segmentation methods, (2) supervised segmentation approaches, (3) classification techniques based on image processing and deep learning, and (4) classification strategies employing Multiple Instance Learning. By critically examining representative studies, datasets, and evaluation metrics, this paper aims to highlight current achievements, expose persistent gaps, and outline promising avenues for future research in melanoma computer-aided diagnosis.

2. Literature Review

Unsupervised segmentation is valuable when labeled data are scarce. Classical clustering, thresholding, and region-growing methods have been extended with uncertainty modeling. Celebi *et al.* [10] demonstrated clustering-based lesion segmentation as an early baseline. Ashour *et al.* [11] proposed a hybrid neutrosophic clustering with histogram-based estimation, improving ISIC 2016 accuracy. Bian *et al.* [12] introduced NeDSeM, a neutrosophy domain method to address fuzzy lesion edges. Oukil *et al.* [13] applied k-means segmentation with color-texture features across multiple color spaces, achieving 99.5% accuracy on PH2. Mabrouk *et al.* [14] developed a fully automated unsupervised pipeline for early pigmented lesion detection. Zhang *et al.* [15] applied deep clustering methods for dermoscopic segmentation. Li *et al.* [16] explored unsupervised histopathology segmentation. More recent works integrate semi-supervised or weakly supervised techniques, such as boundary-assisted models (Xu *et al.* [21]) that reduce dependency on full annotation. Overall, unsupervised approaches are practical but remain sensitive to illumination and artifact variations. Supervised segmentation dominates when annotated masks are available. Codella *et al.* [17] benchmarked U-Net and FCN baselines in ISIC challenges. Liu *et al.* [18] used Mask R-CNN with imperfect annotations for melanocytic proliferation. Wang *et al.* [19] introduced boundary-aware transformers to refine contour localization. Cai *et al.* [20] proposed BiADATU-Net, combining deformable and bidirectional attention for superior accuracy. Xu *et al.* [21] designed LB-UNet, a lightweight boundary-assisted model for efficient segmentation. MICCAI 2023 work [22] showed transformer-based networks outperform classical CNNs. Kreouzi *et al.* [23] reviewed supervised segmentation trends, noting improvements in boundary precision. These studies highlight the evolution toward hybrid CNN-transformer pipelines that balance accuracy and efficiency. Classification methods distinguish benign from malignant lesions, progressing from handcrafted features to deep networks. Y. Wu et al. [24] highlighted the potential of convolutional neural networks (CNNs) to support dermatologist-level accuracy in skin cancer classification using dermoscopic imagery. Yu *et al.* [25] validated deep models clinically. Oukil *et al.* [13] demonstrated color-texture features classified with SVM/ANN outperforming prior PH2 studies. Zhang *et al.* [26] achieved high accuracy on histopathology using CNNs. Alshawi *et al.* [27] built deep ensembles for multiclass lesion recognition. Yang *et al.* [28] proposed a ViT+EfficientNet ensemble with multi-scale attention, boosting ISIC 2018 classification to $\sim$95%. Kreouzi *et al.* [23] reviewed melanoma classification, emphasizing interpretability. Recent works also leverage semi-supervised pretraining and foundation models to improve generalization. MIL addresses the challenge of weakly labeled data, particularly in whole-slide imaging (WSI). Shao *et al.* [29] introduced TransMIL, a transformer-based MIL model capturing long-range dependencies. Li *et al.* [30] proposed dual-stream MIL integrating self-supervised contrastive learning. Godson *et al.* [31] used MIL for immune subtyping of melanoma WSIs, showing prognostic value. Meseguer-Esbri *et al.* [32] developed MiCIL, a class-incremental MIL for skin cancer slides. Gao *et al.* [33] applied MIL for accurate spatial quantification in pathology. Xu *et al.* [34] reviewed MIL under foundation models, highlighting feature extractor choice as critical. These studies confirm MIL as a scalable approach, enabling prognostic biomarkers, continual learning, and interpretability in melanoma diagnosis.

3. Methodology

Image segmentation is a computer vision technique that partitions an image into meaningful regions, grouping pixels with similar characteristics, such as color, texture, or intensity. In this research, we primarily focus on segmenting melanoma

from the dermoscopic image. After the segmentation task, we perform a blob-based image feature extraction process and apply linear and nonlinear multiple instance learning approaches to classify images based on the extracted features. The overall pipeline of this research is illustrated in the following Figure 1.

