SPADE: Superpixel Adjacency Driven Embedding for Three Class Melanoma Segmentation

Pablo Ordóñez , Ying Xie * , Xinyue Zhang , Yixin Chloe Xie , Santiago Acosta Rodriguez , Issac Gutierrez

Posted Date: 16 July 2025

doi: 10.20944/preprints202507.1380.v1

Keywords: melanoma; deep learning; SLIC; border; embeddings; decoder; encoder; attention

 doi:10.20944/preprints202507.1380.v1

Article

SPADE: Superpixel Adjacency Driven Embedding for Three Class Melanoma Segmentation

Pablo Ordóñez *, Ying Xie †, Xinyue Zhang, Chloe Yixin Xie, Santiago Acosta and Issac Guitierrez

College of Computer Science and Software Engineering, Kennesaw State University, Marietta, GA, USA
* Correspondence: pordonez@kennesaw.edu

Abstract

Melanoma remains one of the most lethal forms of skin cancer. In clinical practices, primary care physicians typically rely on the ABCDE criteria (Asymmetry, Border, Color, Diameter, and Evolution), alongside dermoscopic examination and scoring systems, to assess lesion malignancy. However, these assessments are inherently subjective and often influenced by the clinician's level of experience. To address this limitation, Computer-Assisted Diagnosis (CAD) systems have been developed to provide more objective and reproducible evaluations. CAD algorithms either extract ABCD-related features or directly classify lesions from dermoscopic images. In both approaches, accurate lesion segmentation is critical. Yet, approximately 30% of skin lesions exhibit fuzzy or poorly defined borders, complicating the task of drawing a single, definitive contour. In this work, we identify three distinct classes in dermoscopic images (background, border, and lesion core) based on superpixels generated via the Simple Linear Iterative Clustering (SLIC) algorithm. Our contributions are fourfold: (1) redefining lesion borders as regions rather than lines; (2) generating superpixel-level embeddings using a transformer-based autoencoder; (3) incorporating these embeddings as features for classification; and (4) integrating neighborhood information to construct enriched feature vectors. Unlike pixel-level CNN algorithms that often overlook fine-grained boundary contexts, our pipeline fuses global class context with local spatial relationships, significantly improving precision and recall in challenging border regions. Extensive evaluation on the HAM10000 melanoma dataset demonstrates that our superpixel-RAG-transformer pipeline achieves exceptional performance in classifying background, border, and lesion core superpixels.

Keywords: melanoma; deep learning; SLIC; border; embeddings; decoder; encoder; attention

1. Introduction

Melanoma is an aggressive malignant tumor that originates from melanocytic cells, which are primarily located in the skin, but can also be found in the mucosa, uvea, and meningeal membranes. The development of melanoma is influenced by several key risk factors, including a history of excessive sun exposure and sunburns, Fitzpatrick skin types I and II, and the presence of atypical nevi. There are several subtypes of melanoma as shown on Figure 1. Superficial spreading melanoma (SSM) is the most common, accounting for approximately 70% of cutaneous melanomas. Nodular melanoma (NM) represents about 15-20%, while lentigo maligna melanoma (LMM) accounts for 5-10%. Other less common subtypes include acral lentiginous melanoma (ALM), amelanotic melanoma, desmoplastic melanoma (DM), mucosal melanoma, and uveal melanoma. [1–3]

According to epidemiological data, the incidence rate of melanoma in the United States was reported at 21.2 cases per 100,000 individuals in 2022 [4]. Projections from the American Cancer Society estimate that 104,960 new cases will be diagnosed in 2025 [5]. Despite significant advancements in therapeutic approaches, melanoma continues to exhibit a mortality rate of approximately 8% [6,7]. The clinical diagnosis of melanoma relies on the assessment of lesion characteristics based on the

 doi:10.20944/preprints202507.1380.v1

ABCDE criteria, as depicted in Figure 2, which evaluate asymmetry, border irregularity, color variation, diameter greater than 6 mm, and evolution over time [8]. Other than the diameter, the ABC criteria are subject to subjective interpretation. Furthermore, physicians with more clinical experience tend to demonstrate higher diagnostic accuracy [9].

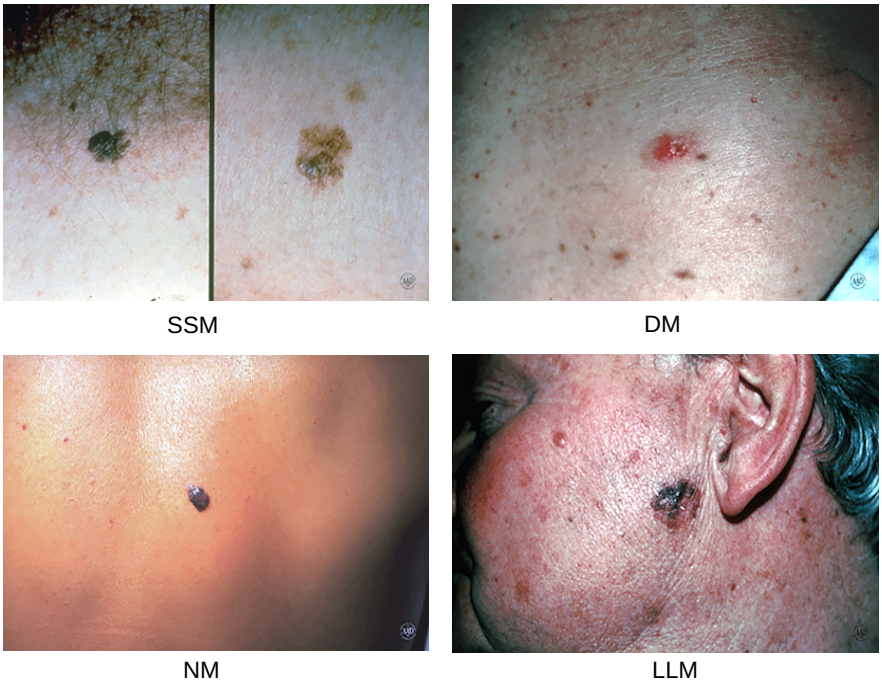


Figure 1. Melanoma Types. Courtesy of American Academy of Dermatology Association.

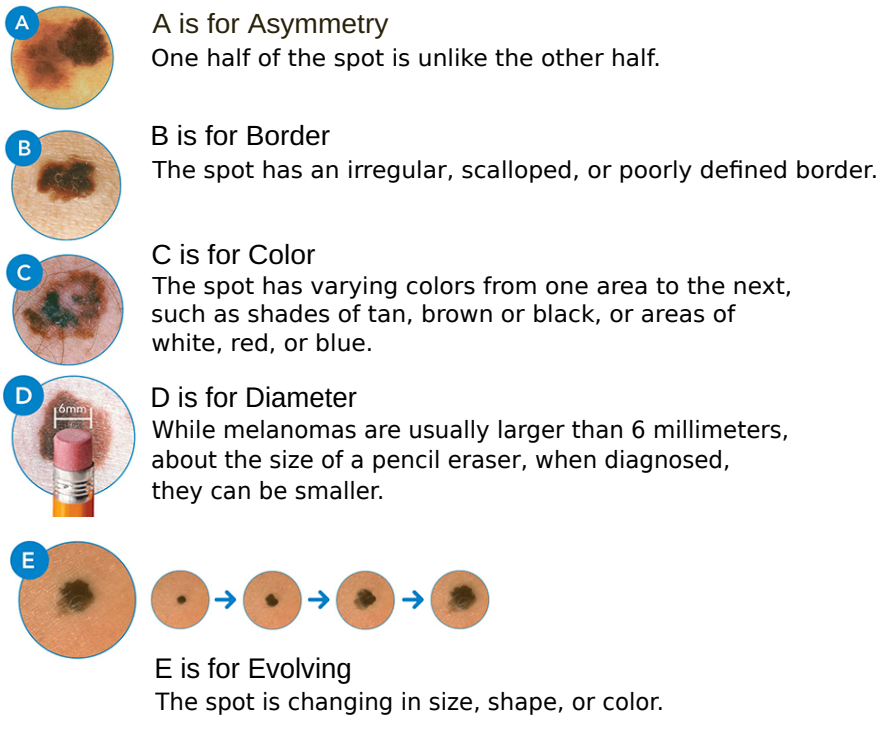
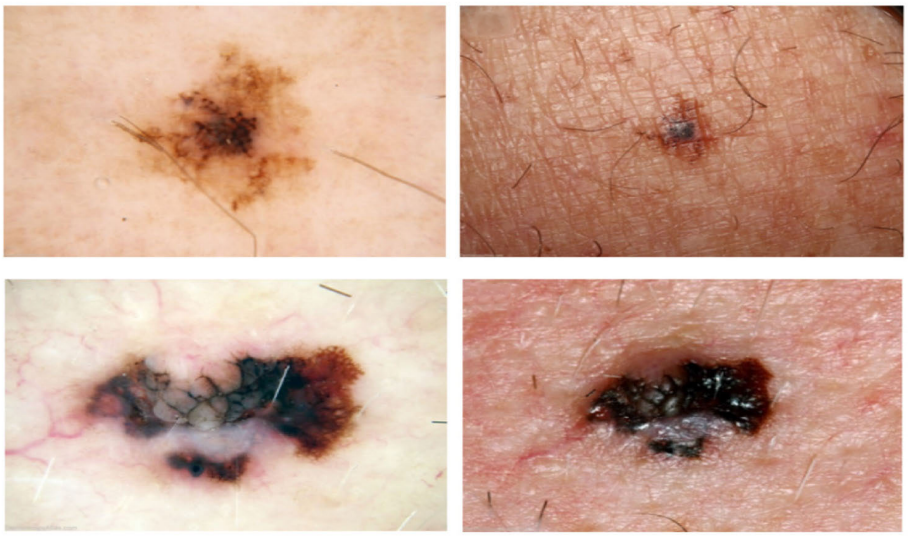


