Breast Cancer Detection using Deep Learning

Author
Goron Paula-Melisa

Reasons

CANCER

- Disease of the century
- A medical condition that can affect any part of the body
- May be caused by an abnormal growth of cells

DETECTION

- It helps with early cancer detection by identifying anomalies in medical images
- Detection improves patients' health and saves lives

Related work

SOURCE: A COMPETITION, BENCHMARK, CODE, AND DATA FOR USING ARTIFICIAL INTELLIGENCE TO DETECT LESIONS IN DIGITAL BREAST TOMOSYNTHESI

DATASET

- 22032 FILES
  - training 19148
  - validation 1163
  - test 1721
  - 220 images with bounding boxes for malignant or benign lesions.

RESULTS

Team 7: The model's ability to detect correctly is 0.786, meaning it successfully identified 78.6% of individuals with breast cancer. With 95% confidence, there is some level of uncertainty associated with the estimate, and the model's detection capability can range from 0.720 to 0.852."

Table 2. Challenge Results[a]

Ranking	Team name	Affiliations	Methods	Training set	Mean sensitivity for biopsied lesions (95% CI)	Phase where team achieved best performance	Code available
1	NYU B-Team	New York University–Langone Health	Phase 1: EfficientDet, Max-Slice-Selection, and Augmentation and Ensembled Perturbations; phase 2: phase 1 methods with cancer cell prediction head and multilocation crop	Phases 1 and 2: DBTex1 and internal data set	0.957 (0.924-0.984)	2	No
2	ZeDuS	IBM Research–Haifa	Phase 1: RetinaNet ensemble with heatmap NMS; phase 2: phase 1 methods with SWIN[46] and NFNet[47]	Phases 1 and 2: DBTex1 with internal data set	0.926 (0.881-0.964)	2	Yes, both phases[48]
3	VICOROB	VICOROB–University of Girona	Phase 1: Fast R-CNN, ensembled; phase 2: phase 1 methods with FP reduction (no ensemble)	Phases 1 and 2: DBTex1 with OPTIMAM/OMI-DB	0.886 (0.836-0.930)	2	Yes, both phases[49,50]
4	Prarit	Queen Mary University of London–CRST and School of Physics and Astronomy	Unknown	Unknown	0.822 (0.754-0.884)	1	No
5	UCLA-MII	UCLA Medical & Imaging Informatics	Phase 1: Faster R-CNN, FPN,[51] IoSIB, and Blob Detector	Phase 1: DBTex1	0.814 (0.751-0.875)	1	Yes, phase 1[52]
6	Pranjalsahu	Stony Brook–Department of Computer Science	Phase 1: Faster R-CNN with Confidence Peak Finder	Phase 1: DBTex1	0.790 (0.717-0.854)	1	Yes; phase 1[53]
7	Team-PittRad	University of Pittsburgh–Department of Radiology	Phase 1: YOLOv5[54] and Cross Stage Partial Networks	Phase 1: DBTex1	0.786 (0.720-0.852)	1	Yes, phase 1[55]
8	Coolwulf	Unknown	Unknown	Unknown	0.390 (0.301-0.475)	1	No
NA	Baseline model[b]	NA	Faster R-CNN	DBTex1	0.379 (0.304-0.456)	NA	Yes[56]
NA	Data set baseline model[b]	NA	DenseNet[32]	DBTex1	0.444 (0.366-0.523)	NA	Yes[45]

Solution

CNN

YOLO

DATASET

PREPROCESSING

EXPERIMENTS

EVALUATION

CNN

CONVOLUTIONAL LAYER

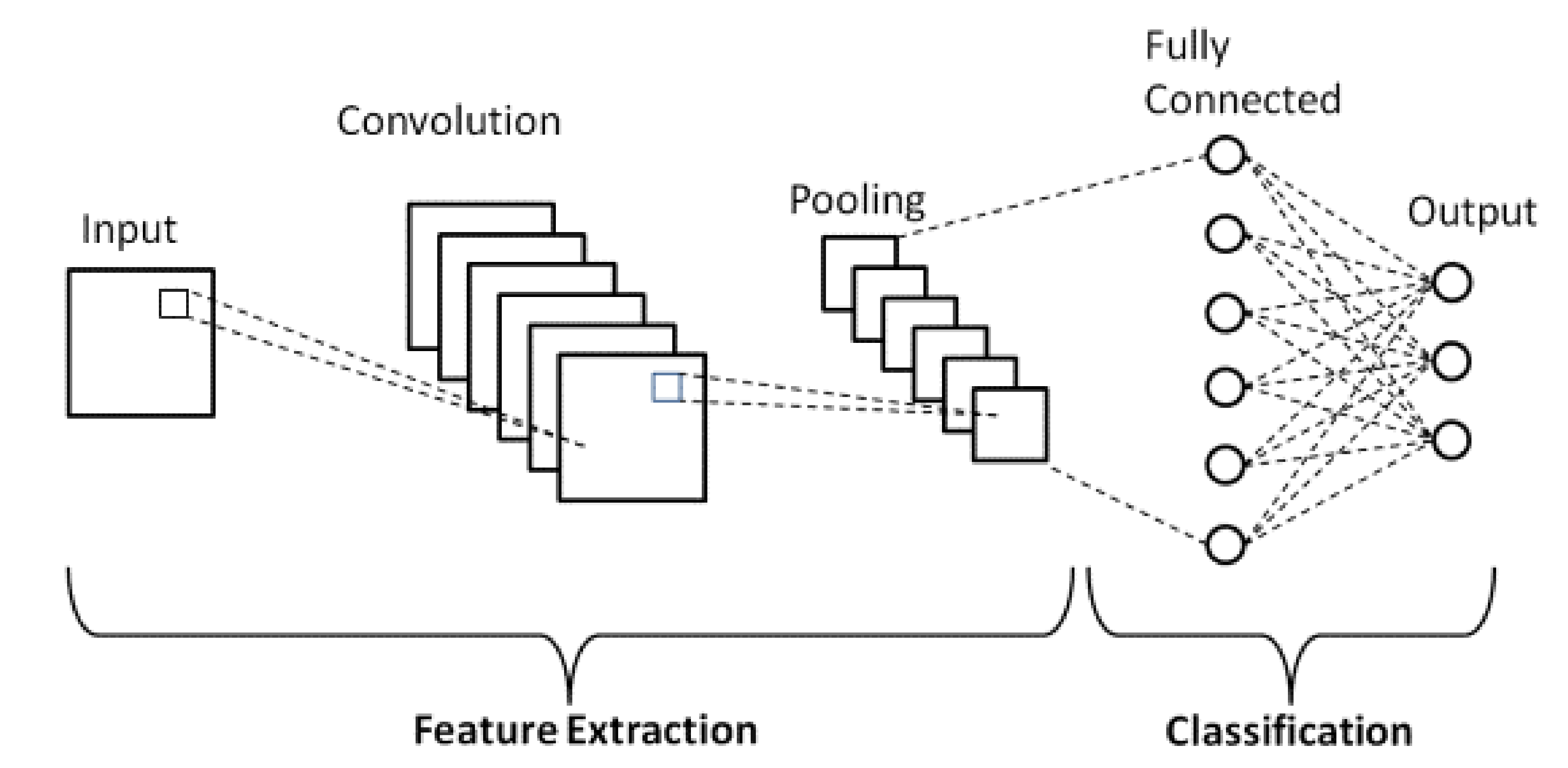
- It uses filters to scan the input image and extract important features.

