Towards Improving Accurate Breast Cancer Diagnosis: Leveraging Pre-trained Convolutional Neural Network for Mammogram Analysis

Marwa Ben Ammar
Research Laboratory of Biophysics and Medical Technologies
Higher Institute of Medical Technologies of Tunis (ISTMT), University of Tunis el Manar
Tunis, Tunisia
marwa.ammar@istmt.utm.tn

Dorra Zaibi
Research laboratory Innov'COM « Innovation of COMmunicant and COoperative Mobiles Laboratory »
Higher School of Communication of Tunis (SUPCOM), University of Carthage
Tunis, Tunisia
e-mail: dorra.zaibi@supcom.tn

Faten Labbene Ayachi
Research laboratory Innov'COM « Innovation of COMmunicant and COoperative Mobiles Laboratory »
Higher School of Communication of Tunis (SUPCOM), University of Carthage
Tunis, Tunisia
e-mail: faten.labbene@supcom.tn

Riadh Ksantini
Department of Computer Science
College of IT, University of Bahrain
Bahrain, Bahrain
e-mail: ksontiniriadh@yahoo.fr

Halima Mahjoubi
Research Laboratory of Biophysics and Medical Technologies
Higher Institute of Medical Technologies of Tunis (ISTMT), University of Tunis el Manar
Tunis, Tunisia
e-mail: halima.mahjoubi@istmt.utm.tn

***Abstract*— Breast cancer poses a significant global health challenge, emphasizing the need for improved diagnostic approaches for early diagnosis and intervention. Mammography, a widely used screening method, provides valuable insights into breast tissue anomalies. Nevertheless, its effectiveness is marred by error-prone interpretations and time-consuming analyses. To address this, our study introduces an innovative strategy to enhance breast cancer diagnosis by employing a Three-Stage One-Class You Only Look Once (YOLO) classification framework, harnessing the power of Deep Learning (DL). By incorporating the YOLO-v8 network, cutting-edge convolutional neural network (CNN) architecture, our proposed methodology aims to mitigate the shortcomings of conventional mammography interpretation. To assess the model's effectiveness, we utilize the Mammography Image Analysis Society (MIAS) dataset, which encompasses inherent data imbalances and intricacies. The framework we present is divided into three stages, each contributing to the refinement of the diagnostic process. Through the application of a one-class classification technique, our model effectively distinguishes between normal and abnormal mammograms. Furthermore, it offers a higher level of granularity by categorizing abnormalities into masses or calcifications. Additionally, the model can differentiate between benign and malignant cases, thereby facilitating precise clinical decision-making.**

Keywords- Breast cancer; mammography; deep learning; YOLO; early diagnosis; one-class classification; three stages methodology; data imbalance

I. INTRODUCTION

In this section, we will provide an overview of the research problem, outline the research questions, and delineate the research objectives. This introductory segment aims to set the stage for a comprehensive understanding of the context and purpose of our research.

A. Research Problem

Breast cancer has taken the lead as the most commonly diagnosed cancer and the fifth cause of cancer deaths among women worldwide. Therefore, only an early and accurate breast cancer diagnosis can significantly improves patient survival rates and paves the way for effective treatment. Mammography remains the most widely utilized by radiologists for accurate breast cancer diagnosis.

 ISBN: 978-1-68558-098-8

Nonetheless, the ever-increasing volume of daily mammograms presents a real challenge for radiologists and physicians, potentially resulting in diagnostic errors and unnecessary biopsies. Two significant types of errors can occur: False-Positive (FP) and False-Negative (FN). False positives carry negative consequences as they misidentify benign areas as cancerous. More critically, false negatives jeopardize patient lives as they occur when radiologists fail to detect abnormalities. Moreover, studies have shown that to reduce FN diagnoses, biopsies are recommended for lesions with a greater than 2% likelihood of malignancy. Consequently, only 15–30% of patients referred for biopsy are ultimately found to have malignancies [1]. To tackle these challenges, our research proposes a computer-aided diagnosis system using the You Only Look Once version 8 (YOLOv8) Convolutional Neural Network (CNN). It operates in three stages, leveraging a One-Class Classification (OCC) approach to effectively detect normal and abnormal mammograms, categorize abnormalities as masses or calcifications, and identify benign and malignant cases for precise clinical decision-making. In the following section, we highlight key Research Questions (RQs) from existing literature that inform our proposed breast cancer diagnosis model.

