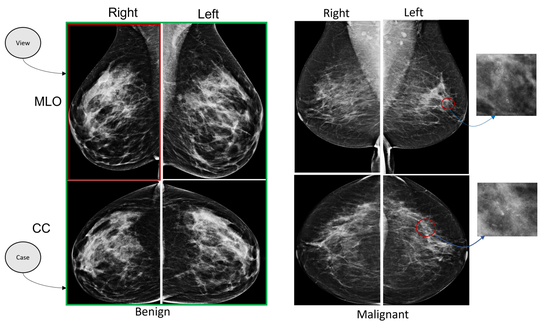
Multi-instance Learning for Breast Cancer Prediction using Mammograms

Shreyasi Pathak [1,2,3] Jörg Schlötterer [2] Jeroen Geerdink [3] Rob Bourez [3] Onno Vijlbrief [3] Maurice van Keulen [1] Christin Seifert [1,2]

[1]University of Twente, Netherlands [2]Institute of Artificial Intelligence in Medicine, University of Duisburg-Essen, Germany [3]Ziekenhuis Groep Twente, Netherlands

Introduction

1. Mammography is used as the first diagnostic test for identifying breast cancer.
2. Multiple views - mediolateral oblique (MLO) and craniocaudal (CC) for both left and right breasts are taken during mammography for careful analysis. Sign of malignancy may be present in either left or right breast, and may be visible in only one of the CC or MLO views.
3. The label of each view is not known, rather, the label is known at the case-level, i.e. a mammogram case will be assigned the malignant label if any one of the views contain malignancy.
4. Further, number of mammogram views taken per patient may vary as it depends on the radiologists' discretion.
5. We aim to propose a breast cancer prediction model that can:
i) work on real-world variable view mammograms
ii) properly consider the view-level information to classify the case.



Dataset & Data Preprocessing

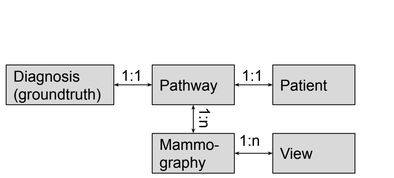


Figure 1. Groundtruth assignment

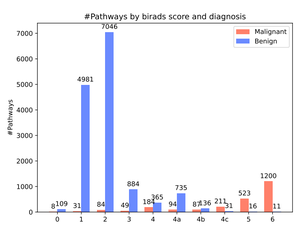
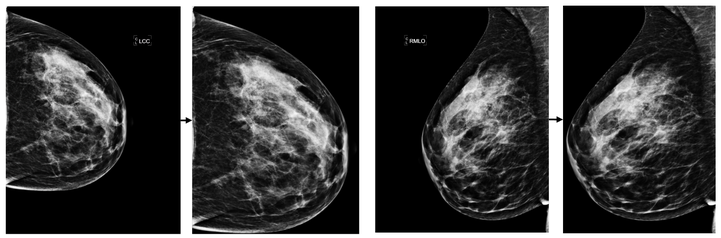


Figure 2. BIRADS score vs diagnosis

Breast cancer data used in our work is from ZGT, Netherlands. We assign groundtruth (malignant/benign) to each mammogram case from the diagnostic pathway followed at the hospital.

	total	benign	malignant	Views (LCC, LMLO, RCC, RMLO) 4	3	2	1
patients	15,988	13524	2464	-	-	-	-
mammogram case	20,979	17,208 (82%)	3,771 (18%)	19629 (93.6%)	276 (1.3%)	693 (3.3%)	381 (1.8%)



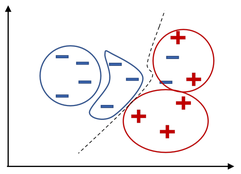
Automatic algorithm developed to remove irrelevant information from images

Problem Definition

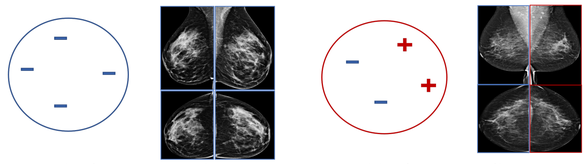
We define our problem of predicting breast cancer from multiple view mammograms as multi-instance learning (MIL).

