

# Mass Detection in Mammogram Images

1<sup>st</sup> Pinshun Wang 2<sup>nd</sup> Fenyu Hsieh 3<sup>rd</sup> Fangyu Hsu 4<sup>th</sup> Yahui Chang 5<sup>th</sup> BingLun Li 6<sup>th</sup> Iting Hsueh

**Abstract**—Mass detection in mammogram images is a critical aspect of early breast cancer diagnosis. This project aims to develop an efficient model for accurately identifying and marking masses in mammograms. The proposed approach involves preprocessing DICOM (Digital Imaging and Communications in Medicine) images and utilizing YOLO (You Only Look Once) for initial mass region identification. Subsequently, a CNN (Convolutional Neural Network) is employed to refine and evaluate the likelihood of each identified region being a mass. Additionally, a user-friendly web interface is designed for medical personnel to upload DICOM images directly.

## I. INTRODUCTION

Breast cancer is one of the most prevalent cancers among women globally. Early detection is crucial for treatment and patient survival. Identifying abnormalities in breast X-ray imaging and determining the presence of masses are essential components of this process. Therefore, the goal of this project is to develop a model capable of accurately identifying and marking breast masses in mammography. Our dataset is sourced from EMBED (Emory Breast Imaging Dataset), comprising a large number of DICOM images that combine breast cancer data from different regions and diagnostic images. We selected approximately 6000 DICOM images meeting specific requirements for model training and evaluation. However, DICOM is a medical imaging format used for storing and transmitting medical images. Therefore, we converted it to PNG format to better meet the subsequent needs of model analysis and we also apply a series of preprocessing to the images before using it in our model. When conceptualizing our model architecture, we found two main approaches from existing research: one utilizing specific algorithms for mass separation followed by CNN analysis, and the other relying solely on YOLO for detection and classification. Ultimately, we decided to combine these two approaches. YOLO is renowned for its fast real-time object detection, quickly identifying regions in the image that may contain masses. Subsequently, CNN is used to further analyze and enhance the accuracy of detection in identified potential regions. The training process involves using 5000 images for YOLO training and 618 for validation, utilizing the YOLOv7 model. CNN training involves 3328 images of masses and backgrounds, using CovNeXt and ResNeXt models. We merged YOLO and CNN for final evaluation, striking a balance between YOLO detection threshold and CNN performance.

## II. METHODS

### A. Overview

The concept of our framework is shown in the Fig. 1. In our training phase, DICOM files undergo an initial preprocessing

step to extract features with enhanced clarity, resulting in PNG-formatted images. Subsequently, these images are separately input into both Convolutional Neural Network (CNN) and You Only Look Once (YOLO) architectures for model training. YOLO takes ROI coordinates from metadata as ground truth to train a model capable of marking the areas that might be masses. CNN is trained to discern whether specific features in the images indicate the presence of masses.

In summary, our approach involves preprocessing the dataset, followed by YOLO identifying potential mass features based on ROI coordinates from metadata. These identified features are then input to CNN, which makes the final determination on whether they correspond to masses.

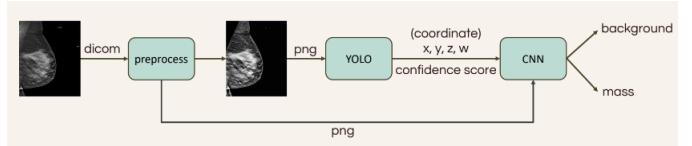


Fig. 1. Overall Design Framework.

### B. Data

1) **Datasets:** We chose EMBED as our dataset for its extensive collection of mammographic exams, which includes detailed Regions of Interest annotations for masses.

The EMBED (Emory Breast Imaging Dataset) utilized in this study is a comprehensive collection of mammographic exams, encompassing a total of 364,000 screenings and diagnostic tests from 110,000 patients across four hospitals over an eight-year period. This dataset predominantly includes 2D and C-view mammographic images. This dataset is instrumental in facilitating advanced studies in breast imaging and cancer diagnosis. [1]

2) **DICOM Preprocessing:** The purpose of preprocessing DICOM is to convert the file format into a format that the model can read and to enhance the efficiency of subsequent preprocessing steps. The following will sequentially introduce how to convert DICOM to PNG, improve the readability of PNG, and synchronize the metadata table within the dataset.

- **Step 1: Transfer DICOM to PNG.**

Below is the information we accessed in the DICOM files

- **To ensure uniform breast positioning:** The attributes Laterality, ViewPosition, and PatientOrientation in DICOM are utilized to determine if the image needs flipping.

- **To obtain the pixel array mapping from DICOM to an 8-bit PNG:** Access to the VOI LUT Function within the DICOM file is necessary.
- **To comprehend pixel value interpretation:** The Photometric Interpretation attribute in DICOM provides relevant information.
- **For windowing in the resulting PNG from DICOM:** It is essential to read the DICOM attributes for window center and window width.

