Table of Contents

1. Introduction

Skin cancer is rapidly escalating global health concern, with melanoma recognized as its most lethal form. Biomedical imaging plays an important role in the early detection and effective treatment of melanoma, a severe type of skin cancer that typically develops on sun-exposed areas of the body (Burada et al., 2024).

Skin lesion segmentation constitutes a fundamental step in skin lesion analysis, forming the basis for subsequent classification tasks. However, it is a challenging process due to the indistinct boundaries of pigmented regions and the visual similarity between lesions and surrounding healthy skin (Liu et al., 2021). As a result, automated skin lesion analysis through image processing and artificial intelligence has emerged as a key research area in computer-aided dermatology.

The segmentation stage is particularly critical, as its accuracy directly affects the performance of later stages. Nonetheless, segmentation is difficult due to variations in lesion size, color, and shape, as well as differences in skin tone and texture. Many lesions have irregular borders, while others blend smoothly into the surrounding skin. Additional obstacles include the presence of hair covering the lesions and specular reflections on the skin surface (Zebari et.al, 2022).

In this study, K-Means clustering for skin cancer images are evaluated. The results demonstrate that K-means exhibits superior capability in segmenting skin cancer lesions. Consequently, a segmentation model is developed that operates through three primary stages: pre-processing, initial segmentation, and post-segmentation.

The model is evaluated using the open-source PH² dataset and compared against ground truth lesion masks.

2. Related Work

Image segmentation for skin lesion analysis has evolved from threshold-based and region-growing approaches to clustering and deep learning paradigms. K-Means clustering remains a cornerstone in unsupervised segmentation due to its simplicity, interpretability, and computational efficiency.

The studies from Burada et.al presents a Fuzzy C-Means (FCM) clustering algorithm has received attention due to its ability to assign pixels to multiple clusters with varying degrees of membership. This soft clustering mechanism enables better handling of noise and overlapping data, often resulting in more precise segmentation than traditional hard clustering methods like K-Means. For instance, prior works have demonstrated that FCM achieves improved robustness in cases where lesion boundaries are ambiguous or gradients are smooth. While Fuzzy C-Means demonstrates superior adaptability in managing fuzzy boundaries, it is computationally intensive and prone to slow convergence, particularly when processing high-resolution dermoscopic datasets such as PH². Its reliance on iterative membership recalculations also increases susceptibility to noise, which can distort cluster assignments in low-contrast regions. In contrast, K-Means clustering offers a more efficient alternative, providing deterministic convergence with lower computational complexity.

The studies of Hierarchical Agglomerative Clustering (HAC) presented by Ramprasad et.al represent another significant advancement in medical image segmentation. HAC constructs a tree-like (dendrogram) representation of data, where each image pixel or superpixel begins as an individual cluster and is iteratively merged with its nearest neighbor

based on similarity metrics such as color, texture, or intensity. This hierarchical structure allows multi-level segmentation, enabling the analysis of both local and global lesion features. Segmenting an image through HAC is a decision-oriented process that divides a source image into non-overlapping, homogenous regions based on similarity assessments. This methodology supports multiple image analysis domains, from medical diagnostics and pattern recognition to remote sensing and computer. However, despite these strengths, HAC's computational overhead and sensitivity to the chosen similarity metric pose challenges for large-scale or high-resolution dermoscopic images. Moreover, the lack of an objective criterion for determining the optimal dendrogram cut level often leads to inconsistent segmentation across different images.

Given these considerations, K-Means offers a strong balance between efficiency, adaptability, and robustness when applied to dermoscopic imagery. K-Means partitions image pixels into k distinct clusters based on color or intensity similarity, minimizing the within-cluster variance through an iterative optimization process. The algorithm's simplicity, convergence stability, and low computational cost make it particularly suitable for baseline segmentation and rapid prototyping of skin lesion detection systems.

3. Methodology

In this section, the researchers presented the details and results of the segmentation method. Image pre-processing was done to enhance the image quality.

The RGB image is first converted to grayscale, followed by histogram equalization to improve contrast and median filtering to reduce noise. Morphological bottom-hat filtering (also as known as black hat filtering) is then used to remove hair artifacts and unwanted void.

Next, the parameter tuning and elbow method is done and evaluated before K-means clustering. The binary conversion is done to obtain the ground truth images. Morphological operations such as dilation, erosion, opening, and closing are subsequently applied to refine the segmented regions and enhance the accuracy of the results.

The performance evaluation is conducted by calculating the Intersection over Union (IoU), performing overlay mask comparisons, and analyzing the confusion matrix. Additionally, key performance metrics such as accuracy, sensitivity, specificity, Jaccard index, and Dice coefficient are computed to assess the effectiveness of the segmentation method.

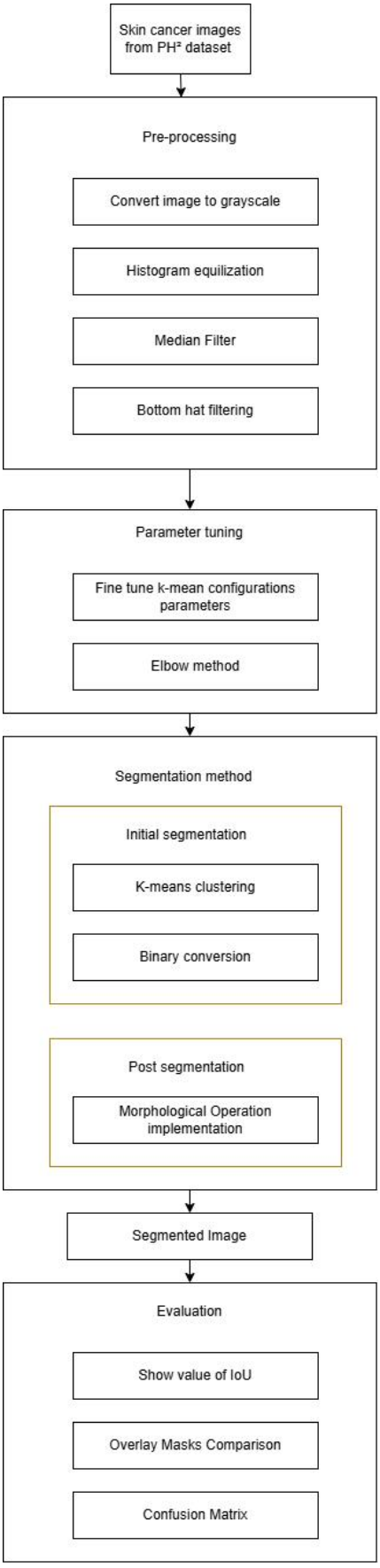


Figure 1. Framework of the Segmentation Method

3.1. Dataset

The PH² database includes the manual segmentation, the clinical diagnosis, and the identification of several dermoscopic structures, performed by expert dermatologists (Mendonça et.al, 2013). 6 out of the collection of 437 images from the dataset was chosen to be included. The images are categories into easy, moderate and difficult. 2 images are chosen for each category.

Category	Description	Initial image chosen
Easy	Clear edge and colour contrast	IMD018, IMD019
Moderate	Minor hair, some colour variation on lesion	IMD002, IMD022
Difficult	Similar colour with skin with hair artifacts and irregular edge	IMD003, IMD006

Table 1. Description of Image Categories Based on Segmentation Difficulty

3.2. Preprocessing

Image enhancement approaches such as greyscale conversion, histogram equalization, median filter and bottom hat filtering are used to prepare for the images for further processing.

The RGB input image was first converted into a grayscale image using the OpenCV function. This conversion simplifies the image by reducing the three-color channels (Blue, Green, and Red) into a single intensity channel, which helps reduce computational complexity for further processing. Both the original and grayscale images were then displayed side by side using Matplotlib to visually compare the difference in color representation and intensity distribution.

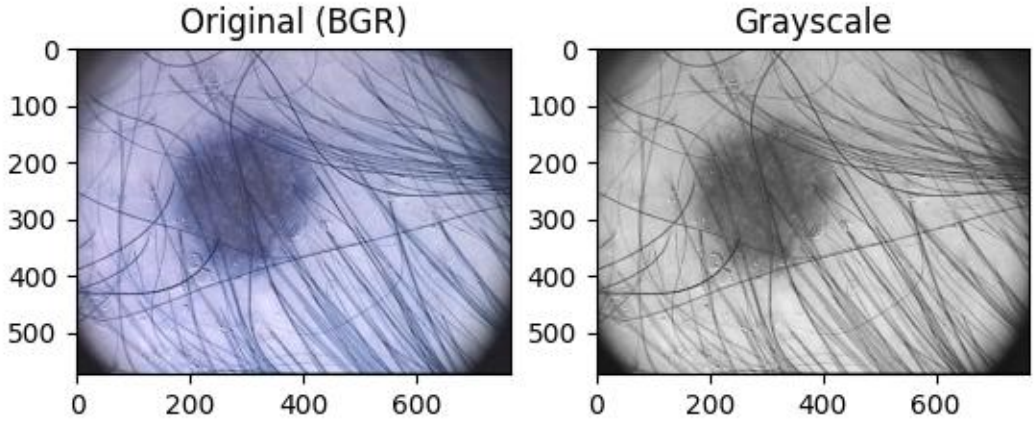


Figure 1. A sample of BGR image and grayscale image.

Histogram equalization was applied to the grayscale image using the OpenCV function cv2.equalizeHist(gray_img) to enhance image contrast. This process redistributes the intensity values of pixels, making darker areas lighter and improving the visibility of details within the image. The original grayscale image and the histogram-equalized image were displayed side by side for comparison. Additionally, their corresponding histograms were plotted to visualize the difference in intensity distribution, showing that the equalized image achieves a more uniform spread of pixel intensities across the grayscale range.

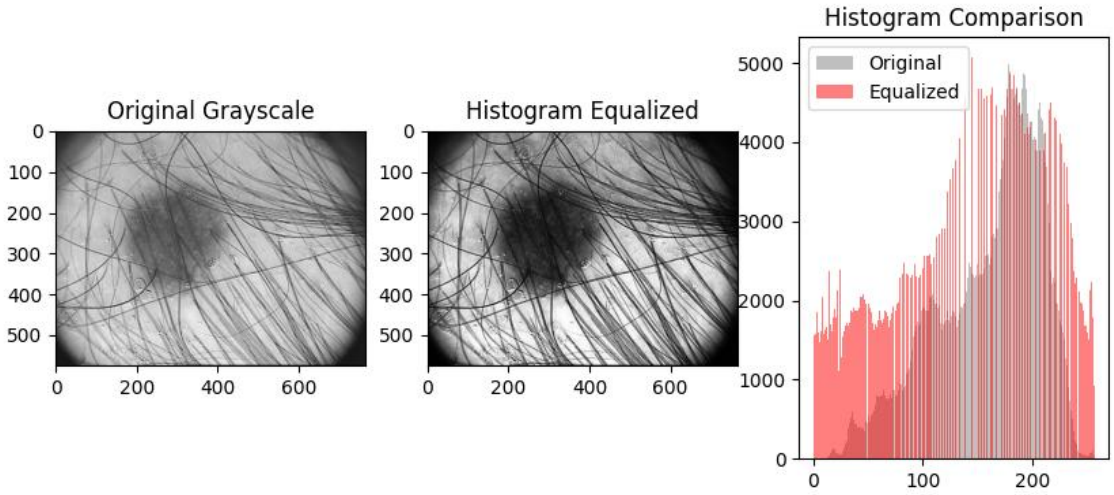


Figure 2. Comparison of the original and histogram-equalized images

In this step, morphological black-hat filtering was applied to remove hair artifacts from the skin lesion images. A rectangular structuring element of size 9×9 was used to detect dark hair strands against the lighter background of the skin. The operation was performed using the OpenCV function cv2.morphologyEx() with the cv2.MORPH_BLACKHAT parameter, which highlights dark regions smaller than the structuring element.

Next, a binary threshold was applied to create a mask that isolates the detected hair pixels. This mask was generated using the cv2.threshold() function, where pixel values above the threshold of 10 were set to 255 (white), forming the hair mask.

Finally, the inpainting technique was applied with the Telea algorithm (Telea, 2004) to reconstruct the regions where hair was detected. This step effectively fills in the removed areas with neighbouring pixel information, resulting in a smooth and clean image.

Figure 3. Original image, hair mask generated by black-hat filtering, and the result after hair removal

3.3. Parameter Tuning

Parameter tuning for the K-means clustering algorithm is implemented to achieve optimal segmentation results for skin lesion images. Three categories of the images, which are easy, moderate, and difficult were used to ensure the tuning process captures varying segmentation challenges.

A grid search is then carried out across different K-means parameters, including initialization methods (random and K-means++), maximum iterations, epsilon values, and the

number of attempts. For each combination, the algorithm evaluates clustering quality using compactness, which measures how tightly the pixels are grouped within clusters. The configuration with the lowest average compactness across all images is selected as the best parameter set for accurate and consistent lesion segmentation.