POOLING

- Reduces the number of network parameters
- Max-pooling selects only the most significant values from a region.
- Preservs important features.
- Creates a more compact representation of the image.

FULLY CONNECTED LAYER

- These layers take all the previous features and process them to generate final results.
- Often followed by a final activation layer (e.g., softmax).



YOLO

```
|- workingdir
        |--yolov5
        |--data
            |--images
                |--train
                |--valid
            |--labels
                |--train
                |--valid
```

- YOLO (You Only Look Once) - object detection
- It utilizes a neural network to predict bounding boxes and object classes.
- Sizes: YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x - Differ in complexity and performance.
- It utilizes a CNN architecture

DBT - BREAST CANCER SCREENING - DIGITAL BREAST TOMOSYNTHESIS (BREAST-CANCER-SCREENING-DBT)

FILE-PATHS.CSV

- pacient id
- study id
- view
- path of the images

LABELS.CSV

- labels
  - benign
  - malign
  - normal

BOXES.CSV (220 RECORDS)

- bounding box
  - X
  - Y
  - Width
  - Height
- Slice

Dataset

RADIOLOGY MANUAL ANNOTATIONS.CSV

- name of the image
- pacient id
- pathologies
  - benign
  - malign
  - normal

RADIOLOGY_HAND_DRAWN_SEGMENTATIONS_V2.CSV

- bounding box : X, Y, Width, Height

CDD-CESM - CATEGORIZED DIGITAL DATABASE FOR LOW ENERGY AND SUBTRACTED CONTRAST ENHANCED SPECTRAL MAMMOGRAPHY IMAGES

800 images

Source: Cancer Imaging Archive (TCIA)

Preprocessing

DATASET

- train - 80%
- validation - 10%
- test - 10%

YOLO

- Each image - label
- txt file:
  - X normalized
  - Y normalized
  - width normalized
  - height normalized

DBT AND CDD-CESM

- (x, y)- (x, y) - the coordinates of the top-left corner of the bounding box,
- width, height of the bounding box.

YOLO

- (x, y) - the normalized coordinates of the center point of the bounding box
- width, height - the normalized width and height of the bounding box

$$nx = (x+Width/2)/width\text{-}of\ \text{-}image$$

Experiments

ID	Weights	Dropout	Optimizer	Epocs	Size	Nb of img	Dataset
1	yolov5s.pt	Nu	SGD	500	320	220	DBT
2	yolov5s.pt	Nu	SGD	100	640	2200	DBT
3	vanilla	Nu	SGD	100	640	2200	DBT
4	vanilla	Nu	SGD	100	640	735	DBT, CDD-CESM
5	yolov5s.pt	0.5	Adam	100	640	735	DBT, CDD-CESM
6	yolov5s.pt	0.5	SGD	100	640	735	DBT, CDD-CESM

WEIGHTS

- pre-trained parameters of the YOLO model (transfer learning)
- obtained by training the model on a large training dataset (e.g., COCO - Common Objects in Context).
- vanilla - no pre-trained parameters are used

DROPOUT

- used to improve the performance of neural network models and prevent overfitting
- it randomly deactivates a number of neurons (or connections between neurons) with a predefined probability

OPTIMIZER

- automatically adjusts the weights to minimize a loss (or cost) function

Evaluation

BOXLOSS
OBJLOSS
CLSLOSS

- measures how well the model locates objects in an image.
- measures the accuracy of the model's object detection within a specific area (grid cell).
- measures the error in classifying objects.

METRICS/MAP_0.5

- a metric that considers the performance of Bbox localization (through Intersection over Union) and the model's classification ability by combining precision and recall

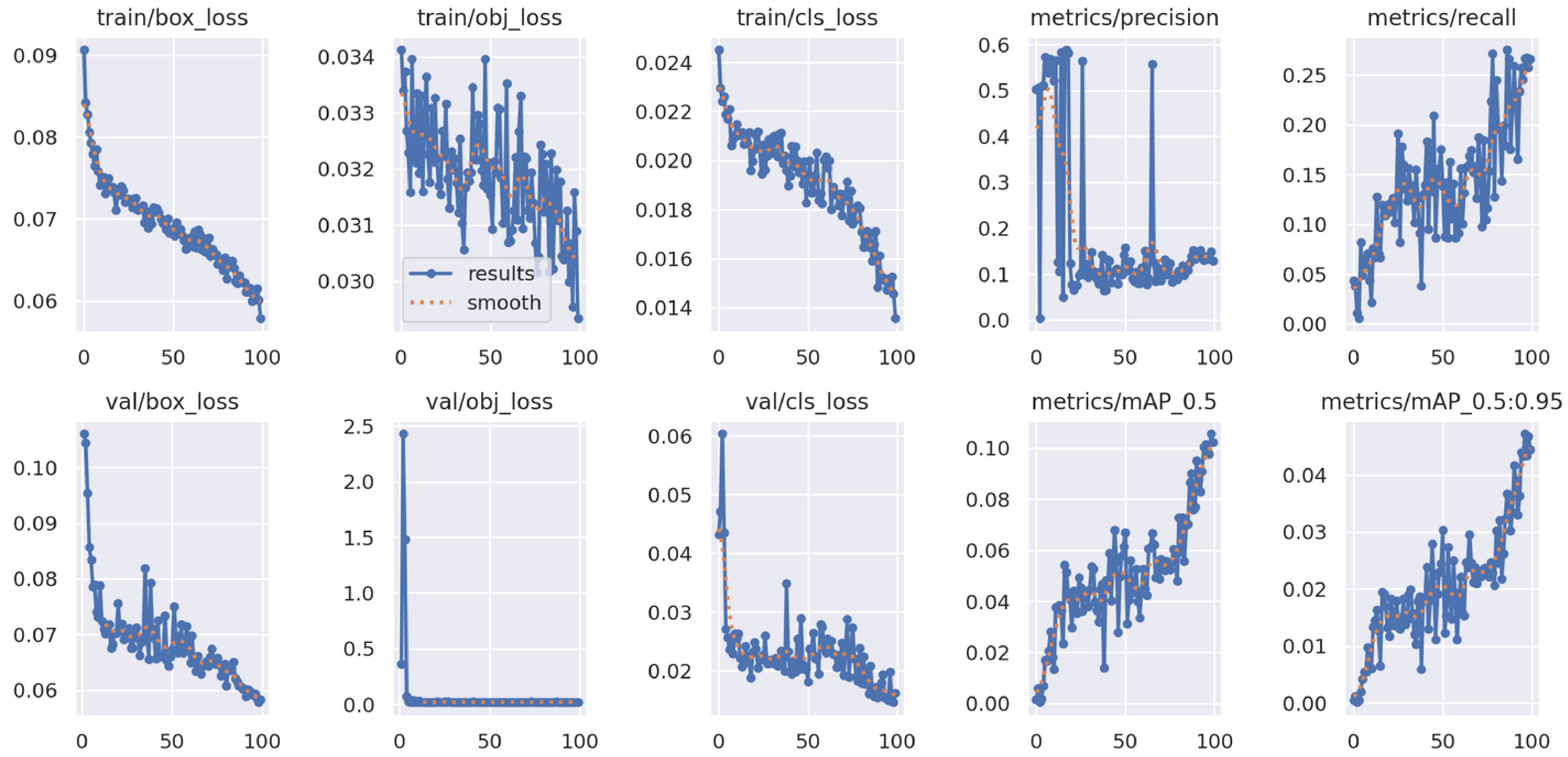
METRICS/RECALL

- measures how many classes were predicted correctly out of the total number of existing classes

