# Tumor Segmentation in Mammographic Images

Medical Image Analysis Assignment 1

Umut Baybece
Student ID: 2743656

18 November 2025

## 1 Introduction

In this report, I present a method and algorithm for automatically detecting tumors in a set of images from the MIAS (Mini Mammography Database). The report provides a step by step explanation of the algorithm used and the reasons for my method choices for each algorithm step. The algorithm first performs various preprocessing and masking operations to remove unnecessary parts from the images. After removing the irrelevant parts, the next step leaves only the breast region in the images. In the third step, the algorithm uses erosion and minimum filters to erode and remove vessels outside the tumor and irrelevant parts. In the next step, I select the brightest point in the processed images. This selected pixel is assigned as the starting point for the region growing algorithm. Therefore, in the next step, I use this starting point for region growing and apply morphological closing to fill in small gaps within the segmented tumor area. Finally, I display the tumor region in the original image and plot the ground truth circle on the images for comparison.

## 2 Segmentation Algorithm Description

### 2.1 Preprocessing Steps

I read only the .pgm files in the provided dataset folder, as .jpg files are compressed and can lose important details. After reading the images, I used a **Gaussian filter**, to remove noise from the mammogram images. Medical images often contain noise from the imaging equipment, and Gaussian blur effectively smooths this noise while preserving important details. I used a 9x9 Gaussian filter. Also, I initially tried a 3 by 3 Gaussian filter, but it didn't smooth the images enough to remove fine details.

In the second step, I used **contrast stretching** to increase the contrast of the white areas, as the tumors and pectoral muscle appeared whiter than the other areas in the image. Therefore, contrast stretching amplified the intensity differences in the image, making it easier to detect the white areas, tumors, and pectoral muscles.

The biggest challenge was the presence of the pectoral muscle in all the images. Because it was the largest and whitest area in all the images, it needed to be removed before any processing requiring intensity change detection in the image. Therefore, I created a **V-shaped binary mask at a 26 degree angle**. This mask creates a V-shaped area that preserves the central breast tissue while removing the muscle area and white labels visible in some images. This mask included two black triangle areas that form a V, completely eliminating the unwanted areas. I initially tried a 30 degree angle, but 26 degrees proved to be optimal because the 30 degree mask also removed some tumors.

After muscle removal, I used **Otsu thresholding** to automatically calculate the best threshold value for separating the breast tissue from the background. The muscle removal with the triangular mask was not perfect. It keeps a V-shaped region, but inside this V-shape there is still background mixed with the actual breast tissue. Otsu thresholding separates the actual breast tissue from the dark background areas and creates a binary image. Then, largest connected component extraction keeps only the biggest white region, which is the breast region in all images. Finally, this binary breast region mask is multiplied with the preprocessed grayscale image to extract only the breast region with its actual intensity values.

## 2.2 Seed Finding Steps

Blood vessels appeared in the breast region as bright thin linear shapes. To prevent them from causing over-segmentation, I applied **morphological erosion** with a disk-shaped structuring element with an 81 pixel radius. I also did several trials with smaller disks but smaller disks didn't erode enough. It removed thin blood vessels while making larger white regions, such as tumors, stand out more clearly. However, even though erosion removed many vessels, some small bright spots remained. Therefore, I applied additional filtering steps. I used a **mean filter** with an 81 by 81 kernel to smooth out the remaining bright pixels after erosion. This filter averages each pixel with its neighbors. Therefore, only big bright regions remain visible in the images. After these steps, I applied a **minimum filter** with a 21 pixel radius to enhance contrast. I tried different sizes for the filter but 21 pixel radius worked well. The minimum filter replaces each pixel with the minimum value in its neighborhood, which lowers the intensity of bright regions. This increased the contrast between different bright areas, making it easier to identify the brightest pixel in the image, which usually corresponded to the tumor center. At the final step, I used "cv2.minMaxLoc()" function to find the brightest pixel within the breast mask. I passed this pixel location info to region growing algorithm to use as the seed value.

## 2.3 Region Growing Steps

Region growing worked well for finding tumors in mammogram images because tumors usually appear as connected areas with similar brightness. Also, region growing only grows from one starting point inside the tumor. This means it only segments the area that is connected to the seed point.