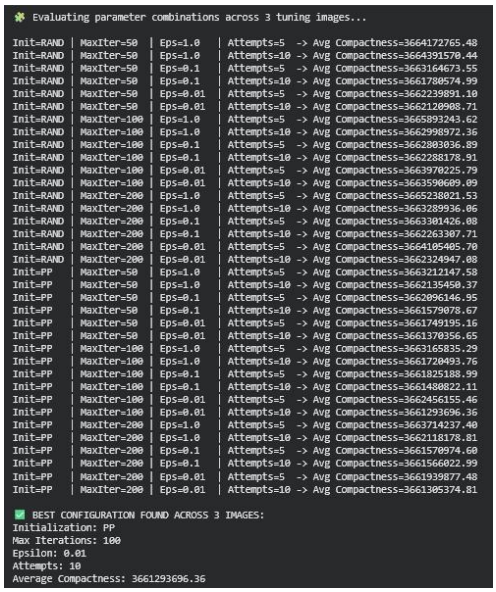


Figure 4. Result of Parameter Tuning

The Elbow Method was used to determine the optimal number of clusters (*K*) for K-Means segmentation. The grayscale image was reshaped into a one-dimensional array of pixel intensities, and K-Means clustering was applied for values of *K* ranging from 1 to 9.

For each *K*, the corresponding distortion was calculated and plotted to form an Elbow Curve. As *K* increases, distortion decreases, but the improvement becomes marginal after a certain point which indicates the optimal *K*.

To automate this selection, the perpendicular distance of each point on the curve from the straight line connecting the first and last points was computed using the formula

$$d_i = \frac{|(y_2 - y_1)x_i - (x_2 - x_1)y_i + x_2y_1 - y_2x_1|}{\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}}$$

where (x_i, y_i)represents each point on the curve. The optimal number of clusters was then determined as

$$K_{\text{optimal}} = \arg\max_i (d_i),$$

which corresponds to the point with the maximum distance, representing the best balance between segmentation accuracy and simplicity.

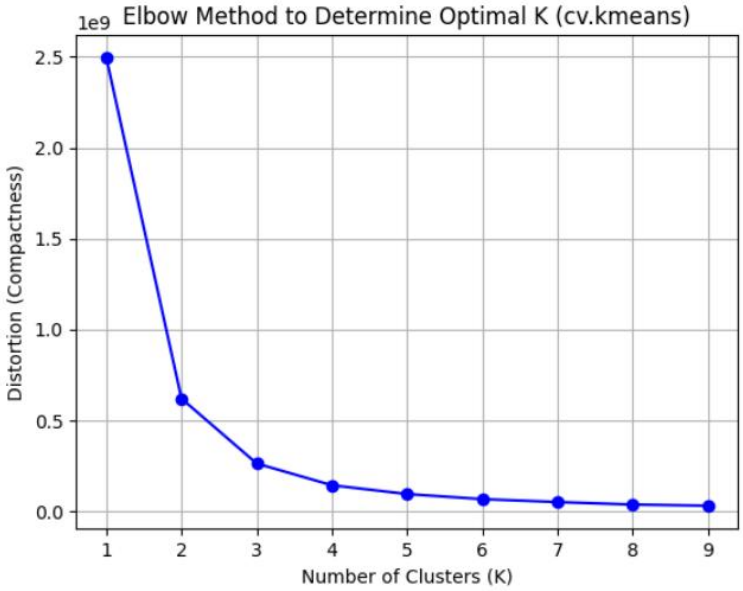


Figure 5. Elbow Method for Optimal Cluster Determinatio

3.4. Segmentation Method

The K-Means clustering algorithm was applied to segment the inpainted grayscale image into distinct regions based on intensity similarity. The image was reshaped into a one-dimensional array and converted to a 32-bit floating-point format before clustering.

Using the optimal number of clusters (k) determined from the Elbow Method, K-Means was executed with a stopping criterion combining a maximum of 100 iterations and a convergence threshold of 0.01. Each pixel was assigned a cluster label corresponding to its nearest cluster center, and the resulting labels were reshaped back into the original image dimensions to produce the segmented output.

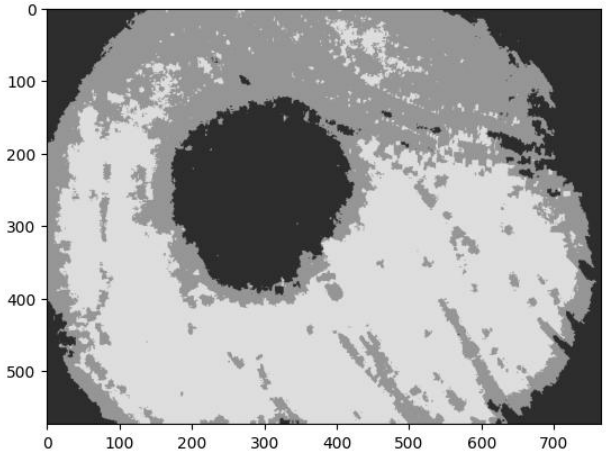


Figure 6. K-Means Segmentation Output

The cluster labels obtained from the K-means algorithm were reshaped to match the original image dimensions, producing a preliminary segmentation map. Depending on the optimal number of clusters (K), different post-processing steps were applied.

When $K = 3$ or above, the mean intensity of each cluster was calculated, and the two brightest clusters were merged to represent the lesion region. For $K = 2$, only the brighter cluster was selected as the lesion area. The resulting binary mask was then inverted so that the lesion appeared in white against a black background.

To enhance segmentation accuracy, a connected component analysis was performed to retain only the largest white region, which corresponds to the main lesion area while removing smaller noise components. The final binary mask effectively highlights the lesion boundary, providing a clear and refined result for further evaluation and analysis.

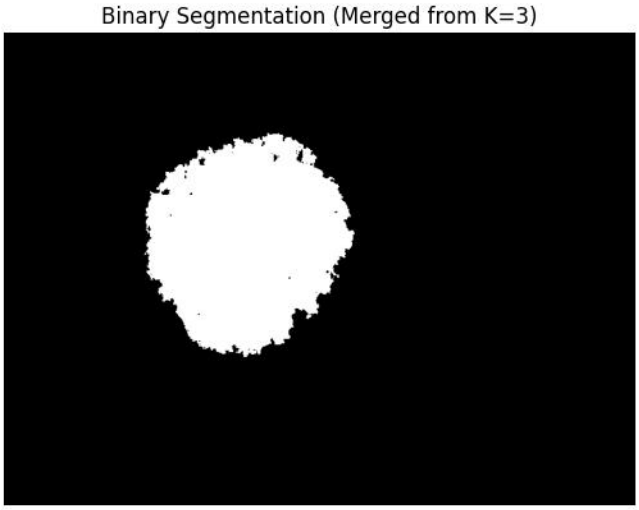


Figure 7. Image of Binary Segmentation

3.5. Evaluation

In this stage, the segmented lesion mask obtained from the K-means clustering process was evaluated against the ground truth mask using the Intersection over Union (IoU) metric.

A series of morphological operations such as erosion, dilation, opening, and closing were applied using different kernel sizes (3×3 to 11×11) to refine the binary mask. For each operation and kernel size, the IoU was calculated to measure segmentation accuracy.

The combination that achieved the highest IoU value was selected as the optimal post-processing configuration. The results were visualized by comparing the ground truth, the initial K-means segmentation, the best-performing morphological result, and the IoU difference map, which highlights mismatched regions between the predicted and actual lesion boundaries.

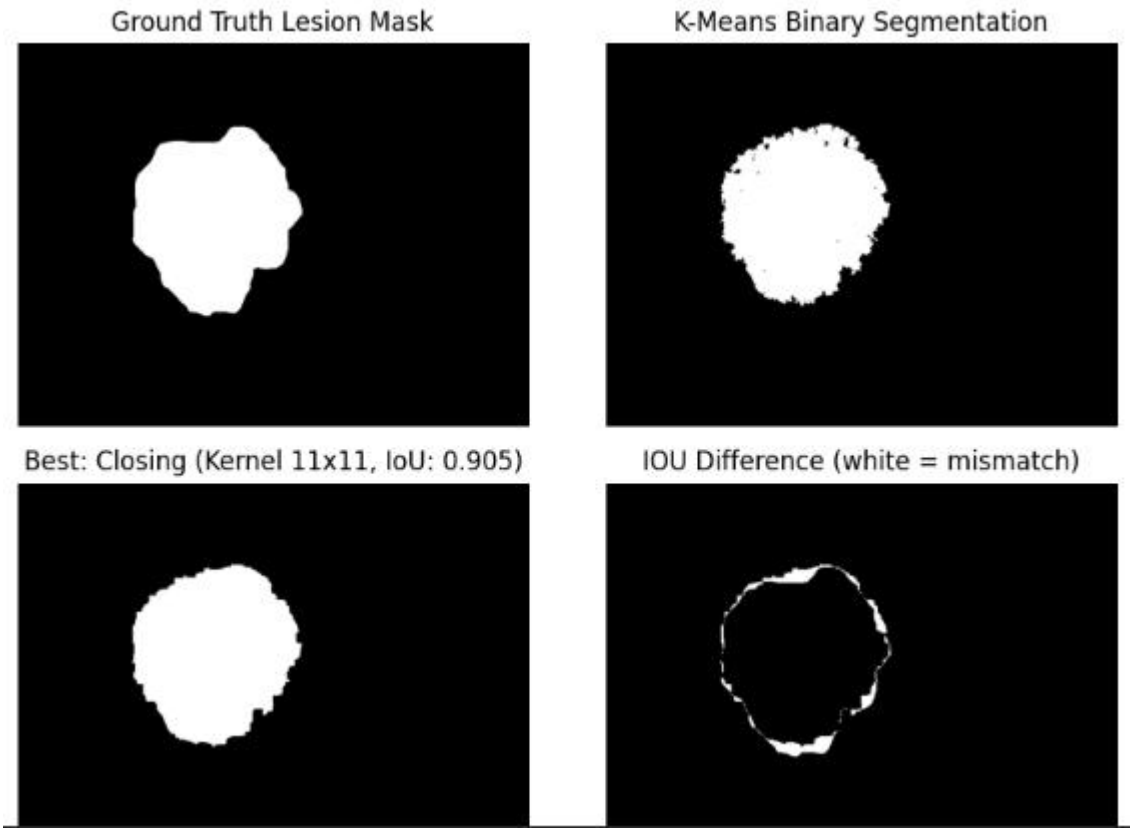


Figure 8. Comparison between ground truth, K-means segmentation, and best morphological result based on IoU.

To visually evaluate the segmentation performance, the ground truth mask and the predicted lesion mask were each overlaid on the original dermoscopic image. The original image was first converted from BGR to RGB format for correct color representation. Both the ground truth and the predicted masks were resized to match the original image dimensions to ensure accurate alignment.

Color-coded overlays were then created for clear visual comparison: the ground truth lesion area was highlighted in green, while the predicted lesion from the optimized segmentation result was shown in red. A semi-transparent blending technique was applied using an alpha value of 0.5, allowing the lesion areas to be visible over the skin background. This visual overlay enables a straightforward comparison between the true and predicted lesion boundaries, helping to assess segmentation accuracy and boundary precision.

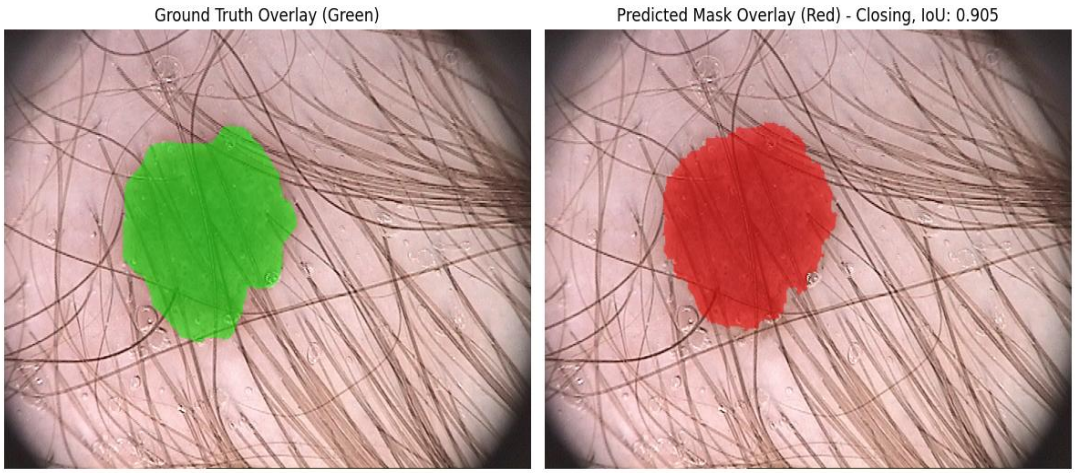


Figure 9. Visualization of ground truth (green) and predicted lesion mask (red) overlaid on the original image for comparison.

