



POLITECNICO
MILANO 1863

**SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE**

EXECUTIVE SUMMARY OF THE THESIS

Improving Breast Cancer Detection through Deep Learning and Digital Breast Tomosynthesis Leveraging Open-Source Data

LAUREA MAGISTRALE IN MATHEMATICAL ENGINEERING - INGEGNERIA MATEMATICA

Author: CHIARA MOCETTI

Advisor: PROF.SSA FRANCESCA IEVA

Co-advisor: DR. FEDERICO COMOTTO - LAIFE REPLY

Academic year: 2022-23

1. Introduction

Breast cancer is a universally challenging public health issue, manifesting as a leading cause of cancer-related mortality among women worldwide. The main counteraction to this complex and heterogeneous disease is a massive screening process, which requires extensive time and human resources.

The development of deep learning models, coupled with the advent of Digital Breast Tomosynthesis (DBT) as a new pseudo-tomographic imaging technique, providing a vertical resolution to breast examinations, is set to revolutionize breast cancer diagnosis. However, deep learning models rely on extensive utilization of data to comply with their complexity.

Health big data, although already emerged as the most significant big data category for its huge potential value of secondary use, is accompanied by serious privacy disclosure concerns. Consequently, conducting research based on consistent and quality data requirements, within an open-source framework, presents substantial impediments. Yet, simultaneously, it marks the critical starting condition for academic and practical investigations.

In this intricate scenario, the thesis is intended

to develop a structured end-to-end deep learning pipeline for breast lesion detection on DBT leveraging exclusively public data. With the vast majority of the studies present in the literature relying on private datasets, the research aims to bridge the existing gap by providing a methodological foundation for future applications on curated private datasets.

2. AI for Breast Cancer Detection in DBT on Open-Source Data

DBT technology employs a sophisticated image acquisition process which offers a pseudo-volumetric reconstruction of the breast. This is achieved by the acquisition of a sequence of stacked 2D X-ray images, from various angles, using a moving X-ray tube around the compressed breast [2]. These projections are then digitally reconstructed into a series of thin slices of the breast, each usually about 1 mm thick, into a pseudo-3D image [4]. The patient's experience during a DBT scan is comparable to 2D X-ray mammography, the standard technique in breast screening.

DBT is particularly beneficial in women with

dense breast tissue, which is characterized by a higher proportion of fibroglandular tissue with respect to fatty tissue. In extremely dense breast tissue, it has been reported that only half of the cancers will be visible [3]. In these cases, DBT accounts for the superimposition breast tissue phenomenon, consisting in the radiographic overlay of breast tissues, which can significantly mask the visualization of underlying carcinomas, complicating the interpretation of the examination images. This effect is particularly common in Digital Mammography (DM), where breast tissues that are separated in the direction of the projection converge onto the same spot in the image. Consequently, distinct normal tissues may mimic a suspicious lesion, and normal tissues may obscure a true malignant lesion. DBT 3D nature, separating breast tissues across individual slices, augments the visibility of anomalies within dense breasts, thereby enhancing the accuracy of early cancer detection among high-risk groups.

The reading time required by the DBT is a critical factor that must be considered in high-volume screening programs, when deciding whether to leverage this technology as imaging modality. In contrast to a traditional DM exam, interpreting a DBT volume involves the comprehensive analysis of up to 80 slices per single DBT examination [1]. This significantly increases the complexity and the implementation time of clinical radiology workflows.

Deep learning models are set to aid the streamlining of this process while, simultaneously, overcome inter-reader variance, the variability in reading times and diagnostic accuracy among different radiologists.

2.1. X-RAIS: the Inspiration for a Further Exploratory Analysis on DBT

The research question has been inspired by Laife Reply's X-RAIS platform, an innovative AI tool designed to support radiologists in mammography interpretation by precisely detecting and characterizing microcalcifications. The successful development of this standardized and observer-independent system has sparked further interest in exploring the capabilities of deep learning models for DBT, which is seen as the 3D advancement of the traditional DM.