B. Research Questions

Taking into account the latest advancements in breast cancer diagnosis using deep learning techniques, our research focused on the following key questions: Which imaging method is most effective in detecting breast cancer? Which algorithm can efficiently and accurately detect and classify breast cancer within a unified framework? How can diagnostic accuracy be improved while minimizing the need for biopsies and reducing errors in identifying malignant cancers? Which model has the highest accuracy rate across all databases? Currently, there is no tool capable of diagnosing breast cancer with both a high degree of accuracy and minimal errors, while also minimizing the number of required biopsies.

C. Outline of Objectives

Our primary research objective is to develop a novel and highly effective YOLOv8-based model for breast cancer diagnosis using mammograms. Our specific focus includes achieving : (1) Reliability: Our model aims for high accuracy, sensitivity, specificity, precision, False-Negative (FN), False-Positive (FP), F1-score, Receiver Operating Characteristic (ROC) curve, Area Under The Curve (AUC), Intersection Over Union (IOU) score, and mean Average Precision (mAP), as these metrics are crucial in medical images analysis [2]; and (2) Transferability: We want our model to be adaptable across different datasets, even when transitioning from analyzing mammograms to other domains like lung X-ray images These objectives are in direct alignment with the research queries outlined in the preceding subsection focusing on research questions.

The subsequent sections of this study are structured as follows: Section 2 delivers a concise review of the current State-Of-The-Art in breast cancer diagnosis using deep learning algorithms on mammography, accompanied by an overview of their results. Section 3 provides an in-depth exposition of our research methodology. Finally, in Section 4, we wrap up the paper by briefly summarizing the expected research outcomes, detailing the present stage of our research, and offering insights for potential future investigations.

II. LITERATURE REVIEW

In this section, we attempt to cover most recent research works that have been done related to diagnosis of breast cancers with applying various techniques of deep learning along with their results. Muduli et al. [3] proposed a deep CNN model for breast cancer classification in mammogram and ultrasound images. This CNN model achieved 96.55%, 90.68%, and 91.28% accuracy on MIAS, DDSM, and INbreast datasets, respectively. Additionally, it reached 100% and 89.73% accuracy on BUS-1 and BUS-2 datasets. Zhao et al. [4] developed three YOLOv3-based models for breast cancer detection and classification using mammograms: a general model, mass model, and microcalcifications model. Their study achieved detection accuracy rates of 93.667%, 97.767%, and 96.870%, and classification accuracy rates of 93.927%, 98.121%, and 97.045%, respectively, using the CBIS-DDSM dataset. Baccouche et al. [5] used an end-to-end YOLO-based fusion model to detect and classify breast lesions (mass, calcification, architectural distortion) in digital mammograms with the UCHC DigiMammo dataset. The approach incorporated prior mammograms for early detection and retrospective prediction. The evaluation achieved detection rates of 93% for mass lesions, 88% for calcification lesions, and 95% for architectural distortion lesions in current mammography. Zebari et al. [6] constructed a breast cancer detection model from mammograms using a pre-trained CNN-based approach. Testing on the mini-MIAS dataset resulted in an impressive accuracy of 95.71%. Alam et al. [7] applied the Unet3+ architecture for semantic segmentation to enhance breast cancer diagnosis in ultrasound images of 309 patients. The Unet3+ model outperformed other models (FCN, Unet, SegNet, DeeplabV3+, and pspNet) with an average accuracy of 82.53%, an intersection over union of 52.57%, a weighted accuracy of 89.14%, and a global accuracy of 90.99%. Boudouh et al. [8] investigated seven pre-trained CNNs for accurate breast tumor detection: Xception, InceptionV3, ResNet101V2, ResNet50V2, ALexNet, VGG16, and VGG19. They gathered data from three distinct databases: MiniMIAS, DDSM, and CMMD. The results were impressive, particularly for ResNet50V2 and InceptionV3, which achieved the highest accuracy rates of 99.9% and 99.54%, respectively. Despite the accomplishments of these computer-aided diagnostic methods, challenges like high memory complexity, practical implementation, and extended runtime persist. Furthermore, these approaches have the following flaws. First, the accuracy of recognizing probable small lesions is quite poor. Second, except for [4] and [5], techniques only identify breast mass lesions, disregarding