Therefore, after finding the seed point, I applied a **4-connectivity region growing algorithm** based on the method described by Adams and Bischof [1]. The algorithm starts from the seed pixel and moves to neighboring pixels with similar intensity values. After several trials with different values, I used a threshold of 12. My experiments showed that a value of 20 worked well for very large tumors but over-segmented small tumors, and a value of 10 worked well for small tumors but under-segmented large tumors. Based on this threshold, the algorithm compares the intensity of a neighboring pixel to the intensity of the seed pixel to determine whether it belongs to the tumor region. If the absolute difference is less than or equal to 12, the pixel is added to the segmented region.

The original algorithm uses a priority queue but I used FIFO queue for simplicity to process pixels and keep a visited array to prevent the same pixel from being used multiple times. For each pixel in the queue, it checks four neighbors (left, right, top, bottom). If a neighbor's intensity is within the threshold and has not yet been visited, it is added to both the queue and the region. This process continues until the queue is empty or reaches maximum limit.

After region growing, I applied **morphological closing** with a 12 by 12 disk to fill small holes in the segmented tumor region. It helped algorithm to have smoother results.

### 2.4 Visualization and Evaluation Steps

After the segmentation process was completed, I created visuals to evaluate the algorithm's performance. First, I extracted the tumor region from the original image by applying a segmentation mask. I put the mask region on image using red color.

I visualized the ground truth information onto the original image by drawing a red circle at the coordinates provided in the info.txt file. The ground truth circle represents the actual tumor location and size, shown by a red circle.

Finally, I created a combined version in which the segmented tumor region is green and the ground truth circle is red. This image helps to compare the algorithm's segmentation results with the tumor's actual position. If the segmentation is accurate, the green segmented region enclosed by the red ground truth circle. The Figure 1 showing the results of each step (using only one image as an example).