To evaluate segmentation performance, morphological operations were applied to refine the binary mask generated from the K-Means segmentation. The optimal operation: erosion, dilation, opening, or closing was selected based on the best-performing combination of operation type and kernel size (*best_k*).

A square kernel of size (*best_k* × *best_k*) was applied to the binary mask using the chosen operation to produce the final segmented result. Both the predicted mask and the ground truth image were converted to binary format for comparison, and the confusion matrix components which are, True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FNFrom these values, the performance metrics were derived using the following formulas:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Jaccard Index} = \frac{TP}{TP + FP + FN}$$

$$\text{Dice Coefficient} = \frac{2TP}{2TP + FP + FN}$$

These metrics provided quantitative insight into the segmentation quality, with accuracy measuring the overall correctness, sensitivity reflecting lesion detection capability, specificity indicating background classification accuracy, and both Jaccard and Dice coefficients assessing the similarity between the predicted mask and the ground truth. The results were

further visualized using a confusion matrix heatmap to illustrate the classification performance of lesion and background regions.

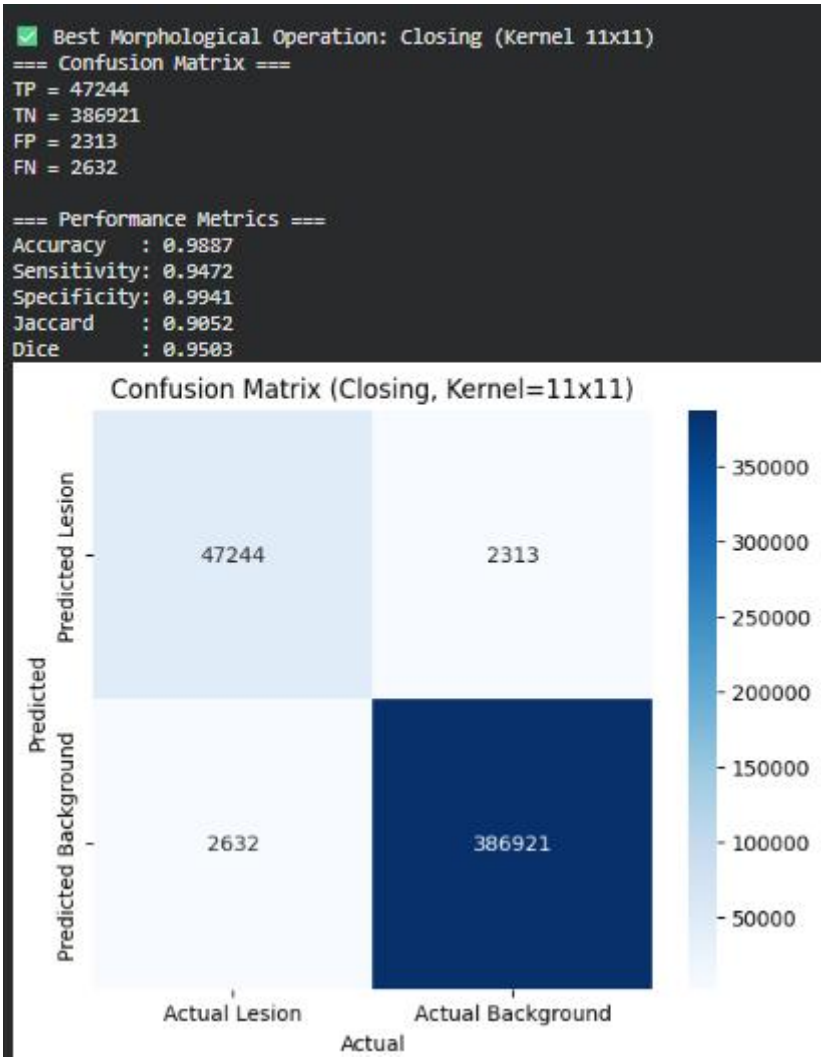
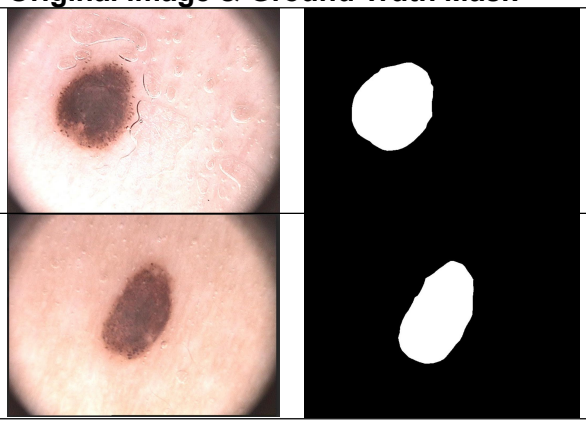
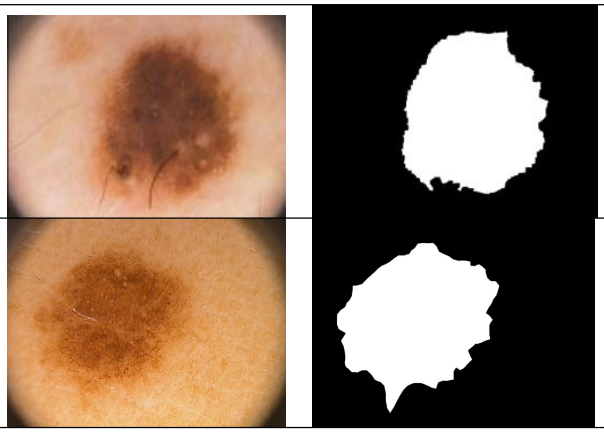
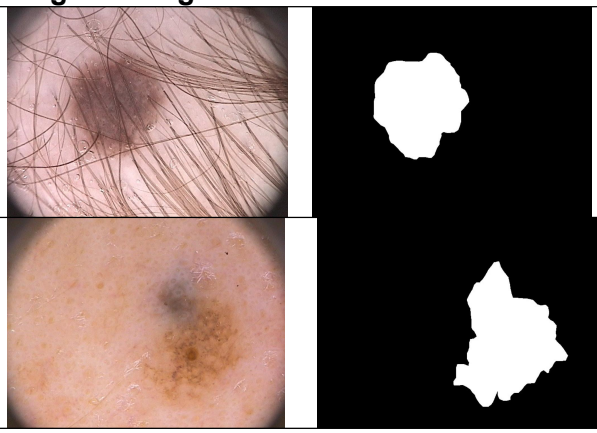


Figure 10. Confusion Matrix Heatmap for the Best Morphological Operation

4. Experimental Results

The experimental results show that K-Means- segmentation effectively identified skin lesion regions with clear boundary separation. The optimal cluster number was determined using the Elbow Method, while morphological operations refined the segmented output. Quantitative evaluation using Accuracy, Sensitivity, Specificity, Jaccard Index, and Dice Coefficient demonstrated high segmentation performance, supported by the confusion matrix visualization showing strong agreement between the predicted and actual lesion areas.

Category: Easy	**Original image & Ground Truth Mask**
IMD019	
IMD018	
Category: Moderate	**Original image & Ground Truth Mask**
IMD002	
IMD022	
Category: Difficult	**Original image & Ground Truth Mask**
IMD003	
IMD006	

4.1 Experimental Results for Easy Category Images

The easy category is defined by images that exhibit clear lesion edges and strong color contrast between the lesion and surrounding skin, making segmentation straightforward and less prone to misclassification. IMD019 and IMD018 are considered as easy category.

4.1.1 Performance Evaluation of IMD018 (Category Easy)

For image IMD018, the Elbow Method determined the optimal number of clusters as K = 3, indicating three distinct intensity regions within the lesion.

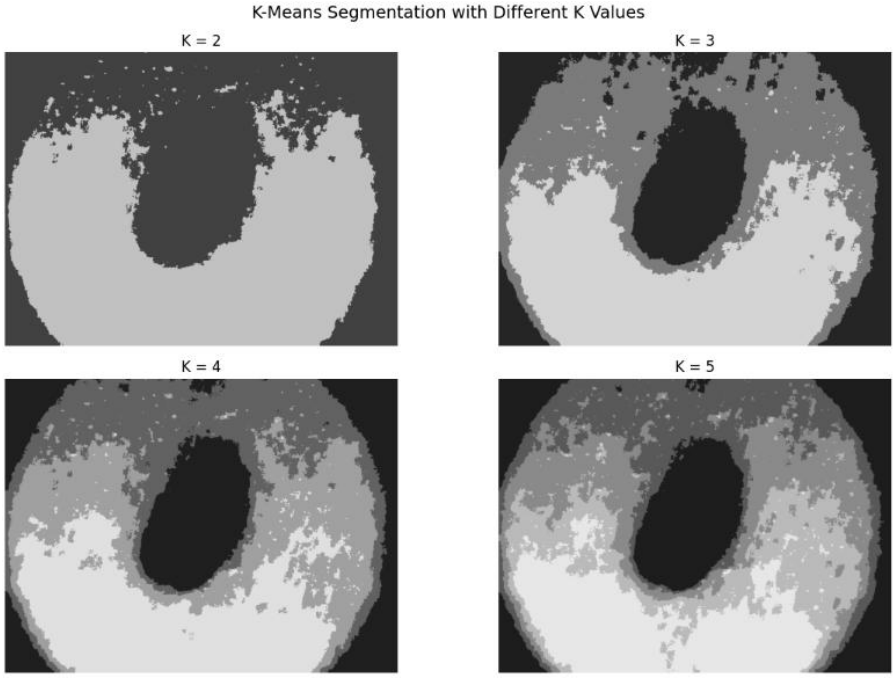


Figure 11. K-Means Segmentation with Different K-values for IMD018

The evaluation was carried out using various morphological operations to enhance segmentation accuracy.

Among these, erosion with a kernel size of 11 × 11 produced the most accurate result, achieving the highest Intersection over Union (IoU) score of 0.8560 for IMD018

```
IoU results for different kernel sizes:
Kernel 3x3 - Erosion: IoU = 0.8049
Kernel 3x3 - Dilation: IoU = 0.7563
Kernel 3x3 - Opening: IoU = 0.7832
Kernel 3x3 - Closing: IoU = 0.7756
Kernel 5x5 - Erosion: IoU = 0.8239
Kernel 5x5 - Dilation: IoU = 0.7379
Kernel 5x5 - Opening: IoU = 0.7883
Kernel 5x5 - Closing: IoU = 0.7730
Kernel 7x7 - Erosion: IoU = 0.8375
Kernel 7x7 - Dilation: IoU = 0.7209
Kernel 7x7 - Opening: IoU = 0.7936
Kernel 7x7 - Closing: IoU = 0.7714
Kernel 9x9 - Erosion: IoU = 0.8480
Kernel 9x9 - Dilation: IoU = 0.7047
Kernel 9x9 - Opening: IoU = 0.7971
Kernel 9x9 - Closing: IoU = 0.7703
Kernel 11x11 - Erosion: IoU = 0.8560
Kernel 11x11 - Dilation: IoU = 0.6892
Kernel 11x11 - Opening: IoU = 0.7997
Kernel 11x11 - Closing: IoU = 0.7687

✅ Best Kernel Size: 11x11
✅ Best Operation: Erosion
✅ Highest IoU: 0.8560
```

Figure 12. Output of Best Morphological Operation and Kernel Size Indicating the Highest IoU Value for IMD018

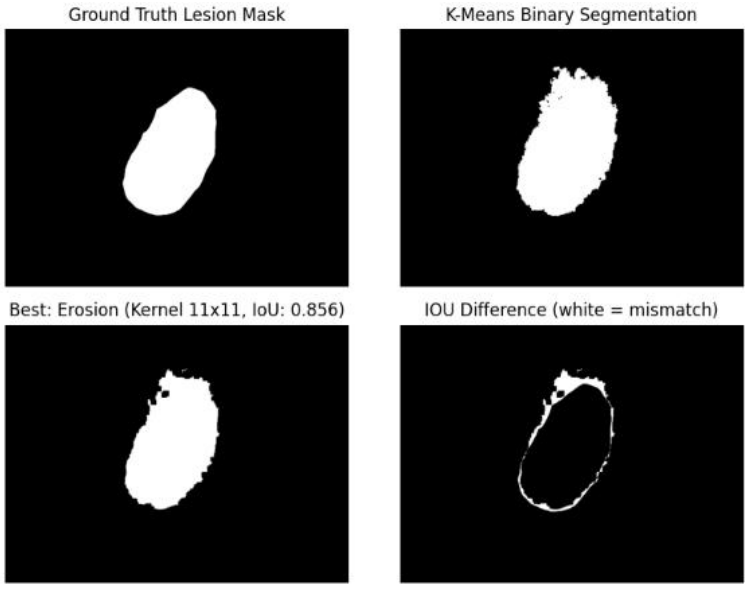


Figure 13. Comparison of Morphological Operations Showing the Best Operation and IoU Differences IMD018

The overlay with the dermoscopic image, shown in red and green, allows for a clear visual comparison between the predicted segmentation mask and the ground truth. The green region represents the actual lesion area from the ground truth, while the red region highlights the predicted segmentation.