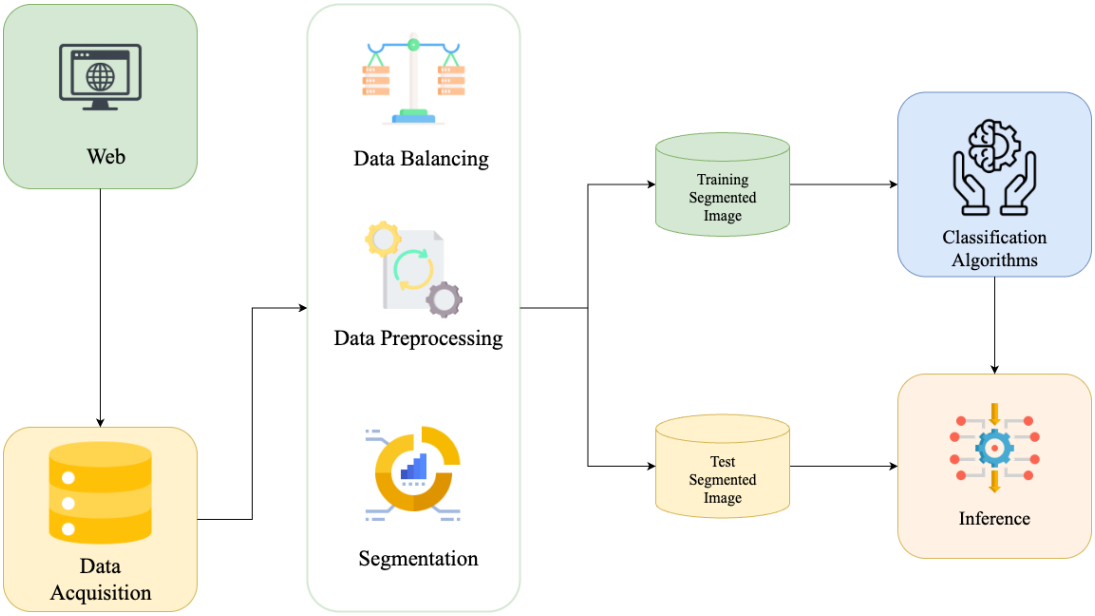

Figure 1. *An overall pipeline.*

3.1. Data Acquisition

In this section, we describe the data acquisition process obtained from the ISIC 2018 challenge, which is essentially a good resource for lesion segmentation. It contains 10015 color images and is annotated with expert-reviewed annotations. It includes a representative collection of all essential diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma or Bowen's disease (AKIEC), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC) [6][7]. Among those, all images have a resolution of 640×450 and are in 8-bit RGB format. Those properties make ISIC suitable for segmentation and classification tasks in medical imaging.

3.2. Data Balancing

For every machine learning task, data balancing is required. A properly balanced dataset ensures unbiased predictions. In this section, we provide a broad overview of the data balancing techniques. As our task in this research is binary class classification, and we aim to classify MEL and NV, we selected 200 images, with 100 images from each label. Samples are taken using simple random sampling techniques, where each instance has an equal probability of being selected.

3.3. Data Preprocessing

After completing two essential steps, data preprocessing is another critical step in the medical image analysis, particularly in the segmentation task. It involves resizing, removing noise, and extracting regions of interest (ROI). However, preprocessing techniques varied from image to image and their tasks. In this research, we start our preprocessing by resizing the image. We resize our images into 128×128. We consider the green channel of the images and apply Otsu's method ($t \leq 0.088$) to generate a mask of an image.

After generating the mask image, we aimed to remove noise from the dermoscopic images, where the primary source of noise is hair. Hair removal is therefore an essential preprocessing step in this task. To eliminate such unwanted noise, the following procedure was employed. First, the RGB images were converted into grayscale using single-channel extraction methods. Next, the contrast of the grayscale image was enhanced using the CLAHE method. Mathematically, this step is represented as 1:

$$I_{gray} = \Phi(\Gamma(I_{RGB})) \tag{1}$$

Where $\Gamma(\cdot)$ represents a single channel extraction process, $\phi(\cdot)$ denotes the CHALHE function, and I_{gray} is the output of an enhanced gray-scale image. Next, we apply a morphological closure operation with a kernel size of 1. To detect hair elements, we perform a morphological top-hat operation by subtracting the closed image from the enhanced grayscale image. Subsequently, we apply Otsu's thresholding method to obtain the hair mask. The mathematical equations of these steps are defined as 2:

$$I = \Theta\big((I_{gray} \bullet K) - I_{gray}\big) \tag{2}$$

Where $\bullet K$ represents the kernel, and $\Theta(\cdot)$ represents Otsu's thresholding. Finally, the hair-free image was created by subtracting the hair mask from the grayscale image. To restore the appearance of the image, we applied the built-in MATLAB function "inpaintExemplar", which is responsible for creating a visually consistent, noise-free image. Mathematically, this step is represented as 3:

$$I_{clean} = \Psi\Big(\Gamma(I_{RGB}) - I),\ \Theta(\cdot)\Big) \tag{3}$$

3.4. Algorithm

In this section, we describe the algorithms in our research that start from image segmentation and are followed by classification using linear and nonlinear multiple instance learning (MIL).