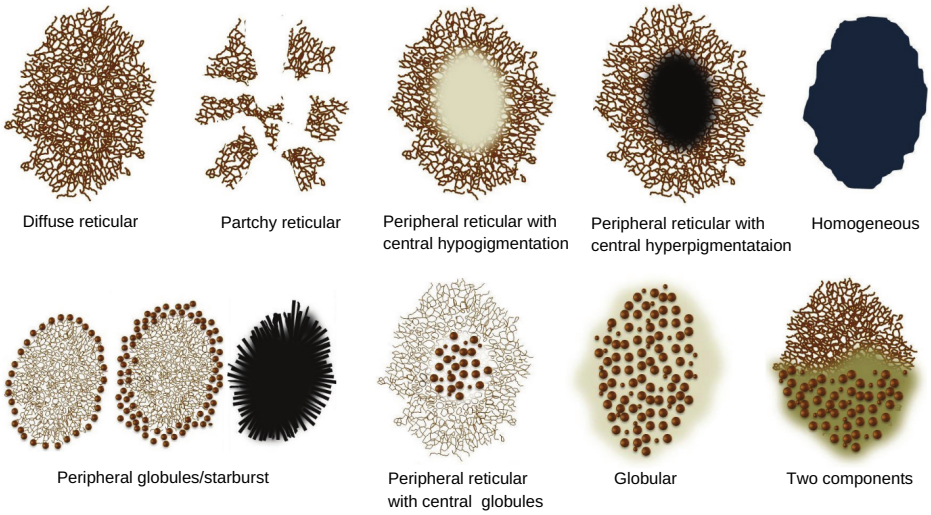
Figure 2. ABCD Features. Courtesy of American Academy of Dermatology Association.

 doi:10.20944/preprints202507.1380.v1

Dermoscopy is a non-invasive diagnostic technique widely used by general practitioners (GP) and dermatologists to improve the detection of skin lesions suspected of malignancy. A dermatoscope, in simple terms, is a magnifying lens equipped with illumination and, in many cases, digital imaging capabilities. It facilitates the visualization of subsurface skin structures that are not visible to the unaided eye as shown in part (a) of Figure 3. Studies have demonstrated that the incorporation of dermoscopy into the diagnostic workflow enhances accuracy; it increases accuracy from 54% to between 76% and 79%, compared to visual inspection alone [10]. Dermoscopy captures unique and mixed basic structural patterns such as reticular, streaked, globular, and homogeneous, as illustrated in part (b) of Fig 3. These structures may be distributed centrally, peripherally, or may be diffused across the lesion, and some of these patterns are observed exclusively in malignant lesions. Additionally, dermoscopy can reveal vascular patterns, which provide further diagnostic value. Although some dermoscopic patterns are clearly identifiable, more complex ones pose a significant challenge, especially for primary care physicians [11]. This also introduces subjectivity into the diagnostic process as pattern identification relies on clinical experience.



(a) Dermoscopy vs Standard Photography



(b) Patterns Types

Figure 3. Dermoscopy Patterns. Courtesy American Family Physician.

Scoring systems, such as the Total Dermoscopy Score (TDS) and the 7-Point Checklist (7PCL), incorporate these features to enhance the diagnostic accuracy of melanoma detection. [12] These systems

 doi:10.20944/preprints202507.1380.v1

integrate asymmetry, border irregularity, color variation, and diameter to provide a comprehensive risk assessment. Table 1: outlines the procedures for calculating the most common scoring systems.

Specifically, TDS classifies lesions as follows:

- TDS $\leq$ 4.75: Benign lesion
- TDS between 4.8 and 5.45: Suspicious lesion
- TDS $\geq$ 5.45: Malignant lesion

TDS and 7PCL are valuable tools in melanoma diagnosis, however, their reliance on subjective assessments can lead to variability in interpretations. [13]

Table 1. Scoring System.

Scoring System	Calculation
ABCD Rule	Total Dermoscopy Score (TDS): $$\text{TDS} = (A \times 1.3) + (B \times 0.1) + (C \times 0.5) + (D \times 0.5)$$ where: • A = Asymmetry score (0–2) • B = Border score (0–8) • C = Color score (1–6) • D = Dermoscopic structures score (1–5)
7-Point Checklist	Weighted scoring: • Major criteria (2 points each): – Change in size – Irregular shape – Irregular color • Minor criteria (1 point each): – Inflammation – Crusting or bleeding – Sensory change – Diameter $\geq$ 7 mm Lesions scoring $\geq$ 3 points warrant further investigation.

For GP, the initial step involves distinguishing melanocytic lesions from non-melanocytic ones through clinical evaluation to ensure accurate preliminary classification. The second step focuses on differentiating nevi from melanoma by applying clinical criteria such as the ABCDE rules, interpreting dermoscopic findings, and utilizing scoring systems. Suspicious lesions are typically referred to a specialist (dermatologist) for further assessment. We have described how human intelligence, through visual inspection, dermoscopic tools, and scoring systems, is used to determine whether a lesion is benign or malignant. Nonetheless, these evaluation methods rely heavily on human judgment, whether through clinical assessment, the use of tools such as dermoscopy, or the application of clinician-derived scoring systems, making them inherently prone to subjectivity.

The diagnostic subjectivity and inter-observer variability associated with melanoma underscore the need for automated and objective computer-assisted diagnosis (CAD) systems. Two primary approaches have been developed for melanoma classification using CAD.

The first approach aligns with the clinical methodology and employs scoring systems based on the ABCD rule. To ensure objectivity, automated methods have been proposed for detecting shape asymmetry, color variegation, and lesion diameter. Among these, Fourier Descriptors are considered the most robust techniques for quantifying asymmetry due to their mathematically rigorous formulation and demonstrated alignment with expert dermatological assessments, achieving up to 92% concordance with dermatologist evaluations [14,15]. For border irregularity, combined shape descriptors including fractal dimension, Zernike moments, and convexity coupled with convolutional

 doi:10.20944/preprints202507.1380.v1

neural network (CNN) classifiers have achieved state-of-the-art results, with reported classification accuracies reaching 93.6% [16]. In assessing color variegation, the CIELAB color space combined with the Minkowski distance metric has proven the most effective, offering perceptual uniformity that closely matches human color vision and handling wide color variations reliably [17]. Regarding lesion diameter, Feret's diameter has been identified as the most accurate and consistent measure, particularly for lesions with irregular boundaries [17].

The second approach focuses on direct skin lesion classification, which can be performed using either full images or segmented lesion regions. Studies have shown that classifiers trained on segmented images generally outperform those trained on whole images [18]. In classical CAD systems, lesion segmentation is typically performed using three main approaches: edge-based, region-based, and threshold-based methods [19]. For classification, methods such as score averaging (AVGSC), linear SVM, and non-linear SVM using a histogram intersection kernel have demonstrated acceptable performance levels [20].

The advent of deep learning has significantly enhanced both segmentation and classification accuracy by leveraging CNNs and transformer-based architectures [21]. Deep supervised learning has led to the development of increasingly sophisticated architectures beyond traditional CNNs.

Popular CNN-based models for melanoma classification include ResNet [22], VGG16 [23], MobileNet [24], DenseNet [24] and Inception [25]. Transformer-based architectures have also demonstrated strong performance [26–28]. Reported classification performance across these models varies, with accuracy ranging from 80% to 98%, sensitivity from 60% to 90%, and specificity from 86% to 98%. Among these, EfficientNet-B7 [29] and transformer-based models [26–28] exhibit the highest diagnostic accuracy.