METRICS/PRECISION

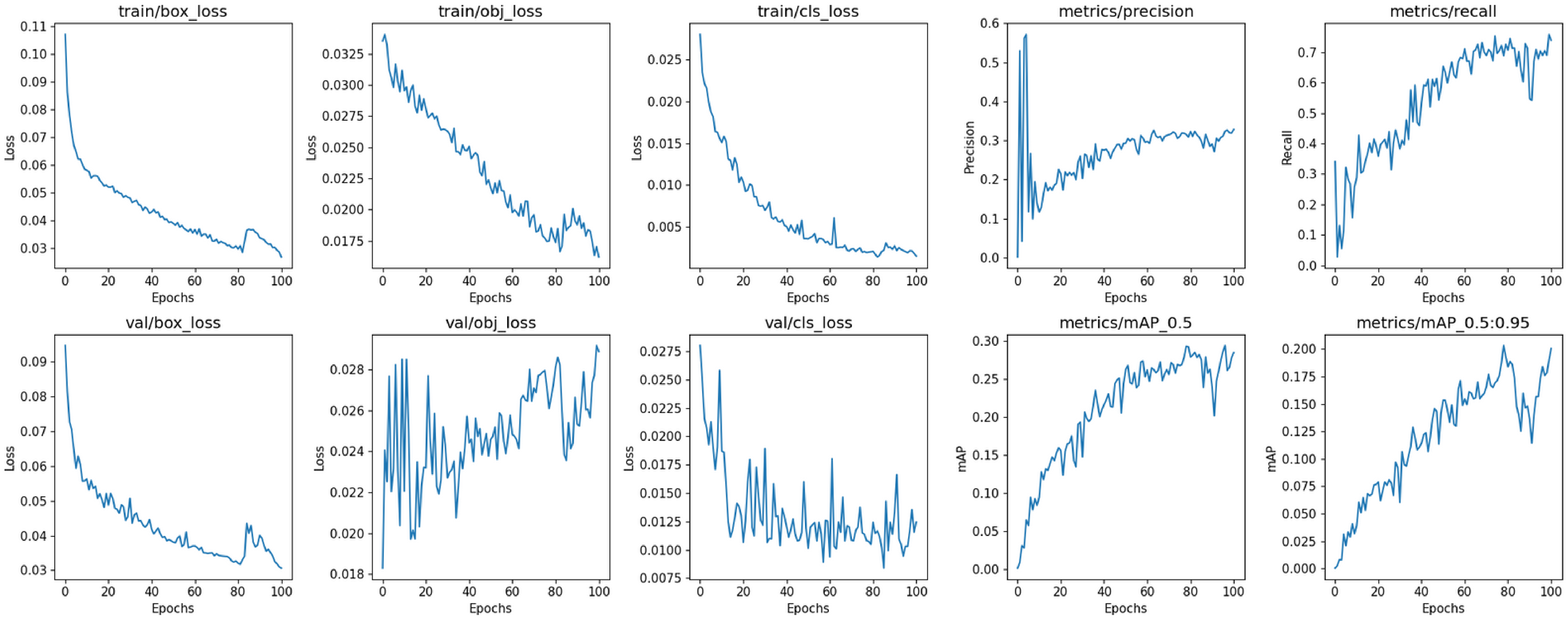
- measures how many classes were predicted correctly out of the total number of predicted classes

