

Classification of Pigmented Skin Lesions Using Dermoscopic Images

María Fernanda Vargas-Martínez, María Fernanda Cruz-Mahecha and Luz Helena Camargo-Casallas*

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

Abstract

Automatic classification of pigmented skin lesions from dermoscopic images requires the integration of lesion segmentation, feature representation, and supervised classification. This study presents a modular image-processing workflow that combines Otsu thresholding and encoder-decoder semantic segmentation with morphological post-processing, local color-shape descriptors, Bag of Visual Words encoding, and support vector machine classification. Semantic segmentation is formulated as a binary pixel-level task that distinguishes lesions from background, whereas the SVM performs image-level classification into carcinoma, melanoma, and melanocytic nevus. The classification model was trained using 1,050 balanced dermoscopic images. Its performance was evaluated on 300 observations using six color representations: CIE Lab, NTSC, Opponent Space, XYZ, HSV, and RGB. A 250-word visual vocabulary was constructed from local descriptors using k -means, and each image was represented by a fixed-length histogram. The best observed performance was obtained using the HSV representation and a Gaussian-kernel SVM, achieving an accuracy of 78.3% on the evaluation subset. These results indicate the technical feasibility of the proposed modular workflow as a computational research prototype.

Keywords

Dermoscopy, Pigmented skin lesions, Semantic segmentation, Otsu's method, SURF descriptors, Bag of Visual Words, Support vector machine, HSV color space, Image classification

1. Introduction

Skin cancer is an important public-health problem, and the early assessment of suspicious skin lesions is relevant to timely clinical decision-making. Among the different types of skin cancer, melanoma is particularly important because its prognosis is strongly associated with the stage at which it is detected [1, 2]. Computational analysis of dermoscopic images has therefore been investigated as a means of supporting reproducible image assessment.

Dermoscopy improves the visual examination of pigmented lesions by reducing surface reflection and revealing structures that are not easily observed in conventional photographs [3, 4]. Dermoscopic interpretation may involve the analysis of global patterns, local structures, color variation, asymmetry, border characteristics, and other morphological criteria. Several diagnostic approaches have been proposed, including pattern analysis, the ABCD rule, the three-point checklist, the seven-point checklist, and the Menzies method [5, 6, 7]. Although these approaches can support the assessment of pigmented lesions, visual overlap may occur among benign and malignant lesions. In particular, the differentiation of melanoma from other pigmented lesions remains challenging when images contain irregular pigmentation, low contrast, artifacts, or overlapping dermoscopic structures [8, 9].

Computer-vision methods can transform dermoscopic images into quantitative representations for automated image analysis. A typical computational workflow involves lesion segmentation, feature extraction, image representation, and supervised classification. Segmentation isolates the lesion from the surrounding skin and image artifacts, whereas feature extraction converts the segmented region into numerical descriptors. A classifier can then assign the image to a predefined category. Deep-

ICAIW 2026: Workshops at the 9th International Conference on Applied Informatics 2026, October 13–17, 2026, Orlando, United States

*Corresponding author.

✉ lhcamargoc@udistrital.edu.co (L. H. Camargo-Casallas)

ORCID 0000-0002-3416-018X (L. H. Camargo-Casallas)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

learning methods have become prominent in dermatological image analysis; nevertheless, differences in image sources, diagnostic labels, class definitions, preprocessing procedures, data partitions, evaluation metrics, and validation settings affect the interpretation of reported results [10, 11, 12].

Color and shape are important components of dermoscopic image interpretation. Structures such as irregular pigmentation, blue-whitish areas, asymmetry, and border irregularities may provide information relevant to the analysis [13, 9]. For this reason, computational representations that combine color and local shape information may provide a useful description of dermoscopic appearance. Bag of Visual Words representations provide a way to describe an image through the distribution of local visual patterns. In this approach, local descriptors are extracted from an image and grouped into a finite visual vocabulary using an unsupervised clustering algorithm. Each image is then represented by a histogram that records the frequency of occurrence of the visual words [14, 15]. This representation provides a fixed-length feature vector that can be used by a supervised classifier such as a support vector machine. The use of explicit local descriptors and a visual vocabulary also facilitates the analysis of individual stages of the computational workflow [16].

Publicly available dermoscopic-image resources have contributed to the development and comparison of computational methods for skin-lesion analysis. The International Skin Imaging Collaboration has promoted the use of dermoscopic images for research and comparative evaluation of automated analysis methods [13, 17, 18]. The present study uses dermoscopic-image resources associated with ISIC to construct and evaluate a modular image-processing workflow. Comparisons among computational configurations require consideration of differences in image sources, diagnostic categories, class distributions, preprocessing procedures, image partitions, and evaluation metrics [11, 19].

