Erasmus Mundus Joint Master in Medical Imaging and Applications

Breast mass detection system based on a multi-scale approach

Students:

Luisana Álvarez Monsalve

Clara Lisazo

Xavier Beltrán Urbano

Professor:

Alessandro Bria

2022-2023

Abstract

This report presents the development of a breast mass detection system using image processing techniques. The study utilized the INbreast dataset of 410 mammograms. The whole process involved preprocessing, multi-scale grayscale morphology and segmentation steps. Preprocessing included downsampling, cropping, and applying Contrast Limited Adaptive Histogram Equalization (CLAHE) for enhanced contrast. A multi-scale approach was adopted, employing Top-Hat and morphological opening operations with varying structuring elements. Superpixels were generated using the SLICO algorithm, and K-means clustering was used for segmentation. Various features, including texture and shape descriptors, were extracted to train a machine learning classifier in order to reduce false positives. The system achieved high sensitivity but had a high number of false positives due to the numerous scales used, that generated several candidate regions. Future research should focus on refining the segmentation algorithm by carefully adjusting the number of scales to achieve a more balanced and effective detection system.

Introduction

Breast cancer is currently one of the most commonly diagnosed cancers and the 5th leading cause of cancer-related death, with an estimated 2.3 million new cases worldwide [1]. Early identification is critical to improving survival rates and the efficacy of treatment since it helps catch problems at their earliest stages [2].

Several imaging modalities are utilised in the examination of the breast in order to allow early detection and diagnosis. One of the most widely used techniques is mammography, due to its high sensitivity. Nevertheless, even when using potent imaging techniques such as mammography, the accurate and speedy detection of masses continues to be a difficult process.

In response to this, the aim of this project is to develop an accurate breast mass detection and segmentation system through the implementation of several image processing and analysis techniques. In order to achieve proper results for a highly variable dataset, a multi-scale approach will be carried out.

Materials and Methods

1 Dataset

The developed mass detection system used the INbreast dataset, which contains 410 mammograms. These include 115 masses belonging to 107 positive images, and 303 negative images, that were free of abnormalities.

2 Preprocessing

First, the original mammograms were downsampled by a factor of 4 using bi-cubic interpolation, in order to reduce processing times. The second step involved cropping the downscaled images to keep only the breast region. For the positive images, the corresponding ground truth images were also cropped to match the same region of interest, which was required for the posterior evaluation of the method. For the cropping of the ground truth images, the coordinates of the bounding rectangle were adjusted to correspond to the original scale rather than the downsampled scale. This adjustment was necessary to ensure accurate alignment with the original mammograms during the posterior evaluation of the method. Once the images were adapted to fulfill the needed requirements of size and shape, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied. The purpose of this was to enhance the contrast of the mammogram by redistributing the intensity values, taking into account the local intensity differences of the image, which helped in highlighting possible lesions. A relatively small tile size of 6 was chosen to ensure that local details and texture characteristics specific to the breast tissue were preserved effectively. Additionally, the clip limit parameter was set to 10 to prevent over-amplification of noise or other irrelevant features. Finally, mean-shift filtering was performed using small spatial and color window radii of 5. The reason for this was to reduce noise in the image while preserving edges and important structures. By smoothing the image, noise artifacts that could interfere with the following grayscale morphology operations were removed. This filtering helped to improve the connectivity of the image, ensuring that the subsequent morphology operations produced more consistent and accurate results.