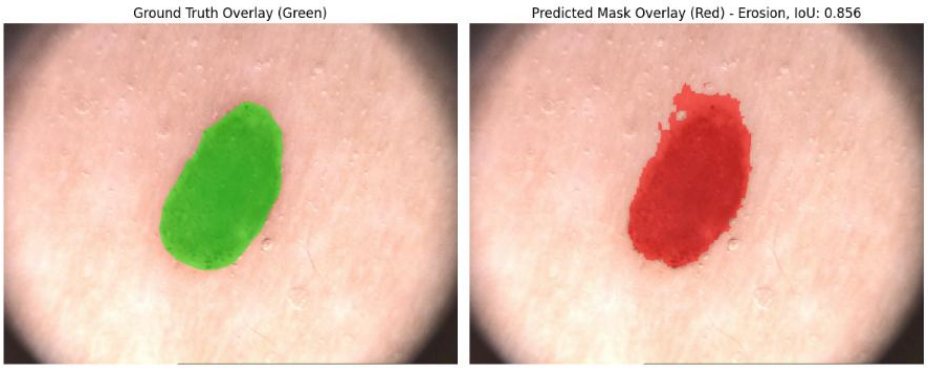


Figure 14. Overlay comparison between ground truth (green) and predicted segmentation (red) on IMD018.

For image IMD018, the segmentation process achieved optimal performance using the Erosion morphological operation with a kernel size of 11 × 11. This configuration provided the most accurate refinement of the lesion mask, effectively preserving lesion boundaries while minimizing noise and over-segmentation. The resulting confusion matrix recorded TP = 38,307, TN = 393,018, FP = 5,528, and FN = 1,492. From these values, the calculated performance metrics were Accuracy = 0.9840, Sensitivity = 0.9625, Specificity = 0.9861, Jaccard Index = 0.8451, and Dice Coefficient = 0.9161. These results indicate excellent segmentation performance, demonstrating strong agreement between the

predicted mask and the ground truth, with the erosion operation effectively enhancing lesion region precision.

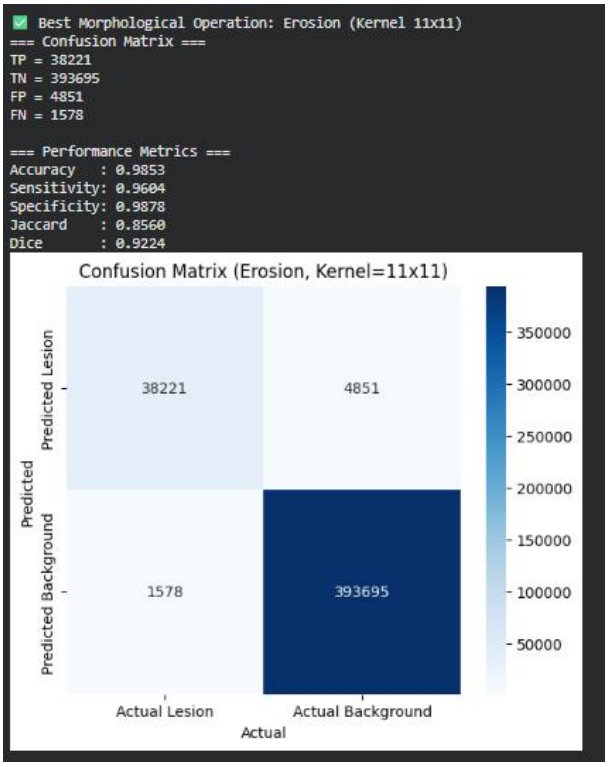


Figure 15. Confusion Matrix and Performance matrix for IMD018

4.1.2 Performance Evaluation of IMD019 (Category Easy)

Similar to image IMD018, image IMD019 also obtained the best IoU value using a kernel size of 3, suggesting that smaller kernels are effective for images with clearer lesion boundaries

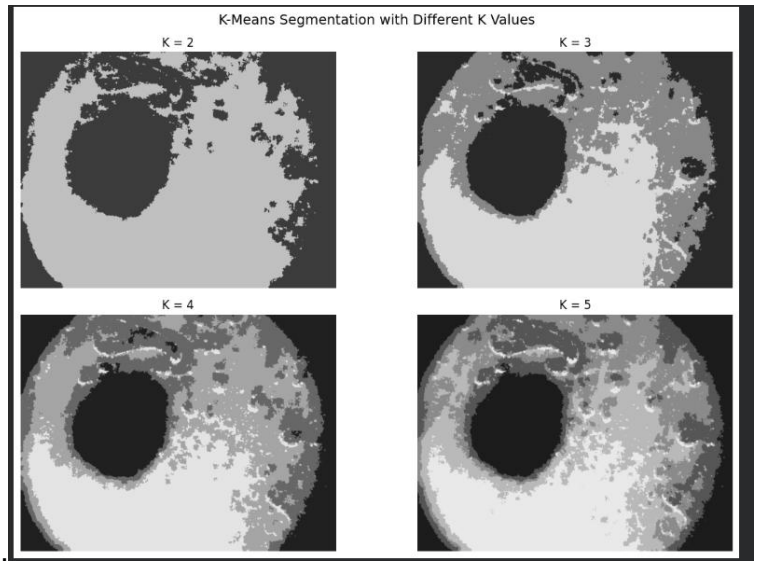


Figure 16. K-Means Segmentation with Different K-values for IMD019

Among these operations, erosion with a kernel size of 11 × 11, which is the same as IMD018 has produced the most accurate result, achieving the highest IoU score of 0.8748 for IMD019. The similarity in optimal kernel size suggests that both images share comparable lesion characteristics, such as edge sharpness and region size, allowing the same morphological parameters to yield effective segmentation.

```
IoU results for different kernel sizes:
Kernel 3x3 - Erosion: IoU = 0.8311
Kernel 3x3 - Dilation: IoU = 0.7846
Kernel 3x3 - Opening: IoU = 0.8108
Kernel 3x3 - Closing: IoU = 0.8050
Kernel 5x5 - Erosion: IoU = 0.8496
Kernel 5x5 - Dilation: IoU = 0.7650
Kernel 5x5 - Opening: IoU = 0.8146
Kernel 5x5 - Closing: IoU = 0.8020
Kernel 7x7 - Erosion: IoU = 0.8632
Kernel 7x7 - Dilation: IoU = 0.7478
Kernel 7x7 - Opening: IoU = 0.8167
Kernel 7x7 - Closing: IoU = 0.7990
Kernel 9x9 - Erosion: IoU = 0.8713
Kernel 9x9 - Dilation: IoU = 0.7320
Kernel 9x9 - Opening: IoU = 0.8205
Kernel 9x9 - Closing: IoU = 0.7976
Kernel 11x11 - Erosion: IoU = 0.8748
Kernel 11x11 - Dilation: IoU = 0.7170
Kernel 11x11 - Opening: IoU = 0.8260
Kernel 11x11 - Closing: IoU = 0.7959

✅ Best Kernel Size: 11x11
✅ Best Operation: Erosion
✅ Highest IoU: 0.8748
```

Figure 17. Output of Best Morphological Operation and Kernel Size Indicating the Highest IoU Value for IMD019

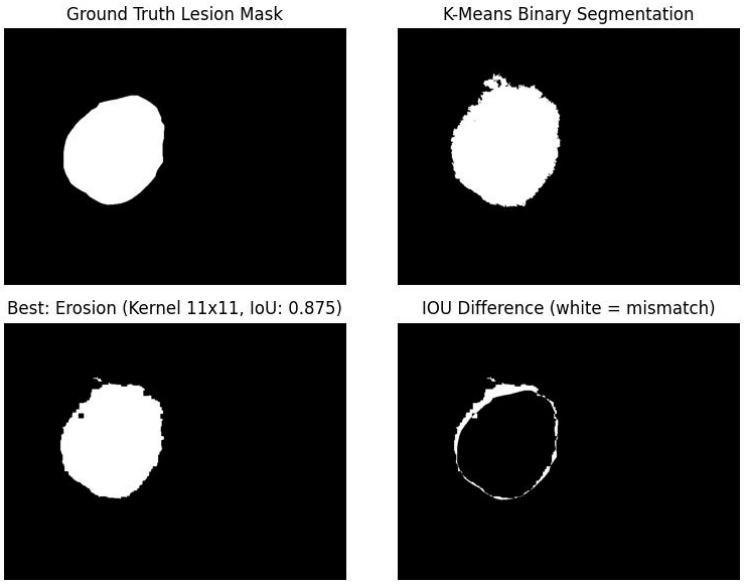


Figure 18. Comparison of Morphological Operations Showing the Best Operation and IoU Differences IMD019

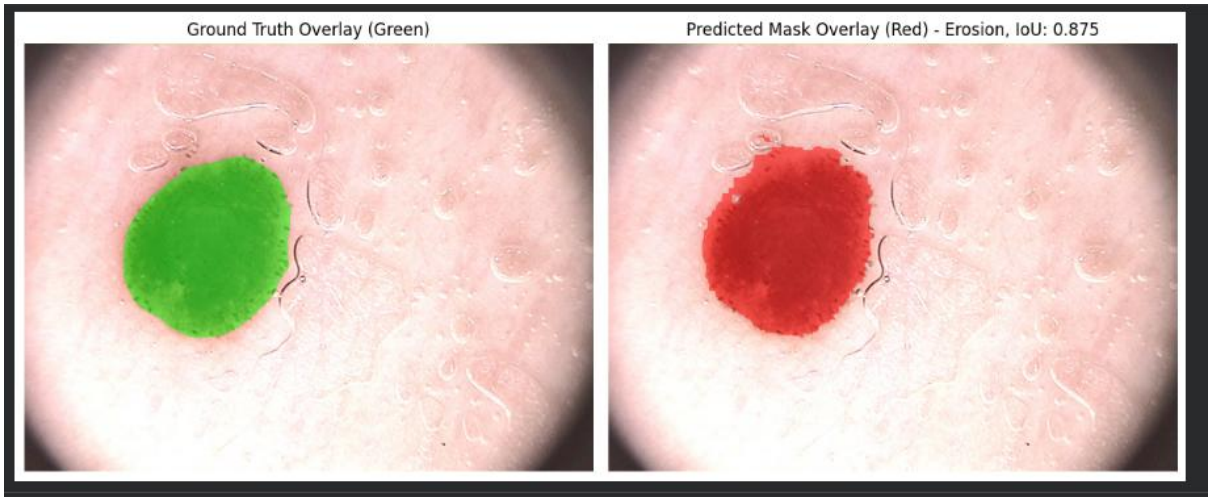


Figure 19. Overlay comparison between ground truth (green) and predicted segmentation (red) on IMD019

For image IMD019, the best-performing morphological operation was erosion with a kernel size of 11 × 11. The confusion matrix indicates 40,175 true positives, 392,613 true negatives, 4,258 false positives, and 1,490 false negatives. Based on these values, the model achieved an accuracy of 0.9869, sensitivity of 0.9642, and specificity of 0.9893, reflecting highly reliable segmentation performance. The Jaccard index of 0.8748 and Dice coefficient of 0.9332 further demonstrate strong overlap between the predicted and ground truth lesion regions. The effectiveness of erosion in this case suggests it helped eliminate minor artifacts and isolate the lesion area more precisely.