- **Step 2: Windowing**

By utilizing window width (WW) and window center (WC), it is possible to map the pixel array of DICOM images to an 8-bit PNG pixel array, highlighting the crucial color regions of the image. This facilitates subsequent preprocessing and model operations for improved analysis and interpretation.

- **Step 3: Update The Metadata Table**

Standardizing the format of the 'ROI coord' field in the unified metadata table to enhance the smooth retrieval and utilization of data by subsequent models and preprocessing procedures.

3) *PNG Preprocessing:* Preprocessing of images is an essential step in readying data for both training and application in the model. This process ensures uniformity in the input to the neural network, enhances image clarity. We used the following methods to preprocess raw mammogram X-ray images from the EMBED database. [2]

- **Step 1: Normalization of Breast Orientation**

Aligning all PNG files to a uniform orientation for enhanced overall consistency, facilitating subsequent model training and preprocessing.

- **Step 2: Handle Images with Paddle**

When capturing breast images, different types of paddles can be utilized to focus on specific tissues for observation. Among them, spot compression and magnification are the paddle types we aim to address. Spot compression involves compressing a specific tissue area to enhance its visibility, allowing for targeted observation of that region; magnification, on the other hand, involves enlarging a specific tissue area, making it the focal point in the imaging. Since the paddles used in breast imaging are made of metal, the pixels in the paddle can be particularly intense, leading to potential confusion in the results of subsequent models and preprocessing. Therefore, it is necessary to employ techniques to remove the paddles and extract the underlying tissue information from within the paddles.

- **Step 3: Remove Noise on Mammogram Images**

Raw mammogram X-ray image has noise and small holes in the image. Using bilateral filtering is able to reduce the noise and fill in the holes in the image. We import bilateral filter from OpenCV. Subsequent to the application of the bilateral filter algorithm, an attempt the switching bilateral filter was made. However, this endeavor was unsuccessful. It requires processing each

pixel individually, however our images include extensive pixel count resulting in a prohibitively time-consuming computation.

- **Step 4: Remove Artifacts by Selecting Breast Area**

In a raw mammogram X-ray image, labels manually added by individuals may affect the accuracy of the model in detecting masses. Therefore, it is necessary to remove these labels and retain only the breast area. Since labels are typically smaller compared to the breast area, we can use Morphological Transformations to eliminate them. For this purpose, we import the "opening" function from OpenCV to effectively remove the labels. The results of the mammogram image after breast tissue enhancement is illustrated in Fig. 2.

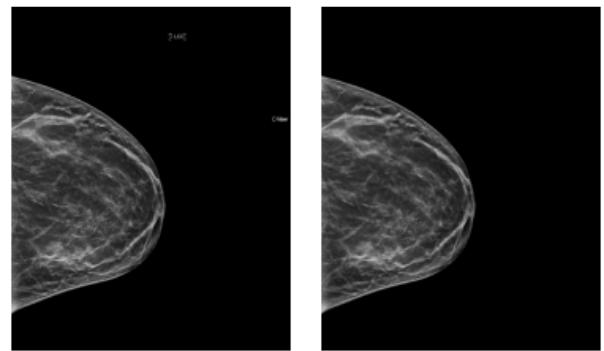


Fig. 2. Example of a mammogram image result of removing artifacts by selecting breast area: the original image and the image that the artifacts removed. (from left to right).

- **Step 5: Enhancing the Contrast of Breast Tissue**

This step involves the amplification of breast tissue contrast. We achieved this through using CLAHE algorithm, designed to intensify the contrast in grayscale images. CLAHE operates by dividing the image into small blocks, known as tiles, and then applying histogram equalization to each of these independently. The results of the mammogram image after breast tissue enhancement is illustrated in Fig. 3.

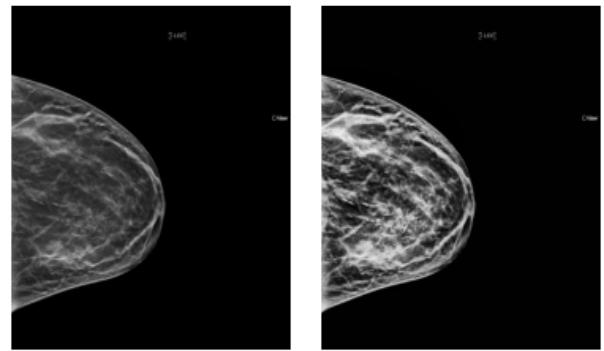


Fig. 3. Example of a mammogram image result of the breast-tissue using CLAHE algorithm: the original image and the image with contrast enhancement (from left to right).

- **Step 6: Removing Pectoral Muscle**

The step incorporates a series of algorithms to eliminate the pectoral muscle representation in mammogram images. Removing pectoral muscle is a crucial step in the image processing sequence, since it displays a similar intensity to anomalies. The result of this step is shown in Fig. 4 [3].