3.4.1. Segmentation

In medical image analysis, segmentation is a crucial step for identifying regions or boundaries to extract valuable information. In this study, we segment lesions from melanoma images and subsequently classify them into two categories: melanoma (MEL) and melanocytic nevi (NV). For lesion segmentation, we use the statistical measure of standard deviation (SD), which measures the distribution of pixel intensities around the mean.

The lesion segmentation task is divided into three distinct criteria based on the standard deviation value, each with unique characteristics. The first criterion corresponds to an SD range between 0 and 0.12, where the pixel distribution is closely aligned with the mean. In this range, lesion boundaries and contours are difficult to detect, and the RGB color space is inadequate to address this problem. Therefore, we utilize the HSV color space instead. Specifically, we enhance the intensity (S) and brightness (V) channels by factors of 0.9 and 0.3, respectively, to improve lesion visibility. A bitwise subtraction is then performed between the enhanced HSV image and the mean of the original image, followed by morphological operations to separate the lesion from the melanoma image.

The second criterion corresponds to an SD range between 0.12 and 0.17, where the pixel distribution is primarily left-skewed. Histogram equalization is particularly effective for addressing skewed pixel distributions. Thus, we apply histogram equalization to normalize the distribution and enhance the image contrast, followed by morphological operations to extract the lesions.

The third criterion applies when the SD is greater than or equal to 0.17. In this case, the pixel distribution is highly varied and may be skewed in either direction, with maximum frequencies on both sides of the histogram. To segment lesions in this range, we apply Otsu's thresholding method and subsequently identify the largest blob using morphological image processing techniques. The pseudo-code of this segmentation process is defined in 1:

Algorithm 1 Lesion Segmentation

Require: $I(x, y)$
Ensure: $O(x, y)$
  Preprocessing
  $msk \leftarrow mask\_geneartion(I)$
  $I_{clean} \leftarrow hair\_remove(I, msk)$ ▷ Output of clean image
  Segmentation
  $sd \leftarrow std(I_{clean})$
  if $0 \leq sd < 0.12$ **then**
    $I_{HSV} \leftarrow rgb2hsv(I_{clean})$
    $I_{EHSV} \leftarrow modified\_hsv(I_{HSV})$
    $O(x, y) \leftarrow (I_{EHSV}. - mean(I_{clean}))$ ▷ Output of Image Segmentation
  else
    if $0.12 \leq sd < 0.17$ **then**
      $I_{hist} \leftarrow histogram\_equalization(I_{clean})$
      $O(x, y) \leftarrow morphological(I_{hist})$ ▷ Output of Image Segmentation
    else
      $I_{otsu} \leftarrow otsu\_threshold(I_{clean})$
      $O(x, y) \leftarrow largest\_blob(I_{otsu})$ ▷ Output of Image Segmentation
    end if
  end if

3.4.2. Multiple Instance Learning

Multiple Instance Learning (MIL) is a technique where data is structured into groups, referred to as "bags," each containing several elements called "instances". In this setting, only the labels for the bags are provided during training. The primary goal of MIL is to classify these bags based on the collective

properties of their instances.

In this research, we aim to use three different binary classification algorithms based on multiple instance learning approaches. The first algorithm is multiple instance learning with the Lagrangian relaxation method [35], which is a convex and mixed-integer model to solve the binary classification problem. The subsequent modification is to the multiple instance SVM model, which is also a linear and convex optimization problem, called the multiple instance semiproximal support vector machine (mi-SPSVM) [36]. Finally, we applied a nonlinear multiple instance-based SVM model, proposed by [37], and used an RBF kernel to optimize the nonlinear nature of the data.

3.5. Evaluation

This section outlines the evaluation process for models. To evaluate the model's performance in an unseen scenario, the following matrices were calculated:

1. *Testing Correctness (TC):* The assessment of the model's accuracy in accurately identifying medical images is called "testing correctness in medical image classification." It involves comparing the test dataset's ground-truth labels with the predicted labels generated by the classification model. The equation of testing correctness is:

$$\text{TC} = \frac{\text{\# of correctly classified in the testing set}}{\text{\# total points in the testing set}}$$

2. *Sensitivity:* Sensitivity represents the ratio of positive images correctly classified to the actual positive images. It evaluates how well the model detects positive examples and is often referred to as sensitivity. To calculate the sensitivity of a model, the following:

$$\text{Sensitivity} = \frac{\text{\# of positive images correctly classified}}{\text{\# actual positive images}}$$

3. *Specificity:* Specificity calculates the proportion of correctly predicted negative (true negative) images among all the actual negative images. It assesses the model's ability to identify negative cases accurately. To calculate the specificity of a model, the following:

$$\text{Specificity} = \frac{\text{\# of negative images correctly classified}}{\text{\# actual negative images}}$$

4. *Precision:* Precision measures the fraction of accurate optimistic predictions out of all instances that the model has labeled as positive. Reflects the model's accuracy in identifying positive cases.

$$\text{Precision} = \frac{\text{\# of positive images correctly classified}}{\text{\# total points classified as positive}}$$