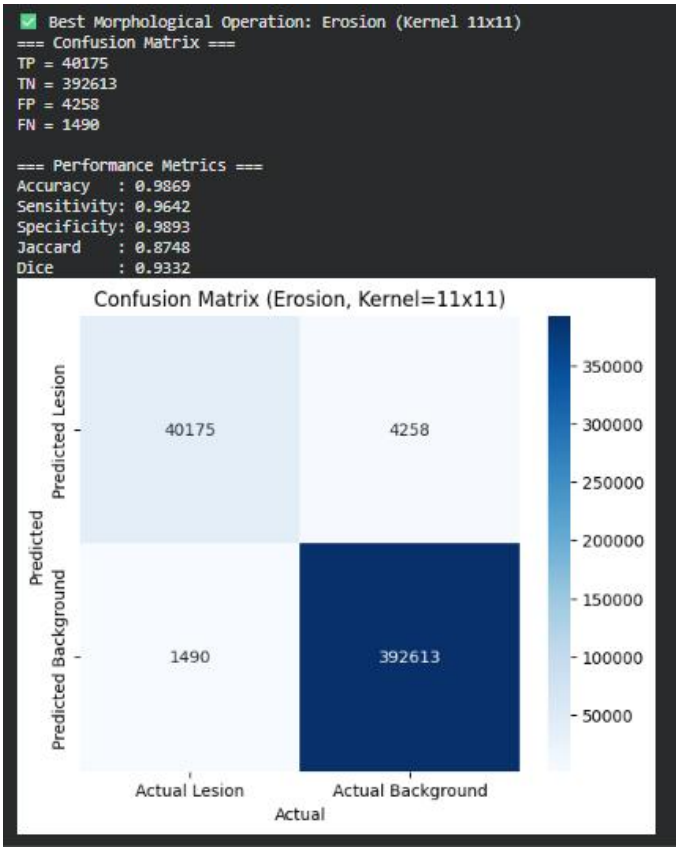


Figure 20. Confusion Matrix and Performance matrix for IMD019

4.2 Experimental Results for Moderate Category Images

The moderate category has a blurred edge and slight color variation between the lesion and the surrounding skin. IMD002 and IMD022 are the examples of moderate category.

4.2.1 Performance Evaluation of IMD002 (Category Moderate)

The moderate image showed a different segmentation result, where using K=3 gave the best outcome by clearly separating the lesion area from the surrounding skin.

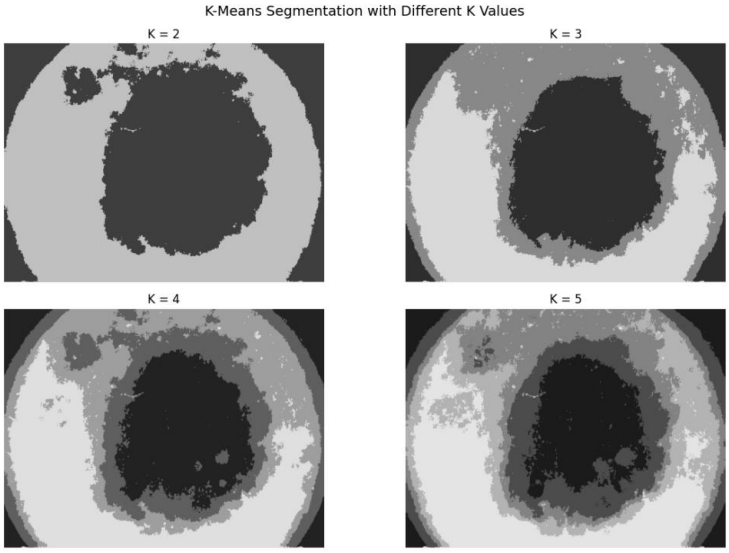


Figure 21. K-Means Segmentation with Different K-values for IMD002

For IMD002, the best result was achieved using the closing morphological operation with a kernel size of 11 × 11, which produced the highest IoU value of 0.8976 among all tested combinations.

```
IoU results for different kernel sizes:
Kernel 3x3 - Erosion: IoU = 0.8776
Kernel 3x3 - Dilation: IoU = 0.8954
Kernel 3x3 - Opening: IoU = 0.8889
Kernel 3x3 - Closing: IoU = 0.8915
Kernel 5x5 - Erosion: IoU = 0.8637
Kernel 5x5 - Dilation: IoU = 0.8962
Kernel 5x5 - Opening: IoU = 0.8865
Kernel 5x5 - Closing: IoU = 0.8938
Kernel 7x7 - Erosion: IoU = 0.8502
Kernel 7x7 - Dilation: IoU = 0.8942
Kernel 7x7 - Opening: IoU = 0.8832
Kernel 7x7 - Closing: IoU = 0.8950
Kernel 9x9 - Erosion: IoU = 0.8368
Kernel 9x9 - Dilation: IoU = 0.8900
Kernel 9x9 - Opening: IoU = 0.8802
Kernel 9x9 - Closing: IoU = 0.8968
Kernel 11x11 - Erosion: IoU = 0.8238
Kernel 11x11 - Dilation: IoU = 0.8840
Kernel 11x11 - Opening: IoU = 0.8767
Kernel 11x11 - Closing: IoU = 0.8976

✅ Best Kernel Size: 11x11
✅ Best Operation: Closing
✅ Highest IoU: 0.8976
```

Figure 22. Output of Best Morphological Operation and Kernel Size Indicating the Highest IoU Value for IMD002

The results show that applying the **closing** operation produced a smoother lesion edge and effectively filled small gaps within the segmented region.

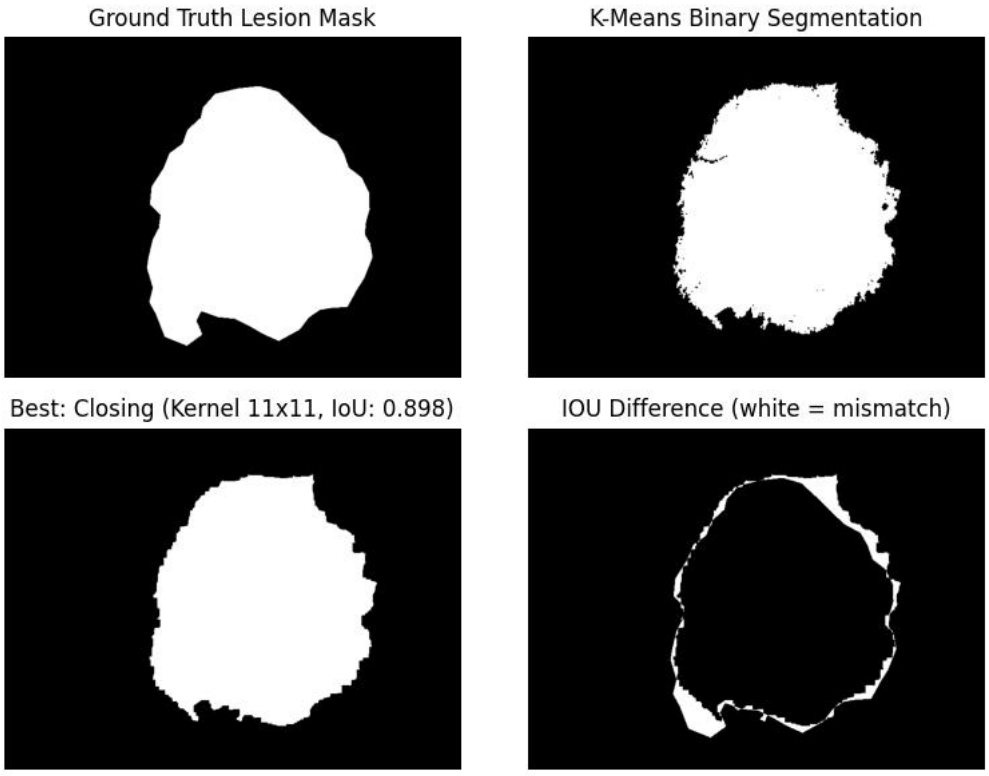


Figure 23. Comparison of Morphological Operations Showing the Best Operation and IoU Differences IMD002

The red overlay clearly shows that the presence of hair does not affect the segmentation result, indicating that the hair removal process was effective.

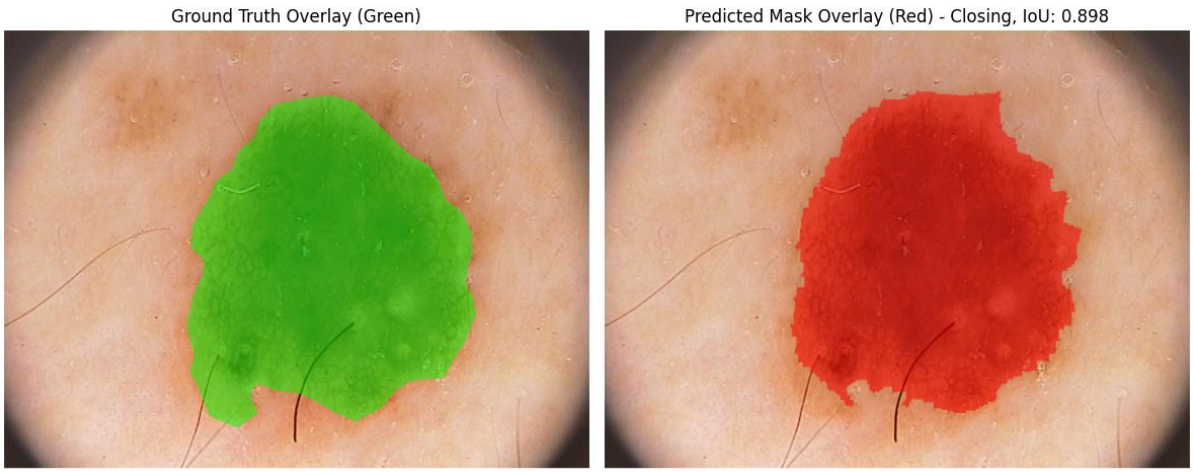


Figure 24. Overlay comparison between ground truth (green) and predicted segmentation (red) on IMD002

The confusion matrix and performance metrics were evaluated for the image IMD002. The results show 11,318 true positives, 311,220 true negatives, 6,084 false positives, and 6,858 false negatives. Based on these values, the segmentation achieved an accuracy of 0.9704, sensitivity of 0.9430, and specificity of 0.9808, indicating strong detection capability.

Additionally, the Jaccard index of 0.8976 and Dice coefficient of 0.9460 demonstrate a high level of overlap between the predicted segmentation and the ground truth mask.

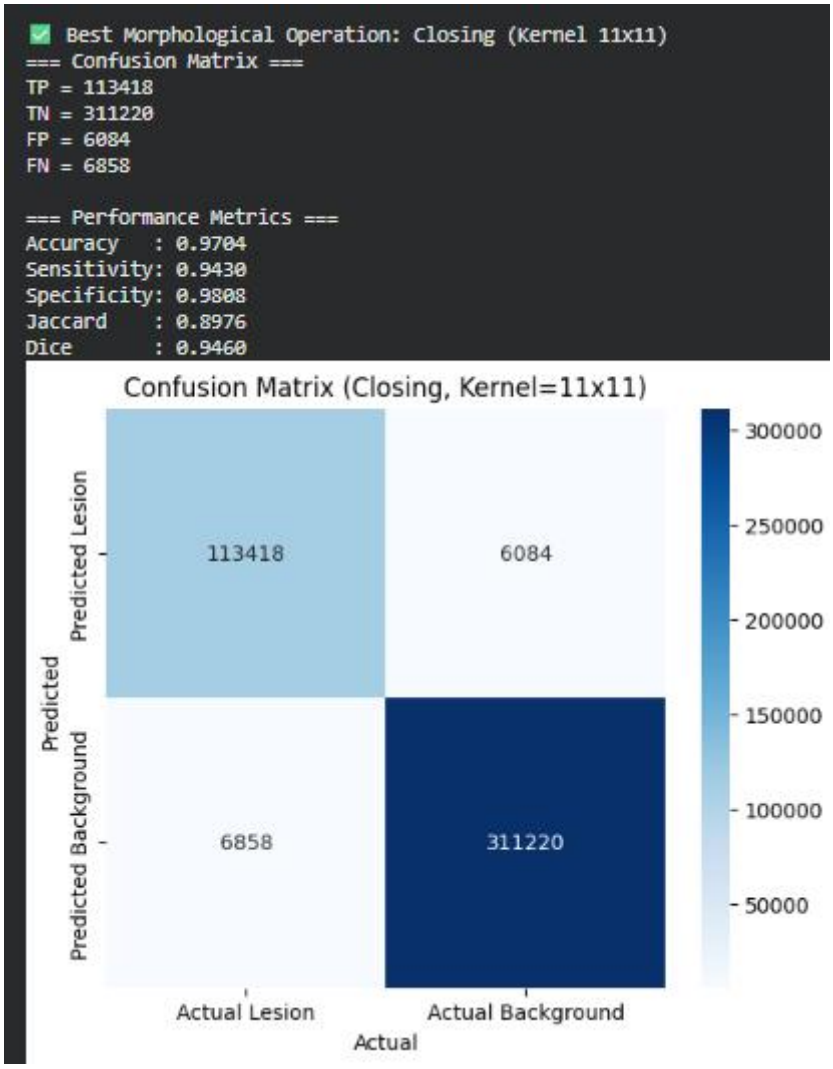


Figure 25. Confusion Matrix and Performance matrix for IMD002

4.2.2 Performance Evaluation of IMD022 (Category Moderate)

For the moderate image IMD022, the segmentation produced a distinct result, with K=3 providing the best outcome by clearly distinguishing the lesion region from the surrounding skin.