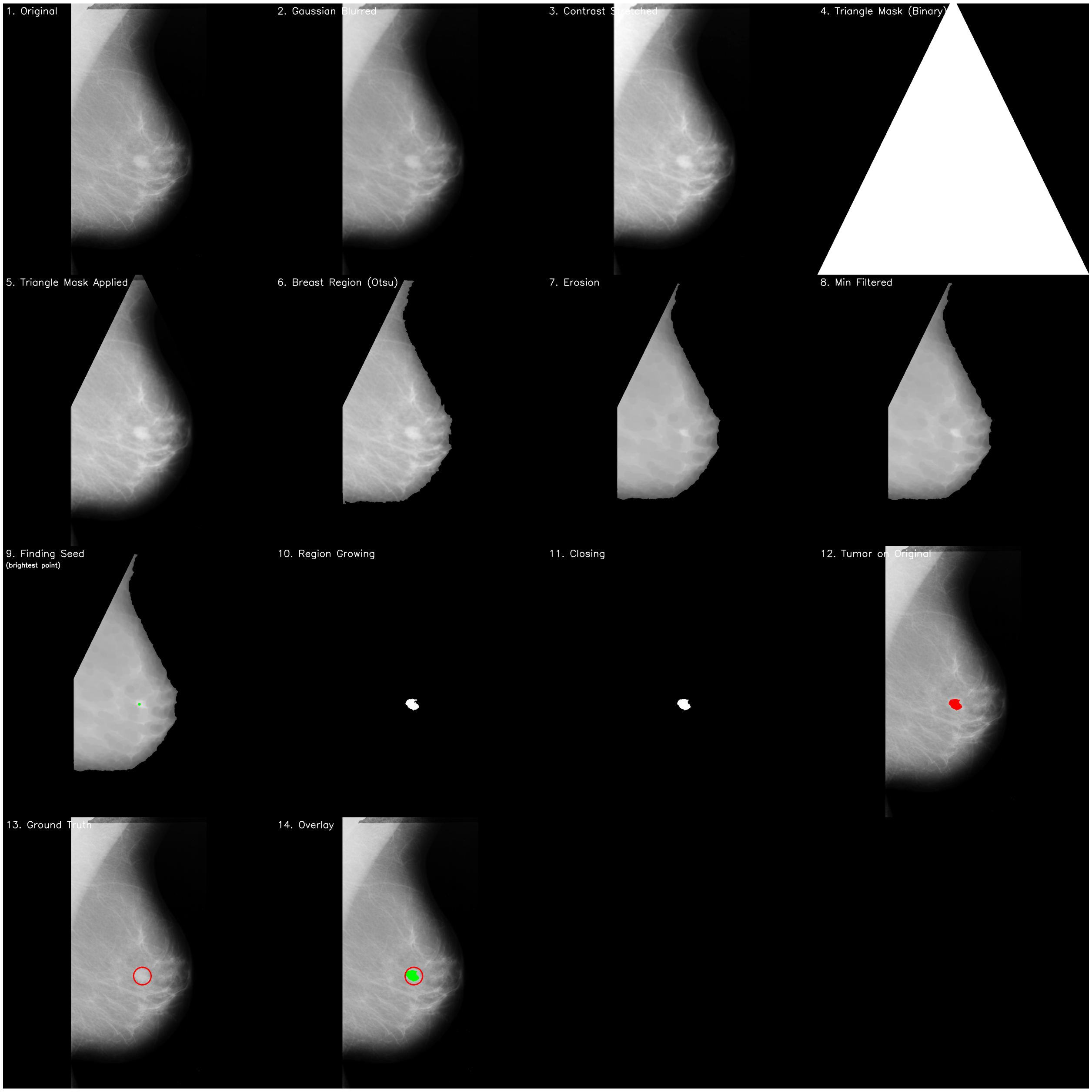


Figure 1: The sequence shows the results of each step for a single image

# 3 Segmentation Results

The proposed algorithm was tested on all 17 images in the dataset, and the results are presented in the Figure 2 and Figure 3 Each row shows the original mammography image, the ground truth annotation containing a red enclosing circle indicating the tumor location, and the segmentation output of the algorithm. To clarify the detected tumor boundaries, the segmented regions are presented as red masks placed on the original images.