 ISBN: 978-1-68558-098-8

other types like microcalcifications in mammography. As a result of these issues, the method's limitations have been discovered. To address the constraints noted above, we focus on the detection and classification of mammograms while also addressing the issues of different types of lesions and small-sized lesions, and we propose a YOLOv8-based computer-aided diagnostic system for mammograms.

Section III summarizes our thorough contributions to the field of breast cancer diagnosis using deep learning techniques and mammography image analysis.

III. METHODOLOGY

Our research aims to develop an accurate breast tumor detection and classification model while reducing FP and FN outcomes. We utilize the publicly available Mammographic Image Analysis Society Digital Mammogram (MIAS) dataset, which can be conveniently accessed through a user-friendly online interface [10]. Our methodology comprises four steps, as illustrated in Fig. 1: data acquisition and splitting, image preprocessing, YOLOv8 deep learning model deployment with OCC approach, and comprehensive performance evaluation. In the subsequent subsections, we will provide detailed insights into each step.

A. Dataset Description

The MIAS dataset, established by a UK research consortium, comprises 322 single-slice digital mammograms from 161 patients. The dataset covers various breast abnormalities, as detailed in Table 1. However, The MIAS dataset has limitations: it's small, potentially causing overfitting in deep learning models; it's imbalanced, with 207 normal and 115 abnormal cases, impacting classification algorithms; and it requires preprocessing to remove extraneous data outside the mammary area.

TABLE I. COMPREHENSIVE DATA DISTRIBUTION OVERVIEW

Total number of Patients	161
Total number of Mammograms	322
Total number of Mammograms with Pathology	115
Total number of Mammograms without Lesions (Normal)	207
Number Of Mammograms With Mass Lesions	57
Number of Mammograms with Calcification Lesions	25
Number of Mammograms with Architectural Distortion Lesions	18
Number of Mammograms with Asymmetry Lesions	14

In this section, we've introduced the mammography dataset that forms the basis of our study. The subsequent sections will provide detailed insights into the techniques utilized for image data preprocessing, a clear explanation of our classification methodology, a discussion on model selection, and an overview of our performance evaluation process.

B. Image Data Preprocessing

To enhance our dataset's quality and applicability. Our approach includes various techniques like noise removal, contrast enhancement, data augmentation, resizing, and normalization for each mammogram breast image within the MIAS dataset. This dataset contains different noises and imaging artifacts, such as tape artifacts and high-intensity rectangular labels, as illustrated in Fig. 2, which need to be removed. Additionally, MIAS mammograms have limited contrast, prompting us to consider contrast enhancement techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE). Besides, we perform image resizing, data augmentation, and normalization to align input images with CNN requirements and address the small dataset size challenge. Data augmentation involves random transformations like rotation and flipping to diversify and expand the training data.

C. Image Data Classification

Our study focuses on developing a YOLOv8-based pipeline for breast mammogram detection and classification. It includes three key stages: detecting abnormalities as normal or abnormal, distinguishing masses from microcalcifications, and classifying benign or malignant cases.

