International Academy, Research, and Industry Association

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

Article Authors

Marwa Ben Ammar, Dorra Zaibi, Faten Labbene Ayachi, Riadh Ksantini, Halima Mahjoubi

Presenter's

Marwa Ben Ammar

Higher Institute of Medical Technologies of Tunis

marwa.ammar@istmt.utm.tn

Brief Presenter's Resume

Marwa Ben Ammar is a PhD student in Biophysics and Medical Imaging at the Laboratory of Research in Biophysics and Medical Technologies (LRBTM) at the Higher Institute of Medical Technologies of Tunis (ISTMT), University of Tunis El Manar.

Her research focuses on using Artificial Intelligence techniques to develop an advanced breast cancer diagnosis system for Tunisian women.

Marwa holds a Bachelor's degree in Experimental Science (2010), an Applied Licence in Biomedical Engineering (2013), and completed a Professional Master's in Biomedical Instrumentation (2015) along with a Research Master's in Biophysics, Radiophysics, and Medical Imaging (2017) at the Higher Institute of Medical Technology of Tunis.

Her expertise encompasses medical devices, programming languages, Business Process Management (BPM), medical image processing, and artificial intelligence (AI) techniques.

Contact information at the end of this presentation...

Outlines of the Presentation

- **Global Research Project**
- **Area of interest**
- **Current Research Focus**
    - *Research Problem*
    - *Research Questions*
    - *Research Objectives*
    - *Literature Review*
    - *Proposed Method*
    - *Conclusion and Future Work*

Global Research Project:

Diagnosis and Early Detection of Breast Cancer in Women

Target Business Process(BP)	**Diagnosis Of Breast Cancer In Women.**
Research Objectives	• Understand and Model Existing Breast Cancer Diagnosis Process in Women. • Analyze Diagnostic Data. • Design a Data Warehouse for Tunisian Women with Breast Cancer. Published Work: "**Data Warehouse for Machine Learning: Application to Breast Cancer Diagnosis**" (Authors: Marwa Ben Ammar, Faten Labbene Ayachi, Riadh Ksantini, Halima Mahjoubi; Published in: https://doi.org/10.1016/j.procs.2021.12.065). • Assess Current Diagnostic Methods. • Evaluate Deep Learning (DL) Techniques. • Develop a Breast Cancer Diagnostic Aided System.
Methodological Support	• Business Process Management and Notation (BPMN) Approach. • Relational Schema Paradigm. • Deep Learning (DL).
Contribution	Total Framework: Data Warehouse & Decision-Aid Platform for Breast Cancer Diagnosis.
Where to deploy this System ? ***For whom ?***	In hospitals. Doctors and Radiologists.
Why ?	• Enhanced Diagnostic Precision. • Comparative Analysis Across Patient Pathways. • Anticipation of Lengthy Procedures. • Long-term Data Preservation. • Tailored Solutions for Target patient populations.

Area of interest:

Hospital Business Processes

Published Work:	**Title: A reverse-engineering methodology for medical enhancement processes.** **Authors: Faten Labbene Ayachi, Hanen Boussi Rahmouni, Marwa Ben Ammar, Halima Mahjoubi.** **Published in: DOI :10.1016/j.procs.2019.12.240**
Master's Research Project:	"Towards a Quality Metric for Hospital Business Processes and a Business Benchmark for Tunisian HIS" (Higher Institute of Medical Technologies of Tunis - 2016/2017)
Target Business Process	**Hospitalized Patient Trajectory With Complementary Acts.**
Context of the Study	● **Habib Thameur University Hospital** (considered as a reference institution). ● **Charles Nicolle University Hospital.**
Major Goals	● Define and analyze the existing business processes (BP) in different Tunisian hospital sites. ● Identify malfunctioning and shortcomings in BP. ● Assess the degree of integration of ICT technologies in BP. ● Evaluate the processes' quality when deployed on different health care sites. ● Maintain and disseminate PROCESS CARTOGRAPHY.

Current Research Focus:

Deep Learning (DL)-driven Breast Cancer Diagnosis

- Leverage Deep Learning (DL) to revolutionize the diagnostic process and ultimately contribute to more accurate and efficient breast cancer diagnosis.