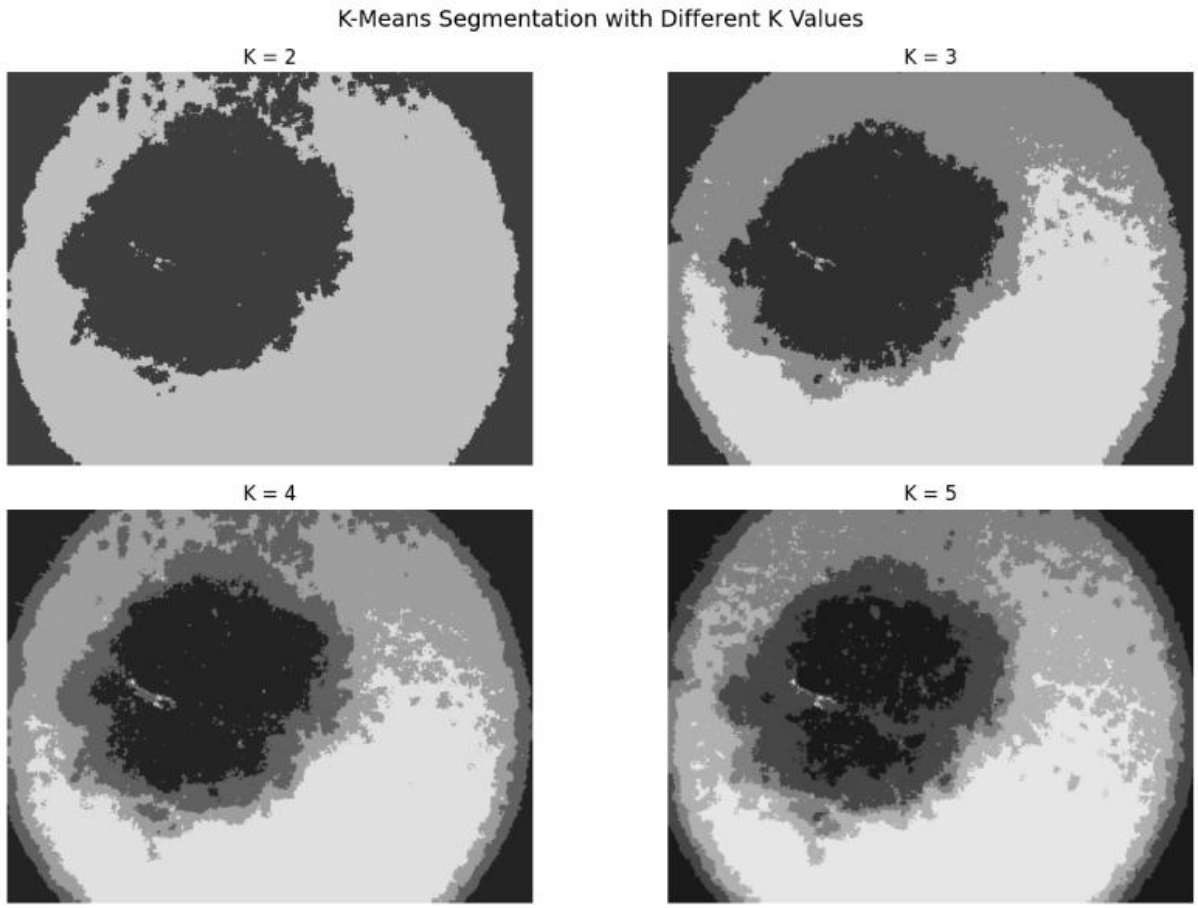


Figure 26. K-Means Segmentation with Different K-values for IMD022

The results in IMD022 shows the Intersection over Union (IoU) scores obtained from different morphological operations and kernel sizes. Among the tested configurations, the dilation operation with a kernel size of 11 × 11 achieved the highest IoU score of 0.8456. This indicates that dilation effectively enhanced the segmented lesion area by expanding and connecting nearby regions, resulting in improved overlap with the ground truth mask.

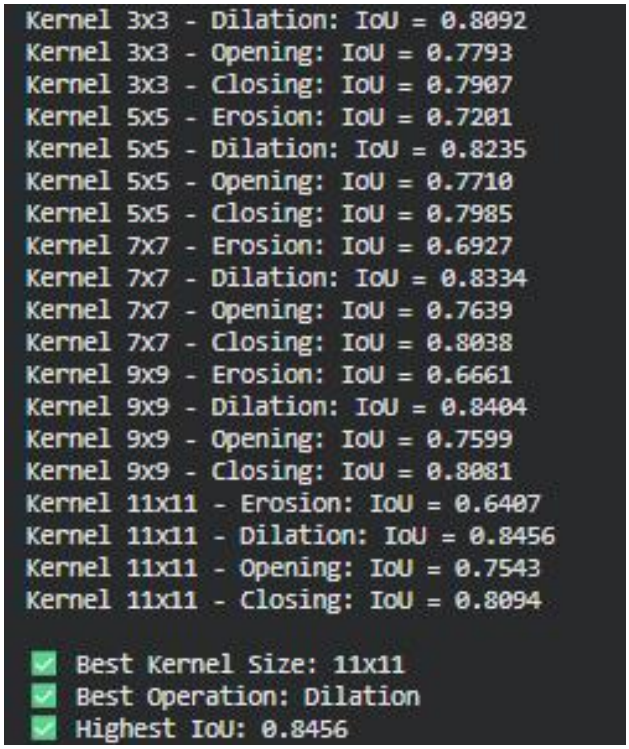


Figure 27. Output of Best Morphological Operation and Kernel Size Indicating the Highest IoU Value for IMD022

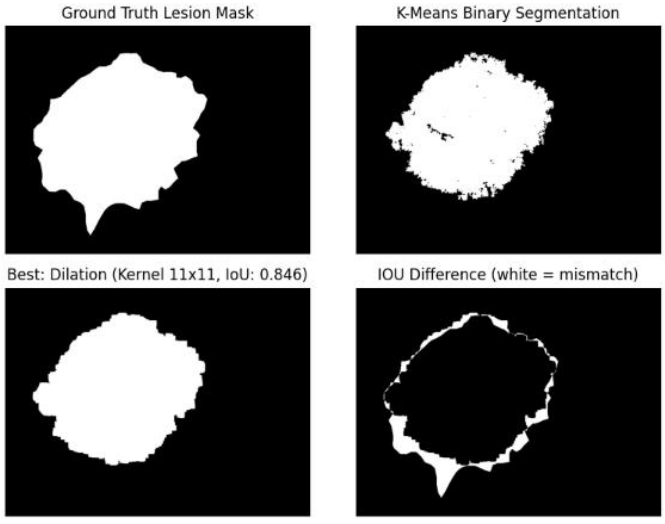


Figure 28. Comparison of Morphological Operations Showing the Best Operation and IoU Differences IMD022

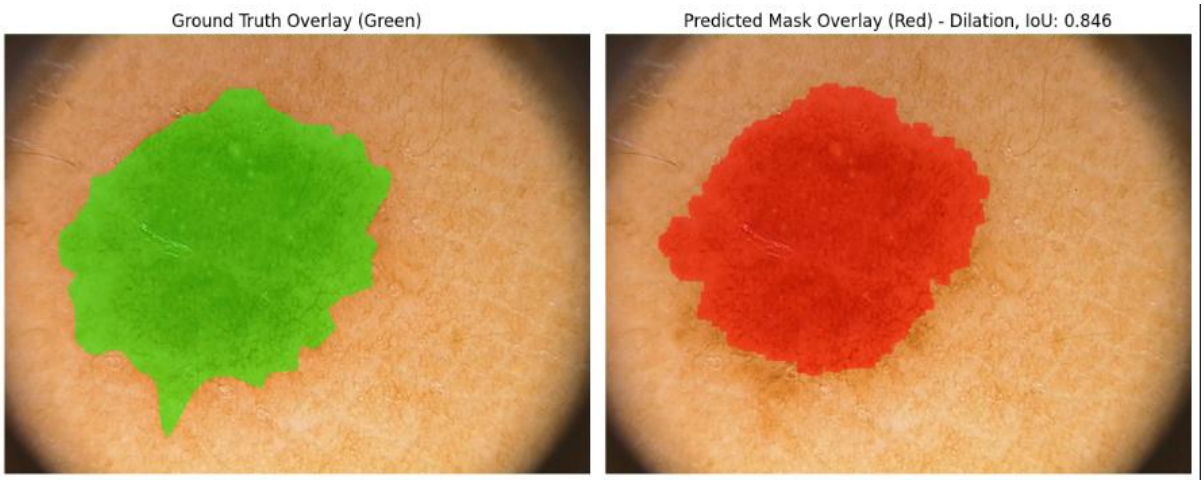


Figure 29. Overlay comparison between ground truth (green) and predicted segmentation (red) on IMD022

The confusion matrix and performance metrics indicate that the dilation operation with a kernel size of 11 × 11 produced strong segmentation results. The model achieved 104,671 true positives, 314,566 true negatives, 3,726 false positives, and 15,382 false negatives. Based on these values, the segmentation reached an accuracy of 0.9564, sensitivity of 0.8719, and specificity of 0.9883, demonstrating reliable detection performance. Furthermore, the Jaccard index of 0.8456 and Dice coefficient of 0.9164 confirm a high degree of similarity between the predicted mask and the ground truth lesion region.

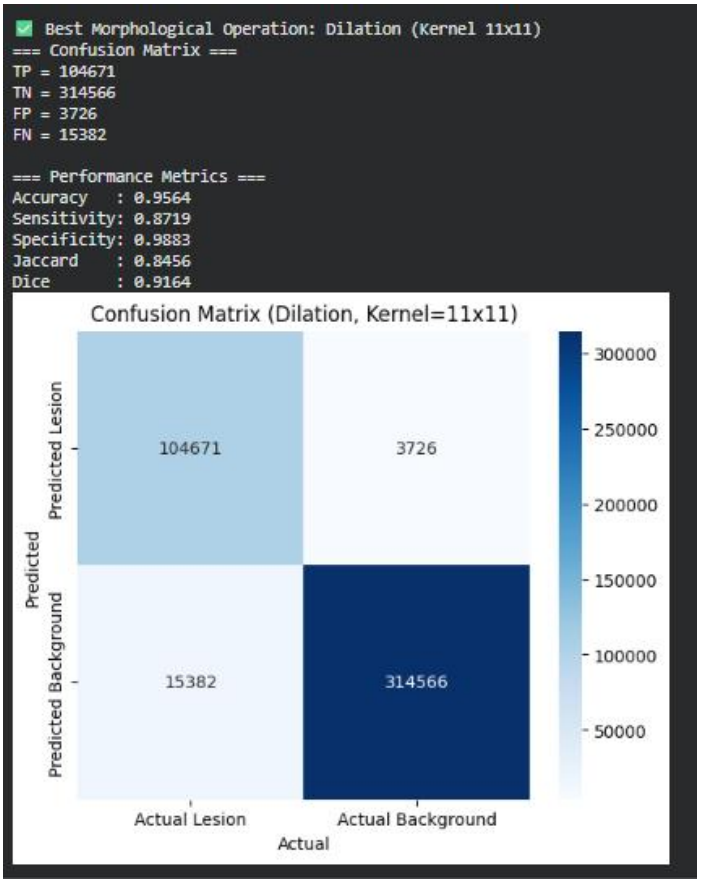


Figure 30. Confusion Matrix and Performance matrix for IMD022

4.3 Experimental Results for Difficult Category Images

The difficult category is characterized by the presence of hair artifacts and highly indistinct lesion boundaries, making segmentation particularly challenging. Both IMD003 and IMD006 are classified as images under difficult category.

