Automatic anomalies detection and labeling from complete mammographies: a retrospective study

Xavier Lessage
Faculty of Engineering
UMONS
Mons, Belgium
xavier.lessage@gmail.com
*Corresponding author

Bilal Hammas
Faculty of Engineering
UMONS
Mons, Belgium
bil.hms22@gmail.com

Saïd Mahmoudi
Faculty of Engineering
UMONS
Mons, Belgium
said.mahmoudi@umons.ac.be

Salvatore Murgo
Regional Hospital Centre
HELORA
Mons, Belgium
salvatore.murgo@gmail.com

Sidi Ahmed Mahmoudi
Faculty of Engineering
UMONS
Mons, Belgium
sidi.mahmoudi@umons.ac.be

***Abstract*—Mass localization in mammography is a critical task for early detection and effective treatment of breast cancer, a prevalent health concern worldwide. Computer Aided Diagnosis systems (CADx) can assist radiologists in their difficult diagnosis task, and play a key role in detecting abnormalities and treating breast cancer. In this paper we propose an innovative method for automated labelling and localization of mammographic masses and microcalcifications. The aim of our method is to detect the presence of masses and microcalcifications on the mammography. Hence, we propose to use Yolo Framework in order to locate tumours in a dataset with complete mammograms. Malignant masses or microcalcifications are usually annotated and analysed by a medical expert. As there are only few medical experts devoted to this annotation in each hospital, this task become too time-consuming. Malignant tumours, on the other hand, are nine times more numerous on average, and require far too much time from radiologists to annotate them manually. Our innovative method consists of automatically extracting all abnormalities, whether benign or malignant, from the full mammogram. Our experiments were carried out with a dataset from a Belgian hospital (HELORA) thanks to a retrospective study containing 800 malignant images and 90 000 benign images classified in two directories, positive and negative respectively. Addressing the challenge posed by the abundance of tumors relative to the limited availability of expert annotators, our approach demonstrates proficiency in reducing the time burden on radiologists.**

***Index Terms*—breast cancer, masses and microcalcifications detection, localisation, automated labelling, convolutional neural networks.**

I. INTRODUCTION

Breast cancer screening is a public health issue. Given that one woman in nine will be affected by the disease, it is vital to detect the first signs of this disease. Numerous campaigns are urging women to undergo regular screening and to have a diagnosis as soon as any clinical abnormality becomes apparent. The number of mammograms to be carried out is therefore fairly high, creating a significant workload for doctors. What's more, interpreting a mammographic image is a difficult task, requiring verification by a second or even a third reader (in the event of a discrepancy) in order to reduce the number of false negatives. It is the most common cancer in women and the second most common cause of death. Indeed, one woman in nine will develop breast cancer before the age of 75. Also, 80% of breast cancer cases occur after the age of 50. Although breast cancer in men is rare, there are nevertheless around 80 new cases every year (+/- 1% compared with women). Screening, which is based on mammography, can detect cancer at an early stage, even if there are no symptoms.

Screening or diagnosing breast cancer on the basis of mammography is a difficult problem, which cannot be dealt with using a conventional image classification method. In fact, we are faced with three major difficulties:

- A major imbalance between the small number of images available showing an abnormality and the large number of healthy images;
- The malignant character of an entire image composed of 4000x3000 pixels is determined by a few small regions, of the order of 100x100 pixels;
- Inter-operator variability: the same image read by two specialists may lead to two different diagnoses, requiring a third reader to make the correct diagnosis.

Knowing this situation, a number of classification techniques specific to mammography images have been developed and are constantly being improved to provide increasingly effective diagnostic tools. However, these approaches require large, properly annotated datasets to train the model. To achieve this, the data must be manually annotated by experienced radiologists, which presents major challenges. Manual annotation is time-consuming, tedious and subjective. Obtaining the resources to annotate large amounts of data is costly, due to the need for qualified radiologists. In addition, annotations can vary from one radiologist to another. Therefore, the main challenge lies in the creation of correctly annotated medical image datasets.

In order to save doctors' valuable time, we propose in this paper an extraction method based on Deep lerning, to automatically extract masses and microcalcification from mammograms previously marked by doctors as positive or negative. This is particularly useful when the number of images to be processed is very high, especially for the majority class containing benign mammograms. Once these anomalies (masses and/or microcalcifications) have been extracted, the patches (crops) formed in this way can be used to train an AI more effectively.