The algorithm applied for the removal of the pectoral muscle uses the Hough transformation in the following steps:

Step 1: Find the region of interest. Since our images are uniformly oriented, with the pectoral muscle consistently positioned in the upper left corner, we specifically target this area for analysis.

Step 2: Apply the median blur filter. Since the images may contain some noise, therefore we choose to use the median blur filter to reduce noise. We also tried out some other filters but found out the median blur filter best fit our expectation.

Step 3: Apply the Canny filter for contour detection, the objective of which is to identify the edges and contours. The Canny filter achieves this by looking for places where the brightness changes sharply to find edges. Then, it will isolate strong and relevant edges.

Step 4: Line detection using the Hough Transform. In this step, it utilizes the Hough Transform to detect lines in the image that has been processed by the Canny filter. It identifies potential lines, then calculates their positions and orientations.

Step 5: Find the shortlist relevant lines. This step involves filtering the lines detected in the previous step to identify the most relevant ones for our analysis. The function begins by setting angle and distance threshold. It selects those whose distance and angle align with our predefined threshold. This would narrows down lines to those that meet our specific geometric criteria.

Step 6: Apply a Hough mask to the region of interest. First, it takes the list from the previous step and sorting these lines based on their distance from the image origin. Then, it selects the line that is closest, identifying it as the pectoral line and used it to mask and remove the pectoral muscle from the image.

### C. Model

*1) Innovative Model Structure:* Inspired by this paper [4], which utilized morphology method to propose candidate regions of masses, we introduced an innovative two-stage model structure, leveraging YOLOv7 as a stronger preliminary mass detector and employing a CNN model to validate the output generated by YOLO. This strategic combination aims to achieve superior performance compared to relying solely on YOLO. Moreover, we anticipate that leveraging the high accuracy in object detection achieved by YOLO will contribute to further improving the overall results.

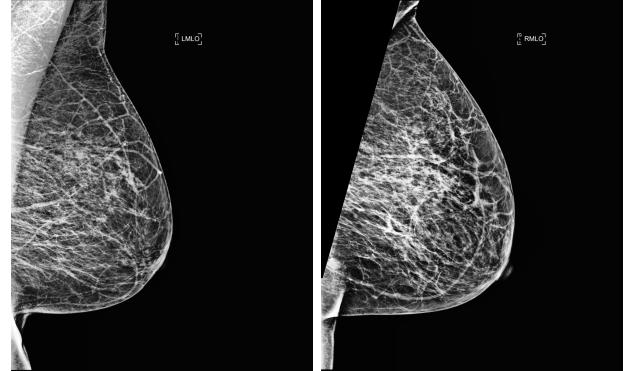


Fig. 4. the image after png preprocessing step 1 to 5 and the image removing pectoral muscle (from left to right).

As depicted in the Fig. 5, we input preprocessed images into YOLO to identify objects resembling a mass, along with associated confidence scores and coordinates. Subsequently, these information (coordinate, confidence scores) is transmitted to the CNN, which ultimately produces the final prediction based on the provided data.

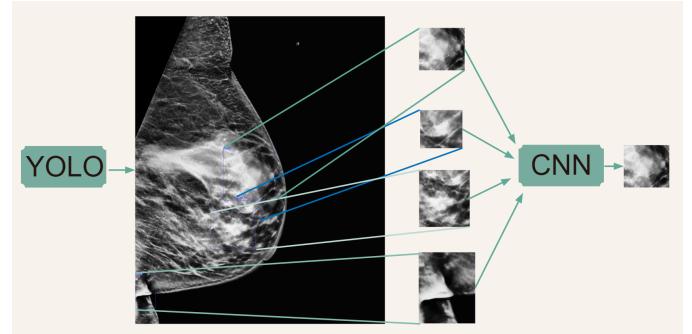


Fig. 5. the model structure

#### 2) YOLOv7:

- **Purpose of YOLO**

As mentioned earlier, we consider YOLO as a preliminary mass detector in this paper. Consequently, our focus is not on achieving high accuracy but rather on maximizing recall. As a result, we plan to adjust the confidence threshold after training to produce more objects resembling a mass in the output. This intentional adjustment is expected to lead to lower precision but higher recall.

- **Training Preparation**

Prior to initiating our training process, several preparatory steps are imperative for refining YOLO's labels and filtering out undesirable data. Initially, we will eliminate duplicate Regions of Interest (ROIs) by establishing a specific threshold for the Intersection Over Union (IOU) in the datasets, set up of the threshold will be discussed later in CNN's section. Subsequently, any ROIs that are entirely black will be excluded. Lastly, ROIs with an area exceeding 1000\*1000 pixels will undergo removal as a third step. Implementing these three measures not only

enhances the performance of YOLO but also speed up the training process.

#### • Train Validation Data Split

After removal of undesirable data in the original datasets, we obtain 5618 Mammogram images in the end. We split 5000 images as training data and 618 images as validation data in YOLO.