Models based solely on CNNs have been increasingly outperformed by architectures incorporating encoder-decoder frameworks, reflecting a shift towards more effective hierarchical feature representations. CNN-based segmentation models fall into two primary categories: pixel-wise upsampling techniques, such as fully convolutional networks (FCNs), and spatial and pyramidal upsampling approaches, such as the DeepLab family of models [30]. Notably, DeepLabV3+ introduces a decoder module designed to enhance segmentation accuracy and refine feature extraction, particularly at object boundaries [31]. Encoder-decoder architecture like U-Net and V-Net utilize symmetric upsampling at each layer to preserve spatial resolution during reconstruction. This improves the fidelity of the segmented output [32] [33,34]. Further advancements have been realized through attention-based models, which improve both feature selection and contextual awareness. Prominent architectures in this category include the Vision Transformer (ViT) [35,36] and Visual Attention Networks [37], which leverage self-attention mechanisms to capture long-range dependencies within an image. The segmentation accuracy of deep learning models for skin lesion analysis ranges from 80% to 98%, with recent architectures consistently surpassing their predecessors in both precision and generalization capacity.

Accurate lesion segmentation is essential for the reliable extraction of ABCD features and for effective classification. In most publicly available datasets, lesion boundaries are manually annotated by dermatologists, residents, medical students, or trained personnel. This process is not only time-consuming but also inherently subjective. For lesions with fuzzy or ambiguous boundaries, substantial inter-annotator variability exists. A study by Kittler et al. reported discrepancies of up to 20-30% due to ambiguous lesion edges and variations in annotator expertise [12,38,39]. To address this, some organizations provide standardized segmentation protocols. For instance, the ISIC archive employs multi-rater consensus in cases with ambiguous lesion boundaries [20]. These practices underscore the degree of subjectivity that remains embedded in CAD systems dependent on manual segmentation for classification or ABCD rule application.

A comprehensive meta-analysis conducted within the medical community evaluated the effectiveness of AI models in skin cancer detection. Following the PRISMA methodology, the review included

 doi:10.20944/preprints202507.1380.v1

272 publications from an initial pool of 14,224 studies, spanning the years 2000 to 2021. The analysis reported a mean F1 score of 0.807 for melanoma detection, with scores ranging from 0.732 to 0.882 [40].

In summary, we have presented insights from both computer science and medical perspectives. Despite significant advancements in deep learning based melanoma detection, current AI methods remain insufficient for widespread clinical adoption. There exists a discrepancy in reported accuracy results between the medical and engineering communities, which warrants further investigation, particularly with respect to melanoma subtypes, patient demographics, image quality, and dataset provenance. Additionally, it is important to explore the bias introduced by relying on annotated lesion borders, especially in cases involving irregular or ambiguous boundaries.

1.1. Contributions

Although AI models have significantly advanced in recent years, delineating lesions remains a subjective task even for specialists. We propose that the lesion border should not be treated as a single, well-defined line, but rather as a transitional region that separates the background from the core of the lesion. Part (b) of Figure 5 presents a blended image composed of the ground truth mask overlaid on the original lesion. In our work, the border region is defined as the area between the outer edge of the mask and the visible boundary of the lesion. This region can be more precisely delineated using superpixels, which provide spatially coherent segmentation aligned with local image features. Our objective is to develop a model capable of accurately identifying all three regions. In this paper, we propose a new method called SPADE (Superpixel Adjacency Driven Embedding) for three class melanoma segmentation. The main contributions of SPADE are as follows.

- Definition of Three Anatomical Zones: We define three distinct zones in dermoscopy images, the background, the border, and the core of the lesion.
- Transformer-based Superpixel Embedding: We train a transformer autoencoder to generate embeddings for each superpixel, enabling the model to capture contextual relationships between spatially distant regions belonging to the same class.
- Context-Aware Region Adjacency Graph (RAG): We construct a Region Adjacency Graph using superpixels obtained via the SLIC algorithm to model spatial relationships.
- Input Definition Based on the Region Adjacency Graph: Each input vector consists of the embedding of a given superpixel along with the embeddings of its immediate neighbors, effectively capturing local spatial context.
- Semantic Representation of Dermoscopy Images: The system effectively captures the semantic content of dermoscopy images, supporting improved lesion characterization.

1.2. Related Work

Fine-tuning AI models for accurate border delineation in blurry or low-contrast dermoscopic images has been addressed through methods that identify and label regions adjacent to lesions. Adegun et al. employ probabilistic models in which each pixel is assigned a value between 0 and 1, representing the likelihood of belonging to the lesion [41]. The algorithm ultimately produces a lesion boundary represented as a line. Halil et al. utilize region of interest (ROI) bounding, where an initial bounding box or elliptical region is drawn around the lesion, followed by consensus among multiple annotators [42]. Although the initial stage localizes the lesion broadly, the final boundary is refined using the GrabCut algorithm, resulting in a sharply delimited line. Zahra et al. propose an approach based on multiple expert-generated masks, from which a consensus mask is derived through fusion [43]. The final output is again a single mask outlined by a distinct boundary line. Li et al. [44,45] introduce a method that explicitly marks regions of uncertainty rather than enforcing a hard boundary. They use class activation maps to generate pseudo-labels and apply binary thresholding to create segmentation masks. This approach allows the model to learn from ambiguous regions, thereby improving robustness in uncertain contexts. In contrast to these methods, which ultimately

 doi:10.20944/preprints202507.1380.v1

produce a line-based lesion boundary, our approach focuses on defining lesion regions, providing a more granular and context-aware representation.

Graph-based methods also commonly leverage superpixels by constructing a RAG, where each node represents a superpixel and edges encode similarity metrics between neighboring regions such as color, texture, or boundary strength. Once the graph is constructed, segmentation is performed using algorithms such as spectral clustering, normalized cuts, or min-cut/max-flow, and neural networks, which partition the graph into semantically meaningful regions.[46]. RAGs naturally integrate with Graph Neural Networks (GNNs) and have been employed for image classification tasks. Nazir et al. applied Graph Convolutional Neural Networks (GCNNs) using both spatial and spectral convolution techniques, demonstrating that spectral-based models outperform spatial-based models and classical CNNs while requiring less computational cost [47]. Avelar et al. explored attention-based GNNs for classification and showed that the feature space can be enhanced by weighting the edges of a superpixel graph using a learned function based solely on geometric information [46]. Nowosad et al. apply graph-based segmentation to non-imagery geospatial data by grouping superpixels based on dissimilarity measures and pruning the resulting graph to a Minimum Spanning Tree (MST) [48]. Additionally, Qin et al. propose a superpixel-based and boundary-sensitive CNN model specifically designed for liver segmentation, demonstrating the effectiveness of combining region-based pre-processing with deep learning techniques [49]. Similar to our work, these studies employed superpixels and RAG to construct graph representations. However, unlike prior work, the current phase of our study does not incorporate graph neural networks or edge attributes for classification.

2. Methods and Algorithm Design

The SPADE pipeline proceeds as follows: the images are first resized; superpixels are then generated using the Simple Linear Iterative Clustering (SLIC) algorithm; a region adjacency graph is constructed; each superpixel is labeled by class (background, border, lesion); class-specific embeddings are generated using a transformer; and finally, a transformer is trained to predict the three classes.

2.1. Resizing Images

All images were initially adjusted to a landscape orientation and resized to a width of 1024 pixels and a height of 768 pixels using a bilinear interpolation algorithm from the OpenCV framework.

2.2. Superpixel Generation Strategy

Superpixel segmentation partitions an image into regions composed of similar and spatially connected pixels providing perceptually meaningful atomic regions.

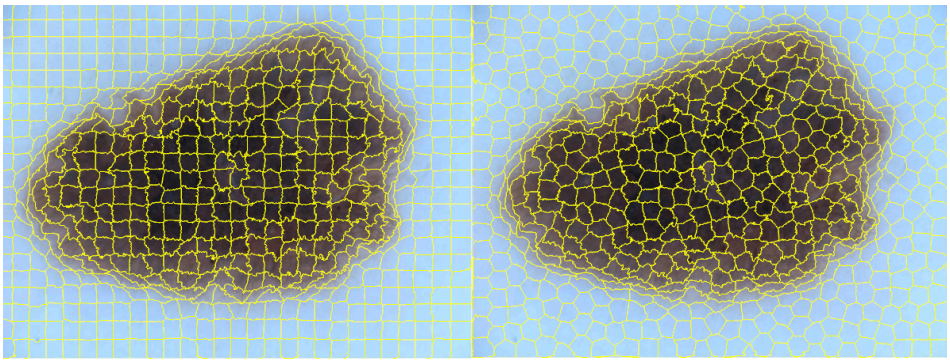
SLIC is classified as a neighborhood-based clustering algorithm. SLIC is an adaptation of the k-means clustering algorithm, specifically designed for efficient superpixel generation. It introduces two key parameters: a search radius that limits the number of distance computations, and a weighted distance metric that balances color similarity and spatial proximity S. The distance function is defined as: $D_s = d_{rgb} + \frac{m}{S} d_{xy}$ where d_{rgb} is the Euclidean distance in the RGB color space, d_{xy} is the geometric distance between pixels, and m is a compactness parameter that controls the trade-off between color and spatial proximity. These parameters enable control over the size and compactness of the resulting superpixels. Furthermore, SLIC's linear computational complexity contributes to its widespread adoption in superpixel generation tasks [50].