- In particular, Leveraging Pre-trained Convolutional Neural Network (CNN) for Mammogram Analysis.

Current Research Focus - **Research Problem (½)**

- Breast Cancer (BrC) stands as the most prevalent form of cancer among women, constituting a significant global health challenge and the primary contributor to cancer-related fatalities on a global scale.

- The statistics underscore the magnitude of this issue, with **2.3 million new cases** reported and a staggering **685,000 documented deaths** in the year 2020 alone.[1]

➢ These figures illuminate the urgent need for enhanced diagnostic methodologies and interventions to mitigate the impact of Breast Cancer on women's health worldwide.

[1] Lei S, Zheng R, Zhang S, Wang S, Chen Ru, Sun K, et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Commun. 2021;1–12. https://doi.org/10.1002/cac2.12207.

Current Research Focus - **Research Problem (2/2)**

- **Mammography stands out as the gold standard for BrC detection and diagnosis.**
  - Pivotal in women's healthcare, leading to a 20% reduction in breast cancer-related mortality.
  - Cost-Efficient.

- **Mammography, while crucial, faces challenges with rising volumes and potential errors.**
  - Types of Errors: False-Positives (FP) and False-Negatives (FN).
  - Consequences of Errors: FP misidentifies benign areas; FN jeopardizes patient lives.

➢ **Integrating Deep Learning-Computer-Aided Diagnosis (DL-CAD), particularly CNNs, offers a promising solution by significantly reducing diagnostic errors in breast cancer diagnosis.**

➢ Significant Reduction in Errors: 85% decrease in human error rates for breast cancer diagnoses.

Current Research Focus - **Research Questions (RQs)**

Taking into account the latest advancements in breast cancer diagnosis using deep learning techniques, our research is centered on addressing the following key questions:

- **RQ1: Understanding Breast Cancer**
  What essential facts are crucial for comprehending breast cancer?
- **RQ2: Optimal Breast Cancer Detection**
  Which imaging method stands out as the most effective in the detection of breast cancer?
- **RQ3: Mammography Datasets for Diagnosis**
  Which publicly available mammography datasets are frequently used in breast cancer diagnosis research?
- **RQ4: Efficient Breast Cancer Classification**
  Which algorithm offers an efficient and accurate approach to detecting and classifying breast cancer within a unified framework?
- **RQ5: Improving Diagnostic Accuracy**
  How can we enhance diagnostic accuracy while reducing the necessity for biopsies and minimizing errors in identifying malignant cancers?
- **RQ6: Model Accuracy Across Databases**
  Which model demonstrates the highest accuracy rate when applied to different databases?
- **RQ7: Evaluation Metrics in Breast Cancer Diagnosis**
  What are the current evaluation metrics employed to assess the performance of computer-aided diagnosis systems for breast cancer?

Current Research Focus - **Research Objectives**

Main Objective

Develop Pre-trained CNN-Based Breast Cancer Model

- **Focus:** High efficacy and novelty.
- **Tool:** YOLOv8 model for mammogram analysis.

Specific Objectives

Emphasizing Reliability and Ensuring Transferability

(1) Establishing Reliability Metrics:

- **Ensure Precision:** Accuracy, Sensitivity, Specificity, Precision.
- **Minimize Errors:** False-Negative (FN), False-Positive (FP), F1-score.
- **Visualize Performance:** ROC curve, AUC, IOU score, mean Average Precision (mAP).

(2) **Enabling Transferability Across Diverse Datasets:**

- Adaptation Beyond Mammograms:
    - Seamless transition to other medical domains (e.g., lung X-ray images).

Current Research Focus - **Literature Review (½)**

Authors and Reference	Methodology and Results
Muduli et al. [2]	Proposed a deep CNN model for breast cancer classification in mammogram and ultrasound images. Achieved accuracy of 96.55%, 90.68%, and 91.28% on MIAS, DDSM, and INbreast datasets, respectively. Reached 100% and 89.73% accuracy on BUS-1 and BUS-2 datasets.
Zhao et al. [3]	Developed three YOLOv3-based models for breast cancer detection and classification using mammograms. Achieved detection accuracy rates of 93.667%, 97.767%, and 96.870%, and classification accuracy rates of 93.927%, 98.121%, and 97.045% using the CBIS-DDSM dataset.
Baccouche et al. [4]	Used an end-to-end YOLO-based fusion model to detect and classify breast lesions in digital mammograms. Achieved detection rates of 93% for mass lesions, 88% for calcification lesions, and 95% for architectural distortion lesions.