#### • Training Process

The effectiveness of transfer learning in deep neural network training is demonstrated. We utilized a pre-trained weight provided by YOLOv7 to train our YOLO, employing 3 epochs as a warm-up phase and continuing training for a total of over 100 epochs. Despite the training duration, precision and recall stop improving around the 100th epoch, prompting us to early stop the training process.

To expedite the training process and reduce memory usage (that is, having larger batch size), we compressed training images to sizes of 640\*640 and 1280\*1280 pixels. However, not only the worse performance in precision and recall at the larger size, the training time for 1280\*1280 images was also prolonged. Consequently, we decided to terminate the training process with the 1280\*1280 size.

#### • Preliminary Mass Detector

YOLO will provide a confidence score for each identified object, such scores indicate that the model's confidence in its prediction that an object is present within that bounding box. Consequently, we can set up a certain threshold to filter out some wrongly predicted mass. A lower threshold will increase the number of identified objects output by YOLO, thereby improving recall but decreasing precision. Conversely, a higher threshold will yield fewer identified objects, leading to higher precision but lower recall. As a result, we conduct an experiment in order to fine-tune the confidence threshold and achieve better overall performance.

#### 3) CNN:

##### • Purpose of CNN

As previously indicated, following the preliminary detection of masses by YOLO, a secondary-stage classification is performed using a Convolutional Neural Network (CNN). The objective is to enhance the precision of YOLO while concurrently preserving its recall, thereby improving the overall performance of the final system.

##### • Data labeling

The data generated by YOLO includes the coordinates and confidence scores for each image, representing the location of masses in the mammograms. In this section, we discuss the labeling process for each Region of Interest (ROI) image as either "mass" or "background." Initially, we compare the Intersection over Union (IOU) between the ground truth provided by the EMBED dataset and the ROI coordinates provided by YOLO in each mammogram. An observation reveals that the ROI areas

provided by EMBED are consistently larger than those identified by YOLO. Consequently, an appropriate IOU threshold is crucial to prevent mislabeling.

To determine the optimal threshold, we systematically analyze the ROI images within various IOU intervals. Fig. 6 illustrates successfully detected masses when the IOU is between 0.1 and 0.15, while Fig. 7 demonstrates instances where masses are not correctly cropped when the IOU is between 0.05 and 0.1. The upper images in the figures depict the ground truth, while the lower images represent the output generated by YOLO. Consequently, by observation from every intervals of IOU, we decide on an IOU threshold of 0.1. If the IOU between the ground truth and the data generated by YOLO exceeds 0.1, the data is labeled as "mass"; otherwise, it is labeled as "background."

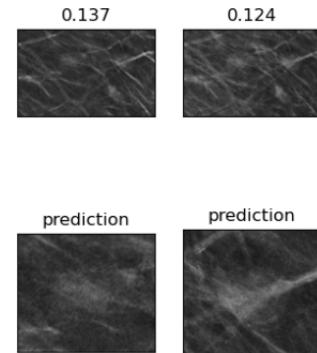


Fig. 6. IOU between 0.1-0.15 ROI images

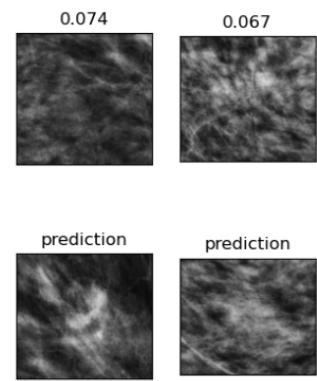


Fig. 7. IOU between 0.05-0.1 ROI images

##### • Data splitting

618 mammogram images from YOLO's validation data are utilized as testing data for the final evaluation of our overall performance. Moreover, 5000 mammogram images from YOLO's training data are partitioned, with 80 percent allocated for training and 20 percent for validation in the context of the CNN model. Finally, we identify 5326 detected ROI images in the training dataset,

comprising 2663 mass images and 2663 background images. Additionally, the validation dataset consists of 1332 detected ROI images, encompassing 666 mass images and 666 background images.

- **Five different methods applied in training CNN**

To improve the performance of the CNN model, we investigated five distinct approaches for classifying masses, outlined as follows. Additionally, the experiments were conducted using PyTorch version 2.1 and CUDA version 12.1 within the Anaconda environment with Jupyter Notebook. The models were trained on Nvidia RTX3060ti GPU.

- **Transfer learning on a single CNN model**

Initially, we employed a CNN model to classify masses detected by YOLO, specifically opting for the ConvNeXt-tiny and ResNeXt50\_32x4d models from PyTorch. These models were initialized with pretrained weights "IMAGENET1K\_V1." The selection of these models was driven by the advantageous utilization of transformers in mass detection, enabling the capture of long-range dependencies within medical images. Unlike traditional convolutional neural networks, transformers leverage self-attention mechanisms, facilitating the efficient consideration of global context information. This capability proves particularly beneficial in mass detection tasks, where the spatial relationships between masses and their surroundings play a crucial role. However, despite the anticipated advantages, our testing results did not meet the expectations, yielding only a 65 percent accuracy for both ConvNeXt-tiny and ResNeXt50\_32x4d models.