Our selection of the YOLOv8 model for our breast cancer diagnosis study using mammograms is based on several key considerations that collectively make it exceptionally well-suited for this task. Firstly, YOLOv8 is renowned for its efficiency and speed in object detection and classification tasks, aligning perfectly with our goal of providing an efficient model for breast cancer diagnosis. Its real-time capabilities are essential for swift and accurate diagnosis. Additionally, YOLOv8 has demonstrated outstanding accuracy in object detection, a crucial aspect for identifying abnormalities in mammograms, which is fundamental in the context of breast cancer diagnosis.

Moreover, YOLOv8's architectural versatility is a significant advantage. The model is capable of handling various object detection tasks, a valuable trait considering the diverse abnormalities and conditions that can be present in mammograms. In the realm of breast cancer diagnosis, this flexibility is highly advantageous.

Furthermore, our study emphasizes the importance of achieving high transferability across different datasets. YOLOv8's adaptability to varying data distributions and its ability to generalize well across diverse datasets ensure that our model can maintain consistent performance, regardless of the specific dataset it is applied to.

Last but not least, the YOLOv8 model is part of a well-established family of models with a substantial user base and ongoing research efforts. This provides us with access to valuable resources, pre-trained models, and a vibrant community of researchers continually working on model improvements and adaptations. In conclusion, our choice of the YOLOv8 model is well-founded in its remarkable efficiency, accuracy, versatility, and transferability, making it an ideal candidate for enhancing breast cancer diagnosis through deep learning techniques. Its real-time capabilities,

 ISBN: 978-1-68558-098-8

combined with its adaptability to different datasets, position it as a robust choice for our mission of advancing breast cancer diagnosis.

The imbalance within the MIAS dataset, consisting of 207 normal cases and 115 abnormal cases, significantly impacts the classification process. Imbalanced datasets often pose challenges for traditional binary or multi-class classification methods, as they tend to favor the larger class, making it difficult to accurately detect the minority class. One effective approach to address these issues is the utilization of OCC. This approach is particularly valuable in domains such as medical image diagnosis, where acquiring data from both healthy and unhealthy patients can be impractical due to high costs or rarity.

In the context of OCC approach, the primary objective is to classify data when information is available for only one group of observations. OCC methods operate with a single dataset, referred to as the "target class," typically representing the class with fewer instances. The aim is to distinguish data belonging to the target class from other potential classes. OCC can be viewed as a specialized form of the two-class classification problem, where only data from one class is considered during the training and validation phases. However, during inference, the classifier encounters data from both the target class and classes outside the target.

In our study, we adopt the terminology of "target" and "outside the target" to differentiate between abnormalities and normal cases, masses and calcifications, and malignant and benign cases. This approach allows us to effectively address the classification challenges presented by the MIAS dataset's class imbalance.

Our choice to primarily employ the MIAS dataset for evaluation was motivated by several key factors. First and foremost, the MIAS dataset stands as one of the most renowned and widely utilized datasets in the realm of mammography image analysis. With a substantial number of mammogram images accompanied by annotations, it serves as a valuable benchmark for our model's performance.

In this study, our principal goal was to construct and benchmark our YOLOv8-based breast cancer diagnosis model within the context of a well-recognized dataset, the MIAS dataset. This approach allows us to assess the model's performance within a known benchmark and aligns with the core objectives of our research.

Moreover, by concentrating our initial evaluation on the MIAS dataset, we intended to establish a solid baseline for our model's performance. Once this robust baseline is achieved, we are fully prepared to extend our evaluation to encompass the other datasets referenced in our literature review.

It is important to acknowledge that evaluating a model on multiple datasets can be resource-intensive and time-consuming. Therefore, it is a customary practice in research to commence with a specific dataset to validate the model's viability before proceeding to a broader spectrum of datasets.