In the literature, several articles present solutions based on the YOLO (You Only Look Once) architecture, including the work of S. Ramachandran et al. that proposes the localisation of pulmonary nodules which caught our attention [7]. In this work proposed in 2018, this team of researchers uses convolutional neural networks to learn nodule detection features, replacing the traditional method based on geometric shape or texture. In this architecture, object detection is treated as a regression problem with a single convolutional network simultaneously predicting several bounding boxes and class probabilities for these boxes. The accuracy obtained with a database containing 2132 cases is 93%, with a sensitivity of 89% and 6 false positives per image. In view of the results obtained by this team, we will use the YOLO Framework to locate medical anomalies, and more specifically the latest version (v6) because of its speed, performance and better ratio between execution time and accuracy.

II. DATASET

A. Private raw dataset

We used a database containing 4,000 mammograms, including 800 positive images and 3,200 negative images. This database, which was anonymised, came from a hospital in Belgium thanks to a convention with UMONS university.

B. Pre-processing phase

In the medical field, almost every database we have contains DICOM images. A DICOM image is more than just an image; it contains information about the pixels, the patient, etc. The grey levels range from 0 to 255 and from -32768 to 32767 for 8-bit and 16-bit DICOM images respectively. In order to view the image with conventional software and use it with our machine learning algorithms, it needs to be saved in PNG or JPG format.

This conversion to PNG and JPG format is a crucial stage, and one for which particular attention has been paid to windowing. Windowing, also known as greyscale mapping, contrast stretching, histogram modification or contrast enhancement, is the process by which the greyscale component of an image is manipulated. This changes the appearance of the image to highlight particular structures. The brightness of the image is adjusted via the window level. The contrast is adjusted by the width of the window. In order to obtain images of the best possible quality, the choice of windowing parameters was entrusted to our experienced radiologist.

C. Region of interest

In order to train our model correctly (binary classifier), an expert senologist framed all the malignant anomalies in our possession. This was an extremely important task so as not to disrupt the training of our model with incorrect annotations. However, the benign tumours could not be framed due to the large quantity present. To do this, we automatically extracted all benign abnormalities from benign mammograms thanks to YOLOv5 and YOLOv6 algorithms.

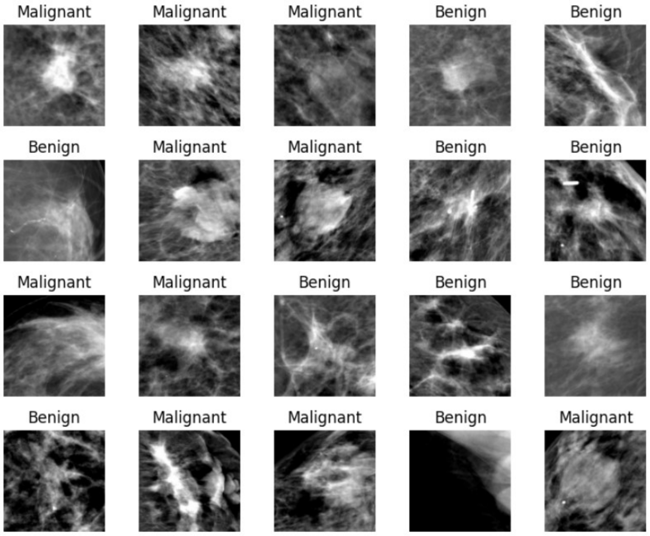

Fig. 1. Extraction of ROIs

III. EXPERIMENTAL RESULTS

A. *Localisation*

1) Tumors: The figure 2 below shows a comparison between the images annotated by the medical expert (ground truth) and our model. It should be noted that out of 224 mammograms, our model based on Yolo v6 was able to correctly locate lesions on 70% of mammograms and incorrectly on 5% of images. A confidence rate is also displayed for each anomaly detected. The first number given corresponds to the class of the abnormality (in this case, class 2 indicates that the tumour is malignant). We note that for the fourth image, our model has more difficulty with a confidence index of 49%.