4.3.1 Performance Evaluation of IMD003 (Category Difficult)

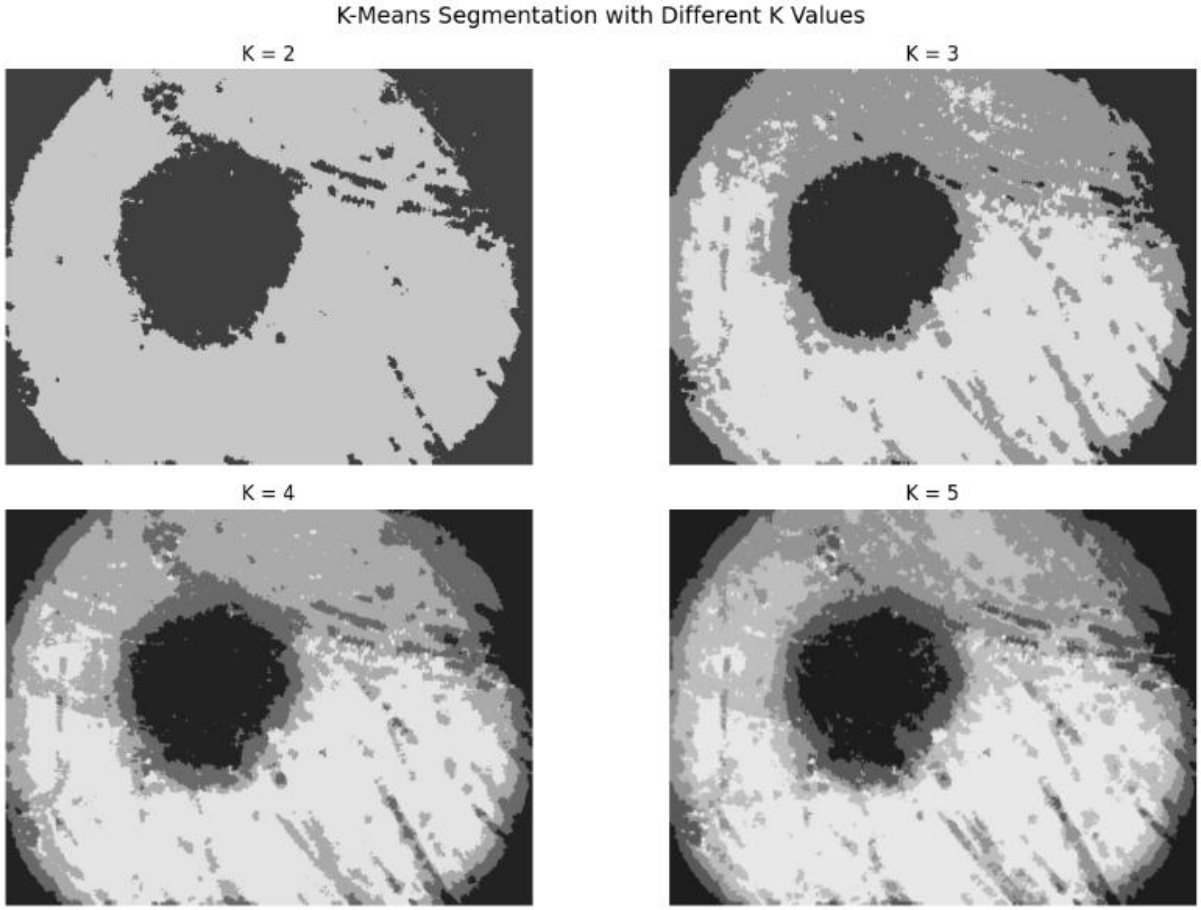


Figure 31. K-Means Segmentation with Different K-values for IMD003

The IoU evaluation across different kernel sizes and morphological operations shows that the closing operation with a kernel size of 9 × 9 achieved the highest IoU score of 0.9043. This result indicates that closing effectively refined the segmented lesion by smoothing boundaries and filling small gaps, leading to a more accurate overlap between the predicted mask and the ground truth.

```
IoU results for different kernel sizes:
Kernel 3x3 - Erosion: IoU = 0.8833
Kernel 3x3 - Dilation: IoU = 0.8980
Kernel 3x3 - Opening: IoU = 0.8960
Kernel 3x3 - Closing: IoU = 0.8971
Kernel 5x5 - Erosion: IoU = 0.8654
Kernel 5x5 - Dilation: IoU = 0.8932
Kernel 5x5 - Opening: IoU = 0.8958
Kernel 5x5 - Closing: IoU = 0.9015
Kernel 7x7 - Erosion: IoU = 0.8440
Kernel 7x7 - Dilation: IoU = 0.8840
Kernel 7x7 - Opening: IoU = 0.8941
Kernel 7x7 - Closing: IoU = 0.9035
Kernel 9x9 - Erosion: IoU = 0.8209
Kernel 9x9 - Dilation: IoU = 0.8734
Kernel 9x9 - Opening: IoU = 0.8922
Kernel 9x9 - Closing: IoU = 0.9043
Kernel 11x11 - Erosion: IoU = 0.7967
Kernel 11x11 - Dilation: IoU = 0.8619
Kernel 11x11 - Opening: IoU = 0.8924
Kernel 11x11 - Closing: IoU = 0.9040

✅ Best Kernel Size: 9x9
✅ Best Operation: Closing
✅ Highest IoU: 0.9043
```

Figure 32. Output of Best Morphological Operation and Kernel Size Indicating the Highest IoU Value for IMD003

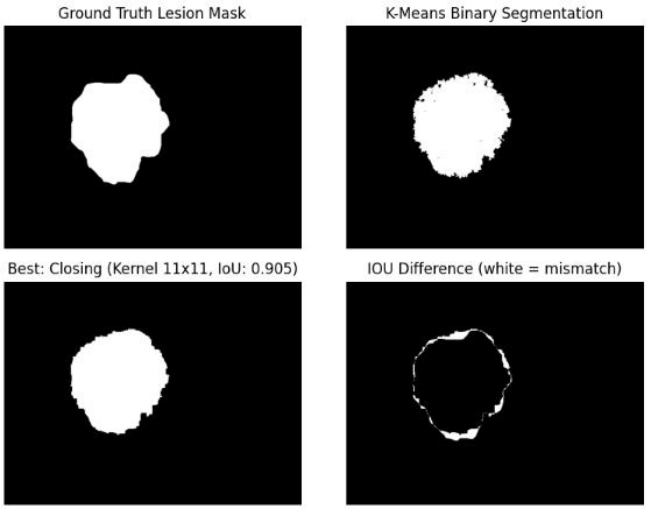


Figure 33. Comparison of Morphological Operations Showing the Best Operation and IoU Differences IMD003

Figure 34. Overlay comparison between ground truth (green) and predicted segmentation (red) on IMD003

The confusion matrix shows 47,336 true positives, 386,765 true negatives, 2,469 false positives, and 2,540 false negatives. These values correspond to an accuracy of 0.9886, sensitivity of 0.9491, and specificity of 0.9937, indicating precise and reliable lesion detection. Moreover, the Jaccard index of 0.9043 and Dice coefficient of 0.9497 further confirm a high level of similarity between the predicted segmentation mask and the ground truth, demonstrating that the closing operation effectively enhances lesion boundaries and fills small imperfections.

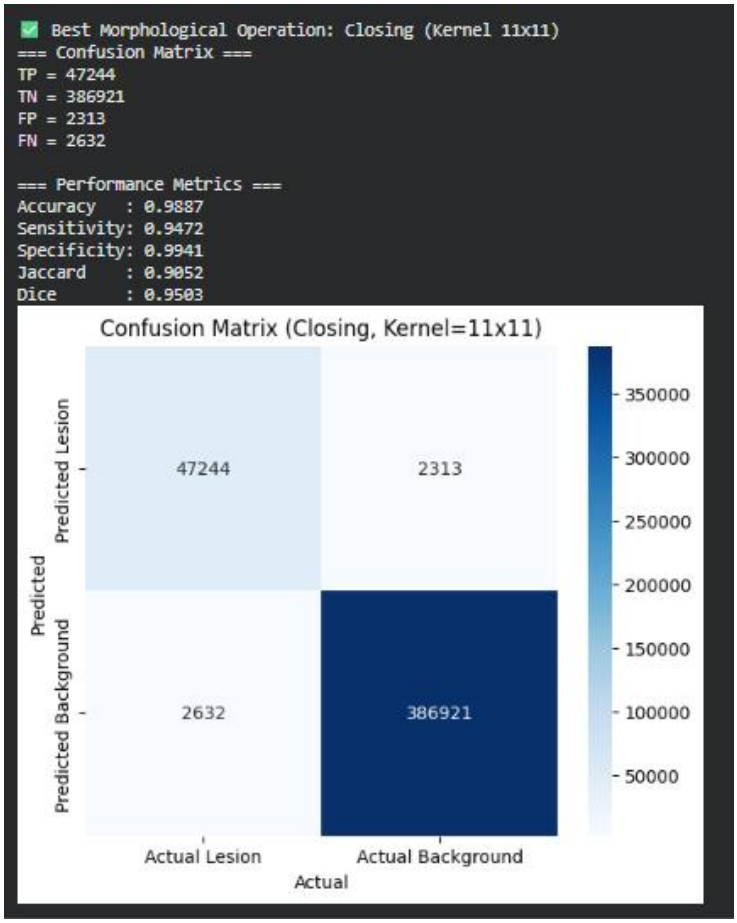


Figure 35. Confusion Matrix and Performance matrix for IMD022

The confusion matrix shows 47,336 true positives, 386,765 true negatives, 2,469 false positives, and 2,540 false negatives. These values correspond to an accuracy of 0.9886,

sensitivity of 0.9491, and specificity of 0.9937, indicating precise and reliable lesion detection. Moreover, the Jaccard index of 0.9043 and Dice coefficient of 0.9497 further confirm a high level of similarity between the predicted segmentation mask and the ground truth, demonstrating that the closing operation effectively enhances lesion boundaries and fills small imperfections.

4.3.2 Performance Evaluation of IMD006 (Category Difficult)

For IMD006, the segmentation performed best with K = 3

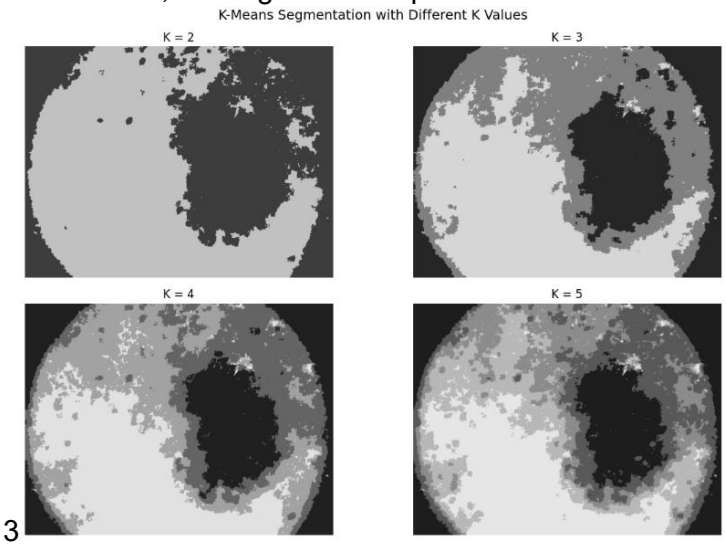


Figure 36. K-Means Segmentation with Different K-values for IMD006

The IoU evaluation for different kernel sizes and morphological operations revealed that the dilation operation with a kernel size of 7 × 7 achieved the highest IoU score of 0.7483. This indicates that dilation was the most effective in enhancing the segmented lesion area by expanding the detected regions and improving the overlap between the predicted mask and the ground truth.

```
IoU results for different kernel sizes:
Kernel 3x3 - Erosion: IoU = 0.7338
Kernel 3x3 - Dilation: IoU = 0.7461
Kernel 3x3 - Opening: IoU = 0.7408
Kernel 3x3 - Closing: IoU = 0.7420
Kernel 5x5 - Erosion: IoU = 0.7248
Kernel 5x5 - Dilation: IoU = 0.7482
Kernel 5x5 - Opening: IoU = 0.7404
Kernel 5x5 - Closing: IoU = 0.7436
Kernel 7x7 - Erosion: IoU = 0.7146
Kernel 7x7 - Dilation: IoU = 0.7483
Kernel 7x7 - Opening: IoU = 0.7402
Kernel 7x7 - Closing: IoU = 0.7442
Kernel 9x9 - Erosion: IoU = 0.7022
Kernel 9x9 - Dilation: IoU = 0.7466
Kernel 9x9 - Opening: IoU = 0.7402
Kernel 9x9 - Closing: IoU = 0.7446
Kernel 11x11 - Erosion: IoU = 0.6882
Kernel 11x11 - Dilation: IoU = 0.7437
Kernel 11x11 - Opening: IoU = 0.7403
Kernel 11x11 - Closing: IoU = 0.7440

✅ Best Kernel Size: 7x7
✅ Best Operation: Dilation
✅ Highest IoU: 0.7483
```

Figure 37. Output of Best Morphological Operation and Kernel Size Indicating the Highest IoU Value for IMD006

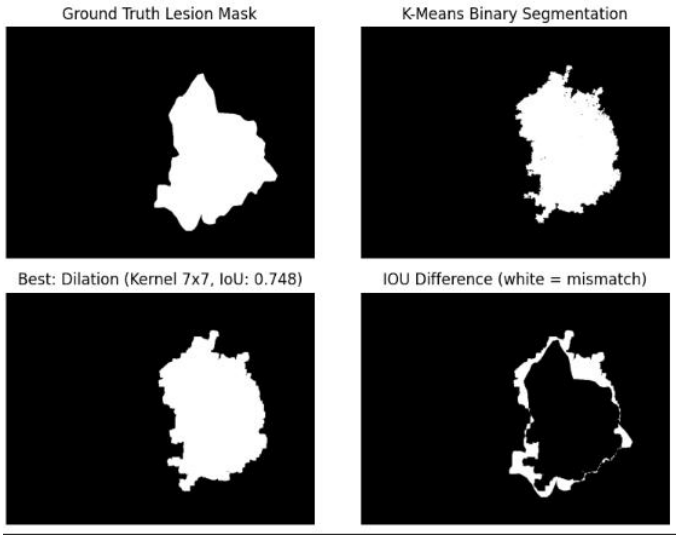


Figure 38. Comparison of Morphological Operations Showing the Best Operation and IoU Differences IMD006