This study is conducted within an entirely open-source framework, with the intent of establishing a foundation that may induce company's partnerships to collect and provide structured and comprehensive private data on which fine-tuning the pipeline designed and implemented. This would enable the enhancement of X-RAIS functionalities with a DBT-based model that improves the solution diagnostic accuracy and efficiency.

3. Datasets

The thesis is based on the utilization of two annotated publicly available datasets.

The primary data source is the Breast Cancer Screening - Digital Breast Tomosynthesis (BCS-DBT), a radiologists annotated tomosynthesis dataset comprising normal and cancer cases. For the development of the research models, we retain a portion of non-lesioned volumes and the entire lesioned part of the dataset, which comprises 431 examination associated to 435 lesion annotations, provided in the bounding box format. We, additionally, employ the information of the slice index, which, according to the radiologists' indication, contains the most evident sign of the lesion.

The second data source is the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM), which comprises 2,620 lesioned scanned film mammography studies, articulated in DM full images, cropped images centered on the lesion, and mask images, providing the binary localization of the lesion on the full image.

4. Methods

We conceptualize the problem within a 2D scope, with the intent of subsequently further extending the model capabilities to embrace the intrinsic 3D nature of DBT images, thereby leveraging their diagnostic potential to the fullest.

The problem is tackled by two different perspectives.

4.1. Approach 1: Detectron

In the first approach we perform lesion detection on full images using Facebook Research object detection model Detectron as a backbone, with a two-stage strategy, composed of the pre-training

on mammography images and the fine-tuning on tomosynthesis images.

Firstly, we obtain the DBT volumes in DICOM format, applying an image enhancement process consisting in the rescale of the pixel values guided by the image window width and window center, which suggest the proper displaying of medical images, providing the optimal brightness and contrast. Restricting to the selected slices of interest, we apply the image pre-processing procedure, which consists of image cropping and flipping. To ensure compatibility with Convolutional Neural Networks (CNNs), we remove the non-informative black band on the breast opposite side, which does not confer diagnostic value, leveraging connected component analysis. Moreover, images are flipped when necessary to maintain a consistent pattern across the dataset; in particular, we ensure that all the slices adhere to a uniform left orientation. In order to apply transfer learning from the CBIS-DDSM we need to replicate on it the data preparation and image pre-processing steps applied to the BCS-DBT. However, the structure of this second dataset brings further complications. The CBIS-DDSM, indeed, does not present direct annotations. Moreover, each full image should be associated with one or more binary masks, although, due to its intricate and confuse nature of open-source dataset, this correspondence is not always established and, moreover, not easily retrievable.

We base masks identification on the analysis of black pixel counts within the images. This step necessitates the establishment of a suitable threshold for the sum of pixel values, due to the variability in lesion sizes. Subsequently, we proceed with the extraction of rectangular bounding boxes outlining the lesion contours. This step is performed extracting the minimum and maximum coordinates of the pixels valued at 255 - white, indicating the lesion, across both axes. These coordinates serve to the construction of a rectangular bounding box identifying the lesion, recorded in the standard format $(x_{min}, y_{min}, x_{max}, y_{max})$, ensuring consistency with the BCS-DBT lesion annotation.

The image pre-processing procedure applied to the CBIS-DDSM dataset conceptually mirrors the one of the DBT images. However, this second dataset comprises digitized screen-film

mammographies, which are marred by noise and artifacts. The successful implementation of this procedure needs some caution. The exclusion of the top and bottom 5% of pixels allows the connected component analysis for image cropping to be effective in the great majority of cases. Similarly, excluding the left and right 5% of pixels is used to determine the image laterality for flipping.