While our preliminary evaluation centers on the MIAS dataset, we are fully cognizant of the significance of future research endeavors that will encompass a wider array of datasets outlined in the literature review. This expansion is vital for a comprehensive assessment of the model's robustness and adaptability across diverse data sources, as envisaged in our research question. We are dedicated to advancing our research to address this aspect thoroughly, ensuring our model's performance is rigorously validated across a broader range of datasets, in line with the goals of our research.

D. Performance Evaluation

Our research will primarily center on evaluating the YOLOv8-based model's capacity to accurately identify the position of breast lesions in mammograms. We will employ two key metrics for this assessment: the IOU score and the mAP.

Following that, we will shift our focus to gauge the performance of the YOLOv8 model. In practical terms, the effectiveness of deep learning-based image classification is determined through a range of metrics, including accuracy, sensitivity, specificity, precision, FP and FN rates, the ROC curve, the AUC, and F1-score.

Consequently, as depicted in Fig. 1, our research work will incorporate a total of ten essential metrics for a comprehensive evaluation. The true positive (TP) represents the number of positive classes that have been correctly classified as positive. The true negative (TN) is the number of negative classes that that have been correctly classified as negative. The false positive (FP) represents the number of negative classes that have been misclassified as the positive class. The false negative (FN) represents the number of positive classes that have been misclassified as negative.

Below, we briefly outline the calculation formulas for the evaluation metrics used.

- mean Average Precision (mAP)

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{APi}. \tag{1}$$

- Intersection Over Union score (IOU)

$$\text{IOU score} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \tag{2}$$

- Accuracy (Acc)

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

- Precision (Pr)

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

- Sensitivity (Sn)

$$\mathrm{Sn} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{5}$$

- Specificity (Sp)

$$Sp = \frac{TN}{(TN + FP)} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

- F1-Score

$$F1 - Score = 2 \times \frac{Recall \times Precision}{(Recall + Precision)} \quad (7)$$

- ROC-AUC

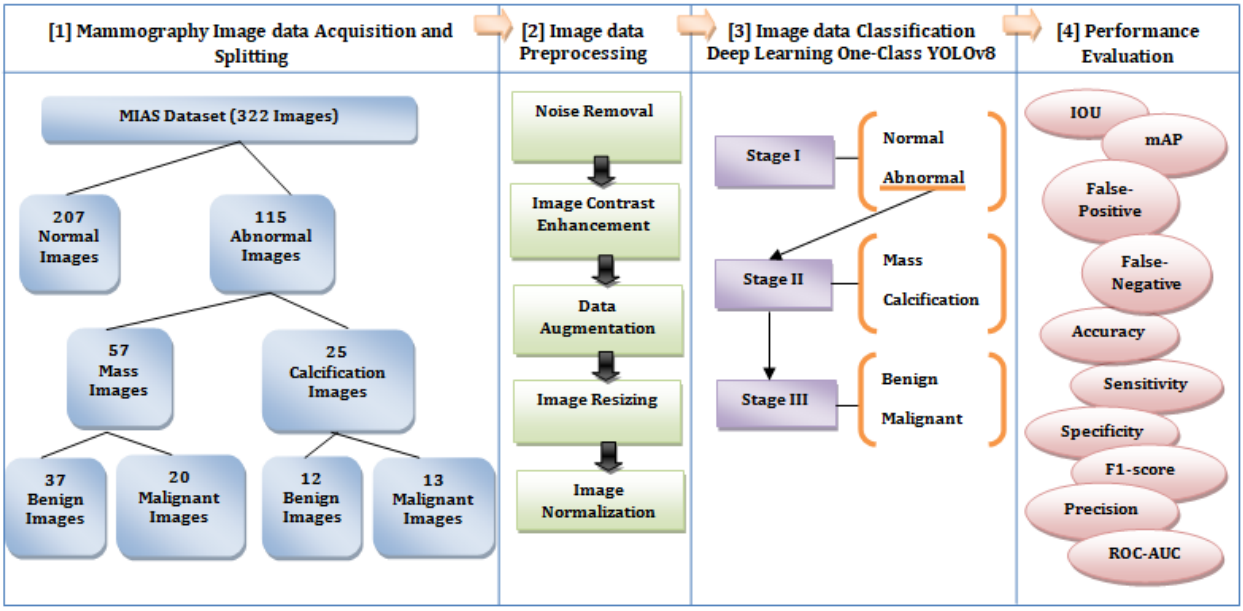

Figure 1. Framework for breast lesions detection and classification using deep learning one-class YOLOv8.