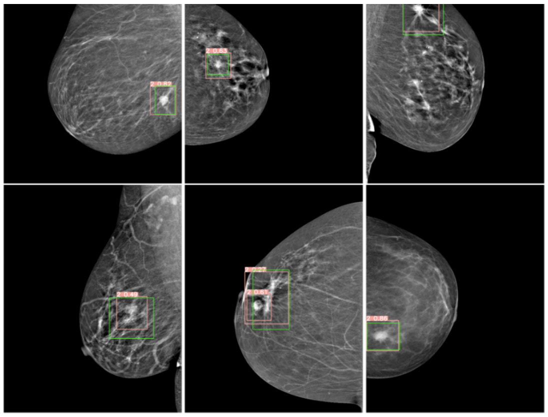

Fig. 2. Comparison between ground truth and prediction

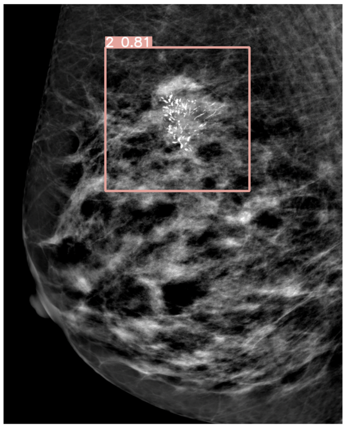

Fig. 3. Micro-calcifications (malignant) with a probability of 81%.

2) Niples: The figure 4 illustrates the location of the nipple.

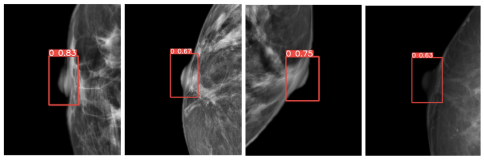

Fig. 4. Nipple localization.

It is important to be able to locate the nipple in order to determine the position of the tumour in relation to the nipple. In practice, to specify the location of the lesion for screening or diagnosis, each breast is divided into four quadrants with reference to the position of the nipple [1] as illustrated in 5. The majority of the mammary gland is located in the upper external quadrant (QSE).

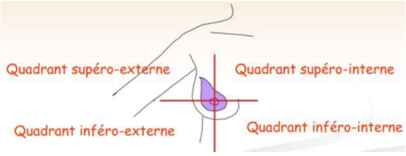

Fig. 5. Localization per quadrant [1]

B. *Experiments with publics datasets*

Our approach has been to work with an expert to annotate the malignant areas. Consequently, we were able to train a detection model and test it on the test dataset. Our second application was to study the diversification capability of this model on completely different mammography datasets. Indeed, the public datasets CMMD [3] and Inbreast [4] were examined with the model trained using annotations from a medical center's images.

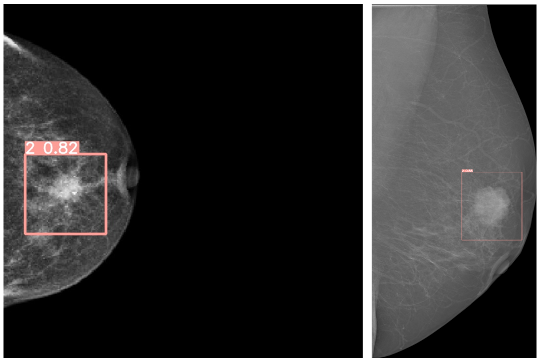

Fig. 6. Localization with CMMD and inbreast datasets

Unfortunately, less than 10% of the images from Inbreast were detected by the algorithm. Additionally, the confidence scores are below 0.6. This could be explained by the lower pixel intensity compared to our dataset, which may account for the lack of performance. However, two better results were obtained with CMMD, which is closer in terms of pixel intensity to our private dataset. Better confidence scores were achieved with a confidence rate of 82% and 2492 crops out of 4146 raw images.

C. *XAI*

In Figure 7, it is possible to observe the doctor's annotation as well as the detection of the model with its confidence rate. However, a detection that converges with that of the doctor is not sufficient to be able to affirm that our model is indeed based on the doctor's diagnosis. It is therefore necessary to justify the reasons for a detection in order to help the diagnosis but also to study the presence of potential bias. To build a breast cancer diagnosis aid for hospitals, the study of the explainability (XAI) of detection models. LRP was used to analyse the detection zone and assign importance to the source pixels of the detections. This explicability method introduced in 2015 by Binder et al [2] is used to find relevance scores for individual features in input data by decomposing CNN output predictions. It works on the basis of retropropagation, the aim of which is to highlight the pixels responsible for the classification. It is then possible to perform the inverse calculation of the output (which is a vector giving the probability of a class) and to obtain the importance of the combinations of pixels and therefore features giving the prediction.

To achieve this, [6] was used as a basis to apply LRP on a detection model in the medical field. Figure 3 whose class is malignant due to a set of microcalcifications. The model had correctly identified the malignant zone. It was therefore necessary to understand on which characteristics the model relied to identify this zone. Using the LRP algorithm applied to the detection model, it was possible to show that the model does indeed rely on microcalcifications for diagnosis.