In this work, a modular workflow is implemented to classify dermoscopic images into three image-level categories: carcinoma, melanoma, and melanocytic nevus. The workflow includes two alternative lesion-segmentation strategies: global thresholding using Otsu's method and supervised semantic segmentation using an encoder-decoder network [20, 21]. Both segmentation approaches produce a binary mask that distinguishes lesion from background. After segmentation and morphological post-processing, local color-shape descriptors are extracted using SURF-based representations in six color spaces: CIE Lab, NTSC, Opponent Space, XYZ, HSV, and RGB [22]. The descriptors are grouped into a 250-word visual vocabulary using k -means, and each image is represented by a fixed-length Bag of Visual Words histogram [14]. Finally, an SVM with a Gaussian kernel performs image-level classification into carcinoma, melanoma, and melanocytic nevus.

The proposed workflow distinguishes two different machine-learning tasks. Semantic segmentation is performed at the pixel level and produces a binary mask with the classes lesion and background. The subsequent SVM performs classification at the image level and assigns one of three categories: carcinoma, melanoma, or melanocytic nevus. These tasks are related within the complete pipeline, but their classes, outputs, and evaluation procedures are different and must not be treated as interchangeable [21, 10].

The proposed workflow provides an interpretable organization of lesion segmentation, morphological post-processing, local color-shape description, visual-word encoding, and image-level classification. This modular structure allows individual components to be analyzed, compared, or replaced without redesigning the complete pipeline. It also provides a technical baseline for future experiments involving alternative descriptors, segmentation metrics, data partitions, and classification models.

The objective of this study is to assess the technical feasibility of a modular computational workflow for image-level classification of dermoscopic images into carcinoma, melanoma, and melanocytic nevus. The analysis focuses on the relationship between lesion segmentation, explicit color-shape representation, Bag of Visual Words encoding, and SVM classification.

2. Methodology

2.1. Image datasets and experimental setup

The dermoscopic images used in this study were obtained from image resources associated with the International Skin Imaging Collaboration (ISIC) [13, 17]. The images were acquired using different dermoscopic systems and originated from multiple clinical institutions. They were stored in JPEG format and normalized to a spatial resolution of 450×600 pixels. Images corresponding to mucosal or nail regions were excluded. Each image contained one primary lesion, although some images could also include hair, marker lines, small secondary lesions, or other pigmented regions that had to be excluded from the analysis.

The complete computational workflow comprised three main stages: lesion segmentation, feature extraction, and image-level classification. Figure 1 illustrates the sequence of processing steps, including the two alternative segmentation routes, morphological post-processing, local color–shape descriptor extraction, Bag of Visual Words encoding, and final SVM classification. Two image resources were used for different methodological purposes.

For semantic segmentation, 2,594 dermoscopic images with pixel-level annotations were used. The annotations were provided with the source images and served as reference masks for training the encoder–decoder network. These images were resized to $192 \times 192 \times 3$ pixels to provide a fixed input size for the segmentation model.

For image-level classification, 1,050 balanced images were used for training, comprising 350 images per class: carcinoma, melanoma, and melanocytic nevus. An independent evaluation subset of 300 observations was used to assess the classifier. The training and evaluation subsets were treated as separate image sets. The training images were used to construct the visual representation and fit the SVM, whereas the evaluation observations were reserved for measuring classification performance.

No new patient recruitment, intervention, or additional clinical data collection was performed. The study used previously acquired dermoscopic images made available for research purposes through the documented source resources. No personal identifiers or additional clinical data were accessed by the authors. The images were analyzed according to the data-use conditions established by their original sources.

The experiments were implemented in MATLAB (MathWorks, Natick, MA, USA) using the Image Processing Toolbox, Computer Vision Toolbox, Statistics and Machine Learning Toolbox, and Deep Learning Toolbox [23].