- **Ensemble method with soft-voting**

During our investigation of mass pathology classification tasks in the literature, we observed that some researchers employed ensemble methods for mass classification. Given that our existing model is a two-stage architecture, involving substantial training and inference time, we opted for a simpler approach. Instead of training new neural networks for ensemble purposes, as demonstrated in [5], we opted to leverage the probabilities generated by two pre-trained models, ConvNeXt-tiny and ResNeXt50\_32x4d, as discussed in previous approach. Our aim was to enhance accuracy through a straightforward averaging of probabilities. However, the achieved improvement in accuracy was merely 0.01-0.02 percent, falling short of our expectations.

- **Ensemble method with hard-voting**

As both previously mentioned models yielded suboptimal performance, even when employed in an ensemble approach, our focus shifted towards extracting more information from ROI images by leveraging the confidence scores provided by YOLO. To enhance performance, we refined the soft-voting strategy used earlier, incorporating the YOLO confidence scores.

Upon analyzing the training data, we observed that mass images received an average confidence score of 0.163, while background images only received 0.11 on average. Consequently, we opted to classify ROI images based on their YOLO confidence scores. Specifically, we designated ROI images with confidence scores greater than 0.15 as mass and those with scores equal to or below 0.15 as background. This information was then combined with the pre-trained ConvNeXt-tiny and ResNeXt50\_32x4d models using a hard-voting approach. Contrary to our expectations, the results deteriorated, suggesting that this approach may not be effective for combining YOLO confidence scores in this context.

- **Fusion network for CNN and YOLO**

Since the YOLO threshold determined in the previous hard-voting approach was evidently suboptimal, while determining a fair and effective threshold proved challenging. Consequently, we pursued an alternative strategy by training a fusion network to learn the classification of masses using both the YOLO confidence score and the probabilities output by the pre-trained ConvNeXt-tiny and ResNeXt50\_32x4d models. The rationale was that a machine learning model might better discern these features than manually selecting a YOLO threshold.

The fusion network comprised a fully connected layer with dimensions 5x32 and a ReLU activation function, followed by a 32x2 fully connected layer with a sigmoid activation function. The Adam optimizer with a learning rate of 0.001 was employed, and the loss function was set to crossentropy loss. During the training process, we observed that the fusion network achieved a maximum accuracy of only 67 percent. This outcome suggested that the YOLO confidence score did not provide significantly more information about mass images for CNN models. It implies that certain masses are inherently challenging to classify for both CNN models and YOLO, making the combination of YOLO confidence scores an ineffective method.

- **Concatenate CNN model with SVM**

Recognizing the limitations of the YOLO confidence score, we explored an alternative approach inspired by traditional R-CNN methods [6]. This involved utilizing a CNN model as a feature extractor and employing SVM for classification. Recognizing that relying solely on CNN-extracted features might be insufficient, we extended our methodology to include a radiomics approach. This involved extracting handcrafted features from ROI images, which were then combined with the CNN-derived features to enhance overall performance. For the handcrafted features, 83 features are extracted through a radiomics approach as Table I shows [7] [8]. Subsequently, feature selection, involving 4 steps in total. Step 1, features are scaled using Z-score for uniformity in magnitude. Step 2, is conducted via

ANOVA F-test, seeking the most significant features for the models. Step 3, a Pearson Redundancy-Based Filter (PRBF) is applied to mitigate multicollinearity by identifying and removing features that are highly correlated with one another beyond a certain threshold. Step 3, Backward Elimination using Ordinary Least Squares (OLS) regression for feature selection. It starts with all candidate variables and systematically removes the least significant variables until all variables in the model are statistically significant. Step 4, using Recursive Feature Elimination (RFE) with XGBoost as the base estimator to select the ten most important features from a set that has already been pruned by backward elimination. Subsequent to the identification of the ten most significant features, we visualized them using a plot diagram. As illustrated in Fig. 8, it was observed that the distributions of 'mass' and 'background' classes for four features are indicative of the similar distribution patterns found across all selected features. To further our analysis, we proceeded to plot the remaining features. However, this extended examination revealed that none of these additional features demonstrated noticeably distinct distribution patterns. The observed similarity in feature distribution for 'mass' and 'background' categories in our study, as opposed to the distinct distributions reported in other works, may be attributed to the differences in methodological application. The referenced papers utilized these methods for discerning benign from malignant masses, a task for which the features are presumably more discriminating. In contrast, our study aimed to differentiate between 'mass' and 'background' as determined by YOLO's outputs, which may inherently bear closer feature similarities.