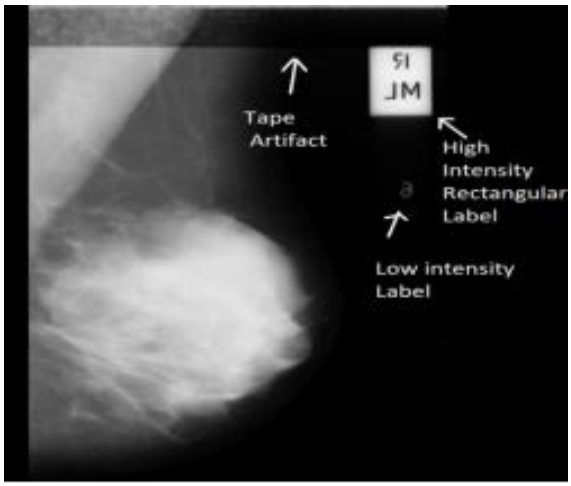

Figure 2. Examples of Noise Artifacts in MIAS Dataset.

IV. CONCLUSION AND FUTURE WORK

In the course of our research endeavor, our primary objective is to develop a state-of-the-art end-to-end learning model designed to diagnose breast cancer by analyzing mammogram images. Our vision for this innovative model centers on two critical aspects: (1) ensuring high reliability and (2) facilitating seamless transferability.

To this end, we have successfully reached several key milestones in our research journey. These milestones encompassed:

- Selection of the most appropriate image modality and dataset.
- Identification of widely adopted image preprocessing techniques.
- Determination of the deep learning model for breast cancer diagnosis and the classification approach.
- Selection of evaluation metrics meticulously designed to assess the proposed model's effectiveness.

Moreover, we have made significant progress by developing an initial YOLOv8-based algorithm. This algorithm, constructed using Python within the Google Colab Notebook, is adept at detecting breast lesions by distinguishing between normal and abnormal cases. The preliminary results of this algorithm, as demonstrated in Fig. 3, illustrate the proficiency of the YOLOv8 model in identifying abnormal breast lesions.

 ISBN: 978-1-68558-098-8

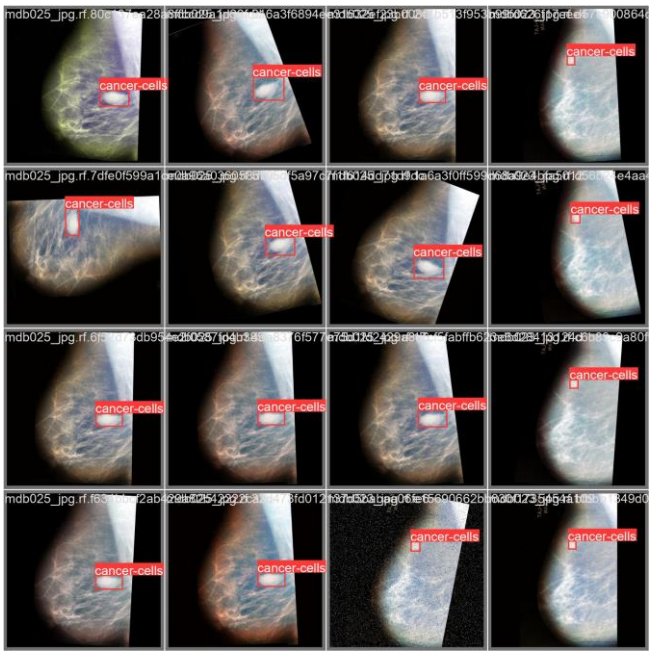

Figure 3. Example of Preliminary Outcomes of the YOLOv8 Model in Detecting Unusual Breast Lesions.