Each pixel in SLIC is represented by a five-dimensional feature vector consisting of three components for color and two for spatial coordinates $C_x = [R_k, G_k, B_k, x_k, y_k]$. The algorithm employs the CIELAB color space, which offers a perceptually uniform representation of color, ensuring a more meaningful similarity measure. The segmentation process in SLIC involves three main steps. First, clusters are initialized based on the desired number of superpixels. Second, an iterative clustering process assigns pixels to the nearest cluster using a distance function that incorporates both color and spatial information. Finally, a post-processing step enforces spatial connectivity to ensure that each superpixel forms a contiguos region [50,51].

 doi:10.20944/preprints202507.1380.v1

SLIC superpixels effectively capture low-level image features like color, position, and depth, while preserving fine-grained boundary details. This is essential for detecting subtle background transitions around lesions, which clinicians often rely on. Moreover, superpixels reduce noise by averaging local pixel variations within each region and help focus computational resources on meaningful structures, since larger segments often carry more semantic information than individual pixels.

The SLIC algorithm was applied to each image to obtain its segmented regions. Specifically, we employed the SuperpixelSLIC algorithm from OpenCV with the following parameters: a desired number of superpixels k set to 600, a area size of 1320, $S = 34$, and a ruler value(m) of 19220 producing the segmentations shown in Figure 4. The software suite provides three different algorithms, SLIC, SLICO, and MSLIC, among which SLIC was chosen due to its balance between algorithmic compactness, boundary recall, and computational efficiency. Each superpixel was assigned a unique identifier, which was used to track and analyze the segmented regions [52]. At the beginning of the algorithm, each superpixel is assigned a unique identifier (ID) when the initial centroids are defined. This ID is stored as an integer value in the database and remains constant throughout the duration of the experiment.

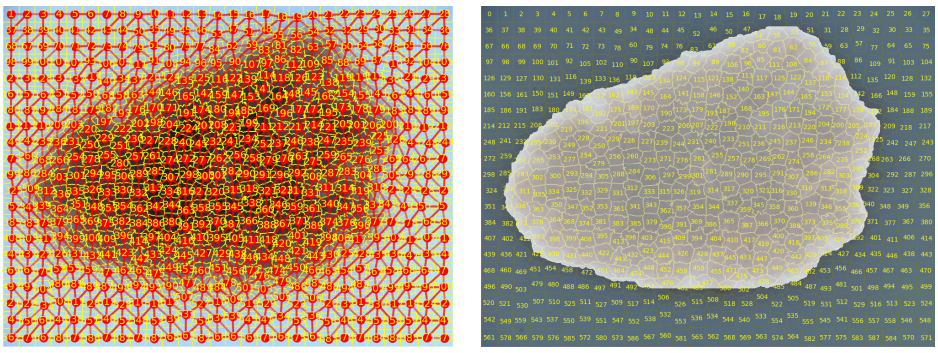


Iteration #1 Iteration #40

Figure 4. SLIC Superpixels Image Sequences.

2.3. Constructing the Region Adjacency Graph (RAG)

The RAG was constructed by traversing the labeled superpixel matrix and establishing undirected edges between adjacent superpixels with differing labels based on 4-and 8-connected neighborhood criteria, capturing spatial connectivity through horizontal, vertical, and diagonal adjacency as depicted in part (a) of Figure 5 [53].



(a) Region Adjacent Graph (RAG) (b) Superpixels Image with Mask

Figure 5. Dermoscopy Image, Superpixels, RAG, and Mask.

We hypothesize that, similar to how transformers leverage the sequential ordering of words to learn language structure, our Region Adjacency Graph (RAG) explicitly defines "neighbor" relationships among superpixels. Unlike Vision Transformers, where all patches are included as input tokens, our RAG-based approach provides meaningful context to the transformer. In this framework, the transformer's self-attention mechanism operates over graph nodes, attending preferentially to

 doi:10.20944/preprints202507.1380.v1

adjacent regions that share meaningful boundaries or semantic relationships. In contrast to Vision Transformers, which may waste computational resources attending to regions lacking relevant context, our RAG-based model focuses attention on superpixels containing semantically relevant information.

For instance, consider an image with a size 224×224 pixels, with patches sizes of 16×16; this results in a total of 196 patches. In a Vision Transformer, each transformer layer computes self-attention with a complexity of $\mathcal{O}(N^2\dot{D})$, where N is the number of patches and D is the embedding dimension. In our superpixel based approach, we observe that each superpixel has, on average, 7 neighbors (ranging from $2-19$), leading to a significantly reduced complexity of $\mathcal{O}(7^2\dot{D})$.

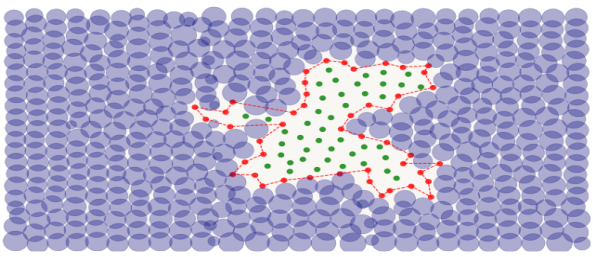
2.4. Creating Features

For each image, we extracted metadata including the filename, height, and width. We also computed the mean and standard deviation for each RGB channel. For each superpixel, we computed the mean, standard deviation, kurtosis, and skewness per RGB channel. Using the segmented region as a binary mask in a 2D image, where pixels belonging to the superpixel were set to 1 and all others to 0, we calculated the area $(m00)$, centroid $X = (m10/m00)$, centroid $Y = (m01/m00)$, and second central moments using the OpenCV moments function. Additionally, neighboring superpixels were identified and included as part of the superpixel-level information.

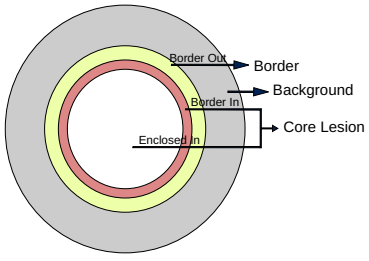
To facilitate further analysis, the superpixels were categorized into three classes (background, border, and core lesion) for a subset of 2,000 images from the HAM10000 dataset ("Human Against Machine with 10,000 training images") [54]. A custom annotation tool was developed to allow manual labeling of superpixels. The supplementary video (Video 1) provides a visual demonstration of how the annotator selects the "border in" and "border out" regions on dermoscopic images. The "border out" class includes superpixels marking the transition between the lesion and the surrounding skin, whereas the "border in" class contains superpixels located just inside the lesion boundary.

To delineate the lesion interior, a concave hull algorithm was applied to enclose all superpixels within the "border in" region. Unlike the convex hull, which may encompass unrelated background superpixels, the concave hull prevents the inclusion of superpixels that do not belong to the lesion, as illustrated in part (a) of Figure 6.

Based on the proposed classification scheme, we define the core lesion as the union of superpixels labeled as "border in" and those enclosed by them. The border class corresponds to the "border out" superpixels and all the remaining superpixels are assigned to the background class, as shown in part (b) of Figure 6



(a) Concave Hull: Red dots indicate the centers of superpixels labeled as "border in", green dots correspond to superpixels labeled as "enclosed in", and blue circles represent "background" superpixels along with those labeled as "border out"



(b) Superpixel Grouping: Superpixels labeled as "border in" are grouped together with those labeled as "enclosed in" to define the core lesion region.

Figure 6. Grouping Superpixels on Classes.

 doi:10.20944/preprints202507.1380.v1

Table 2 presents the total number of superpixels per category, where class labels 0, 1, and 2 correspond to Background, Core Lesion, and Border, respectively.

Table 2. Superpixels Classes Statistics.

Class	Label	Superpixels	%
0	Background	809,891	68.9
1	Lesion	284,116	24.1
2	Border	80,229	6.8

2.5. Transformer Autoencoders for Generating Embeddings

To generate embeddings for each superpixel, we train a transformer autoencoder independently for each semantic class (Background, Border, or Core). Each superpixel is resized to contain 700 pixels per channel, and the latent space dimensionality is set to 256. The model architecture is illustrated in Figure 7. The encoder consists of four multi-head self-attention layers (with four heads each), interleaved with four feed-forward linear layers. This encoder maps the input patch tokens to a 256-dimensional latent vector. The decoder is symmetric, consisting of four self-attention layers and four linear layers, and reconstructs the original patch from the latent representation. Each attention and feed-forward sub-layer is followed by residual connections and layer normalization to stabilize training and improve gradient flow.