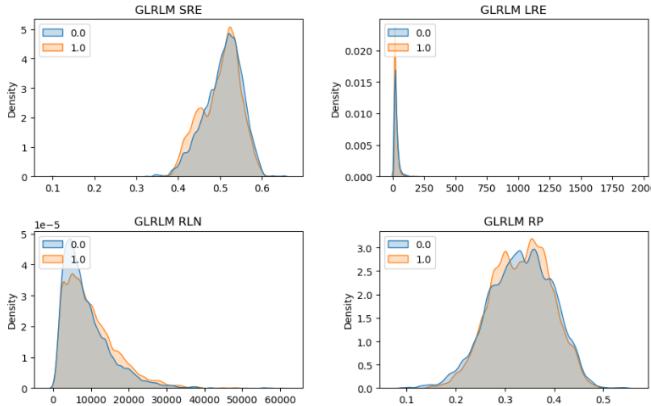


Fig. 8. distribution of four key features for 'Mass' (Class 1 in orange) and 'Background' (Class 0 in blue).

As for the feature extracted from pretrained ConvNeXtiny and ResNeXt50\_32x4d models, concerning the features extracted from these two models, when combined, the total number of features amounted

TABLE I  
STATISTICAL MEASURES OF RADIOMICS FEATURES

| Feature categories     | Feature names (optional)            | Dimensions |
|------------------------|-------------------------------------|------------|
| First-order statistics | Energy, Entropy, ...                | 18         |
| GLCM                   | Autocorrelation, Cluster Shade, ... | 24         |
| GLSZM                  | Gray Level Variance, ...            | 16         |
| GLRLM                  | Gray Level Run Emphasis, ...        | 16         |
| NGTDM                  | Coarseness, ...                     | 5          |
| GLDM                   | Dependence Entropy, ...             | 14         |

to 2816, we employed the feature selection method called VarianceThreshold from scikit-learn to eliminate features with low variance. This process resulted in a reduction from 2816 features to 526 features. However, despite this feature reduction, the accuracy achieved by SVM with linear kernel was only 64 percent, which was lower than using only a single CNN model to classify.

### III. RESULTS

#### A. YOLO

In this section, we will showcase the outcomes of YOLO's predictions under various confidence score thresholds. Given that YOLO serves as our preliminary mass detector, it becomes imperative to fine-tune the threshold in order to optimize recall while maintaining an acceptable level of precision.

In Fig. 9, we note that under YOLO's default confidence threshold of 0.25, the model can detect some evident masses. However, non-round masses are occasionally overlooked. Therefore, we opted to further reduce the confidence threshold to 0.15. In Fig. 9, YOLO's predictions under the 0.15 threshold are depicted, revealing distinctions compared to the predictions under the 0.25 threshold. Unlike the 0.25 threshold, the 0.15 threshold results in the detection of more objects, thereby improving recall, albeit at the risk of misclassifying some tissues as masses. Subsequently, to capture more objects resembling masses, we also conducted tests with thresholds of 0.10 and 0.05, as illustrated in Fig. 10. Furthermore, Table II presents the precision, recall, and f1-score for 618 YOLO's validation images, evaluated by IOU at each confidence threshold.

TABLE II  
PERFORMANCE OF DIFFERENT CONFIDENCE THRESHOLD IN PRELIMINARY MASS DETECTOR (YOLO)

| Threshold | Precision | Recall | F1    |
|-----------|-----------|--------|-------|
| 0.25      | 0.615     | 0.194  | 0.295 |
| 0.15      | 0.442     | 0.381  | 0.410 |
| 0.15      | 0.333     | 0.512  | 0.403 |
| 0.05      | 0.204     | 0.710  | 0.316 |

#### B. CNN

To assess the accuracy of mass predictions, we compute the Intersection over Union (IOU) between Region of Interest