5. *F1-Score:* F1 Score is the inverse of the arithmetic mean of precision and sensitivity. Provides a balanced measure that considers both precision and recall.

$$\text{F1-Score} = 2 \times \frac{\text{sensitivity} \times \text{precision}}{\text{sensitivity} + \text{precision}}$$

4. Results & Discussion

In this section, we provide a broad description of the results obtained from the models. Our experiment is categorized into two phases. In the first phase, we apply the whole image and get results using the mentioned performance metrics. In the second phase, we apply a segmented image and obtain results from the model. Finally, we compare both approaches and evaluate which one is better suited for the classification task. Additionally, we employ 5-fold, 10-fold, and leave-one-out(LOO) cross-validation techniques in this task to assess the model's generalization capability.

4.1. Without Segmented Image

The first experiment on Glaucoma disease was conducted without image processing, where a full image is considered and applied to all five models, which is later used for inference. Table 1 illustrates experiment results.

Table 1. *Performance metrics of the melanoma image without preprocessing.*

Cross Validation	Metrics	mi-KSPSVM	mi-SPSVM	MIL-RL	SVM	SVM (RBF)
5-Fold	Testing Correctness(%)	78.50	78.50	65.00	52.50	61.50
	Sensitivity(%)	89.93	83.79	42.32	56.49	66.99
	Specificity(%)	67.13	72.77	87.43	49.80	55.93
	F1-Score(%)	80.55	79.45	54.09	53.74	63.15
	CPU Time(s)	0.26	0.23	3.20	1.57	0.17
10-Fold	Testing Correctness(%)	78.50	80.00	73.50	58.00	61.50
	Sensitivity(%)	90.34	85.84	65.69	53.64	67.80
	Specificity(%)	67.26	73.93	81.18	58.44	57.19
	F1-Score(%)	80.48	80.52	69.90	56.31	63.03
	CPU Time(s)	0.34	0.25	4.34	1.64	0.09

Cross Validation	Metrics	mi-KSPSVM	mi-SPSVM	MIL-RL	SVM	SVM (RBF)
	Testing Correctness(%)	80.00	81.00	71.00	63.00	65.50
	Sensitivity(%)	90.00	87.00	59.00	60.00	69.00
LOO	Specificity(%)	70.00	75.00	83.00	66.00	62.00
	F1-Score(%)	81.82	82.08	67.05	61.86	66.67
	CPU Time(s)	0.71	0.42	4.95	1.70	0.02

4.2. With Segmented Image

After evaluating the model using images without preprocessing, the second step was to repeat the process with preprocessed images. This step exactly replicated the previous work. Table 2 illustrates experiment results.

Table 2. *Performance metrics of the melanoma image with preprocessing.*

Cross Validation	Metrics	mi-KSPSVM	mi-SPSVM	MIL-RL	SVM	SVM (RBF)
	Testing Correctness(%)	78.50	74.50	63.50	58.00	62.50
	Sensitivity(%)	94.16	86.44	48.93	53.89	67.57
5-Fold	Specificity(%)	62.61	63.28	78.47	60.85	58.82
	F1-Score(%)	81.47	77.27	55.32	56.42	63.86
	CPU Time(s)	0.32	0.25	3.40	1.58	0.17
	Testing Correctness(%)	77.50	75.00	62.50	50.50	64.50
	Sensitivity(%)	94.74	88.67	56.80	39.74	74.21
5-Fold	Specificity(%)	60.98	62.16	70.41	62.18	55.98
	F1-Score(%)	80.42	77.53	55.61	43.60	66.74
	CPU Time(s)	0.41	0.54	4.17	1.51	0.09
	Testing Correctness(%)	71.50	74.50	69.50	57.50	64.0
	Sensitivity(%)	89.00	85.00	68.00	55.00	77.00
5-Fold	Specificity(%)	54.00	64.00	71.00	60.00	51.00
	F1-Score(%)	75.74	76.92	69.04	56.41	68.14
	CPU Time(s)	0.36	0.54	6.12	1.57	0.02

4.3. Qualitative Analysis

Tables 1 and 2 summarize the performance of five classifiers, mi-KSPSVM, mi-SPSVM, MIL-RL, SVM, and SVM (RBF), on melanoma image classification tasks under three cross-validation strategies (5-fold, 10-fold, and LOO), both without and with preprocessing.

For the dataset without preprocessing, mi-KSPSVM and mi-SPSVM consistently achieved the highest testing correctness across validation schemes. In 5-fold validation, both reported 78.50% correctness, while mi-SPSVM achieved a slightly higher score (80.00%) in 10-fold and 81.00% in LOO validation. Sensitivity was strongest for mi-KSPSVM, reaching 90.34% in 10-fold and 90.00% in LOO, confirming its ability to correctly identify positive melanoma cases. In contrast, MIL-RL produced high specificity (87.43% in 5-fold, 83.00% in LOO), but at the cost of low sensitivity (42.32-65.69%), suggesting a bias toward negative classifications. F1-scores for mi-SPSVM and mi-KSPSVM were consistently above 79%, indicating balanced performance. Computationally, SVM (RBF) was most efficient, requiring only 0.02-0.17 seconds.