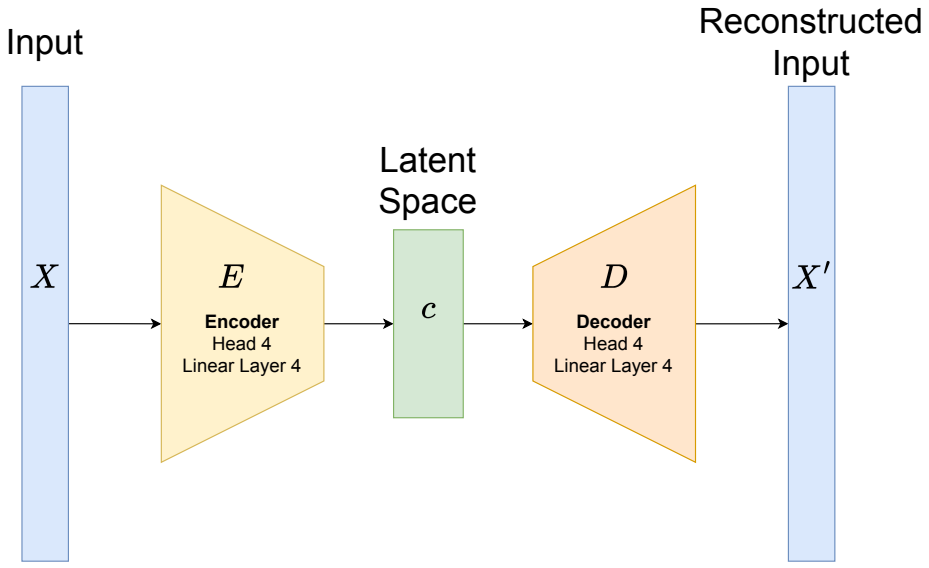


Figure 7. Transformer Autoencoder.

We train a separate autoencoder for each class using only superpixels labeled accordingly. The reconstruction loss is measured using Mean Squared Error (MSE) between the original input and the decoder's output. Optimization is performed using the Adam optimizer with early stopping, based on an 80/20 training-validation split (random seed = 40).

By leveraging the transformer's self-attention mechanism, these autoencoders effectively learn to encode irregular, shape-conforming superpixels. This allows them to capture not only intra-region content, but also inter-region semantic patterns—capabilities that fixed-grid patches or traditional CNN encoders cannot achieve as efficiently. We hypothesize that combining these learned embeddings with our Region Adjacency Graph (RAG) structure (Section II.C) enables a context-aware segmentation framework built upon semantically meaningful, spatially grounded building blocks.

The extracted information was stored in a Neo4j graph database [55], where the data was structured into three main entities Images, Superpixels, and Centroids as shown in part (a,b) Figure 8. The Image entity contained all image-related properties, while the Superpixel entity stored superpixel-specific characteristics including the embeddings generated by the latent space. The Centroid entity

 doi:10.20944/preprints202507.1380.v1

recorded the centroid's x and y coordinates. Within the database, IN_IMAGE relationships linked each Image entity to its corresponding Superpixel entities. NEIGHBOR_OF relationships connected Superpixel entities with their neighboring superpixels, such that edge weights represented the Euclidean distance between RGB mean values. Finally, CENTER_AT relationships associated each Superpixel entity with its respective Centroid.

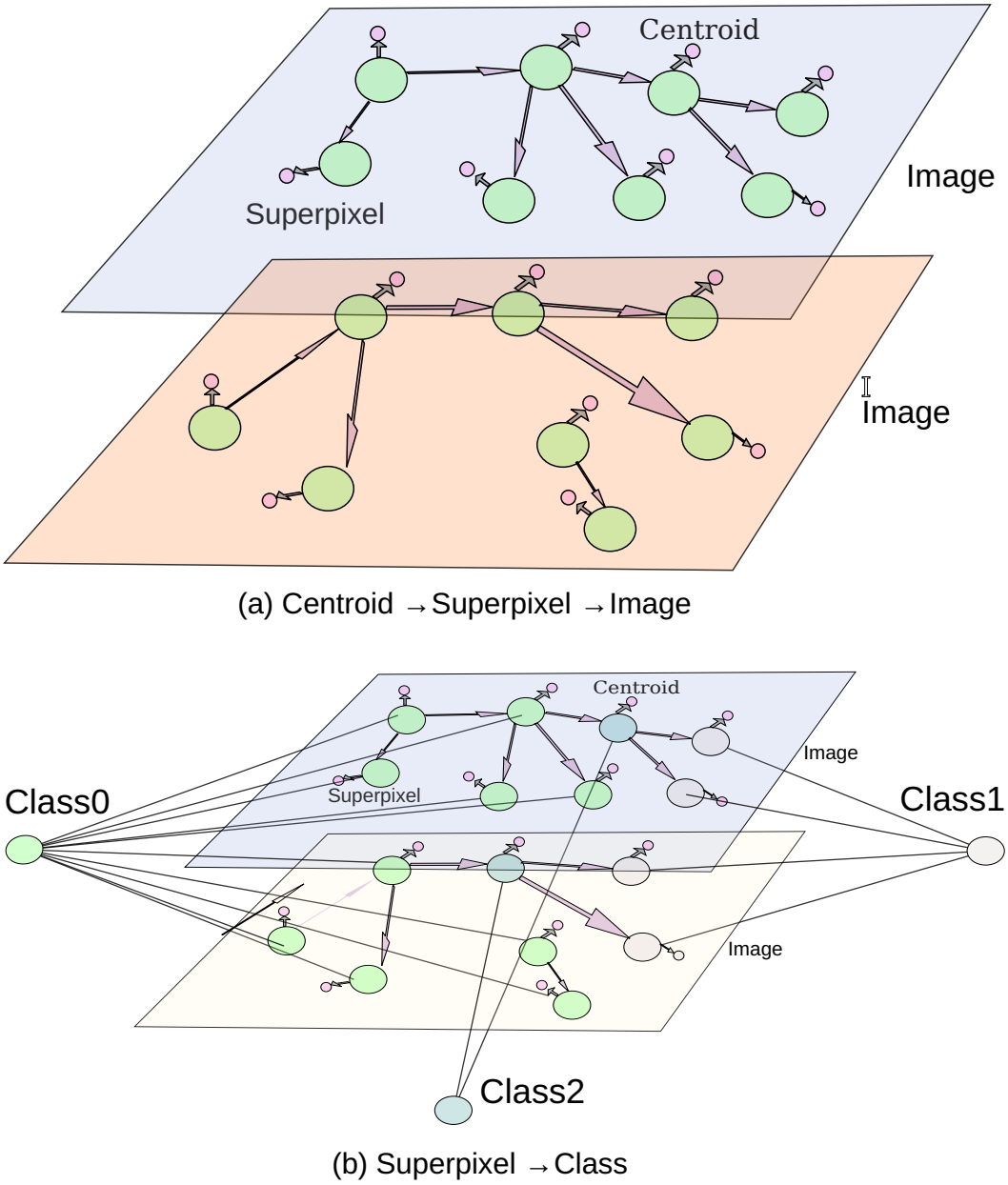


Figure 8. Graph Representation of RAG.

PyTorch version 2.2.2 and CUDA 12 was used to create and run the Deep Learning models for training and inference. The data was normalized and split amongst training and validation sets, 80% and 20% respectively. We set the random seed to 40 for all experiments.

 doi:10.20944/preprints202507.1380.v1

2.6. Algorithm

Algorithm 1: Obtain Embeddings from Superpixels and Neighbors

Input: RGB image I
Output: Embedding vectors for each superpixel and its neighbors

Function ResizeImage(I)
    Apply bilinear interpolation to standardize input size;
    return resized image I';
Function GenerateSuperpixels(I')
    Apply SLIC algorithm to RGB image;
    return set of superpixels S;
Function ConstructRAG(S)
    Build Region Adjacency Graph G using 8-connected neighbors;
    return graph G;
Function LabelSuperpixels(S)
    Manually annotate each superpixel as *background*, *border*, or *core*;
    return labeled set L;
Function TrainAutoencoderPerClass(S, L)
    foreach *class* $c \in \{$*background, border, core*$\}$ **do**
        Resize each superpixel to 700 pixels per channel;
        Set latent space to 256 dimensions;
        Train transformer autoencoder for class c;
    return trained autoencoders $\mathcal{A}_c$;
Function InferEmbeddings($S, \mathcal{A}_c$)
    foreach *superpixel* $s \in S$ **do**
        Use corresponding class autoencoder $\mathcal{A}_c$ to infer embedding;
        Store embedding vector e_s;
        Store statistical features from superpixels s_f
    return $\{e_s, s_f\}_{s \in S}$;
Function CombineWithNeighbors($G, \{e_s\}$)
    foreach *superpixel* s **do**
        Retrieve 8-connected neighbors N_s from G;
        Concatenate embedding e_s, and s_f features , and those of N_s;
    return final embedding set $\{E_s\}_{s \in S}$;
$I' \leftarrow$ ResizeImage(I);
$S \leftarrow$ GenerateSuperpixels(I');
$G \leftarrow$ ConstructRAG(S);
$L \leftarrow$ LabelSuperpixels(S);
$\mathcal{A}_c \leftarrow$ TrainAutoencoderPerClass(S, L);
$\{e_s\} \leftarrow$ InferEmbeddings($S, \mathcal{A}_c$);
$\{E_s\} \leftarrow$ CombineWithNeighbors($G, \{e_s\}$);
return $\{E_s\}$