In MIL, classification is performed on a bag of instances, $X = \{x_1, x_2, .., x_k\}$ with a bag label, $Y = \{0, 1\}$, where the instances in the bag are permutation-invariant and do not have dependency among each other. Each instance in the bag has a label, y_k, which remains unknown and only the label of the bag is known. A positive bag contains atleast one positive instance whereas the negative bag contains all negative instances.

$$Y = \begin{cases} 0, & \text{iff } \sum_k y_k = 0 \\ 1, & \text{otherwise} \end{cases}$$

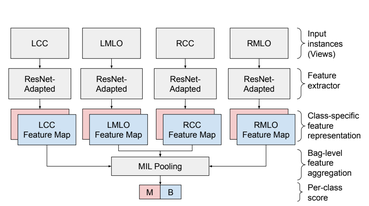


+ Positive Instance
− Negative Instance

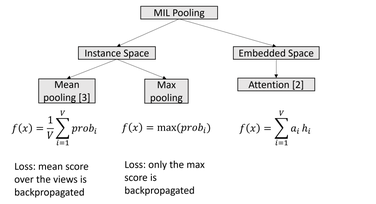


Multi-instance Learning Methods

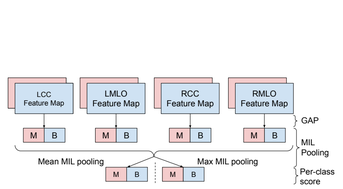
Each view is passed through a feature extractor, ResNet-Adapted, proposed in Kim et al. [3]). The feature map extracted from the last layer of ResNet-Adapted can be converted to class-specific score for each view. Aggregation of information over views (MIL pooling) at the score-level is called instance-space and at feature-level is called embedded-space. We want to compare existing MIL pooling methods - mean [3], max, attention [2] in our work.



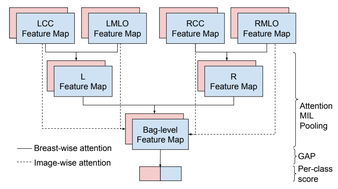
(a) Overall model architecture



(b) Instance-space vs embedded-space MIL pooling



(c) Instance-space model architecture



(d) Embedded-space model architecture

Figure 3. Breast cancer model architecture

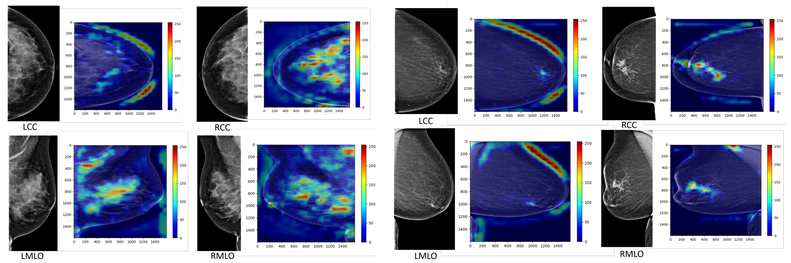
Preliminary Results

MIL-mean stands for MIL pooling using mean over the view-specific scores; MIL-max uses max for MIL pooling; MIL-IWA stands for image-wise attention; MIL-BWA is breast-wise attention. Kim et al. [3] uses MIL-mean method - first row shows performance reported in their paper on their private dataset & second row (MIL-mean) is applying [3] on our dataset.

Model	Type	MIL Pooling	Recall (%)	Specificity (%)	Acc (%)
Kim et al. [3] (on private data)	Instance-space	Mean	75.6	90.2	82.9
MIL-mean	Instance-space	Mean	67.0	74.1	72.1
MIL-max	Instance-space	Max	67.0	74.6	71.6
MIL-IWA	Embedded-space	Attention	70.5	78.8	76.4
MIL-BWA	Embedded-space	Attention	56.4	78.1	71.8

Post-hoc Interpretability

We show Grad-cam++ [1] visualization for a true positive benign case and a true positive malignant case predicted using MIL-IWA model.



(a) Benign case (b) Malignant case

Conclusion

We have defined breast cancer prediction problem as multi-instance learning. These are some preliminary results & conclusions.
1. Attention MIL pooling performed the best compared to max & mean MIL pooling.
2. Post-hoc interpretability showed that MIL pooling is focusing on some correct regions for our example case. In malignant case, the malignant tumor gets highlighted; in benign case, multiple breast tissue region gets highlighted.
3. We don't achieve as high performance as [3] does on their dataset. Apart from the dataset being different, one of the other reasons is that we use mammograms with any BIRADS score (Fig. 2) (harder prediction task), whereas, [3] uses BIRADS score 1 and 6, which are confirmed benign and malignant cases respectively.
4. We are working on gaining better understanding of the model & further improving it.

References

[1] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 839–847. IEEE, 2018.

[2] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.

[3] Eun-Kyung Kim, Hyo-Eun Kim, Kyunghwa Han, Bong Joo Kang, Yu-Mee Sohn, Ok Hee Woo, and Chan Wha Lee. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. *Scientific reports*, 8(1):1–8, 2018.

Shreyasi Pathak, PhD Candidate

s.pathak@utwente.nl