Evaluation



5th Experiment

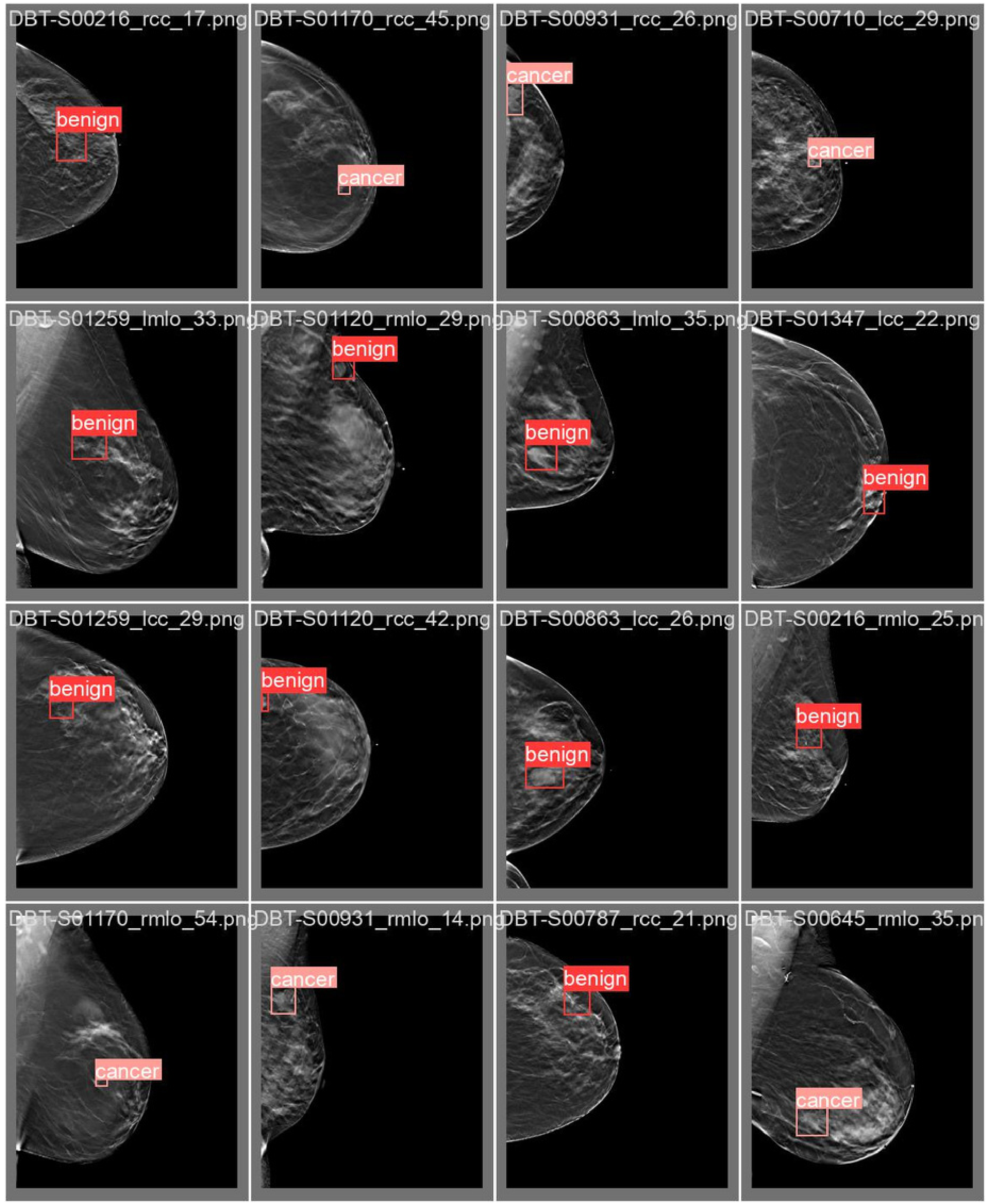
6th Experiment



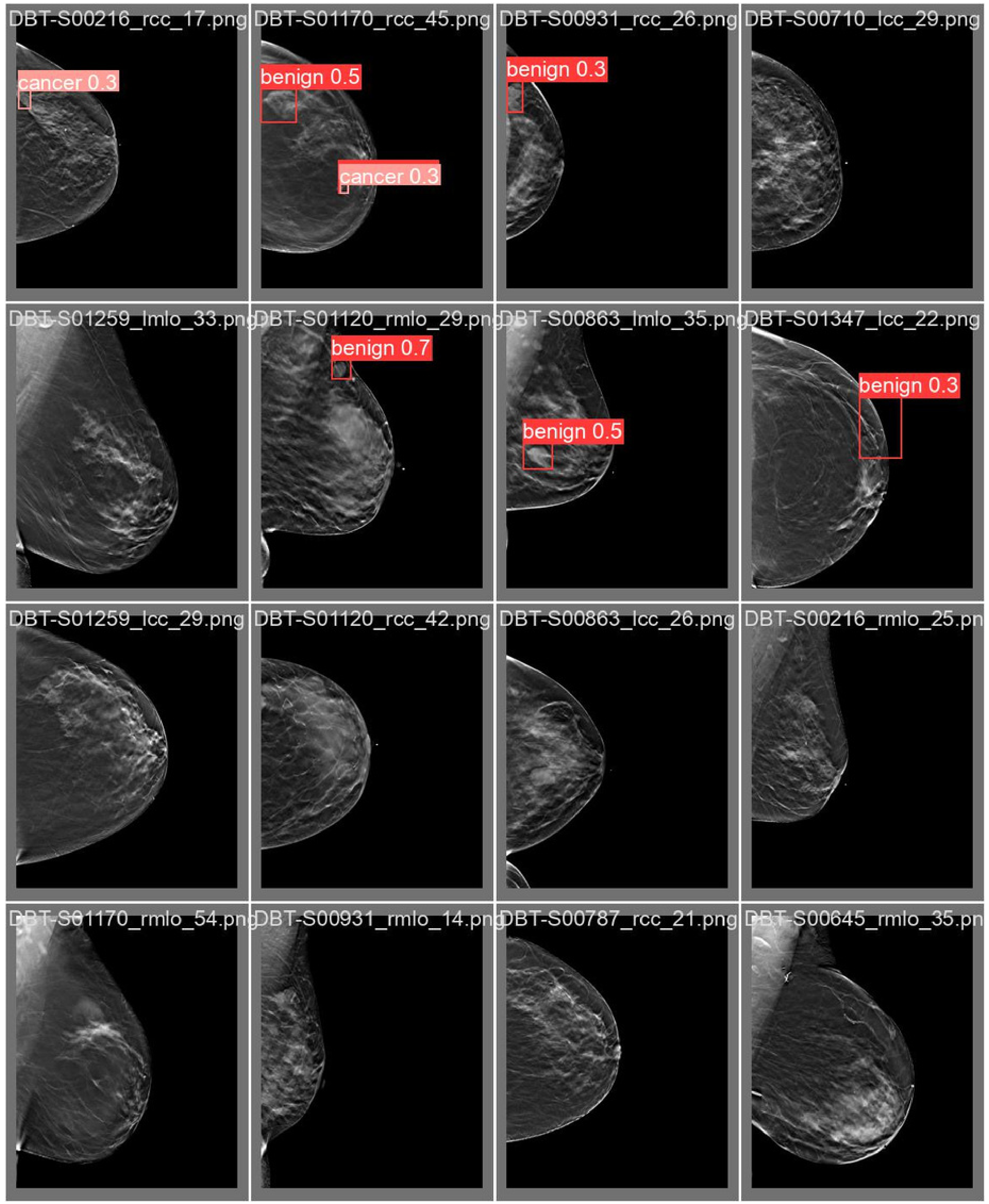
1st Experiment

Evaluation

Labels



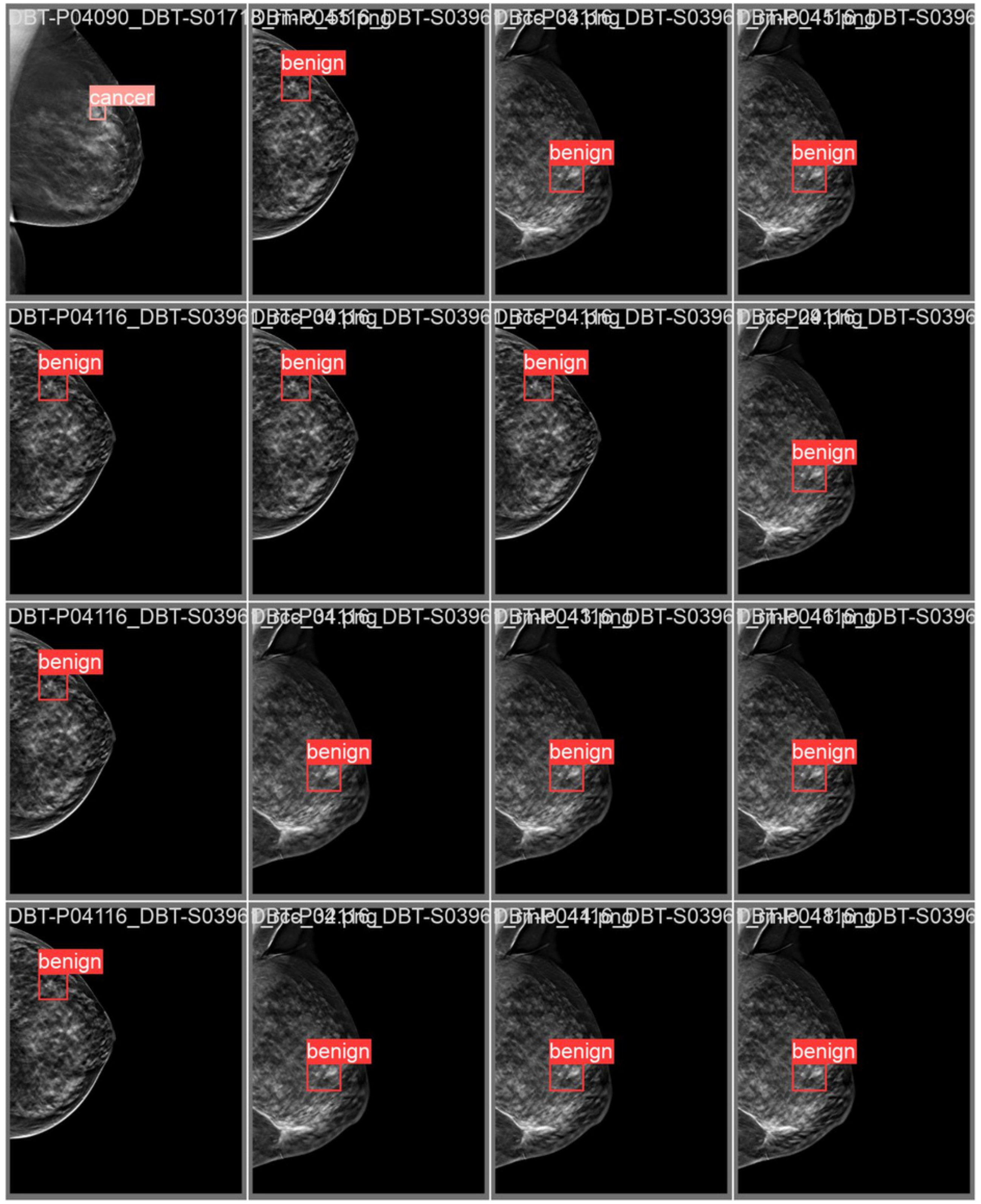
Predictions



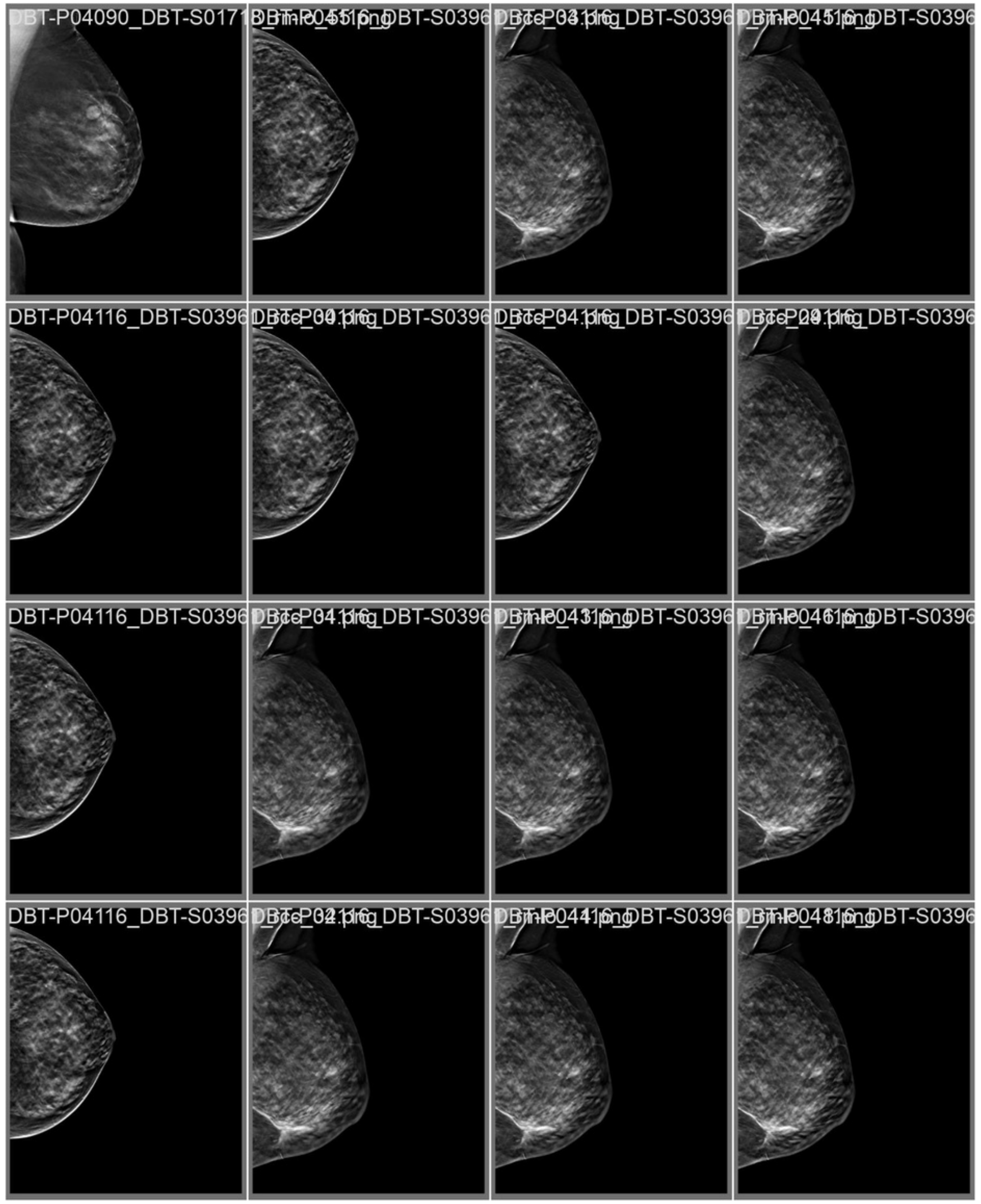
2nd Experiment

Evaluation

Labels



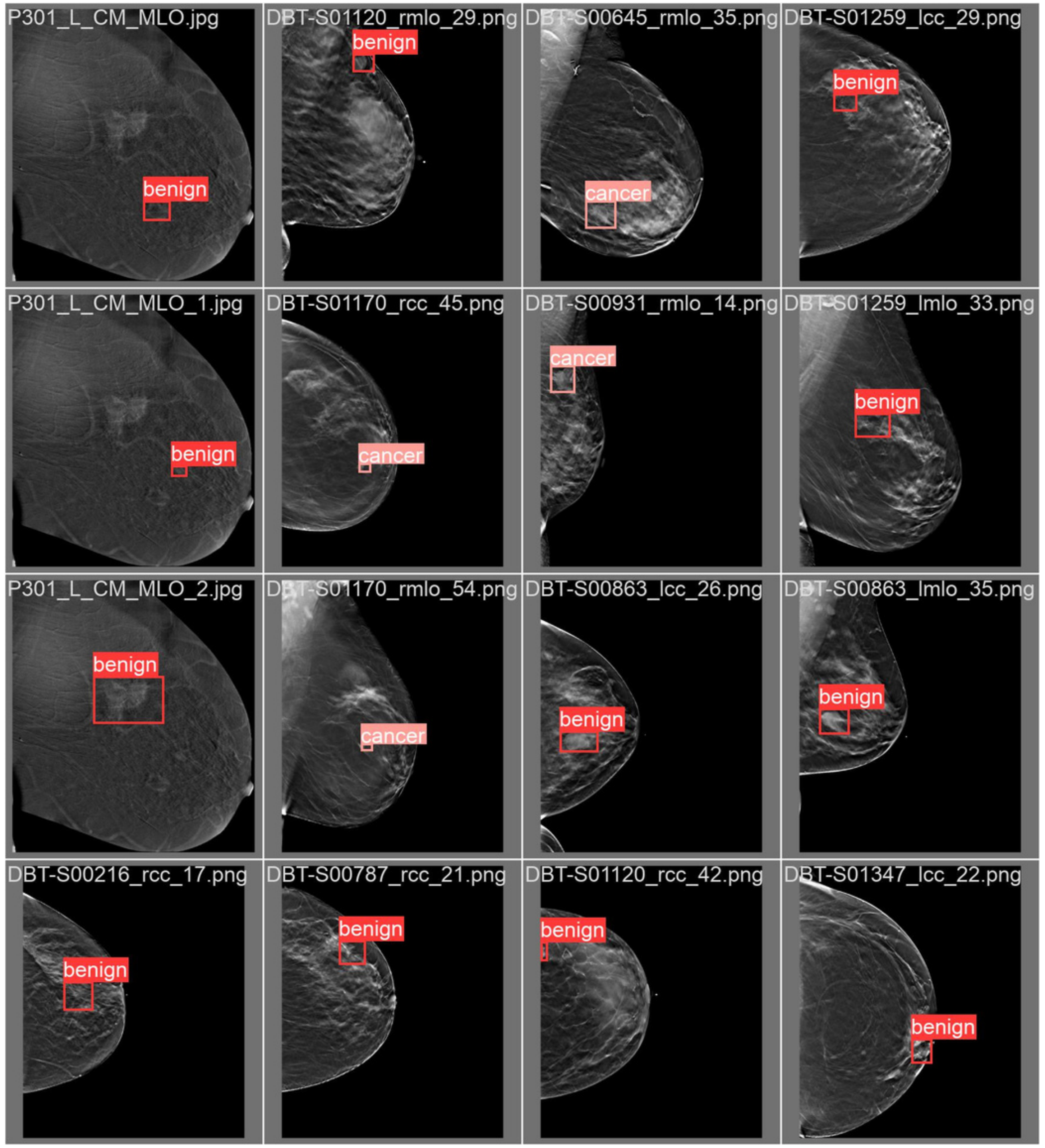
Predictions



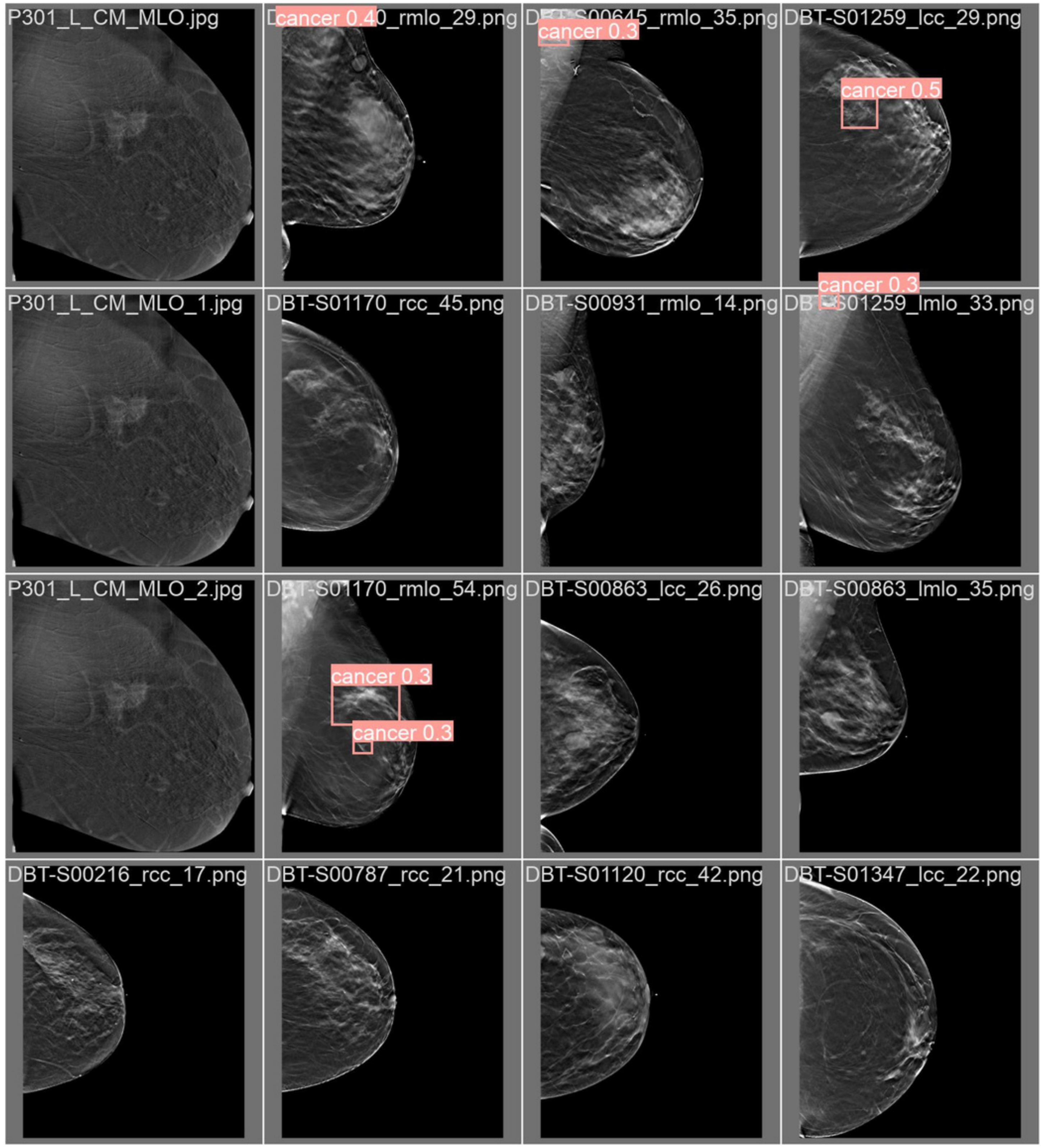
Evaluation

4th Experiment

Labels



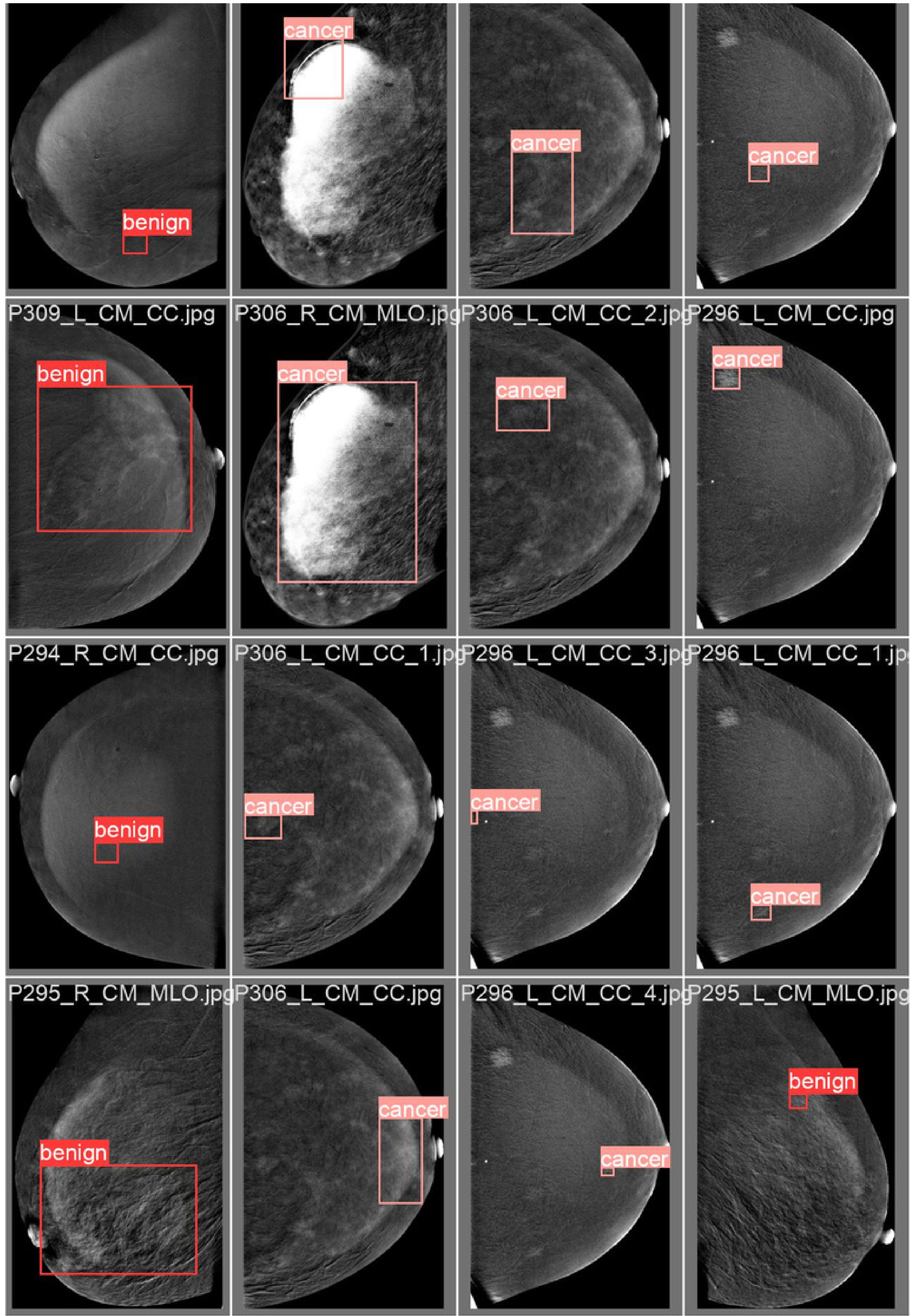
Predictions