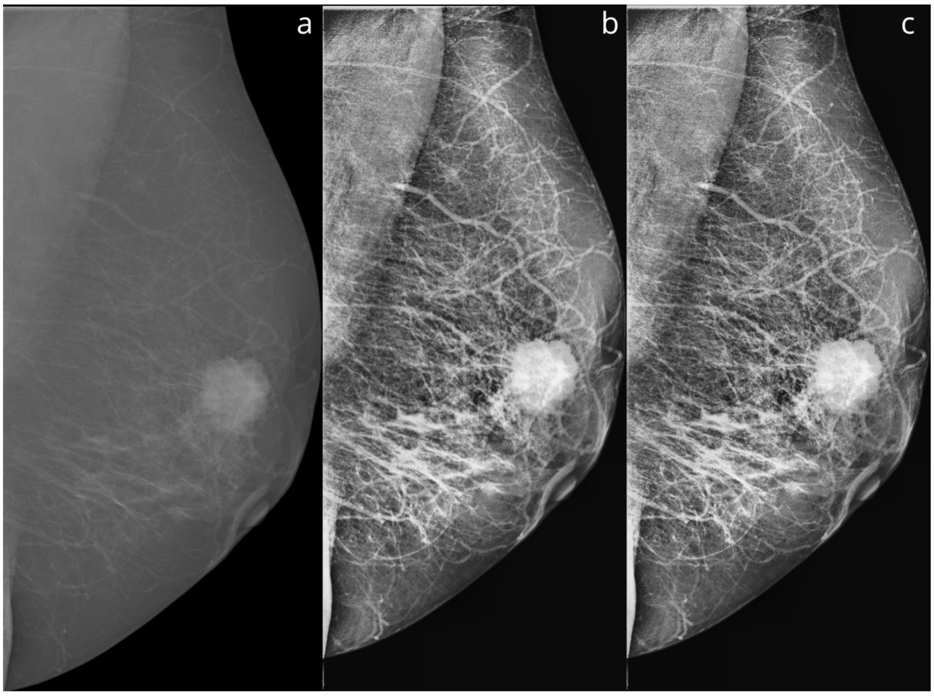

Figure 1: Results of the preprocessing stages: a. Original cropped image. b. CLAHE image. c. Mean-shift filtered image.

3 Multi-scale grayscale morphology and segmentation

Due to the high variability in the sizes of the breast masses in the dataset, a multi-scale approach was needed, in which structuring elements of different sizes were used as proposed by Min et al. [3].

The first step before the segmentation, consisted in applying Top-Hat with disk-shaped structuring elements (DSE) of size R, followed by opening with DSE of size r, being $R > r$. In our case, R and r were vectors containing 11 different values in order to enhance masses of various sizes. For all elements in the vectors, the relationship $R = 2 \cdot r$ was maintained. By applying Top-Hat filtering, bright structures in the image, smaller than the size of the DSE, were highlighted. By using a larger structuring element of size R in the Top-Hat operation, larger and brighter structures, that could correspond to potential regions of interest, were emphasized. After Top-Hat, morphological opening was applied to the resulting image. This was performed to remove small, bright structures or noise while preserving the overall shape and bigger structures in the image. This allowed to separate connected structures and reduce noise artifacts that may have been amplified during the Top-Hat operation. By using a smaller DSE, it was ensured that only relatively smaller structures that could correspond to noise were eliminated, while the larger structures that had been highlighted through Top-Hat remained intact.

Prior to generating superpixels from the resulting images of the morphological operations, Gaussian blurring with a small kernel size ($\sigma = 3$) was implemented in order to reduce noise and smooth out irregularities in the image that could lead to the generation of an unnecessary amount of superpixels. The Simple Linear Iterative Clustering Optimized (SLICO) algorithm was utilised to generate superpixels. This algorithm iterated over all the scales and performed oversegmentation, producing tesselated images with defined superpixel boundaries. The original pixel values were then replaced with the average intensities inside each superpixel, obtaining modified images that highlighted the mean intensity of each region.By replacing the detailed pixel-level information with one grayscale value for each superpixel, the complexity of the image is reduced, improving the robustness of subsequent segmentation steps.

Finally, K-means segmentation was performed on the images obtained in the previous step. Applying K-means clustering to the simplified image, allows for more effective clustering. Since superpixels grouped pixels based on their intensity similarities, the accuracy and stability of the clustering process are enhanced. In order to implement the clustering, it was necessary to flatten the superpixels mean-intensity images to 2D arrays. Then, the K-means clustering algorithm was applied using 4 clusters. The number of clusters was determined empirically and by observing the intensity differences in the superpixel images. Cluster labels were reshaped to the original image shape and the cluster centroids were obtained. A binary segmentation mask was then created and the brightest cluster, corresponding to the lesion, was kept in the mask. It is important to mention that with this approach, one binary segmentation mask per scale was obtained. Contours associated with the segmented images were filtered by area, keeping those that had an area between two thresholds obtained empirically.