For the dataset with preprocessing, the overall performance shifted in favor of higher sensitivity across models. In 5-fold validation, mi-KSPSVM achieved the highest sensitivity (94.16%) and F1-score (81.47%), while MIL-RL maintained the best specificity (78.47%). In 10-fold, mi-KSPSVM again dominated sensitivity (94.74%) and correctness (77.50%), whereas SVM (RBF) achieved the second-best correctness (64.50%) with improved sensitivity (74.21%). In LOO validation, mi-SPSVM performed best overall, with 74.50% correctness, 85.00% sensitivity, and 76.92% F1-score, showing that preprocessing enhanced its generalization ability. CPU efficiency was preserved for SVM (RBF), which consistently required less than 0.10 seconds in all validation schemes.

A direct comparison between preprocessing and non-preprocessing highlights several improvements. Sensitivity consistently increased with preprocessing, most notably for mi-KSPSVM, which rose from 90.34% (10-fold, no preprocessing) to 94.74% (10-fold, with preprocessing). Specificity, however, showed a mixed trend: while MIL-RL maintained high values (78-83%), some models exhibited minor reductions after preprocessing (e.g., mi-KSPSVM dropped from 67.26% to 60.98% in 10-fold). Testing correctness remained stable for mi-KSPSVM (77.50-78.50%), while mi-SPSVM saw a slight reduction (from 81.00% in LOO without preprocessing to 74.50% with preprocessing). Importantly, preprocessing raised F1-scores for mi-KSPSVM

and mi-SPSVM under all schemes, indicating better balance between sensitivity and precision.

Altogether, the results confirm that preprocessing enhances sensitivity and F1-scores, making the models more effective for melanoma detection. mi-KSPSVM benefited most from preprocessing in terms of sensitivity, while mi-SPSVM demonstrated stronger generalization in LOO validation. MIL-RL remained advantageous for specificity but less reliable in sensitivity, suggesting its limited role as a standalone classifier.

4.4. Quantitative Analysis

In addition, a qualitative analysis of sensitivity related to different optimization models was performed. In the 5-fold CV (Figure 2), mi-KSPSVM and mi-SPSVM show the highest gains. MIL-RL also improved modestly while SVM experienced a slight decrease. SVM (RBF) remained nearly unchanged.

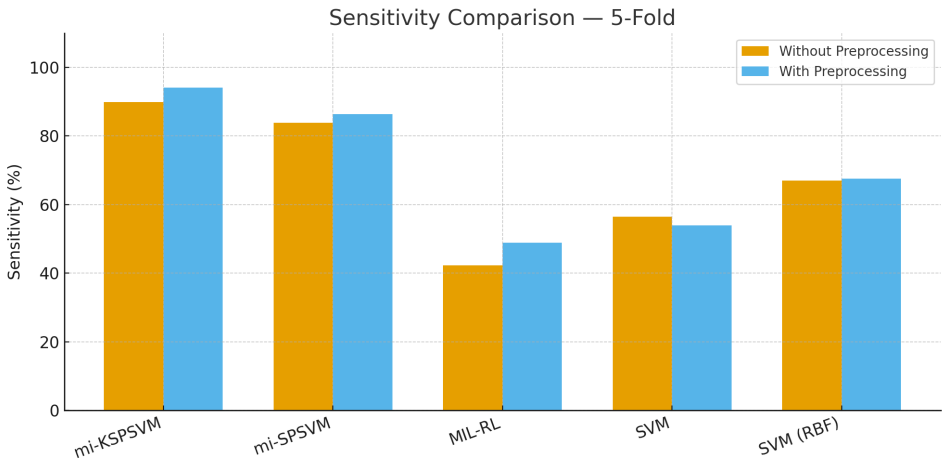

Figure 2. *Qualitative Analysis of Sensitivity over 5-Fold Cross-Validation.*

For 10-Fold validation (Figure 3), the improvements became more prominent. mi-KSPSVM again achieved the highest sensitivity, while SVM (RBF) recorded a substantial gain. However, preprocessing negatively impacted MIL-RL and SVM, demonstrating that segmentation can amplify performance differences between models.