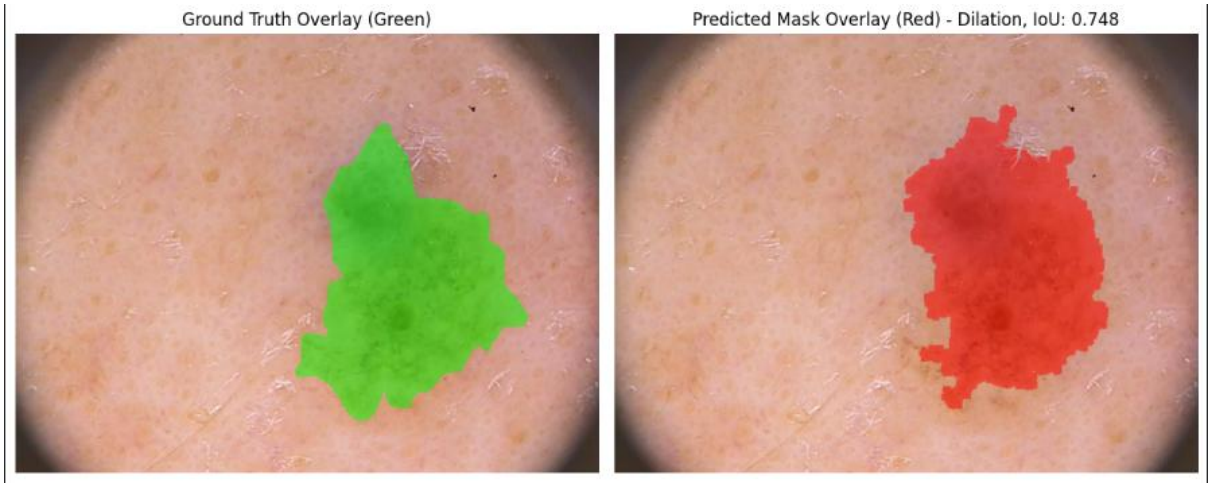


Figure 39. Overlay comparison between ground truth (green) and predicted segmentation (red) on IMD006

The dilation operation with a kernel size of 7 × 7 achieved the best segmentation performance for this image. The confusion matrix results show 56,773 true positives, 363,236 true negatives, 11,315 false positives, and 7,786 false negatives. From these values, the model obtained an accuracy of 0.9565, sensitivity of 0.8794, and specificity of 0.9698, demonstrating strong lesion detection capability. Additionally, the Jaccard index of 0.7483 and Dice coefficient of 0.8560 indicate good agreement between the predicted segmentation mask and the ground truth, confirming that the dilation operation effectively enhanced the lesion boundaries and improved region connectivity.

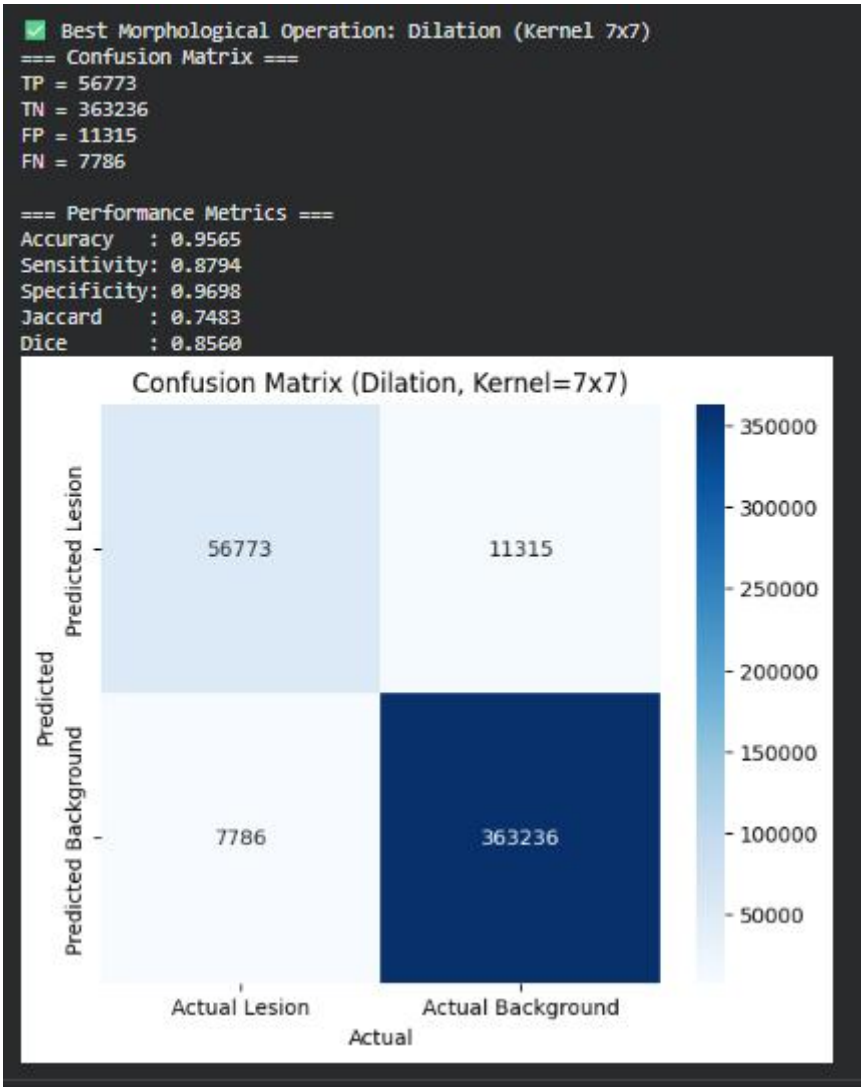


Figure 40. Confusion Matrix and Performance matrix for IMD006

4.4 Summary of the Results.

Category: Easy	**K-means**	**Operation**	**Performance Metrics**
IMD018	3	Erosion	Accuracy 0.9853 Sensitivity 0.9604 Specificity 0.9878 Jaccard 0.8560 Dice 0.9224
IMD019	3	Erosion	Accuracy 0.9869 Sensitivity 0.9642 Specificity 0.9893 Jaccard 0.8748 Dice 0.9332
Category: Moderate	**K-means**	**Operation**	
IMD002	3	Closing	Accuracy 0.9704 Sensitivity 0.9430 Specificity 0.9808 Jaccard 0.8976 Dice 0.9460
IMD022	3	Opening	Accuracy 0.9564 Sensitivity 0.8719 Specificity 0.9883 Jaccard 0.8456 Dice 0.9164
Category: Difficult	**K-means**	**Operation**	
IMD003	3	Closing	Accuracy 0.9887 Sensitivity 0.9472 Specificity 0.9941 Jaccard 0.9052 Dice 0.9503
IMD006	3	Dilation	Accuracy 0.9565 Sensitivity 0.8794 Specificity 0.9698 Jaccard 0.7483 Dice 0.8560

Table 2. Summary result details for IMD018, IMD019, IMD002, IMD022, IMD003, IMD006

5. Conclusion

The experimental results demonstrate that the K-Means clustering approach, combined with appropriate preprocessing, can effectively segment skin lesion regions from dermoscopic images. Through techniques such as grayscale conversion, histogram equalization, and inpainting for hair artifact removal, the preprocessing pipeline substantially enhanced the clarity and uniformity of lesion boundaries. The integration of these steps proved essential in producing cleaner binary masks that closely matched the ground truth segmentation, as verified through visual overlay comparisons using red and green color channels.

One of the findings from this study is that the segmentation performance remained consistent across images of varying complexity, including those with mild blur or uneven illumination. The Intersection over Union (IoU), Jaccard Index, Dice Coefficient, accuracy, and sensitivity collectively validated the reliability of the segmentation outcomes, reflecting strong correspondence between predicted and actual lesion areas. This prooves that unsupervised clustering methods, when supported by systematic parameter tuning and preprocessing, can provide a viable alternative to more complex supervised models for medical image analysis.

The primary strength of this approach lies in its interpretability and computational efficiency. Unlike deep learning methods that require large, labelled datasets, the K-Means algorithm operates effectively on limited data while maintaining transparency in its decision-making process. The simplicity of implementation also allows easy adaptation to different imaging datasets, making it suitable for practical diagnostic applications where annotated data may be scarce. Moreover, the evaluation using multiple images across different difficulty levels supports the robustness of the methodology.

However, certain limitations remain. The algorithm’s dependence on colour and intensity features restricts its ability to distinguish subtle texture variations or pigmentation patterns. While the preprocessing steps mitigate noise and lighting inconsistencies, they may not fully address challenges such as extreme illumination variation or overlapping skin tones.

In future study, researchers could include integrating texture-based descriptors or hybrid models that combine K-Means with advanced learning frameworks like convolutional neural networks (CNNs) to capture deeper semantic features. Expanding the dataset and refining evaluation metrics would further enhance the reliability and clinical relevance of the segmentation model.

6. Reference

Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., Alsaiari, S.A., Saeed, A.H.M., Alraddadi, M.O. and Mahnashi, M.H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. *International Journal of Environmental Research and Public Health*, 18(10), p.5479. doi:https://doi.org/10.3390/ijerph18105479.

Liu, L., Tsui, Y.Y. and Mandal, M. (2021). Skin Lesion Segmentation Using Deep Learning with Auxiliary Task. *Journal of Imaging*, 7(4), p.67. doi:https://doi.org/10.3390/jimaging7040067.

Mendonca, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S. and Rozeira, J. (2013). PH^2 - A dermoscopic image database for research and benchmarking. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. doi:https://doi.org/10.1109/embc.2013.6610779.

Ramprasad, M.V.S., S.S.V. Nagesh, V. Sahith and Rohith Kumar Lankalapalli (2023). Hierarchical agglomerative clustering-based skin lesion detection with region based neural networks classification. *Measurement Sensors*, 29, pp.100865–100865. doi:https://doi.org/10.1016/j.measen.2023.100865.

Sreedhar Burada, B.E. Manjunathswamy and Kumar, M.S. (2024a). Early detection of melanoma skin cancer: A hybrid approach using fuzzy C-means clustering and differential evolution-based convolutional neural network. *Measurement Sensors*, 33, pp.101168–101168. doi:https://doi.org/10.1016/j.measen.2024.101168.

ZEBARİ, N.A. and TENEKECİ, E. (2022). Skin Lesion Segmentation Using K-means Clustering with Removal Unwanted Regions. *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*. doi:https://doi.org/10.54365/adyumbd.1112260.

CML3014N Machine Learning
MARKING RUBRIC
ASSIGNMENT 1
Assignment Weighting (30%)

REPORT COMPONENT (100%)

LEARNING OUTCOME	MARKING CRITERIA	SCALE					
		Fail (0-49)	3rd Class (50-59)	2nd Lower Class (60-69)	2nd Upper Class (70-79)	1st Class (80-100)	YOUR MARKS/COMMENTS
CO2: Critically review and select machine learning methods to solve real-world problems	Choice of Clustering Technique (20%)	• No suitable method chosen or implemented incorrectly.	• Method used but with no justification.	• Method chosen with limited justification.	• Appropriate method with clear justification.	• Well-justified choice of suitable clustering method, demonstrating deep understanding of strengths and limitations.	
	Parameter Tuning (20%)	• No attempt at parameter tuning.	• Very little attempt at tuning, weak explanation.	• Limited experiments with minimal justification.	• Adequate experiments with reasonable justification.	• Extensive experiments with clear reasoning for optimal settings.	
	Results and Evaluation (20%)	• No results or evaluation presented	• Poor results with minimal visuals, weak or incorrect evaluation.	• Acceptable results with limited visuals and superficial evaluation.	• Good results with sufficient visual examples, reasonable evaluation using at least one metric.	• High-quality results with clear, well-presented visuals and comprehensive evaluation using multiple metrics, critically compared with ground truth.	
	Report Quality & Discussion (30%)	• Report missing major sections or lacks originality.	• Report incomplete, poorly written, with little analysis.	• Report covers essentials but lacks depth or clarity.	• Well-written and clear report, with some critical analysis.	• Excellent structure and writing, with insightful, critical discussion and originality.	
	Code quality (10%)	• Very trivial program structure and without code comments.	• Trivial program structure but with some code comments.	• Clear program structure and appropriate comments.	• The program code is well-structured and commented.	• The program code is efficient, well-structured, and commented.	
						Overall score (100%)	