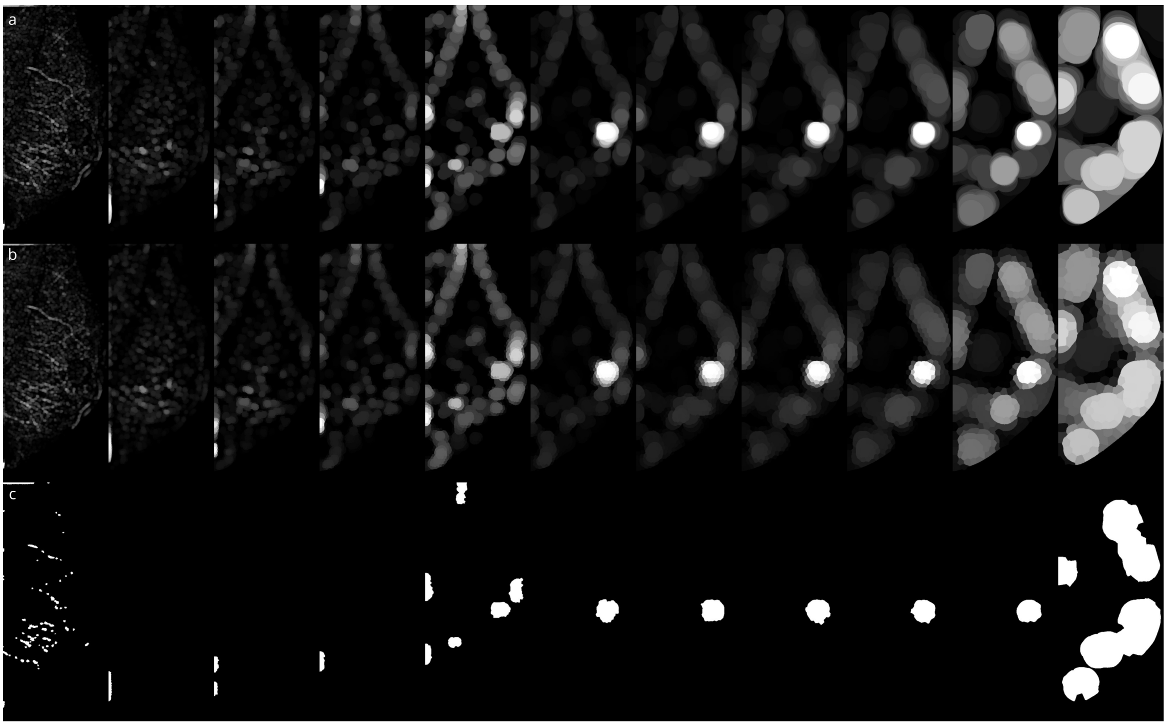

Figure 2: Results of the grayscale morphology and segmentation steps: a. Multi-scale grayscale morphology. b. SLIC superpixels images. c. K-means segmented images.

4 Feature extraction

In order to reduce the number of false positive candidates, a machine learning algorithm was used to discard those that do not belong to true masses. For this, several features associated to all the segmented regions were extracted.

Regarding texture descriptors, these included features derived from the Gray-Level-Co-occurrence-Matrix (GLCM) and Local Binary Patterns (LBP). First, the GLCM was computed for a distance offset of 1 and 4 different angles: 0, $\pi/4$, $\pi/2$, $3\pi/4$. From this matrix, descriptors of contrast, dissimilarity, homogeneity, energy, correlation and Angular Second Moment (ASM) were calculated. Concerning the LBP, the resulting histogram of the LBP image was obtained for 2 sets of parameters: radius equal to 1 and number of neighbours equal to 8, and also for a radius of 2 pixels in a neighbourhood of 16 pixels. By extracting LBP features with varying neighbourhood sizes and radii, it is aimed to capture different levels of texture information from the extracted regions. All these texture features were computed in the bounding rectangles containing the candidates in order to capture and quantify variations between the texture within the segmented area and the background texture.