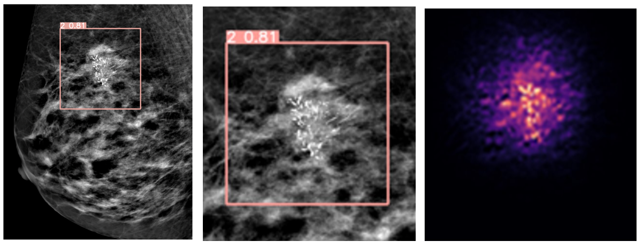

Fig. 7. Visualisation of the pixels responsible for malignant tumours with LRP.

The tumours present different aspects and have been studied using detection and explicability algorithms, as shown in the Figure 8. Indeed, it is important that the model understands that there are different malignant forms. As explained above, tissue intensity, shape and dispersion must be taken into account.

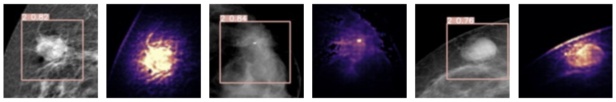

Fig. 8. Visualisation of the pixels responsible for malignant tumours with different aspects.

D. Classification

The results of the binary classification of benign and malignant tumours using the ResNet50 architecture are presented in Figures 9 and 10, while Figure 11 shows the results of the Visual Transformer model. We observe that we have achieved very promising results, with an area under the curve ROC of 92% for both architectures. The ROC (Receiver Operating Characteristic) curve, also known as the sensitivity/specificity curve, is used to calculate the ability of a screening test to distinguish between positive and negative images. The sensitivity of the test, i.e. the true positives, is shown on the y-axis and the false positives, i.e. the specificity, on the x-axis. The optimum value is in the top left-hand corner of the graph, indicating a high proportion of true positives and a low proportion of false positives.

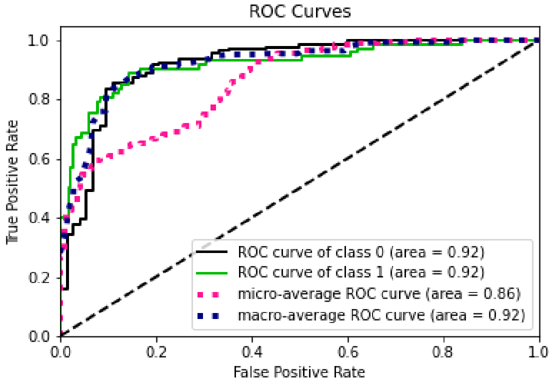

Fig. 9. ROC Curves (ResNet50 Model)

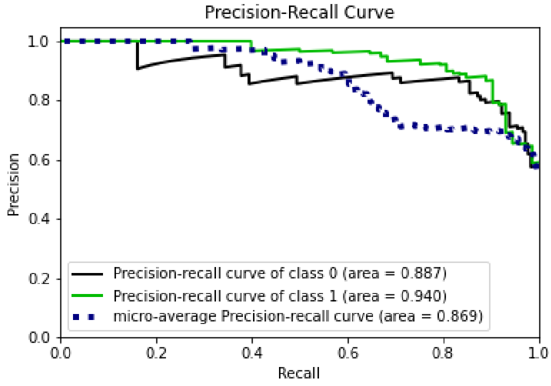

Fig. 10. Recall Precision Curve (ResNet50 Model)

Accuracy (or specificity), which represents the ability of a classifier to classify an image correctly, is defined as the ratio of true positives to the sum of true and false positives:

$$Specificity = \frac{True\ Positive}{True\ Positive\ + False\ Positive} \tag{1}$$

Recall (or sensitivity), which represents the ability of a classifier to identify all positive cases. This corresponds to the percentage of what was correctly classified for all the cases that were actually positive:

$$Sensitivity = \frac{True\ Positive}{True\ Positive\ + False\ Negative} \tag{2}$$

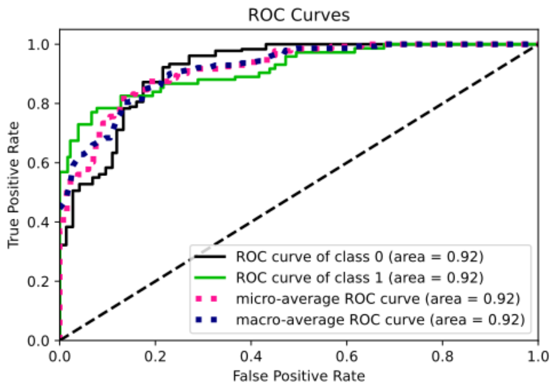

Fig. 11. ROC Curves (Transformer Model).