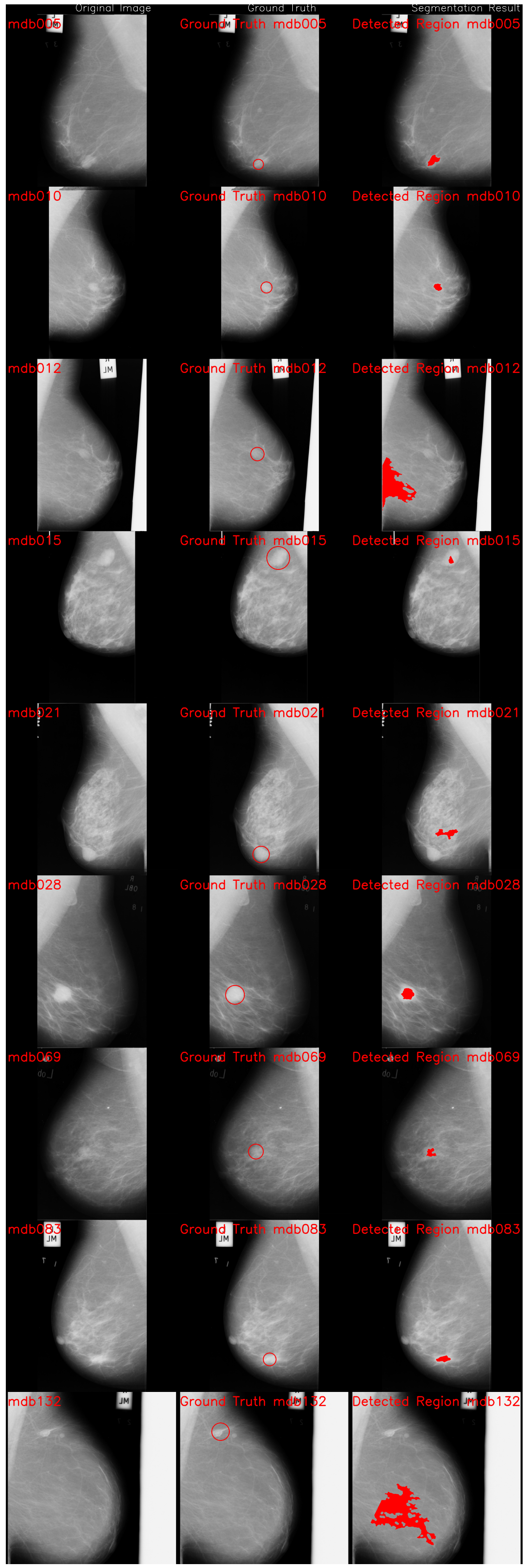


Figure 2: Comparison results for images 1-9

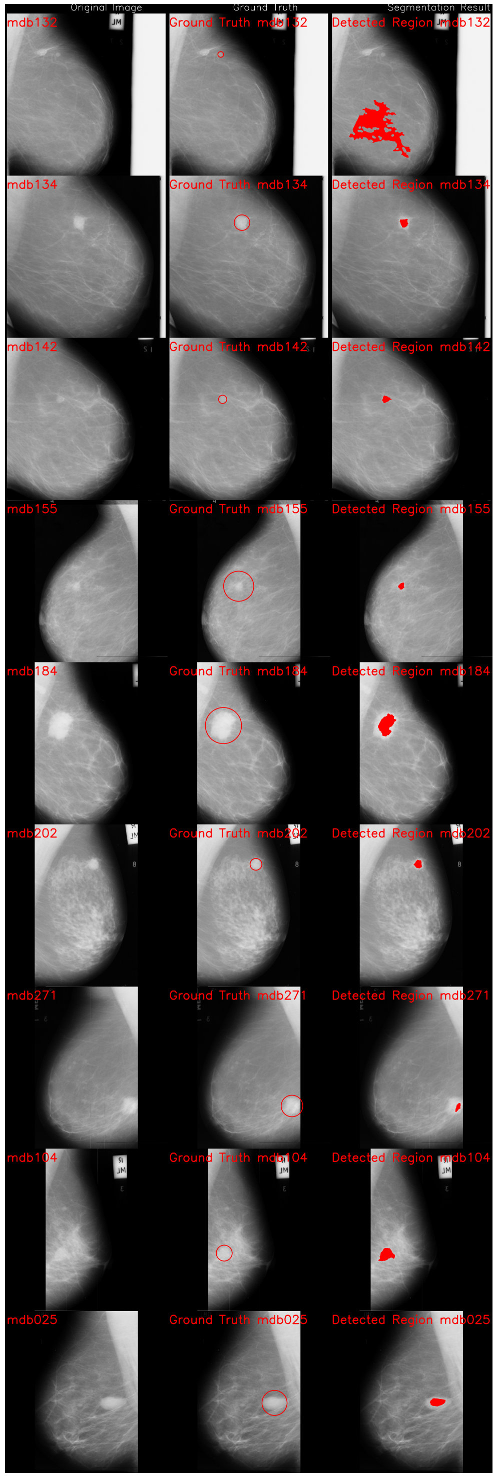


Figure 3: Comparison results for images 10-17

# 4 Performance Discussion

The algorithm successfully detected tumors in 14 out of 17 images, allowing us to conclude that the algorithm performed well on this limited number of images which are from the MIAS dataset. A total of 18 tumors existed across all images; one image contained two tumors and the others had only one tumor. Although the algorithm found tumor regions, it failed to perfectly match the ground truth circle because the circles could not fully cover irregularly shaped tumors. However, these 14 segmented regions were within the true circle, and with the region growing algorithm they clearly revealed the true tumor shapes.

This evaluation revealed some limitations of the current approach. Because the designed algorithm only finds one seed and applies the region growing algorithm to this specific seed, it only detects a single tumor in the images. Since in these set of images only one picture had more than one tumor it can be accepted for this provided image set. In cases where multiple tumors are present, only the most brightest one can be segmented. I also experimented with different parameters and thresholds to select the most effective ones while developing the algorithm. Therefore, these parameters optimized for this dataset may not generalize to images from different mammography datasets.

The 26 degree angle for muscle extraction was effective for this dataset, but this geometric assumption may not be suitable for images from different mammography datasets and may not be applicable to all patient positioning scenarios. Furthermore, tumors located very close to the chest wall are particularly challenging because they may be partially removed during the muscle extraction step. Also, the algorithm relies on intensity based features. Therefore, it is prone to noise which appears as bright regions. This can lead to false positive results in such cases.

Despite these limitations, the algorithm successfully performs fully automated tumor segmentation without requiring manual intervention. The region growing approach fits well to irregular tumor shapes, producing clinically meaningful results.

# References

[1] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, June 1994, doi: 10.1109/34.295913.