Training is approached with the objective of optimizing the performances while, meanwhile, accommodating the constraints typical of medical image analysis and taking into account the computational demands proper of Detectron complex architecture. Data augmentation is intentionally simplified to the application of random vertical flip and 95% of the original image size crop. Given the intensity-based characterization of lesions, transformations affecting brightness, color, and contrast are excluded to prevent potential model confusion. To ensure a solid starting point for training, we initialize the model weights from a base pre-trained on the COCO-dataset. We apply the Stochastic Gradient Descent (SDG) with a step-based learning rate scheduler, providing gradual learning rate adjustments, involving an initial warm-up phase, and the subsequent decade at pre-defined steps. The model is trained for 30,000 iterations. After, we initialize the model weights using those derived from the training on the CBIS-DDSM dataset, and we repeat the training loop for 9,000 iterations on the BCS-DBT dataset.

4.2. Approach 2: Patches Model

To address the challenge posed by the limited image availability, we explore a second approach, in which we develop a classification model that differentiates between lesioned and non-lesioned tissue patches, followed by a phase of lesion localization, performed applying the newly developed classification model on a grid of patches generated from the full images. This methodology aims at augmenting the number of usable images and also streamlining and controlling the modeling process. We exclusively utilize the BCS-DBT extracted slices as data.

4.2.1 Classification

To generate "lesion" class patches we develop a cropping strategy that accounts for the le-

sion dimensions variability. In details, we design an algorithm which, given an image and a corresponding lesion bounding box as input, automatically extracts the maximal number of 224×224 - a selected standard dimension - patches on the basis of the larger dimension of the lesion itself. Also, since the classification model we aim to develop is functional to the subsequent localization phase, we observe that dividing the full images in a grid may result in patches where the lesion is not centrally located. To provide contextual information surrounding the lesion, therefore, we extract an additional number of patches per lesion by cropping the full image around the lesion periphery. Algorithm 1 details the entire procedure.

Algorithm 1 Patches Extraction Algorithm

- 1: **Input:** Image, bounding_box_center, boundix_box_sizes,
 - 2: **Set:** $l = 224$, dimension of the patches,
 - 3: Select the maximum between the height and width of the bounding box.
size = $\max(\text{boundix_box_sizes})$,
 - 4: Depth = $\lfloor \frac{\text{size}-1}{l} \rfloor$,
 - 5: Initialize **center_list** as a list with only the center of the bounding box.
 - 6: **for** i in Depth **do**
 - 7: Initialize **angle_list** as an empty list,
 - 8: **for** **center** in **center_list** **do**
 - 9: angles = $\{(x_{\text{center}} - \frac{l}{2}, y_{\text{center}} + \frac{l}{2}), (x_{\text{center}} - \frac{l}{2}, y_{\text{center}} - \frac{l}{2}), (x_{\text{center}} + \frac{l}{2}, y_{\text{center}} - \frac{l}{2}), (x_{\text{center}} + \frac{l}{2}, y_{\text{center}} + \frac{l}{2})\}$
 - 10: Add these angles to the **angle_list**.
 - 11: **end for**
 - 12: Add the angles of the **angle_list** in the **center_list**,
 - 13: Avoid repetition using a **set**.
 - 14: **end for**
 - 15: **for** **center** in **center_list** **do**
 - 16: Create the patch, cutting the Image given the angles computed from **center** as in line 9,
 - 17: Save the patch as **.png** file.
 - 18: **end for**
-

For the generation of "no lesion" class patches we restore a portion of the normal DBT volumes within the BCS-DBT. From each of their central pre-processed slices we sample 150 random

points and, taking each of them as a reference corner point, we extract a patch of size 224×224 , ensuring, based on the black pixel count, that it is not a background non-informative patch. We target this process to obtain a dataset with an imbalance ratio of 1 over 20, which corresponds to the median imbalance resulting from the analysis of the proportions of lesioned and non-lesioned patches on the dataset of full images on which the classification model will be applied in the detection phase.

In this setting, we are not limited by the annotations constraints; therefore, we employ several data augmentation techniques exclusively addressing the "lesion" class. This process is applied to attain a training imbalance ratio of 1 over 5, which facilitates the model learning process. The validation set, conversely, remains in its original representation.