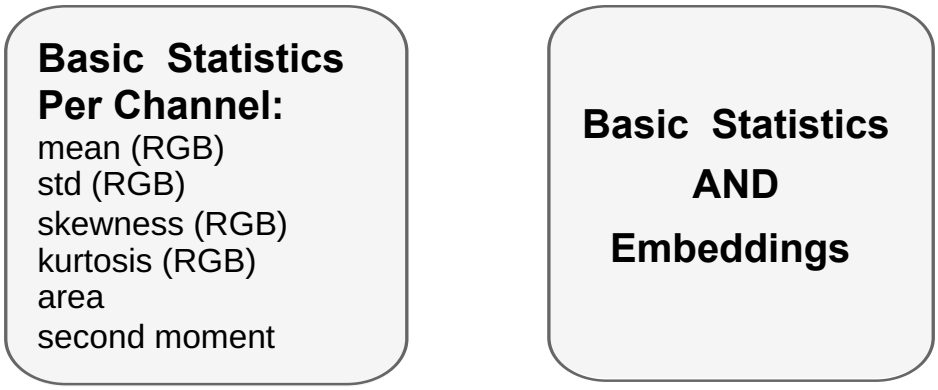
 doi:10.20944/preprints202507.1380.v1

3. Results

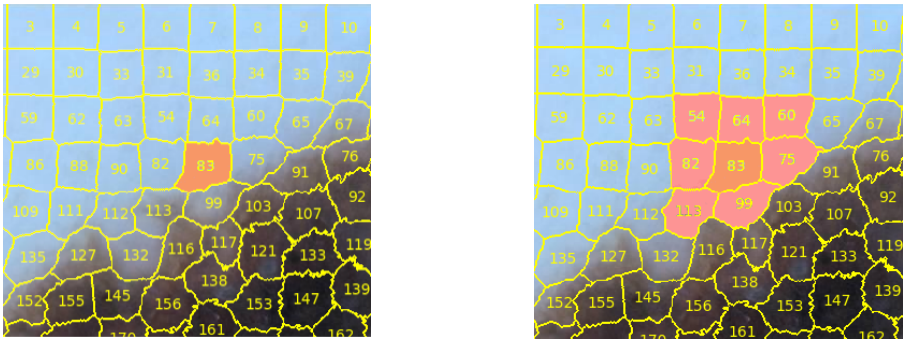
3.1. Experiments Phase One

HAM10000 dataset contains dermatoscopic images of both malignant and benign pigmented lesions, with diagnoses confirmed through histopathology, dermatologist consensus, and clinical evolution.

The features can be originated solely from basic statistics extracted from each superpixel, or from a combination of basic statistics and learned embeddings, as illustrated in part (a) of Figure 9. Additionally, feature vectors were constructed either using information from a single superpixel or from its neighboring superpixels, as shown in part (b) of Figure 9. In the latter case, the order of concatenation was determined by sorting the superpixels according to their ID numbers in ascending order. The combination of these features and their spatial origin forms the basis for constructing our input vectors.



(a) Vector Features



(b) Vector Origin: On the left is composed only from one superpixel. On the righ, involves neighboring superpixels

Figure 9. Building Vectors.

In the first phase of our experiments, we constructed four distinct types of input-output feature vectors to support different modes of training and inference:

Type I: The input vector consists of basic region descriptors mean, standard deviation, skewness, and kurtosis of the RGB channels; area; centroid; and shape moments extracted from the target superpixel. The output is the class label associated with the target superpixel.

Type II: The input vector includes the basic descriptors of the target superpixel as well as those of its neighboring superpixels. The output remains the class label of the target superpixel.

 doi:10.20944/preprints202507.1380.v1

Type III: Similar to Type II, the input concatenates descriptors of both the target and neighboring superpixels. However, the output vector extends to predict the class label for every superpixel included in the input group, encompassing both the target and its neighbors.

Type IV: This configuration consists of two sets of input-output vectors. The first input vector encodes the statistical descriptors of the neighboring superpixels, with the corresponding output vector specifying their class labels. The second input vector captures the descriptors of the target superpixel alone, with a dedicated output vector indicating its class.

The following neural architectures were used to model these input-output relationships:

Linear Neural Network: A fully connected feed-forward network with three hidden layers of sizes 128, 64, and 32, applied to the Type I and Type II feature vectors.

GRU RNN (Many-to-One): A recurrent neural network utilizing four Gated Recurrent Units (GRUs), applied to the sequential embedding of neighbor features (Type II). The final hidden state is passed to a softmax classifier for target class prediction.

Transformer Encoder–Decoder (Many-to-One): The encoder, composed of four self-attention heads, processes the sequence of target and neighbor superpixels (Type IV). A decoder with four attention heads attends to the encoded context and predicts the target superpixel's class.

Transformer Encoder (Many-to-Many): A transformer encoder block (with four self-attention heads) jointly processes the concatenated feature set from multiple superpixels (Type III). It outputs class predictions for each region simultaneously.

In our first experiment, we trained a linear Neural Network on the first vector type to predict the class of each superpixel. As shown in Table 3 the Linear Neural Network performs adequately given the limited feature set, particularly for Classes 0 and 1 with F1 Scores of 0.8333 and 0.8077 respectively. Class 1 exhibits better differentiation from other classes, with the highest accuracy (0.9091) and competitive recall (0.7915). The model performs poorly on class 2 with a F1 Score of 0.2937.

Table 3. Models Performance.

		Class 0				Class 1				Class 2			
		F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall
Linear Neural Network	Single Superpixel (S)	0.8333	0.794	0.944	0.7458	0.8077	0.9091	0.8246	0.7915	0.2937	0.794	0.1919	0.6251
	Neighbors Superpixels	0.8446	0.8087	0.9596	0.7542	0.8279	0.9185	0.8495	0.8074	0.3257	0.8011	0.2114	0.7097
RNN many(S) to one(S)	GRU	0.7985	0.7589	0.9383	0.695	0.7905	0.8941	0.763	0.8201	0.2397	0.7738	0.1559	0.5188
	Transformer Encoder Decoder	0.9377	0.9126	0.9229	0.9529	0.8828	0.9432	0.8823	0.8833	0.3572	0.9271	0.4468	0.2976
RNN many(S) to many(S)	Transformer Encoder	0.9816	0.9685	0.9753	0.988	0.9196	0.9812	0.9199	0.9193	0.4816	0.9718	0.6026	0.4011
	Transformer Encoder with Embeddings	1	1	1	1	1	1	1	1	1	1	1	1

In the second experiment, we trained a linear Neural Network on the second vector type to predict the class of the target superpixel. As depicted in Table 3 compared to the previous model, the values are higher; however, the behaviour is similar. The model is good enough for class 0 and 1, but under-performs for class 2. Class 1 is better differentiated than the other classes. These results suggest that incorporating neighboring superpixel properties enhances classification especially for class 0, and 1.

In the third experiment, we modeled the data as a sequential input of superpixels in ascending order, and trained a RNN, specifically a GRU, on the second vector type to predict the class of the target superpixel. As illustrated in Table 3 this model, compared to the previous Linear Neural Networks experiments, performs marginally lower for classes 0 and 1 implying that sequential modeling may not fully exploit the contextual dependencies in the third vector type.

In the fourth experiment, we implemented a many-to-one RNN architecture using a transformer model with encoder-decoder components, trained on the fourth vector type to predict the class of the target superpixel. The encoder processes multiple superpixels (including their class information), while the decoder focuses on the target superpixel to generate its predicted class. As portrayed in Table 3, Class 0 achieves exceptional performance, with an F1 score of 0.9377, precision of 0.9229, and recall of 0.9529, indicating highly reliable predictions. Class 1 demonstrates strong results (F1: 0.8828, precision: 0.8823) and the highest accuracy (0.9432), suggesting effective utilization of contextual

 doi:10.20944/preprints202507.1380.v1

information from neighboring superpixels. Class 2, however, continues to underperform, with an F1 score of 0.3572 and remarkably low recall (0.2976), despite moderate precision (0.4468).

In the fifth experiment, we employed a many-to-many RNN architecture, specifically a transformer model, trained on the third vector type to predict the classes of all superpixels simultaneously. As detailed in Table 3, Class 0 achieves near-perfect performance, with an F1 score of 0.9816, precision of 0.9753, and recall of 0.9880, demonstrating robust alignment between predictions and ground truth. Class 1 also exhibits strong performance (F1: 0.9196, precision: 0.9199, recall: 0.9193), with accuracy peaking at 0.9812. However, Class 2 remains a challenge, with an F1 score of 0.4816 and low recall (0.4011), despite moderate precision (0.6026). These results underscore the efficacy of many-to-many transformers in leveraging global dependencies for multi-superpixel classification tasks, though Class 2 may require targeted architectural or data-centric interventions.

3.2. Experiment Phase Two

In the second phase of our experiments, we selected the best-performing model from the initial evaluation, the transformer encoder model, as the basis for further analysis. To enhance the feature representation, we expanded the Type III input vectors by incorporating additional embeddings generated by the transformer autoencoders. This augmentation resulted in a total of 270 features per input instance.