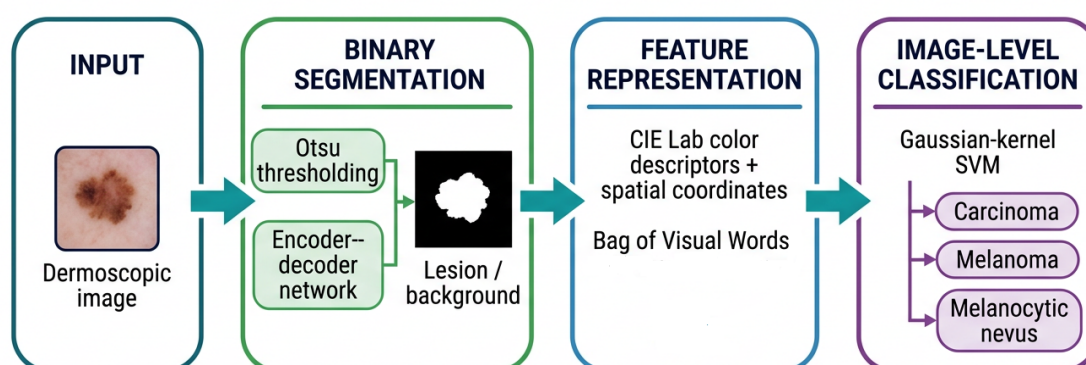


Figure 1: Overview of the proposed image-classification workflow. The input dermoscopic image is processed by one of two alternative segmentation routes: Otsu thresholding or encoder–decoder semantic segmentation. The resulting binary lesion mask is refined through morphological post-processing. Local color–shape descriptors and spatial information are then extracted from the segmented lesion and encoded as a 250-component Bag of Visual Words histogram. Finally, the histogram is classified by the SVM into carcinoma, melanoma, or melanocytic nevus. The segmentation module performs binary pixel-level prediction, whereas the SVM performs three-class image-level classification.

2.2. Lesion segmentation

Lesion segmentation was performed using two alternative approaches: global thresholding based on Otsu's method and supervised semantic segmentation using an encoder–decoder network. Both approaches produced binary masks distinguishing the lesion from the surrounding background. The masks were subsequently refined using morphological post-processing.

2.2.1. Otsu thresholding

Otsu's method is a global thresholding technique that separates a grayscale image into two classes by selecting the threshold that maximizes the between-class variance [20]. In this study, the two resulting classes were interpreted as lesion and background.

Each RGB image was first converted into a luminance image using the conventional digital television coefficients [24]:

$$Y = 0.299R + 0.587G + 0.114B. \quad (1)$$

For a candidate threshold t , pixels were divided into two classes according to their luminance. Pixels assigned to the lesion class were given a value of one, whereas pixels assigned to the background class were given a value of zero. The resulting binary mask was passed to the post-processing stage.

2.2.2. Semantic segmentation

The second segmentation approach used an encoder–decoder semantic segmentation network trained with pixel-level annotations. Encoder–decoder architectures reduce the spatial resolution during feature extraction and subsequently restore the resolution to produce a pixel-level output mask [21, 10]. The network assigned each pixel to one of two classes: lesion or background. Therefore, semantic segmentation was formulated as a binary pixel-level task. This task was different from the subsequent SVM classification, which assigned each complete image to one of three image-level categories.

The network received an RGB image of $192 \times 192 \times 3$ pixels. The encoder consisted of two sequences of convolution, ReLU activation, and max-pooling. These layers reduced the spatial resolution from 192×192 to 96×96 and then to 48×48 pixels. The decoder restored the spatial resolution through two transposed-convolution layers, increasing the representation from 48×48 to 96×96 and finally to 192×192 pixels. A 1×1 convolution produced two output channels corresponding to lesion and background. A Softmax layer converted the outputs into pixel-level probabilities, and the pixel-classification layer assigned one of the two classes to each pixel.

The encoder–decoder architecture used for binary semantic segmentation is illustrated in Figure 2. The figure summarizes the reduction and recovery of spatial resolution and the two-class pixel-level output.

The encoder–decoder architecture was trained using 2,594 pixel-labeled dermoscopic images. Stochastic gradient descent with momentum was used with a momentum of 0.9, an initial learning rate of 0.01, a reduction factor of 0.7 every five epochs, and minibatches of 32 images. Horizontal and vertical reflections were used for data augmentation. Because the lesion and background classes were imbalanced at the pixel level, inverse-frequency class weighting was applied [23].

2.2.3. Mask post-processing

The masks generated by both segmentation approaches were refined using morphological opening, closing, and hole-filling operations [25]. These operations reduced small isolated regions and artifacts that could interfere with feature extraction. A circle-detection procedure with empirically selected parameters was also used to identify the main lesion region when necessary. This procedure reduced the influence of elongated structures such as hair and marker lines.