As we draw our research efforts to a close in this exploration of an advanced framework for mammogram image analysis in breast cancer diagnosis, it's important to note that our experiments are currently underway. The preliminary findings are highly promising, and we are on the brink of realizing the full potential of this groundbreaking technology. We eagerly anticipate sharing the final results, which have the potential to reshape the landscape of breast cancer diagnosis.

Looking forward, we propose a novel avenue for future work: a multimodal fusion architecture that leverages both breast images and tabular non-image data. This architecture incorporates a probability fusion approach, often referred to as "late fusion." It operates on the basis of considering the output probabilities from an image-only model and a non-image-only model, with the aim of yielding a final prediction. The underlying idea is that by incorporating non-image data alongside image data, we can significantly enhance predictive performance compared to a unimodal (single-source) approach.

Broadly, our proposal entails a decision-making pipeline that enables a hierarchical classification of breast cancer. To achieve this, we propose to aggregate the predictions from two models: the Deep Support Vector Data Description (DSVDD) and the One-Class Convolutional Neural Network (OCCNN) through a meta-model, consisting of a simple two-layer Multilayer Perceptron (MLP). This approach is geared towards improving the accuracy of breast cancer diagnosis.

For the training of these models, we consider specific combinations of feature subsets and model architectures. In our pursuit of exploring whether the fusion of image and non-image features can enhance breast cancer prediction, we propose to begin by establishing unimodal baseline models that exclusively employ image data and tabular non-image data. Subsequently, we propose to embark on the development of a multimodal fusion model that jointly learns from both image and non-image data.

REFERENCES

[1] E. A. Sickles, "Periodic mammographic follow-up of probably benign lesions: results in 3184 consecutive cases," Radiology., vol. 179, pp. 463–468, 1991.

[2] E. Mahoro, M. A. Akhloufi, "Applying Deep Learning for Breast Cancer Detection in Radiology," Curr. Oncol., vol. 29, pp. 8767-8793, 2022.

[3] D. Muduli, R. Dash, and B. Majhi, "Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach," Biomed. Signal Process. Control., Vol. 71, ISSN. 1746-8094, 2022.

[4] J. Zhao, T. Chen, and B. Cai, "A computer-aided diagnostic system for mammograms based on YOLOv3," Multimed. Tools Appl., vol. 81, pp. 19257–19281, 2022.

[5] A. Baccouche, B. Garcia-Zapirain, Y. Zheng, and A. S. Elmaghraby, "Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques," Comput. Methods Programs Biomed., Vol. 221, 2022.

[6] D. A. Zebari, H. Haron, D. M. Sulaiman, Y. Yusoff, and M. N. Mohd Othman, "CNN-based Deep Transfer Learning Approach for Detecting Breast Cancer in Mammogram Images," IEEE 10th Conference on Systems, Process & Control (ICSPC)., pp. 256-261, 2022.

[7] T. Alam, W. C. Shia, F. R. Hsu, and T. Hassan, "Improving Breast Cancer Detection and Diagnosis through Semantic Segmentation Using the Unet3+ Deep Learning Framework," Biomedicines., vol. 11, pp. 1536, 2023.

[8] S. S. Boudouh, M. Bouakkaz, "Breast cancer: toward an accurate breast tumor detection model in mammography using transfer learning techniques," Multimed Tools Appl., vol. 82, pp. 34913–34936, 2023.

[9] G. H. Aly, M. Marey, S. A. El-Sayed, M. F. Tolba, "YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms," Comput. Methods Programs Biomed., Vol. 200, ISSN. 0169-2607, 2021.

[10] https://www.kaggle.com/datasets/kmader/mias-mammography.

 ISBN: 978-1-68558-098-8