The updated feature vectors were trained using a many-to-many transformer encoder architecture, consisting of a single encoder block with four self-attention heads. The model demonstrated perfect classification performance, achieving an accuracy and F1 score of 1.0 across all semantic classes.

To assess the robustness and generalizability of this performance, we conducted a five-fold cross-validation. The model consistently produced the same results across all folds, confirming the stability and reliability of the learned representations.

As observed in the transformer many-to-many encoder model, using input vectors without embeddings yields strong classification performance for class 0 and class 1, with F1 scores of 0.98 and 0.91, and accuracies of 0.96 and 0.97, respectively. These results suggest that simple statistical features are sufficient for accurately identifying superpixels corresponding to the background and core of the lesion. In other words, distinguishing these two classes is not particularly challenging for the model. Therefore, incorporating embeddings naturally aligns with the goal of maintaining or improving performance.

In contrast, classification of class 2 (the lesion border) proves significantly more difficult, with an F1 score of 0.48 despite a high accuracy of 0.97, showing the class imbalance and mis-classification When embeddings are added and the input is expanded to include information from neighboring superpixels, performance improves substantially.

These findings strongly suggest that the border superpixels carry the most significant information in the entire image. Moreover, the border regions may be crucial for distinguishing between benign and malignant lesions.

We acknowledge that our dataset consists of 2,000 images, and it is likely that the variation in lesion morphology across diverse skin types and racial backgrounds is limited, an inherent limitation of the dataset.

4. Discussion

At this stage of our research, achieving high performance in either lesion classification (benign vs. malignant) or segmentation is not our primary objective. For the medical community to adopt AI in clinical settings, it is essential to address several challenges related to data quality, methodological rigor, and most importantly interpretability [40].

In our study, unlike previous approaches which focus on segmentation or disease classification, the primary role of superpixels is to capture the semantic of the superpixels. Each superpixel is labeled as a background, border, or lesion, corresponding to the classes identified in the image. Superpixels from each class are used to train transformer-based autoencoders in order to produce meaningful

 doi:10.20944/preprints202507.1380.v1

embeddings that capture the context of all superpixels of each class regardless of distance between them. We proved that these embeddings capture the semantic content of the classes more effectively than relying solely on the statistical properties of the superpixels.

To enhance classification performance, we utilize not only the embedding of the target superpixel but also incorporate embeddings from its neighboring superpixels as input to the classifier. The transformer captures the context of the near superpixels (spatial locality). We consider our approach to be novel, with the main contribution being a model that learns to interpret the semantic structure of an image through superpixel-based embeddings.

We hypothesize that superpixels provide more natural and semantically meaningful context for the classification transformer compared to the fixed 16×16 square patches used in the Vision Transformer (ViT), while maintaining the same number of features [56].

We establish the originality of our methodology in several key aspects:

- Defining the border as a region instead of a well defined line.
- Generating embeddings for every superpixel using a transformer autoencoder
- Incorporating those embeddings as features for further training
- Taking into account the neighborhood to create the input vectors

However, our study has several limitations. These include the dependency on SLIC segmentation quality, the need for manual annotations, and the requirement to train the transformer autoencoder as a separate task for embedding generation.

5. Conclusion

The objective of our work is to classify superpixels extracted from natural skin images to identify background, border, and core lesion regions. A RAG is constructed from the superpixels to determine neighboring relationships for each individual superpixel. We first constructed input vectors based on the image statistics of each superpixel and used them to train linear neural networks. In our second approach, we expanded the input vectors to include features from both the superpixel and its neighbors, and trained RNNs, specifically GRUs and Transformers. Finally, we generated embeddings for each superpixel using transformer-based autoencoders, and used these embeddings as inputs to train transformer-based RNNs. The best results were achieved using the autoencoder-generated embeddings. We proved that these embeddings capture the semantic characteristics of each class more effectively than raw statistical features.

Author Contributions: Pablo Ordoñez was responsible for study design, data collection and analysis, software tool development, experiment execution, manuscript drafting, and revisions. Santiago Acosta contributed to software development, image annotation, and manuscript revisions. Issac Gutierrez participated in software tool development and image annotation. Ying Xie contributed to the study design and provided critical revisions of the manuscript. Chloe Yin Xie and Xinyue Zhang both contributed to the critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding: This research received no grant from any funding agency in public, commercial, or not-for-profit sectors.

Acknowledgments: Add any acknowledgments.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Bradford, P.T.; Goldstein, A.M.; McMaster, M.L.; Tucker, M.A. Acral Lentiginous Melanoma: Incidence and Survival Patterns in the United States, 1986-2005. *Archives of Dermatology* **2009**, *145*, 427–434. https://doi.org/10.1001/archdermatol.2008.609.
2. Becker, F.F. Lentigo Maligna and Lentigo Maligna Melanoma: Recognition and Treatment. *Archives of Otolaryngology* **1978**, *104*, 352–356. https://doi.org/10.1001/archotol.1978.00790060054013.
3. SEER. Types of melanoma, 2025.

 doi:10.20944/preprints202507.1380.v1

4. SEER. Melanoma of the Skin – Cancer Stat Facts, 2025. Online; accessed 2025-01-19.
5. ACS. Melanoma Skin Cancer Statistics American Cancer Society, 2025. [Online; accessed 2025-01-19].
6. Siegel, R.L.; D., M.K.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* **2022**, *72*, 7–33, [https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21708]. https://doi.org/https://doi.org/10.3322/caac.21708.
7. Society, A.C. Cancer Facts and Figures 2022, 2025.
8. Hassel, J.C.; Enk, A.H., Fitzpatrick's Dermatology, 9e. In *Fitzpatrick's Dermatology, 9e*; Kang, S.; Amagai, M.; Bruckner, A.L.; Enk, A.H.; Margolis, D.J.; McMichael, A.J.; Orringer, J.S., Eds.; McGraw-Hill Education: New York, NY, 2019; chapter 116.
9. Boiko, P.E.; Koepsell, T.D.; Larson, E.B.; Wagner, E.H. Skin cancer diagnosis in a primary care setting. *J. Am. Acad. of Dermato.* **1996**, *34*, 608–611. https://doi.org/10.1016/S0190-9622(96)80059-4.
10. Soyer, H.P.; Smolle, J.; Kerl, H.; Stettner, H. Early diagnosis of malignant melanoma by surface microscopy. *The Lancet* **1987**, *330*, 803.
11. Binder, M.; Schwarz, M.; Winkler, A.; Steiner, A.; Kaider, A.; Wolff, K.; Pehamberger, H. Epiluminescence Microscopy: A Useful Tool for the Diagnosis of Pigmented Skin Lesions for Formally Trained Dermatologists. *Archives of Dermatology* **1995**, *131*, 286–291. https://doi.org/10.1001/archderm.1995.01690150050011.
12. Kittler, H.; Pehamberger, H.; Wolff, K.; Binder, M. Diagnostic accuracy of dermoscopy. *The Lancet Oncol.* **2002**, *3*, 159–165. https://doi.org/10.1016/S1470-2045(02)00679-4.
13. Ahnlide, I.; Bjellerup, M.; Nilsson, F.; Nielsen, K. Validity of ABCD Rule of Dermoscopy in Clinical Practice. *Acta Derm Venereol* **2016**, *96*, 367–372.
14. Zahn, C.T.; Roskies, R.Z. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers* **1972**, *C-21*, 269–281. https://doi.org/10.1109/TC.1972.5008949.
15. Toureau, V.; Bibiloni, P.; Talavera-Martínez, L.; González-Hidalgo, M. Automatic Detection of Symmetry in Dermoscopic Images Based on Shape and Texture. In Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2020). Springer, 2020, pp. 591–603. https://doi.org/10.1007/978-3-030-50146-4_46.
16. Ali, A.M.; Li, J.; Yang, G.; O'Shea, S. A Novel Fuzzy Multilayer Perceptron (F-MLP) for the Detection of Border Irregularity in Skin Lesions. *Frontiers in Medicine* **2020**, *7*, 1–14. https://doi.org/10.3389/fmed.2020.00297.
17. Ali.; Abder-Rahman.; Jingpeng, L.; Jane, O.S. Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images. *PLOS ONE* **2020**, *15*, 1–21. https://doi.org/10.1371/journal.pone.0234352.
18. Samuel, N.E.; Anitha, J. Melanoma Detection and Classification based on Dermoscopic Images using Deep Learning Architectures-A Study. In Proceedings of the 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), 2022, pp. 993–1000. https://doi.org/10.1109/ICIRCA54612.2022.9985523.
19. Alsahafi, Y.S.; Elshora, D.S.; Mohamed, E.R.; Hosny, K.M. Multilevel threshold segmentation of skin lesions in color images using Coronavirus Optimization Algorithm. *Diagnostics (Basel)* **2023**, *13*.
20. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H.; 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), 2018, [arXiv:cs.CV/1710.05006].
21. Patel, R.H.; Foltz, E.A.; Witkowski, A.; Ludzik, J. Analysis of Artificial Intelligence-Based Approaches Applied to Non-Invasive Imaging for Early Detection of Melanoma: A Systematic Review. *Cancers* **2023**, *15*, 4694.
22. Rosas-Lara, M.; Mendoza-Tello, J.C.; Flores, A.; Zumba-Acosta, G. A Convolutional Neural Network-Based Web Prototype to Support Melanoma Skin Cancer Detection. In Proceedings of the 2022 Third International Conference on Information Systems and Software Technologies (ICI2ST), 2022, pp. 1–7. https://doi.org/10.1109/ICI2ST57350.2022.00008.
23. Subramanian, B.; Muthusamy, S.; Thangaraj, K.; Panchal, H.; Kasirajan, E.; Marimuthu, A.; Ravi, A. A new method for detection and classification of melanoma skin cancer using deep learning based transfer learning architecture models. *Research Square* **2022**. https://doi.org/10.21203/rs.3.rs-1857063/v1.
24. Sagar, A.; Jacob, D. Convolutional Neural Networks for Classifying Melanoma Images **2020**. https://doi.org/10.1101/2020.05.22.110973.
25. Gangwani, D.; Liang, Q.; Wang, S.; Zhu, X. An Empirical Study of Deep Learning Frameworks for Melanoma Cancer Detection using Transfer Learning and Data Augmentation. In Proceedings of the 2021 IEEE