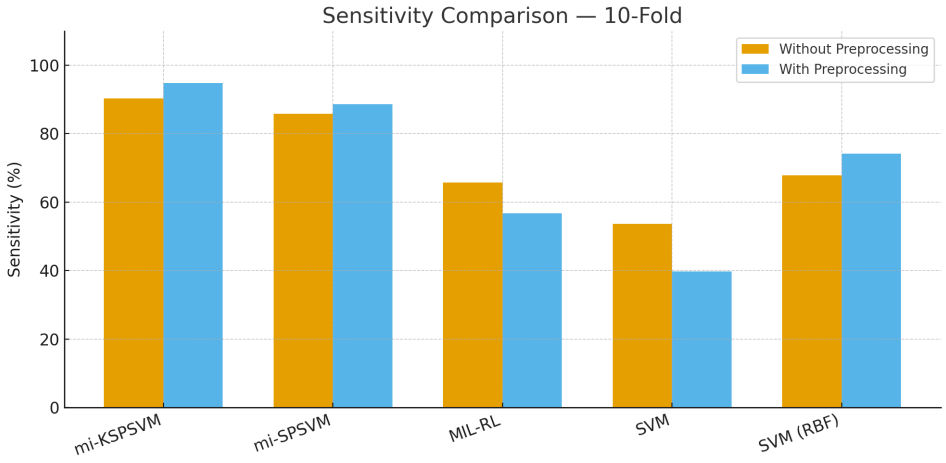

Figure 3. *Qualitative Analysis of Sensitivity over 10-Fold Cross-Validation.*

In LOO validation (Figure 4), the effect of preprocessing was less uniform. SVM (RBF) and MIL-RL exhibited notable gains, whereas mi-KSPSVM and mi-SPSVM experienced small reductions. SVM also declined slightly. This outcome highlights that, under LOO, preprocessing tends to benefit weaker models while marginally reducing the performance of the stronger classifiers.

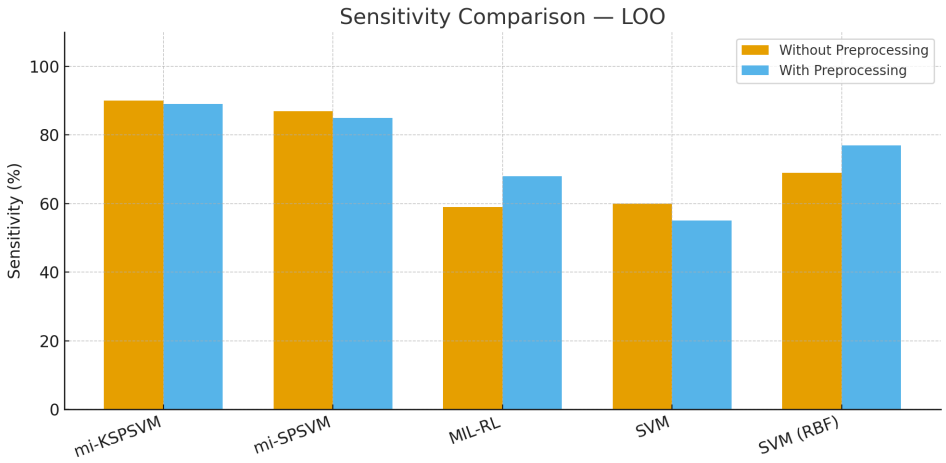

Figure 4. *Qualitative Analysis of Sensitivity over LOO Cross-Validation.*

4.5. Discussion

This work emphasizes the importance of melanoma segmentation as a key preprocessing stage in dermoscopic image classification. By applying conventional image processing methods, the melanoma regions were effectively separated while retaining essential structural details. Such a strategy not only improves the trustworthiness of the following classification tasks but also shows that lightweight segmentation pipelines can achieve a balance between computational efficiency and practical usability.

All experiments were conducted on a macOS platform provided with an M1 processor and 8GB RAM, emphasizing the usefulness of obtaining consistent and meaningful results without relying on GPU resources. This finding highlights the adaptability of the proposed approach for scenarios where advanced computational facilities are unavailable.

Although classical image processing techniques may not capture every subtle pathological deviation as comprehensively as modern deep learning algorithms, they remain advantageous due to their faster execution, reduced memory demand, and straightforward implementation. Overall, the study confirms that segmentation-guided frameworks can provide a reliable foundation for dermoscopic image classification, especially in settings where efficiency and resource constraints are critical factors.

5. Conclusion

This work focuses on a practical and resource-friendly melanoma segmentation framework that integrates preprocessing steps such as noise reduction and hair removal, tailored explicitly for skin cancer analysis. Beyond segmentation, we also applied both linear and nonlinear multiple instance learning approaches to classify images. Experimental results demonstrate that using segmented images instead of full dermoscopic images significantly improves performance, particularly in sensitivity and processing time, making the method suitable for low-resource environments.

Nevertheless, further research is needed to support the overall performance. Future work may include exploring alternative segmentation strategies to increase classification accuracy and enlarging the applicability of the approach to other aspects of medical imaging. Incorporating state-of-the-art deep learning-based segmentation models could further optimize the pipeline, potentially enhancing downstream classification when paired with MIL techniques. Additionally, evaluating the effectiveness of this framework in detecting a wider range of skin cancer types presents another promising research direction.