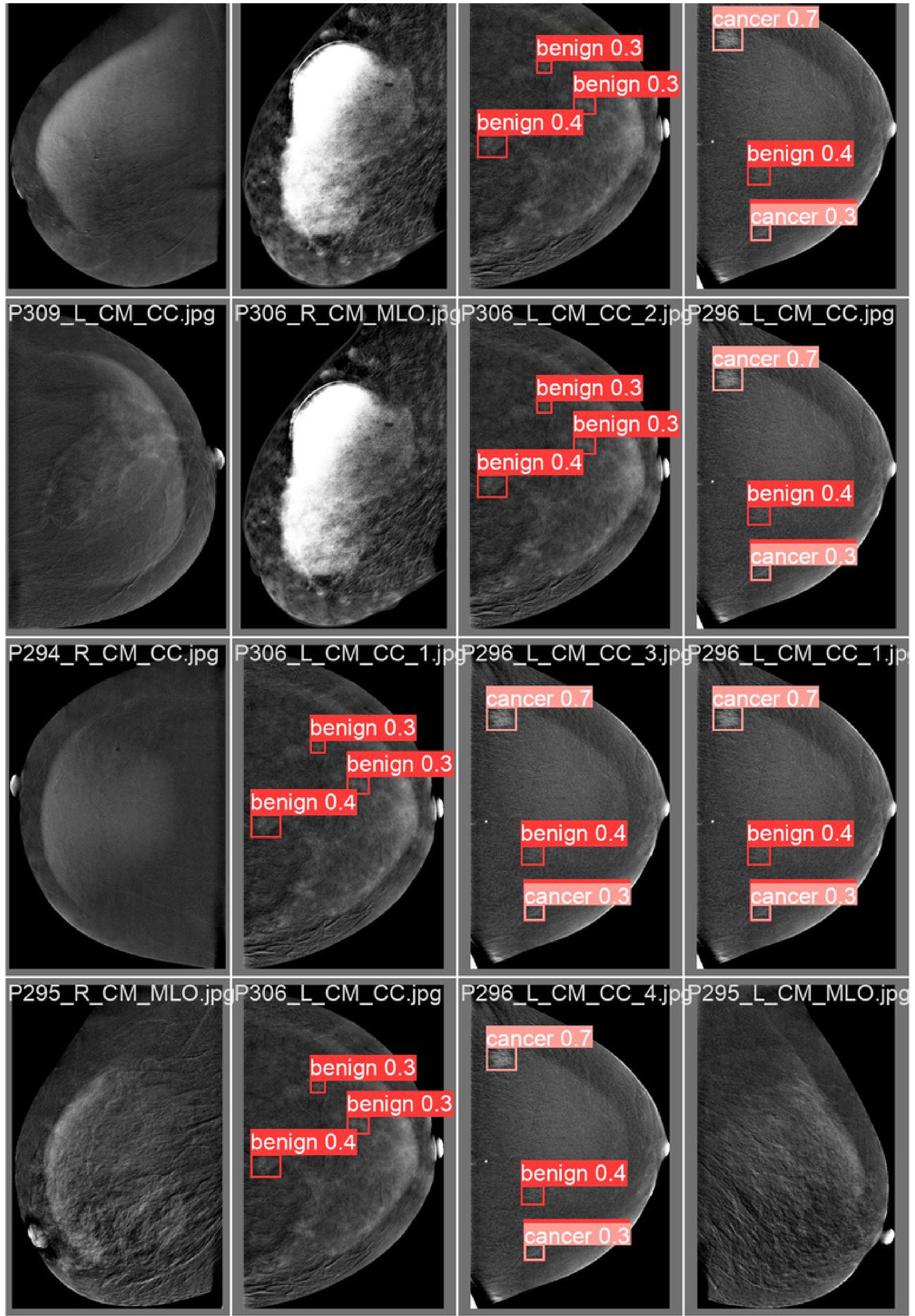
Evaluation

5th Experiment

Labels



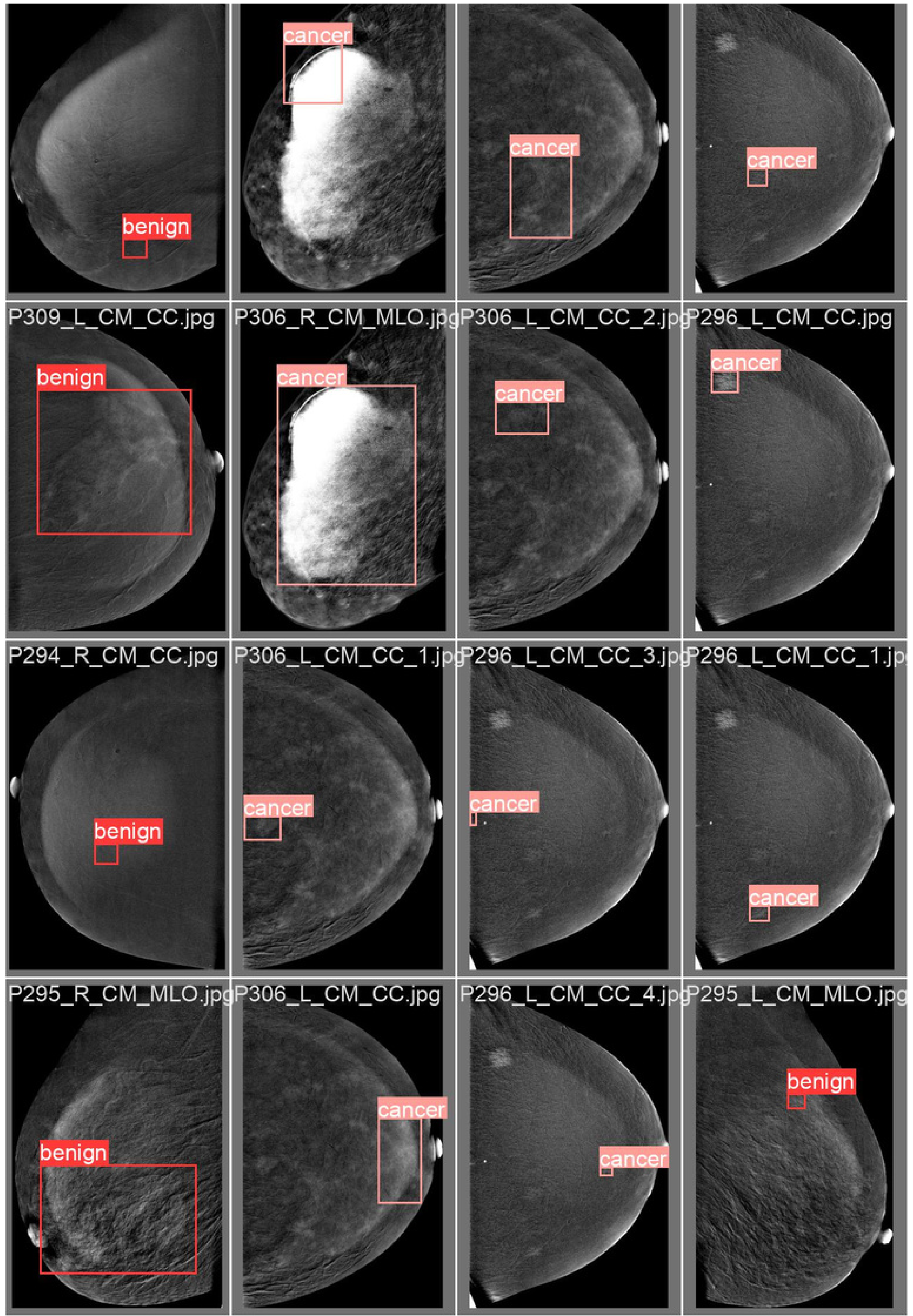
Predictions



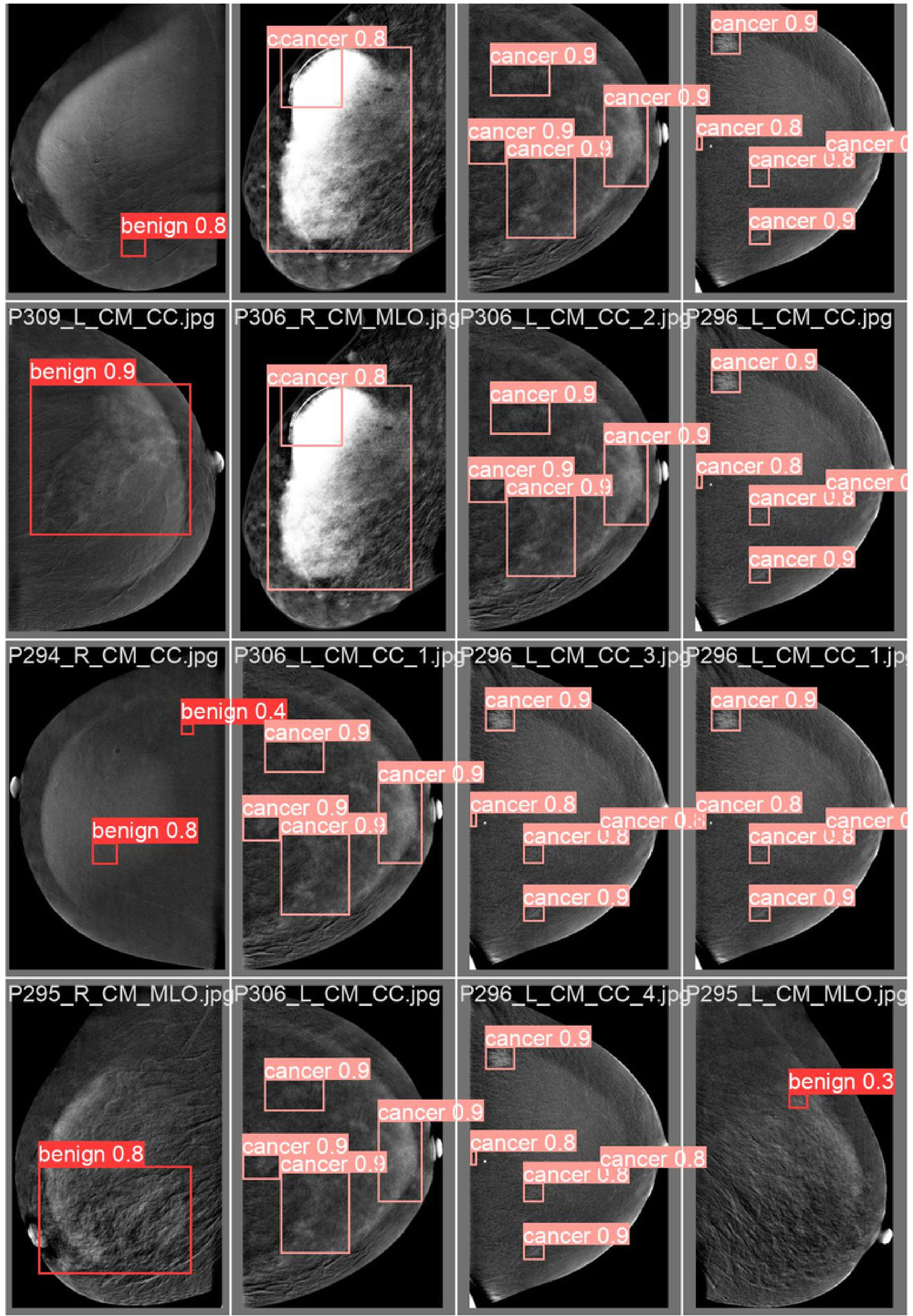
6th Experiment

Evaluation

Labels



Predictions



SWOT Analysis

Strengths:

- good predictions
- portability
- easy detection on other datasets

Weaknesses:

- few images

Opportunities:

- usable in the medical field

Threats:

- similar models

Conclusions

- using SGD as optimiser is better than using Adam