We implement a training process focused on maximizing the recall, while, concurrently, managing to keep the false positives rate in check. We employ a ConvNeXt model, integrating top layers independently selected, with ImageNet weights initialization. We fine-tune this model on our dataset using a reduced learning rate and applying the focal loss function, specifically designed to address class imbalance. The Adam optimizer is selected and a customized learning rate scheduling strategy is implemented, designed to dynamically adjust the learning rate based on the validation loss computed at the end of each training epoch.

4.2.2 Localization

The localization phase is implemented through a Gradient-weighted Class Activation Mapping (Grad-CAM) sliding window approach.

First, we resize the image, and the associated bounding box, to ensure that its dimensions are divisible by the patch size 224×224 . The resized image is then divided into patches of fixed dimensions 224×224 . The classification model is employed to obtain predictions for each of these patches, excluding the non-informative ones, consisting of nearly exclusively black pixels. We calculate the gradients of the predicted class with respect to the output feature map of the last convolutional layer, merging them to highlight the areas of the image most influential for the model prediction. The heatmap gen-

erated from this process individuates the patch regions which most reveal a lesion, providing an insight into the areas where the model concentrates its attention. Upon calculating the heatmaps for each patch, we aggregate them to form a comprehensive visualization overlaid on the original full image.

To determine the predicted bounding boxes based on the heatmap, we first convert it into a binary matrix, by setting to 1 all the pixels contributing to the lesion classification, identified by the heatmap values higher than 0. Subsequently, contours are identified within this binary representation, serving as the outlines for the predicted lesions. For each contour detected, a rectangular bounding box is determined, encompassing its area.

4.3. Performance Metrics and Model Selection

For our initial evaluation process, we employ the Average Precision (AP) metric at 50% of Intersection Over Union (IOU) threshold. This metric computes the precision by classifying detections as true positives when their overlap with the actual ground truth bounding box exceeds 50%.

As a second evaluation strategy, we introduce a tailored metric devised considering the specific context of the research. We consider all the predicted bounding boxes per single image: such predictions are deemed accurate if the intersection over the minimum area between the predicted bounding box and the true bounding box exceeds 50%, and inaccurate otherwise. The average of this metric, across all the test samples, serves as an indicator of the model recall.

In addition, to monitor the occurrence of false positives, we count the number of incorrectly predicted lesions per image. Averaging this value, we obtain the mean amount of false positives per image.

Since Detectron outputs the confidence scores associated to the predictions and in the second approach the determination of a patch being classified as lesioned depends on whether the model prediction score, for that label, exceeds a predetermined threshold, we tune the values for this hyperparameters using the validation set.

Moreover, for the second approach, which is prone to a higher number of false positives, we

delve into the analysis of the incorrect predictions by computing the normalized centers distance of the predicted and the ground truth bounding boxes.

5. Evaluation

5.1. Approach 1: Detectron

We present the outcomes of the metrics detailed in the previous section, in correspondence of the selected value for the confidence threshold at which a given prediction is considered to be valid. The threshold, tuned on the validation set, was established at 0.2 value.

The AP at the 50% of IOU criterion is 0.554.

According to the metric outlined above, tailored to the research question, the recall is 0.837 and the average false positives count is 0.977.

5.2. Approach 2: Patches Model

The second approach is assessed exclusively through the metric we specifically devised. The results are obtained by setting the acceptance threshold for a "lesion" class prediction at 0.6, a value tuned on the validation set.

The outcome for the recall is 0.627, and the average false positives count per image is 1.581.

5.3. Individual Predictions

As representative instances of the results obtained, in the Figures 1 and 2, we present two correct predictions from both the approaches, on the same DBT test image.