 doi:10.20944/preprints202507.1380.v1

International Conference on Big Knowledge (ICBK), 2021, pp. 38–45. https://doi.org/10.1109/ICKG52313.2021.00015.
26. Cirrincione, G.; Cannata, S.; Cicceri, G.; Prinzi, F.; Currieri, T.; Lovino, M.; Militello, C.; Pasero, E.; Vitabile, S. Transformer-Based Approach to Melanoma Detection. *Sensors (14248220)* **2023**, *23*, 5677.
27. Khan, S.; Khan, A. SkinViT: A transformer based method for Melanoma and Nonmelanoma classification. *PLoS ONE* **2023**, *18*, 1–19.
28. Xie, J.; Wu, Z.; Zhu, R.; Zhu, H. Melanoma Detection based on Swin Transformer and SimAM. In Proceedings of the 2021 IEEE 5th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC), 2021, Vol. 5, pp. 1517–1521. https://doi.org/10.1109/ITNEC52019.2021.9587071.
29. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2020, [arXiv:cs.LG/1905.11946].
30. Hosny, K.M.; Elshoura, D.; Mohamed, E.R.; Vrochidou, E.; Papakostas, G.A. Deep Learning and Optimization-Based Methods for Skin Lesions Segmentation: A Review. *IEEE Access* **2023**, *11*, 85467–85488. https://doi.org/10.1109/ACCESS.2023.3303961.
31. Zafar, M.; Amin, J.; Sharif, M.; Anjum, M.A.; Mallah, G.A.; Kadry, S. DeepLabv3+-Based Segmentation and Best Features Selection Using Slime Mould Algorithm for Multi-Class Skin Lesion Classification. *Mathematics* **2023**, *11*. https://doi.org/10.3390/math11020364.
32. Chen, B.; Liu, Y.; Zhang, Z.; Lu, G.; Kong, A.W.K. TransAttUnet: Multi-Level Attention-Guided U-Net With Transformer for Medical Image Segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2024**, *8*, 55–68. https://doi.org/10.1109/tetci.2023.3309626.
33. Akyel, C.; Arıcı, N. LinkNet-B7: Noise Removal and Lesion Segmentation in Images of Skin Cancer. *Mathematics* **2022**, *10*. https://doi.org/10.3390/math10050736.
34. Shreshth, S.; Divij, G.; Anil, K.T. Detector-SegMentor Network for Skin Lesion Localization and Segmentation, 2020, [arXiv:eess.IV/2005.06550v1].
35. Zhang, L. FITrans: Skin Lesion Segmentation Based on Feature Integration and Transformer. *2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), Neural Networks, Information and Communication Engineering (NNICE), 2023 3rd International Conference on* **2023**, pp. 324–329. https://doi.org/10.1109/NNICE58320.2023.10105777.
36. Dong, Y.; Wang, L.; Li, Y. TC-Net: Dual coding network of Transformer and CNN for skin lesion segmentation. *PLOS ONE* **2022**, *17*, 1–18. https://doi.org/10.1371/journal.pone.0277578.
37. Liu, S.; Zhuang, Z.; Zheng, Y.; Kolmaniè, S. A VAN-Based Multi-Scale Cross-Attention Mechanism for Skin Lesion Segmentation Network. *IEEE Access* **2023**, *11*, 81953–81964. https://doi.org/10.1109/ACCESS.2023.3298826.
38. Wighton, P.; Lee, T.K.; Lui, H.; McLean, D.I.; Atkins, M.S. Generalizing Common Tasks in Automated Skin Lesion Diagnosis. *Trans. Info. Tech. Biomed.* **2011**, *15*, 622–629. https://doi.org/10.1109/TITB.2011.2150758.
39. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, 2018, [arXiv:cs.CV/1802.02611].
40. Jones, O.T.; Matin, R.N.; van der Schaar, M.; Prathivadi Bhayankaram, K.; Ranmuthu, C.K.I.; Islam, M.S.; Behiyat, D.; Boscott, R.; Calanzani, N.; Emery, J.; et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. *The Lancet Digital Health* **2022**, *4*, e466–e476. https://doi.org/10.1016/S2589-7500(22)00023-1.
41. Adegun, A.A.; Viriri, S.; Yousaf, M.H. A Probabilistic-Based Deep Learning Model for Skin Lesion Segmentation. *Applied Sciences* **2021**, *11*. https://doi.org/10.3390/app11073025.
42. Ünver, H.M.; Ayan, E. Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. *Diagnostics* **2019**, *9*. https://doi.org/10.3390/diagnostics9030072.
43. Mirikharaji, Z.; Abhishek, K.; Izadi, S.; Hamarneh, G. D-LEMA: Deep Learning Ensembles from Multiple Annotations – Application to Skin Lesion Segmentation, 2021, [arXiv:eess.IV/2012.07206].
44. Li, X.; Peng, B.; Hu, J.; Ma, C.; Yang, D.; Xie, Z. USL-Net: Uncertainty Self-Learning Network for Unsupervised Skin Lesion Segmentation, 2024, [arXiv:cs.CV/2309.13289].
45. Lu, W.; Gong, D.; Fu, K.; Sun, X.; Diao, W.; Liu, L. Boundarymix: Generating pseudo-training images for improving segmentation with scribble annotations. *Pattern Recognition* **2021**, *117*, 107924. https://doi.org/https://doi.org/10.1016/j.patcog.2021.107924.
46. Avelar, P.H.C.; Tavares, A.R.; da Silveira, T.L.T.; Jung, C.R.; Lamb, L.C. Superpixel Image Classification with Graph Attention Networks. In Proceedings of the 2020 33rd SIBGRAPI Conference on Graphics, Patterns

 doi:10.20944/preprints202507.1380.v1

and Images (SIBGRAPI), Los Alamitos, CA, USA, 2020; pp. 203–209. https://doi.org/10.1109/SIBGRAPI51738.2020.00035.

47. Nazir, U.; Wang, H.; Taj, M. Survey of Image Based Graph Neural Networks. *CoRR* **2021**, *abs/2106.06307*, [2106.06307].
48. Nowosad, J.; Stepinski, T.F. Extended SLIC superpixels algorithm for applications to non-imagery geospatial rasters. *International Journal of Applied Earth Observation and Geoinformation* **2022**, *112*, 102935. https://doi.org/https://doi.org/10.1016/j.jag.2022.102935.
49. Qin, W.; Wu, J.; Han, F.; Yuan, Y.; Zhao, W.; Ibragimov, B.; Gu, J.; Xing, L. Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation. *Physics in Medicine and Biology* **2018**, *63*, 095017. https://doi.org/10.1088/1361-6560/aabd19.
50. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods, 2010.
51. Gaur, U.; Manjunath, B.S. Superpixel Embedding Network. *IEEE Transactions on Image Processing* **2020**, *29*, 3199–3212. https://doi.org/10.1109/TIP.2019.2957937.
52. OpenCV team. OpenCV (Open Source Computer Vision Library). https://opencv.org/, 2023. Version 4.9.0.
53. Jaworek-Korjakowska, J.; Kleczek, P. Region Adjacency Graph Approach for Acral Melanocytic Lesion Segmentation. *Applied Sciences* **2018**, *8*. https://doi.org/10.3390/app8091430.
54. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data* **2018**, *5*, 180161. Included in ISIC 2019: https://challenge.isic-archive.com/data/, https://doi.org/10.1038/sdata.2018.161.
55. Neo4j, Inc.. Neo4j: Graph Data Platform. https://neo4j.com/, 2023. [Online; accessed 2025-01-19].
56. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR* **2020**, *abs/2010.11929*, [2010.11929].