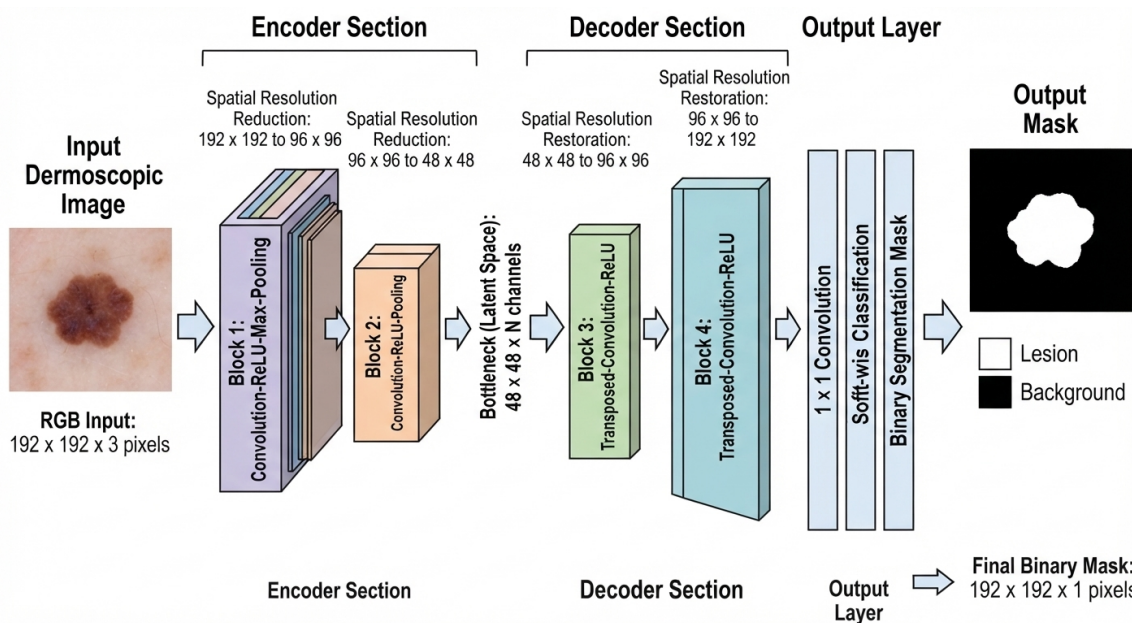


Figure 2: Encoder–decoder architecture for binary semantic segmentation. The network receives an RGB image of $192 \times 192 \times 3$ pixels and produces a binary lesion-versus-background mask. The encoder reduces the spatial resolution through convolution, ReLU activation, and max-pooling layers, whereas the decoder restores the spatial resolution through transposed-convolution layers. The final output contains two pixel-level classes: lesion and background.

2.3. Feature extraction and Bag of Visual Words

The segmented images were transformed into local color–shape representations before image-level classification. The feature-extraction stage was applied to the 1,050 training images and to the 300 evaluation observations using the same descriptor configuration.

The images were represented in six color spaces: CIE Lab, NTSC, Opponent Space, XYZ, HSV, and RGB. SURF descriptors were extracted from local points of interest in each color representation [22]. Each local shape descriptor contained 64 components and was concatenated with the corresponding three-channel color vector.

The local descriptors extracted from the training images were grouped using k -means clustering [14]. A visual vocabulary containing 250 clusters was constructed, with each cluster centroid representing one visual word. The descriptors from each image were assigned to their nearest visual-word centroid, and each image was represented by a 250-component histogram.

The evaluation observations were represented using the same visual vocabulary learned from the training images. The visual-word centroids were not recalculated using the evaluation observations.

2.4. Image-level classification

The 250-component Bag of Visual Words histograms were used as input vectors for image-level classification. The classification task involved three categories: carcinoma, melanoma, and melanocytic nevus.

Decision-tree-based, k -means-based, and SVM classifiers were evaluated within the implemented workflow. The SVM with a Gaussian radial basis function kernel produced the highest observed performance among the evaluated configurations and was selected for the reported analysis.

The SVM was trained using 1,050 images, comprising 350 images per class. Its performance was evaluated on 300 independent observations. The evaluation observations were not used to construct the visual vocabulary or to fit the SVM.

The performance of the SVM was compared across the six color representations. Accuracy was

calculated from the diagonal elements of the corresponding confusion matrix according to:

$$\text{Accuracy} = \frac{\sum_{i=1}^C n_{ii}}{N} \times 100, \quad (2)$$

where n_{ii} denotes the number of correctly classified observations for class i , C is the number of image-level classes, and N is the total number of evaluation observations.

3. Results and Discussion

3.1. Lesion segmentation

Both segmentation approaches produced binary lesion-versus-background masks. Otsu thresholding provided a low-complexity alternative for images with sufficient contrast between the lesion and the surrounding skin. However, because it relies on a global intensity threshold, its output could be affected by uneven illumination, noise, dark borders, hair, marker lines, bubbles, and other structures unrelated to the lesion [20].

The encoder–decoder network generated binary masks for dermoscopic images belonging to the three image-level categories: carcinoma, melanoma, and melanocytic nevus. These masks must be interpreted as pixel-level lesion-versus-background predictions and not as three-class image-level classifications. The network was trained using 2,594 pixel-labeled dermoscopic images. During training, the semantic segmentation network reached an accuracy close to 90%. This value describes the observed training behavior for the binary lesion-versus-background task.