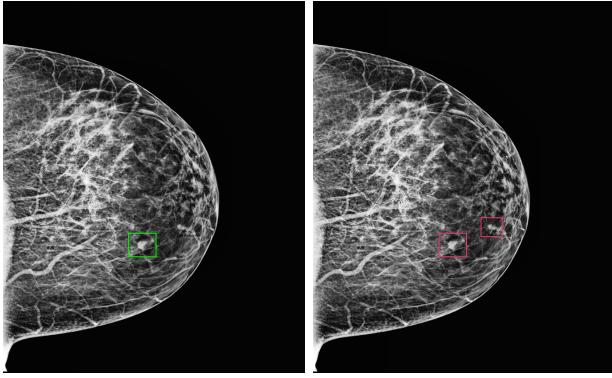


Fig. 9. Left: Threshold 0.25 image; Right: threshold 0.15

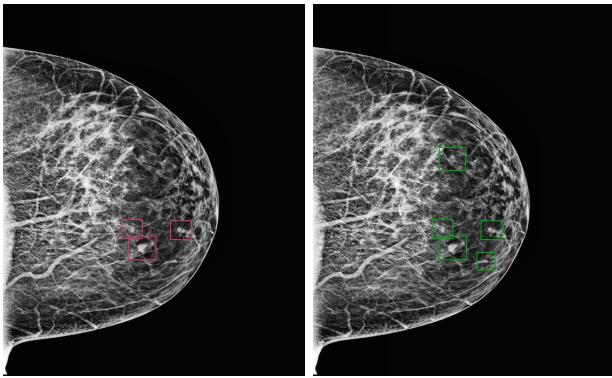


Fig. 10. Left: threshold 0.1; Right: threshold 0.05

(ROI) images and ground truth, as elaborated in Section 2.2.3 during the data labeling process. Table III comprehensively presents the accuracy, precision, recall, and F1-score for the five distinct methods outlined in Section 2.2.3 under the testing data. The fusion network is not included in the table, as the training accuracy remains at 67 percent, and we did not perform additional evaluation. The top two methods in the table indicate the use of the single CNN model, while the other methods remain consistent with the descriptions in Section 2.2.3.

TABLE III  
PERFORMANCE OF DIFFERENT METHODS IN MASS CLASSIFICATION

| Method        | Accuracy | Precision | Recall | F1    |
|---------------|----------|-----------|--------|-------|
| ConvNeXt-tiny | 0.659    | 0.666     | 0.643  | 0.654 |
| ResNeXt50     | 0.652    | 0.651     | 0.655  | 0.653 |
| Soft-voting   | 0.667    | 0.672     | 0.653  | 0.663 |
| Hard-voting   | 0.654    | 0.694     | 0.551  | 0.614 |
| CNN + SVM     | 0.648    | 0.648     | 0.649  | 0.648 |

### C. YOLO+CNN

For the evaluation of our overall model performance, we introduced the ratio R as (1).

$$R = \frac{\text{increase in precision}}{\text{decrease in recall}} \quad (1)$$

This ratio is utilized to compare the original YOLO performance with the performance after concatenating it with CNN models. We anticipate R value larger than one, indicating that the CNN model removes more background images than mass images. In Table IV below, different YOLO confidence scores are employed for the preliminary detection of masses, followed by classification with the pretrained ConvNeXt-tiny model. Since different methods shown in Table III have the similar results when concatenating with YOLO, so we only show the result of using convNeXt-tiny. The ratio R in the table is calculated by comparing the performance between Table III and Table IV. The observed trend reveals that R increases as the YOLO confidence score threshold becomes higher. This suggests that the CNN model performs better on images with higher YOLO confidence scores, indicating that the features learned by YOLO and the CNN model may be similar. Consequently, the YOLO confidence score is proportional to the CNN output logits. Additionally, the results from the fusion network presented in Table III also support this conclusion since combining the confidence score in the fusion network does not improve the model performance.

TABLE IV  
PERFORMANCE FOR YOLO+CNN UNDER VARIOUS YOLO CONFIDENCE SCORE THRESHOLDS

| Threshold | Precision | Recall | F1    | R     |
|-----------|-----------|--------|-------|-------|
| 0.05      | 0.334     | 0.498  | 0.399 | 0.614 |
| 0.1       | 0.480     | 0.376  | 0.422 | 1.084 |
| 0.15      | 0.596     | 0.304  | 0.403 | 1.986 |
| 0.2       | 0.698     | 0.225  | 0.340 | 3.614 |

## IV. APPLICATION

### A. Front end

we design an user interface via a webpage. In order to allow user upload the image without any transformation. Our website enable users to upload DICOM images directly. As the Fig. 11 show, the model will convert these image into PNG format and display them on the left side of the website after preprocessing. And then, our model will identify and mark the masses, displaying the results on the right side.

### B. Back end

After receiving PNG images from the frontend, backend preprocessing is initiated to optimize the images for analysis. The processed results are first transmitted back to the frontend for visual representation. Subsequently, the preprocessed images undergo YOLO analysis using previously trained parameters to identify potential mass locations. The features identified as potential masses are then passed to a Convolutional Neural Network (CNN) for a final determination of their classification as actual masses.

In the concluding phase, the images with annotated features, along with information regarding the classification of each feature as a mass or non-mass, are communicated back to the frontend for visualization.