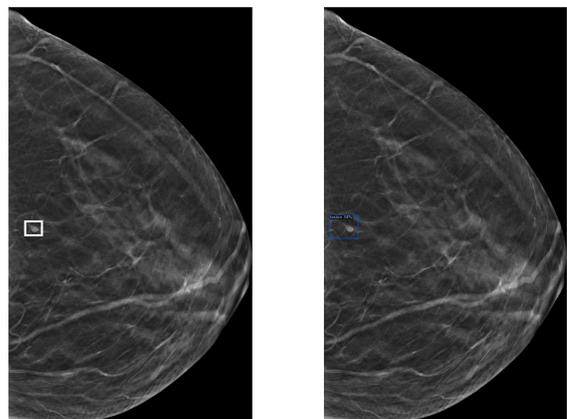


Figure 1: Ground truth and prediction of a lesion correctly detected with Detectron

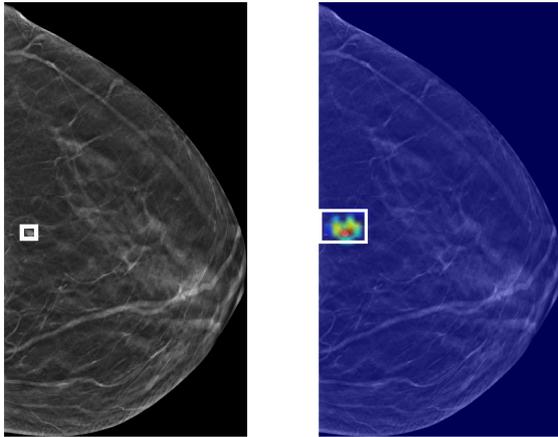


Figure 2: Ground truth and prediction of a lesion correctly detected with Patches Model

Detectron prediction demonstrates higher precision, while the heatmap output generated by the Patches Model offers enhanced interpretability.

6. Conclusion

In the thesis, we delved into the application of deep learning for the automatic detection of breast cancer lesions in DBT images on open-source data, developing an end-to-end pipeline for lesion identification, exploring and comparing two proposed alternative strategies.

The first approach, based on the object detection model Detectron, demonstrated the most impressive performance. Nevertheless, this method poses challenges due to the limited control over the model implementation. The complexity and computational intensity of Detectron architecture restrict the ease of modifications and experimentation with diverse settings, outlining the necessity to strike a balance between computational efficiency and problem-specific implementation requirements.

To circumvent these limitations, a second strategy was devised, built upon a preliminary analysis of image patches. This approach decomposes the task into a dual-phase process, significantly increasing the volume of usable images extracted from the relatively restricted dataset. This method yields highly interpretable results, delivered in the form of heatmaps identifying the image area that most contributed to the lesion prediction.

The comprehensive methodologies developed and validated through the research exhibit distinct merits and generate reliable outcomes. Their comparison enables the verification of results consistency, ensuring confidence in their applicability.

Differently to the prevailing presence, in the literature, of lesion detection algorithms utilizing privately curated datasets, the study achieved respectable outcomes entirely leveraging open-source data, developing two valuable approaches. These methods facilitate comparative analysis and can be adapted to meet the specific requirements of the investigation objectives.

This research addresses the balance between performance and interpretability, still bearing in mind that further enhancements in both aspects can be achieved with the access to more consistent and higher-quality data.

References

- [1] Avi Aizenman, Trafton Drew, Krista A Ehinger, Dianne Georgian-Smith, and Jeremy M Wolfe. Comparing search patterns in digital breast tomosynthesis and full-field digital mammography: an eye tracking study. *Journal of Medical Imaging*, 4(4):045501–045501, 2017.
- [2] Alice Chong, Susan P Weinstein, Elizabeth S McDonald, and Emily F Conant. Digital breast tomosynthesis: concepts and clinical practice. *Radiology*, 292(1):1–14, 2019.
- [3] Stamatia V Destounis, Renee Morgan, and Andrea Arieno. Screening for dense breasts: digital breast tomosynthesis. *American journal of roentgenology*, 204(2):261–264, 2015.
- [4] Mark A Helvie. Digital mammography imaging: breast tomosynthesis and advanced applications. *Radiologic Clinics*, 48(5):917–929, 2010.