- a larger dataset significantly improves the model's performance
- using pre-trained weights provides better results than vanilla

Future work

- expanding/Diversifying the dataset
- adding background images
- increasing the image size for training
- choosing a different optimizer algorithm (e.g., Momentum)

References

- Buda, Mateusz et al., Breast Cancer Screening - Digital Breast Tomosynthesis (Breast-Cancer-Screening-DBT), The Cancer Imaging Archive, 2020, DOI: 10.7937/E4WT-CD02.
- Ferreira, João Otávio N., Orange Trees Detection with YOLOv5 in UAV Imagery, URL: https://joaootavionf007.medium.com/orange-trees-detection-with-yolov5-in-uav-imagery-22ec29db922e.
- Konz, N., & Buda, M. (2023). A Competition, Benchmark, Code, and Data for Using Artificial Intelligence to Detect Lesions in Digital Breast Tomosynthesis. JAMA Network Open, 6(2), e230524. doi:10.1001/jamanetworkopen.2023.0524.
- Li, Fei-Fei s, i Karpathy, Andrej, "Convolutional Neural Networks for Visual Recognition", în Stanford University (2015).
- Nielsen, Michael A., Neural Networks and Deep Learning, Determination Press, 2015.
- Raschka, Sebastian, "A Gentle Introduction to Backpropagation", în Blog post (2018).
- The Cancer Imaging Archive, The Cancer Imaging Archive, National Cancer Institute, URL: https://www.cancerimagingarchive.net/.
- Ultralytics LLC, YOLOv5, https://github.com/ultralytics/yolov5.
- https://www.upgrad.com/blog/basic-cnn-architecture/