A complete evaluation of segmentation performance would require independent metrics such as the Dice coefficient, intersection over union, sensitivity, specificity, and precision. Figure 3 shows qualitative examples of the original images and their corresponding binary masks. White pixels represent the lesion region, whereas black pixels represent the background. These examples illustrate the output of the segmentation stage.

3.2. Mask post-processing

The masks generated by both segmentation approaches were refined using morphological opening, closing, and hole-filling operations. These operations reduced small isolated regions and artifacts that could interfere with feature extraction. A circle-detection procedure with empirically selected parameters was also used to identify the main lesion region when necessary. This procedure reduced the influence of elongated structures such as hair and marker lines.

3.3. Feature representation

The segmented images were transformed into local color–shape representations before image-level classification. The feature-extraction stage was applied to the 1,050 training images and to the evaluation observations using the same descriptor configuration.

The images were represented in six color spaces: CIE Lab, NTSC, Opponent Space, XYZ, HSV, and RGB. SURF descriptors were extracted from local points of interest in each color representation [22]. Each local shape descriptor contained 64 components and was concatenated with the corresponding three-channel color vector.

The local descriptors extracted from the training images were grouped using k -means clustering [14]. A visual vocabulary containing 250 clusters was constructed, with each cluster centroid representing one visual word. The descriptors from each image were assigned to their nearest visual-word centroid, and each image was represented by a 250-component histogram.

The evaluation observations were represented using the same visual vocabulary learned from the training images. The visual-word centroids were not recalculated using the evaluation observations. The resulting histograms were used as fixed-length input vectors for the image-level classifiers.

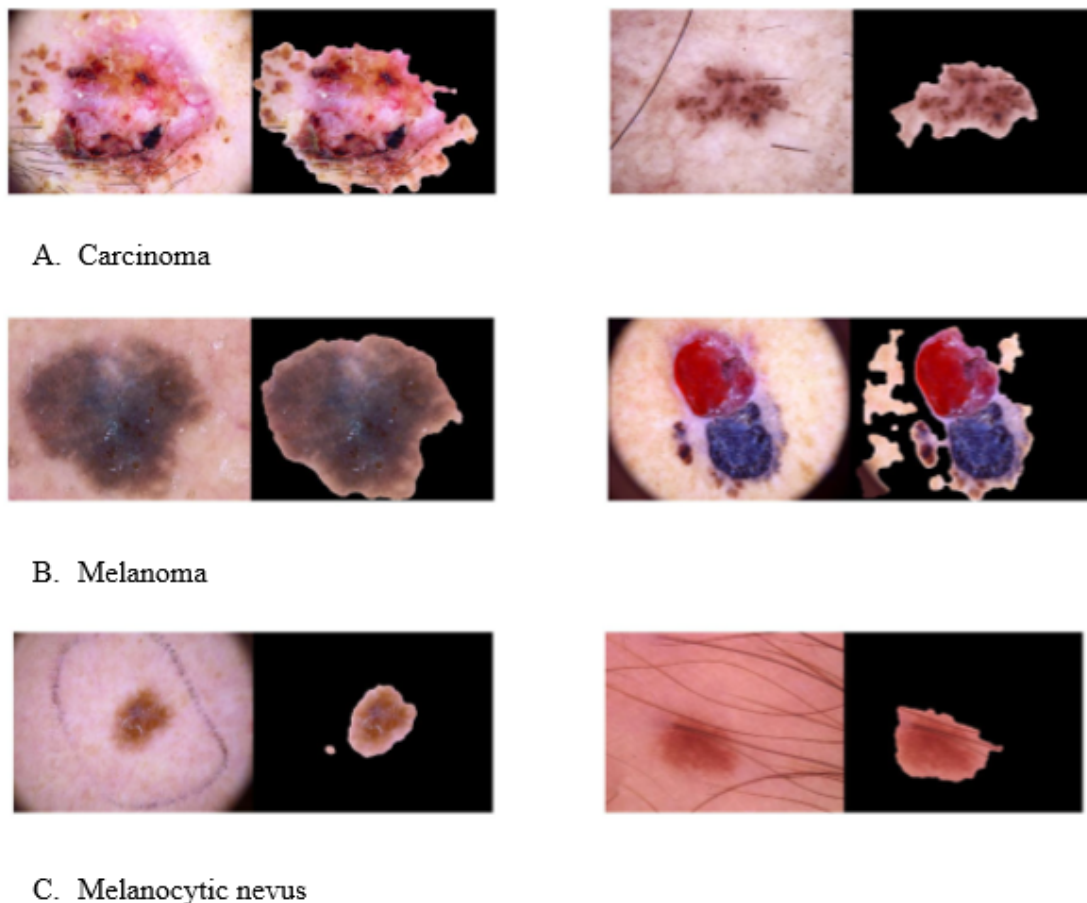


Figure 3: Qualitative examples of binary semantic segmentation for dermoscopic images from the three image-level categories. Each row shows an original image and its predicted lesion-versus-background mask. White pixels represent the lesion region and black pixels represent the background.