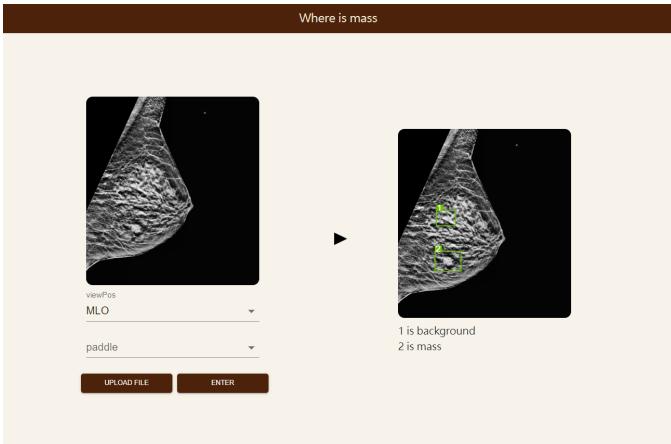


Fig. 11. the application structure

## V. CONCLUSION

Firstly, YOLO's identification of the background, resembling the actual context, poses a challenge for our CNN model. In our system, we initially employ YOLO for object detection. When the background shares visual features with the objects, YOLO might mistakenly identify background regions as objects or extract features insufficient for accurate differentiation. Specifically, when YOLO identifies background regions resembling actual breast cancer lesions, such as certain tissue structures, shadows, or other visual features, it may incorrectly classify these areas as potential objects containing masses. This similarity can lead to inaccuracies in the generated labels or regions, subsequently impacting further analysis conducted by the CNN. For our CNN model, the presence of these similar regions makes it challenging to distinguish which features correspond to genuine breast cancer lesions, thereby reducing the model's ability to accurately detect masses. This predicament adversely affects the overall performance of the model, making it challenging to achieve satisfactory accuracy and recall. Secondly, in our attempt to explore various datasets, we encountered challenges as many datasets were not sufficiently comprehensive. Some did not provide ROI coordinates, while others had insufficient data volume. After finally identifying a dataset that better suited our requirements, we encountered issues during the practical implementation. We observed mislabeling errors in the Regions of Interest (ROIs), with many ROIs outlining completely black areas. These inaccuracies introduced confusion, adversely affecting the training process and yielding suboptimal model results. In future work, addressing the mentioned issues and optimizing our methodology could offer a pathway for improving the performance of the breast cancer mass detection model.

## VI. DATA AND CODE AVAILABILITY

You can get more code information about our project with the following link.

<https://github.com/smilingweixiao/MLTeam28>

## AUTHOR CONTRIBUTION STATEMENTS

- **Pinshun Wang:** Mass detection and classification related paper research, CNN model training and evaluation, Concatenate YOLO with CNN and evaluate the performance, Write CNN and result part in the report
- **Fenyu Hsieh:** Data preprocess method researching, Remove noise on images, Enhance the contrast of breast tissue, Remove pectoral muscle, Find other datasets and do some handle, Handcraft feature extraction and selection, Help to train CNN, Managed front-end and back-end integration of the website, Write dataset, preprocess introduction, step 3, 5, 6, and feature extraction, selection in report
- **Fangyu Hsu:** Data preprocess method researching, remove the artificial labels on the raw mammography. Find other datasets and do some handle. Design the UI of the website. Write the PNG preprocess step 4, Application Front end, model structure figure and do final check in report. Prepare and conclude the information for presentation. Presentation speaker. Prepare meeting minute.
- **Yahui Chang:** Classify the dataset into with mass and without mass. DICOM preprocessing. Fix the error in the metadata table. Extract the tissue in the paddle. Enhance ROI image. Maintain the API for preprocess Maintain website backend. Maintain README in repo. Write overview, DICOM preprocessing, PNG preprocessing step 1 and 2, backend in the report.
- **BingLun Li:** Mass detection and classification related paper research, YOLO model training and evaluation, prepare meeting agenda, write YOLO and result part in the report.
- **Iting Hsueh:** Mass detection and classification related paper research, YOLO model training, write abstract, introduction and conclusion in the report.

## REFERENCES

- [1] A. T. T. R. R. M. G.-I. G. M. C. M. I. J. H. Jiwon Jeong, Brianna L. Vey, “The emory breast imaging dataset (embed): A racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images,” *Radiology: Artificial Intelligence*, 2023.
- [2] K. A. D.-V. N. K. M. G. M. A. N. Ruchaya, V. I. Kobera, “Segmentation of breast masses in digital mammography based on u-net deep convolutional neural networks,” *Journal of Communications Technology and Electronics*, 2022.
- [3] T. J. O. V. D. T. K. Pascal Vagssa, Nafissatou Mallam Doudou, “Pectoral muscle deletion on a mammogram to aid in the early diagnosis of breast cancer,” *International Journal of Engineering, Science and Technology*, 2020.
- [4] W. J. W. S. Z. Y. X. Y. Sun L, Sun H, “Breast mass detection in mammography based on image template matching and cnn,” *Sensors*, 2021.

- [5] A. S. E. Asma Baccouche, Begonya Garcia-Zapirain, “An integrated framework for breast mass classification and diagnosis using stacked ensemble of residual neural networks,” *Scientific reports*, 2022.
- [6] T. D. J. M. Ross Girshick, Jeff Donahue, “Rich feature hierarchies for accurate object detection and semantic segmentation,” 2014.
- [7] Y. Q. B. Z. Meredith A. Jones, Rowzat Faiz, “Improving mammography lesion classification by optimal fusion of handcrafted and deep transfer learning features,” *Physics in Medicine Biology*, 2022.
- [8] G. Z. J. H. Y. J. Y. C. Min Li, Liyu Zhu, “Predicting the pathological status of mammographic microcalcifications through a radiomics approach,” *Intelligent Medicine*, 2021.