Since the shape of the detected ROIs can help distinguish the abnormal regions from the normal ones, some shape descriptors were also obtained. These included the perimeter, area, circularity and density of the candidates.

Finally, some general descriptors were computed, which included the mean and the standard deviation of the intensity within the segmented contours. The mean intensity might help identify variations in brightness that could be indicative of abnormalities. The standard deviation indicates the level of variation of the intensity within the lesion, quantifying its level of homogeneity.

The flowchart representing all the steps of the algorithm is shown in Figure 3.

Results

Examples of the segmentation results obtained with the developed method for some of the INbreast dataset images are shown together with their corresponding ground truths in Figure 4. The final segmentation generated 8012 candidates in the positive images, which is 75 candidates per image, and 22407 candidates in the negative images, which is equivalent to 74 candidates per image. For the analysis of the positive images, a match with the ground truth was considered when the Intersection Over Union (IOU) was above 0.2. Based on this criterion, only 421 of the candidates of the positive images were labeled as positive. The other 7591 candidates were excluded from the analysis. The final ratio between positive and negative candidates was of 1:53. The ROC and FROC curves obtained from the results of the classification are shown in Figure 5. The maximum sensitivity achieved was of 99.07%.

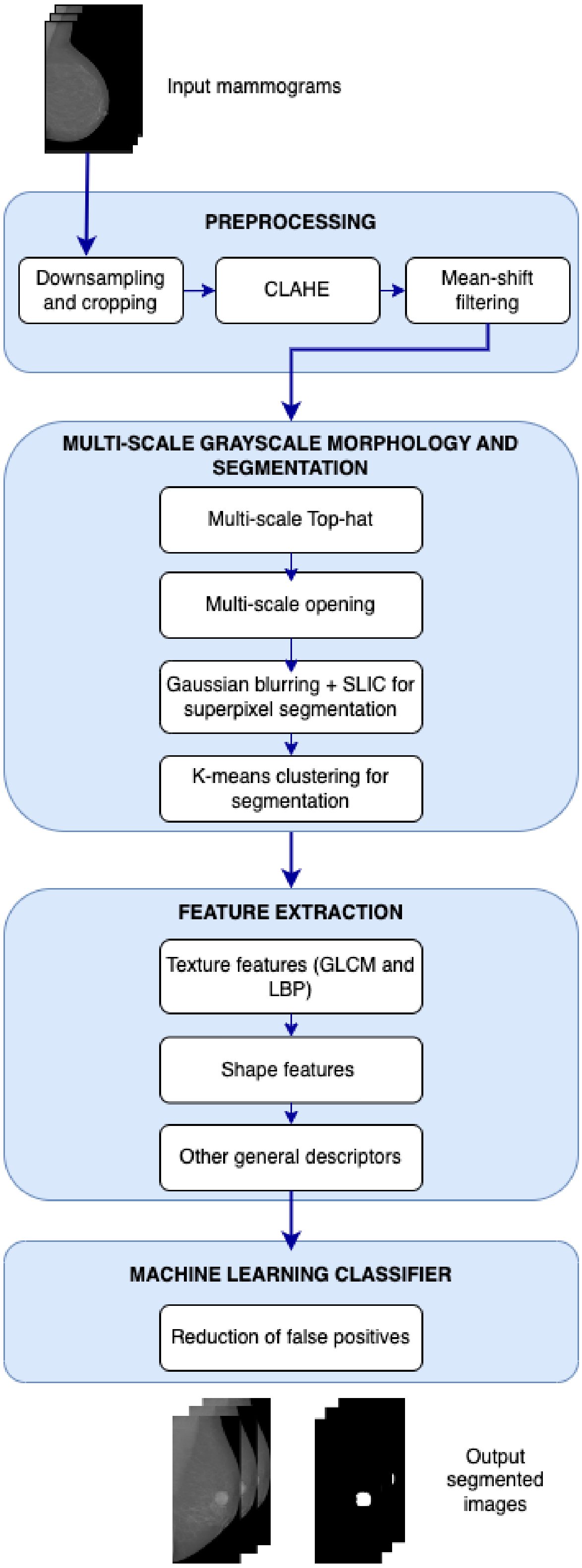

Figure 3: Flow diagram of the developed pipeline.

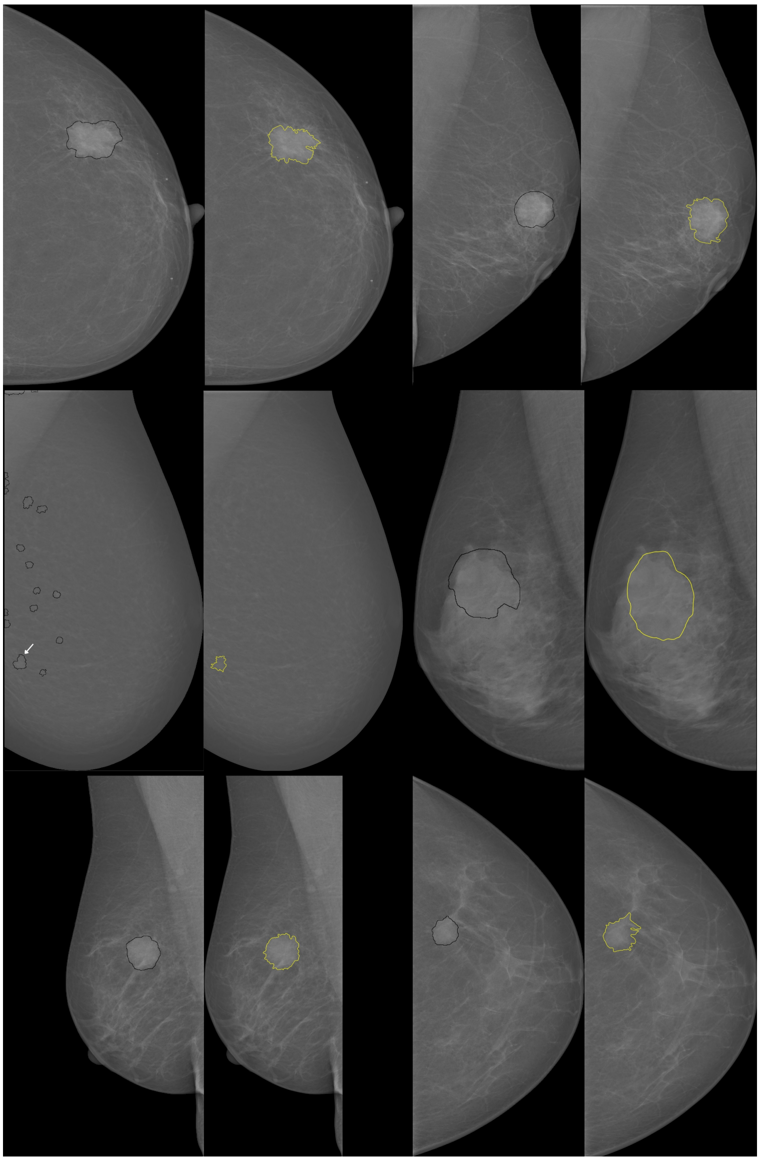

Figure 4: Results of the segmentation. In each pair of images it can be seen the segmented image (left) and the ground truth (right).