3.4. Image-level classification

The SVM was trained using 1,050 images, comprising 350 images per class: carcinoma, melanoma, and melanocytic nevus. Its performance was evaluated using 300 separate observations. The evaluation observations were not used to construct the visual vocabulary or to fit the SVM.

Decision-tree-based, k -means-based, and SVM classifiers were evaluated within the implemented workflow. The SVM with a Gaussian radial basis function kernel produced the highest observed performance among the evaluated configurations and was selected for the reported analysis.

3.5. Comparison of color representations

The color representations were compared using the same SURF-based color–shape descriptors, 250-word Bag of Visual Words vocabulary, SVM framework, three image-level classes, and 300-observation evaluation subset. The observed results are summarized in Table 1.

Among the evaluated three-class configurations, HSV produced the highest observed accuracy at 78.3%, followed by RGB at 73.3%, Opponent Space and XYZ at 70.3%, NTSC at 68.3%, and CIE Lab at 65.3%. All configurations were evaluated using the same three image-level categories and 300 observations.

Table 1

Observed SVM performance for the evaluated three-class color representations.

Color representation	Observations	Accuracy (%)
CIE Lab	300	65.3
NTSC	300	68.3
Opponent Space	300	70.3
XYZ	300	70.3
HSV	300	78.3
RGB	300	73.3

3.6. Confusion matrix

The confusion matrix for the HSV configuration is presented in Table 2. The matrix contains 300 evaluation observations. Rows represent the true classes and columns represent the predicted classes. The diagonal entries represent correct classifications, whereas the off-diagonal entries represent classification errors.

Table 2

Confusion matrix for the HSV configuration.

True class	Carcinoma	Melanoma	Melanocytic nevus
Carcinoma	89	13	12
Melanoma	7	72	14
Melanocytic nevus	4	15	74

The HSV configuration produced 235 correct classifications. Its overall accuracy was calculated as follows:

$$\text{Accuracy} = \frac{89 + 72 + 74}{300} \times 100 = 78.3\%. \quad (3)$$

The corresponding class-specific recalls were 78.1% for carcinoma, 77.4% for melanoma, and 79.6% for melanocytic nevus. The confusion matrix indicates relatively similar recall values across the three classes under the HSV configuration. The diagonal values were 89 correct classifications for carcinoma, 72 for melanoma, and 74 for melanocytic nevus. The off-diagonal values represent misclassified observations among the three image-level categories.

3.7. Discussion

The implemented workflow produced an observed accuracy of 78.3% for the HSV configuration on the three-class evaluation subset. This result was obtained using SURF-based local color–shape descriptors, a 250-word Bag of Visual Words vocabulary, and an SVM with a Gaussian radial basis function kernel.

The distinction between the segmentation and classification tasks is essential for interpreting the results. The encoder–decoder network performed binary pixel-level segmentation by assigning each pixel to either the lesion or background class. The SVM operated at the image level and assigned each complete image to one of three categories: carcinoma, melanoma, and melanocytic nevus. Consequently, the approximately 90% segmentation accuracy observed during training and the 78.3% classification accuracy describe different tasks and should not be compared directly. The segmentation value is a training result, whereas the classification accuracy was calculated from the corresponding evaluation observations.

The two segmentation approaches served complementary purposes. Otsu thresholding provides a simple and computationally efficient method for separating lesion and background when their intensity distributions are sufficiently different [20]. However, its dependence on a global threshold makes it sensitive to illumination changes, hair, marker lines, dark borders, bubbles, and other artifacts. The

encoder–decoder network provides a more flexible alternative because it learns pixel-level representations from annotated images. Its behavior depends on the quality and representativeness of the annotations, the training configuration, and the evaluation protocol [21, 10].

Morphological post-processing and circle detection were used to reduce artifacts and refine the lesion region. Opening, closing, and hole-filling operations can remove small structures, connect nearby regions, and smooth binary boundaries [25]. These operations are useful for preparing the region of interest for feature extraction, but their parameters can also alter the resulting segmentation mask. Future experiments should therefore document these parameters systematically and evaluate their individual contribution through controlled ablation experiments.

The feature representation combined local color information with local shape information. SURF descriptors provided a description of local image structure, while the color channels supplied complementary chromatic information [22]. The descriptors were aggregated using a 250-word visual vocabulary constructed with *k*-means. This Bag of Visual Words representation converted a variable number of local descriptors into a fixed-length histogram suitable for SVM classification [14, 15].