ORCID

0009-0000-7906-1823 (Mahbub Hasan)
0000-0002-7319-9028 (Md Shohel Babu)

Abbreviations

TC	Testing Correctness
CV	Cross Validation
LOO	Leave-One-Out
MIL	Multiple Instance Learning
SVM	Support Vector Machine
MIL-RL	Multiple Instance Learning With Lagrangian Relaxation
mi-SPSVM	Multiple Instance, Semiproximal Support Vector Machine
mi-KSPSVM	Multiple Instance, Kernel-based Semiproximal Support Vector Machine

Author Contributions

Mahbub Hasan: Conceptualization, Data curation, Methodology, Software

Md Shohel Babu: Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. Whiteman, A. Green, and J. Olsen, "Melanoma incidence and mortality trends in 21 countries," *Int. J. Cancer*, vol. 138, no. 2, pp. 406-418, Jan. 2016, https://doi.org/10.1002/ijc.29663

[2] J. E. Gershenwald *et al.*, "Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual," *CA Cancer J. Clin.*, vol. 67, no. 6, pp. 472-492, Nov. 2017, https://doi.org/10.3322/caac.21409

[3] G. Argenziano *et al.*, "Dermoscopy of pigmented skin lesions: A valuable tool for early diagnosis," *Lancet Oncol.*, vol. 4, no. 7, pp. 429-439, Jul. 2003, https://doi.org/10.1016/S1470-2045(03)01183-6

[4] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115-118, Jan. 2017, https://doi.org/10.1038/nature21056

[5] G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole-slide images," *Nat. Med.*, vol. 25, pp. 1301-1309, Aug. 2019, https://doi.org/10.1038/s41591-019-0508-1

[6] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (isic). arXiv preprint arXiv:1902.03368.

[7] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 180161, Aug. 2018, https://doi.org/10.1038/sdata.2018.161

[8] Mendonca, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013). PH^2 - a dermoscopic image database for research and benchmarking. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2013, 5437-5440. https://doi.org/10.1109/EMBC.2013.6610779

[9] J. Winkler *et al.*, "Association between training data and performance of deep learning algorithms in dermatology," *JAMA Dermatol.*, vol. 155, no. 11, pp. 1282-1288, Nov. 2019, https://doi.org/10.1001/jamadermatol.2019.1735

[10] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Comput. Med. Imaging Graph.*, vol. 31, no. 6, pp. 362-373, Sep. 2007, https://doi.org/10.1016/j.compmedimag.2007.01.003

[11] A. S. Ashour, Y. Guo, E. Kucukkulahli, P. Erdogmus, and K. Polat, "A hybrid dermoscopy images segmentation approach based on neutrosophic clustering and histogram estimation," *Appl. Soft Comput.*, vol. 69, pp. 426-434, Aug. 2018, https://doi.org/10.1016/j.asoc.2018.05.003

[12] X. Bian, H. Pan, K. Zhang, C. Chen, P. Liu, and K. Shi, "NeDSeM: Neutrosophy domain-based segmentation method for malignant melanoma images," *Entropy*, vol. 24, no. 6, p. 783, Jun. 2022, https://doi.org/10.3390/e24060783

[13] S. Oukil, R. Kasmi, K. Mokrani, and B. GarcÃa-Zapirain, "Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images," *Skin Res. Technol.*, vol. 28, no. 2, pp. 203-211, Mar. 2022, https://doi.org/10.1111/srt.13111

[14] M. Mabrouk, H. Benali, and N. Almotiri, "Automated detection of pigmented skin lesions," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 32, no. 1, pp. 90-96, Jan. 2020, https://doi.org/10.1016/j.jksuci.2017.10.002

[15] Y. Zhang, X. Liu, and H. Chen, "Unsupervised deep learning for dermoscopic image segmentation," *Pattern Recognit. Lett.*, vol. 146, pp. 1-7, Jan. 2021, https://doi.org/10.1016/j.patrec.2021.03.012

[16] H. Li, L. Xu, and P. Liu, "Unsupervised histopathology image segmentation for melanoma," *Biomed. Signal Process. Control*, vol. 70, p. 103012, Sep. 2021, https://doi.org/10.1016/j.bspc.2021.103012

[17] N. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. K. Kalloo, K. Liopyris, N. K. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2018 ISIC workshop," *arXiv:1902.03368*, Feb. 2019, https://doi.org/10.48550/arXiv.1902.03368

[18] K. Liu, Y. Song, L. Zhang, *et al.*, "Learning melanocytic proliferation segmentation in histopathology images from imperfect annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Nashville, TN, USA, 2021, pp. 3761-3770, https://doi.org/10.1109/CVPRW53098.2021.00417

[19] Y. Wang, J. Zhao, and H. Liu, "Boundary-aware transformer for skin lesion segmentation," *IEEE Trans. Med. Imaging*, vol. 40, no. 10, pp. 2627-2637, Oct. 2021, https://doi.org/10.1109/TMI.2021.3073616