IV. CONCLUSION AND FUTURE WORK

In this paper, we proposed an innovative method to extract the anomalies (malignant and benign tumours) from a dataset of mammograpy labelized. This makes it possible to train an AI directly on the anomalies and not the entire mammogram.

One of these objectives would be to develop diagnostic aid software applied to mammographies. By combining two AI models, one to locate anomalies using our method and the other to perform a binary classification of the localised lesion (negative or positive) thanks to a model that will have been trained on positive and negative lesions. The model built based on the physician's annotations could be improved by using a greater number of annotations and by diversifying the sources of images from several hospitals. However, care must be taken to normalise the images from the different machines during pre-processing because, as can be seen in the Figure 12 below, the grey levels and contrasts are significantly different.

One solution to diversify our binary classification model with regard to the suspicion of malignant and benign tumours, and to have more data, would be to use federated learning for breast cancer in collaboration with several hospitals.

Finally, it is crucial to apply XAI algorithms, particularly in the context of diagnostic assistance software. In the medical field, a thorough understanding of the diagnosis is necessary for its validation, and there is also a risk of bias that can lead to positive results without taking essential features into account. These algorithms are of significant importance because they are able to focus on tiny, potentially malignant features that can be difficult to spot.

ACKNOWLEDGMENT

This research was partially funded by Infortech of UMONS and CETIC.

REFERENCES

[1] URL: https://univ.ency-education.com/uploads/1/3/1/0/13102001/anato2an-glande_mammaire2019grine.pdf (visited on 08/17/2023).

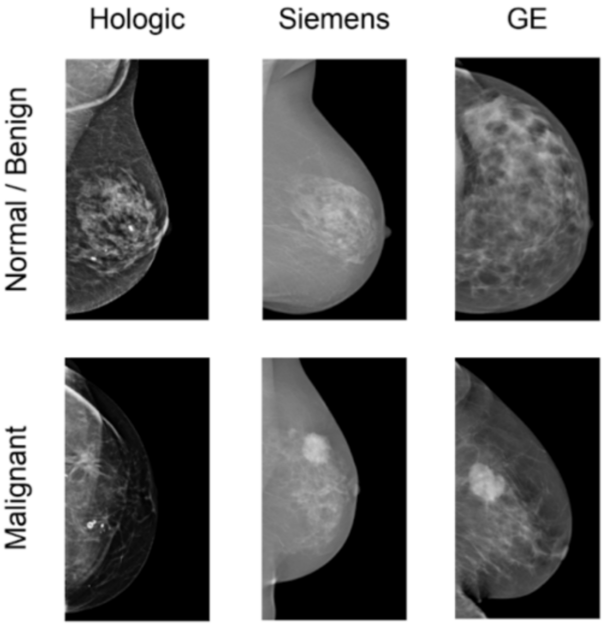

Fig. 12. Mammograms from differents machines [5]

[2] Binder Alexander et al. "Layer-wise Relevance Propagation for Neural Networks with Local Renormalization Layers". In: (2016). DOI: 10.48550/arXiv.1604.00825.

[3] Cui et al. "The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast". In: (2021).

[4] Moreira IC et al. "INbreast: toward a full-field digital mammographic database". In: (2012). DOI: 10.1016/j.acra.2011.09.014.

[5] Amelia Jiménez-Sánchez et al. "Memory-aware curriculum federated learning for breast cancer classification". In: Computer Methods and Programs in Biomedicine (2022), p. 107318. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2022.107318. URL: https://www.sciencedirect.com/science/article/pii/S016926072200699X.

[6] Karasmanoglou et al. "Heatmap-based Explanation of YOLOv5 Object Detection with Layer-wise Relevance Propagation". In: June 2022, pp. 1–6. DOI: 10.1109/IST55454.2022.9827744.

[7] Ramachandran S. et al. "Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans". In: Medical Imaging 2018: Computer-Aided Diagnosis. Ed. by Nicholas Petrick and Kensaku Mori. Vol. 10575. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Feb. 2018, 105751I, p. 105751I. DOI: 10.1117/12.2293699.