The representation is explicit and computationally interpretable, which facilitates the analysis of individual stages of the workflow. However, histogram encoding can discard relationships between distant image regions and may not preserve the complete spatial organization of a lesion. This limitation is relevant when the interpretation of a lesion depends on the arrangement of multiple local structures.

The comparison of color representations was conducted using the same three image-level categories, the same 300-observation evaluation subset, the same descriptor configuration, the same visual vocabulary size, and the same classifier family. HSV produced the highest observed accuracy at 78.3%, followed by RGB at 73.3%, Opponent Space and XYZ at 70.3%, NTSC at 68.3%, and CIE Lab at 65.3%.

The observed differences among color representations indicate that the chromatic representation affected the behavior of the complete feature-extraction and classification pipeline under the experimental configuration used in this study. The HSV representation combines hue, saturation, and intensity information, which may provide a useful description of color variation in dermoscopic images. These results must nevertheless be interpreted in relation to the fixed descriptor configuration, vocabulary size, data partition, evaluation subset, and classifier family used in the experiments.

The confusion matrix provides information that is not captured by overall accuracy alone. For the HSV representation, the diagonal entries were 89 for carcinoma, 72 for melanoma, and 74 for melanocytic nevus. The corresponding recalls were 78.1%, 77.4%, and 79.6%, respectively. These values indicate relatively similar behavior across the three classes under the evaluated configuration. The off-diagonal entries show that classification errors occurred among all three image-level categories.

The modular organization of the workflow facilitates the analysis of individual processing stages. The segmentation method, morphological operations, local descriptor configuration, visual vocabulary, and classifier can be modified independently in future experiments. This structure also makes it possible to study the contribution of each component through controlled comparisons.

Several methodological aspects should be considered when interpreting the results. The evaluation was performed using a specific image partition, a fixed number of visual words, one local descriptor family, and an SVM with a Gaussian kernel. Changes in the data partition, image preprocessing, descriptor parameters, vocabulary size, or classifier configuration may produce different results. Additional experiments using repeated partitions, independent datasets, alternative descriptors, and complementary evaluation metrics could provide a broader characterization of the workflow.

Overall, the results show that the proposed combination of lesion segmentation, local color–shape description, Bag of Visual Words encoding, and SVM classification can be organized into a modular computational workflow for the analysis of dermoscopic images. The results also illustrate the importance of maintaining consistent classes, observations, preprocessing procedures, and evaluation conditions when comparing different color representations.

4. Conclusions

This study implemented a modular computational workflow for classifying dermoscopic images into three image-level categories: carcinoma, melanoma, and melanocytic nevus. The workflow integrated two alternative lesion-segmentation strategies, morphological post-processing, local color–shape descriptors, Bag of Visual Words encoding, and support vector machine classification.

The semantic segmentation stage was formulated as a binary pixel-level task that distinguished lesion from background. The encoder–decoder network was trained using 2,594 pixel-labeled dermoscopic images and reached an observed training accuracy close to 90%. This result corresponds to the segmentation task and should be distinguished from the subsequent image-level classification evaluation.

For image-level classification, the SVM was trained using 1,050 balanced images and evaluated using 300 separate observations. Six color representations were compared under the same three-class configuration: CIE Lab, NTSC, Opponent Space, XYZ, HSV, and RGB. A 250-word Bag of Visual Words vocabulary was constructed from local SURF-based color–shape descriptors, and the resulting histograms were used as input vectors for classification.

The results show that the proposed combination of lesion segmentation, local color–shape representation, Bag of Visual Words encoding, and SVM classification can be organized into a modular workflow for the computational analysis of dermoscopic images. The separation of the processing stages also facilitates the modification and evaluation of individual components in future experiments.