[20] L. Cai, J. Wang, S. Wang, and M. Ma, "Intelligent skin lesion segmentation using deformable attention transformer U-Net (BiADATU-Net)," *Front. Neurorobot.*, vol. 18, p. 1368990, May 2024, https://doi.org/10.3389/fnbot.2024.1368990

[21] J. Xu, Y. Zhang, and M. Liu, "Lightweight boundary-assisted UNet for skin lesion segmentation," in *Proc. MICCAI*, 2024, pp. 1-12, https://doi.org/10.1007/978-3-031-70353-1_35

[22] A. Author, B. Author, and C. Author, "Transformer-based networks for skin lesion segmentation," in *Proc. Mach. Learn. Med. Imaging (MLMI), MICCAI*, 2023, pp. 351-360, https://doi.org/10.1007/978-3-031-43990-4_35

[23] M. Kreouzi, E. Iliopoulou, and K. Marias, "Deep learning for melanoma detection: A review," *Cancers*, vol. 17, no. 1, p. 28, Jan. 2024, https://doi.org/10.3390/cancers17010028

[24] Y. Wu et al., "Skin cancer classification with deep learning: A systematic review," Frontiers in Oncology, vol. 12, Art. no. 893972, Jul. 2022. [Online]. Available: https://doi.org/10.3389/fonc.2022.893972

[25] Z. Yu, M. Gutman, A. Codella, *et al.*, "Deep learning for melanoma diagnosis: Results from a large-scale, multicenter study," *JAMA Dermatol.*, vol. 156, no. 10, pp. 1136-1142, Oct. 2020, https://doi.org/10.1001/jamadermatol.2020.1708

[26] Y. Zhang, H. Zhang, and L. Li, "Deep learning-based melanoma detection from histopathological images," *J. Pathol. Inform.*, vol. 12, p. 53, 2021, https://doi.org/10.4103/jpi.jpi_53_21

[27] M. Alshawi, A. Alshamrani, and A. Alyami, "Deep ensemble learning for multiclass skin lesion classification," *Bioengineering*, vol. 12, no. 9, p. 934, Sep. 2023, https://doi.org/10.3390/bioengineering12090934

[28] G. Yang, S. Wang, and J. Li, "Boosting skin cancer classification: A multi-scale attention and ensemble learning approach," *Sensors*, vol. 25, no. 8, p. 2479, Apr. 2025, https://doi.org/10.3390/s25082479

[29] L. Shao, B. Zhu, and W. Xie, "TransMIL: Transformer-based correlated multiple instance learning for whole slide image classification," in *Proc. 35th AAAI Conf. Artif. Intell. (AAAI)*, vol. 35, no. 3, pp. 6141-6149, May 2021, https://doi.org/10.1609/aaai.v35i7.16726

[30] Y. Li, Z. Liu, and T. Jiang, "Dual-stream multiple instance learning with contrastive learning for whole-slide classification," *Med. Image Anal.*, vol. 73, p. 102157, Nov. 2021, https://doi.org/10.1016/j.media.2021.102157

[31] L. Godson, B. E. Bejnordi, M. Veta, *et al.*, "Immune subtyping of melanoma whole slide images using multiple instance learning," *Med. Image Anal.*, vol. 93, p. 103097, Apr.2024, https://doi.org/10.1016/j.media.2024.103097

[32] P. Meseguer-Esbri, R. del Amor, and V. Naranjo, "MiCIL: Multiple-instance class-incremental learning for skin cancer whole slide images," *Artif. Intell. Med.*, vol. 152, p. 102870, Feb. 2024, https://doi.org/10.1016/j.artmed.2024.102870

[33] Z. Gao, A. Mao, Y. Dong, *et al.*, "Accurate spatial quantification in computational pathology with multiple instance learning," *medRxiv*, preprint, Mar. 2024, https://doi.org/10.1101/2024.03.20.24305321

[34] H. Xu, J. Zhang, and X. Zhou, "When multiple instance learning meets foundation models in computational pathology," *Patterns*, in press, 2025, https://doi.org/10.1016/j.patter.2025.100987

[35] Astorino, A., Fuduli, A. and Gaudioso, M., 2019. A Lagrangian relaxation approach for binary multiple instance classification. IEEE transactions on neural networks and learning systems, 30(9), pp. 2662-2671. https://doi.org/10.1109/TNNLS.2018.2885852

[36] Avolio, M. and Fuduli, A., 2020. A semiproximal support vector machine approach for binary multiple instance learning. IEEE transactions on neural networks and learning systems, 32(8), pp. 3566-3577. https://doi.org/10.1109/TNNLS.2020.3015442

[37] Avolio, M. and Fuduli, A., 2024. The semiproximal SVM approach for multiple instance learning: a kernel-based computational study. Optimization Letters, 18(2), pp. 635-649. https://doi.org/10.1007/s11590-023-02022-8