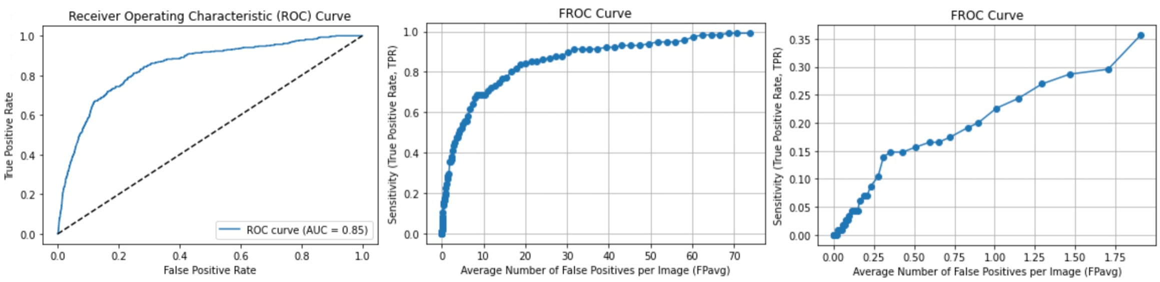

Figure 5: ROC curve and FROC curves illustrating the performance of the classifier and the developed segmentation pipeline. The ROC curve demonstrates an AUC value of 0.85. The FROC curve in the middle depicts the full range of false positives per image, while the FROC curve in the right highlights performance up to a maximum of 1 false positive per image.

Discussion and conclusion

The findings of this analysis revealed that the segmentation algorithm employed in the detection process resulted in a high number of false positives per image. The primary focus during the development of the algorithm was to maximize the sensitivity, which led to the use of a multi-scale approach with numerous scales. This approach resulted in a large number of extracted candidates, leading to an elevated number of false positives per image.

Despite the classifier's success in accurately identifying positive and negative samples (ROC AUC = 0.85), the abundance of negative candidates contributed to a high occurrence of false positives. However, due to the utilization of multiple scales and low IOU threshold for matching a candidate with a ground truth, the sensitivity achieved was considerably high, reaching 99.07%.

Additionally, alternative approaches were also investigated, considering a match between a candidate and ground truth when the IOU exceeded 0.5. For this case, the maximum sensitivity obtained was 90.6%. However, including the features of the negative candidates from positive images resulted in an increased number of false positives and a subsequent decrease in specificity, for both thresholds of 0.2 and 0.5 IOU. Although the inclusion of these negative candidates improved the classifier's performance, as indicated by a higher AUC in the ROC curve, due to the increased sample size, it also led to a higher number of misclassified samples among the negative candidates (because of the considerably higher total number of negative regions).

To improve the results and address the issue of the high false positive rate, it would be recommendable to reconsider the number of scales used in the segmentation process. By reducing the number of scales, it is expected that the sensitivity will decrease, but this adjustment would also lead to a reduction in the number of false positives per image. This would balance better the sensitivity and specificity, ensuring a more accurate and reliable detection system.

In addition, there are other improvements that can be made in the future that could improve the efficiency and performance. These improvements might include normalizing the size of the bounding rectangles from which the GLCM-derived features were extracted, as well as implementing feature selection to keep only the most meaningful features. Also, the efficiency of the algorithm could be improved by feeding only the centroids of the superpixels to the K-means clustering algorithm, instead of all the pixel values in the superpixel means image.

In conclusion, this analysis showed the trade-off involved in optimizing the segmentation algorithm for breast lesion detection. While maximizing sensitivity can yield good results in terms of detecting positive samples, it comes at the cost of an increased false positive rate. Future research can explore alternative approaches, including a reduction in the number of scales, as well as an improvement in the feature extraction pipeline, to achieve a more balanced and effective detection system.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, pp. 209–249, feb 2021.

[2] J. K. Birnbaum, C. Duggan, B. O. Anderson, and R. Etzioni, “Early detection and treatment strategies for breast cancer in low-income and upper middle-income countries: a modelling study,” *The Lancet Global Health*, vol. 6, pp. e885–e893, aug 2018.

[3] H. Min, S. S. Chandra, N. Dhungel, S. Crozier, and A. P. Bradley, “Multi-scale mass segmentation for mammograms via cascaded random forests,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, IEEE, apr 2017.