Further work should examine the effect of segmentation parameters, morphological operations, descriptor configurations, visual vocabulary size, classifier settings, and data partitions. Repeated evaluations, additional image resources, and complementary performance measures could provide a broader characterization of the proposed workflow.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly for grammar and spelling checks. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] P. P. Naik, Cutaneous malignant melanoma: A review of early diagnosis and management, *World Journal of Oncology* 12 (2021) 7–19. doi:10.14740/wjon1349.
- [2] Instituto Nacional de Cancerología, Análisis de la situación del cáncer en Colombia 2015, 2017.
- [3] D. S. Corralo, *Dermatoscopia digital*, 2015. Online resource.
- [4] K. Harrison, The accuracy of skin cancer detection rates with the implementation of dermoscopy among dermatology clinicians: A scoping review, *The Journal of Clinical and Aesthetic Dermatology* 17 (2024) S18–S27.
- [5] G. Argenziano, C. Catricalà, M. Ardigo, P. Buccini, P. De Simone, L. Eibenschutz, G. Ferrari, G. Mariani, V. Silipo, I. Sperduti, I. Zalaudek, Seven-point checklist of dermoscopy revisited, *British Journal of Dermatology* 164 (2011) 785–790. doi:10.1111/j.1365-2133.2010.10194.x.
- [6] P. Zaballos, C. Carrera, S. Puig, J. Malvehy, H. Kittler, Criterios dermatoscópicos para el diagnóstico del melanoma, *Medicina Cutánea Ibero-Latino-Americana* 32 (2004) 3–17.
- [7] Instituto Mexicano del Seguro Social, Abordaje diagnóstico de melanoma maligno, 2010. Clinical guideline.
- [8] C. Carrera, M. A. Marchetti, S. W. Dusza, G. Argenziano, R. P. Braun, A. C. Halpern, N. Jaimes, H. J. Kittler, J. Malvehy, S. W. Menzies, G. Pellacani, S. Puig, H. S. Rabinovitz, A. Scope, H. P. Soyer, W. Stolz, R. Hofmann-Wellenhof, I. Zalaudek, A. A. Marghoob, Validity and reliability of dermoscopic criteria used to differentiate nevi from melanoma: A web-based international

- dermoscopy society study, *JAMA Dermatology* 152 (2016) 798–806. doi:10.1001/jamadermatol.2016.0624.
- [9] O. Yélamos, R. P. Braun, K. Liopyris, Z. J. Wolner, K. Kerl, P. Gerami, A. A. Marghoob, Usefulness of dermoscopy to improve the clinical and histopathologic diagnosis of skin cancers, *Journal of the American Academy of Dermatology* 80 (2019) 365–377. doi:10.1016/j.jaad.2018.07.072.
- [10] H. K. Jeong, C. Park, R. Henao, M. Kheterpal, Deep learning in dermatology: A systematic review of current approaches, outcomes, and limitations, *JID Innovations* 3 (2023) 100150. doi:10.1016/j.xjidi.2022.100150.
- [11] M. P. Salinas, et al., A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis, *npj Digital Medicine* 7 (2024). doi:10.1038/s41746-024-01103-x.
- [12] Y. Liu, et al., Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease, *npj Digital Medicine* 6 (2023) 184. doi:10.1038/s41746-023-00914-8.
- [13] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern, Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging, hosted by the international skin imaging collaboration, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, IEEE, 2018, pp. 168–172. doi:10.1109/ISBI.2018.8363547.
- [14] C. S. Venegas-Barrera, J. Manjarrez, Visual categorization with bags of keypoints, *Revista Mexicana de Biodiversidad* 82 (2011) 179–191.
- [15] H. Gao, L. Dou, W. Chen, J. Sun, Image classification with bag-of-words model based on improved sift algorithm, in: 2013 9th Asian Control Conference, 2013, pp. 27–31. doi:10.1109/ASCC.2013.6606177.
- [16] S. Erdemir, M. Turkoglu, Weighted bag of visual words with enhanced deep features for melanoma detection, *Expert Systems with Applications* 238 (2024) 122033. doi:10.1016/j.eswa.2023.122033.
- [17] International Skin Imaging Collaboration, Isic 2018: Skin lesion analysis toward melanoma detection, <https://challenge.isic-archive.com/>, 2018. Consulted 2026-09-20.
- [18] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Scientific Data* 5 (2018) 180161. doi:10.1038/sdata.2018.161.
- [19] B. C. R. S. Furriel, B. D. Oliveira, R. Próa, J. Q. Paiva, R. M. Loureiro, W. P. Calixto, M. R. C. Reis, M. Giavina-Bianchi, Artificial intelligence for skin cancer detection and classification for clinical environment: A systematic review, *Frontiers in Medicine* 10 (2024) 1305954. doi:10.3389/fmed.2023.1305954.
- [20] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1979) 62–66. doi:10.1109/TSMC.1979.4310076.
- [21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.
- [22] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, *Computer Vision and Image Understanding* 110 (2008) 346–359. doi:10.1016/j.cviu.2007.09.014.
- [23] MathWorks, Introducing deep learning with matlab, MathWorks documentation, 2018.
- [24] International Telecommunication Union, Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios, Recommendation ITU-R BT.601-7, 2011. Available at: https://www.itu.int/dms_pubrec/itu-r/rec/bt/r-rec-bt.601-7-201103-i!!pdf-e.pdf.
- [25] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2 ed., Springer, Berlin, Heidelberg, 2013. doi:10.1007/978-3-662-05088-0.