[2] 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.
[3] 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.
[4] 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.

Current Research Focus - **Literature Review (2/2)**

Authors and Reference	**Methodology and Results**
Zebari et al. [5]	Constructed a breast cancer detection model from mammograms using a pre-trained CNN-based approach. Achieved an impressive accuracy of 95.71% on the mini-MIAS dataset.
Alam et al. [6]	Applied the Unet3+ architecture for semantic segmentation to enhance breast cancer diagnosis in ultrasound images. Outperformed other models with an average accuracy of 82.53%.
Boudouh et al. [7]	Investigated seven pre-trained CNNs for accurate breast tumor detection: Xception, InceptionV3, ResNet101V2, ResNet50V2, ALexNet, VGG16, and VGG19. Achieved high accuracy rates, with ResNet50V2 and InceptionV3 leading with rates of 99.9% and 99.54%, respectively.

[5] 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.
[6] 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.
[7] 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.

Current Research Focus - **Proposed Method (⅕)**

Objective:

Develop an accurate breast tumor detection and classification model, aiming to reduce False Positives (FP) and False Negatives (FN) outcomes.

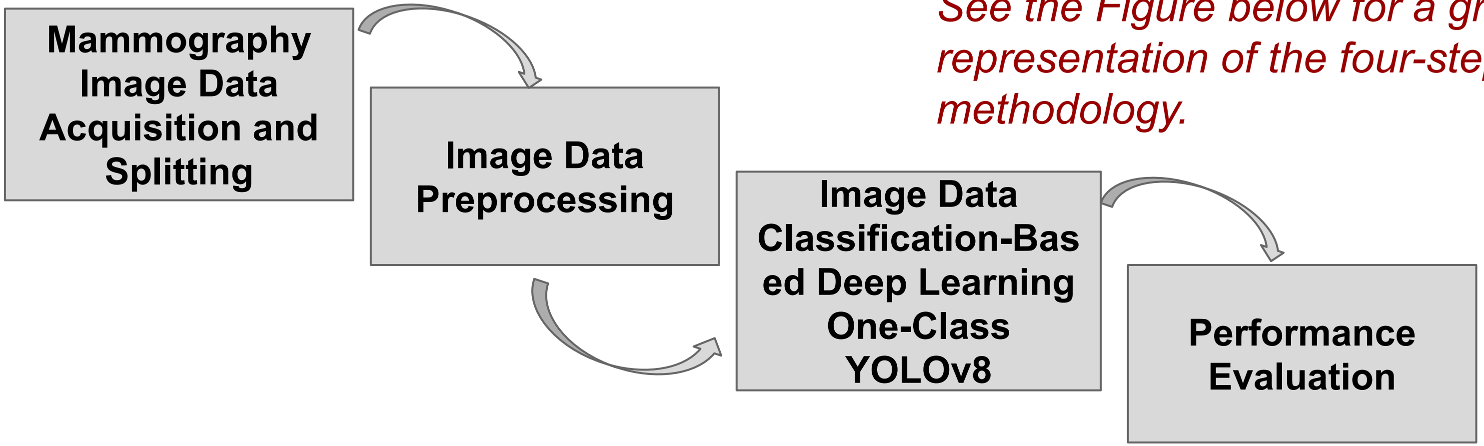
Dataset:

Utilizing the MIAS dataset from the "Mammographic Image Analysis Society".

Methodology:

1. Data Acquisition and Splitting
2. Image Preprocessing
3. YOLOv8 Deep Learning Model Deployment with OCC Approach
4. Comprehensive Performance Evaluation

See the Figure below for a graphical representation of the four-step methodology.



Current Research Focus - **Proposed Method (⅖)**

1. Data Acquisition and Splitting

MIAS Dataset Overview:

- Established by a UK research consortium.
- Contains 322 single-slice digital mammograms from 161 patients.
- Released in 1994 and accessible at https://www.kaggle.com/datasets/kmader/miasmammography.
- No registration required (accessed on 10 November 2023).

Data Details:

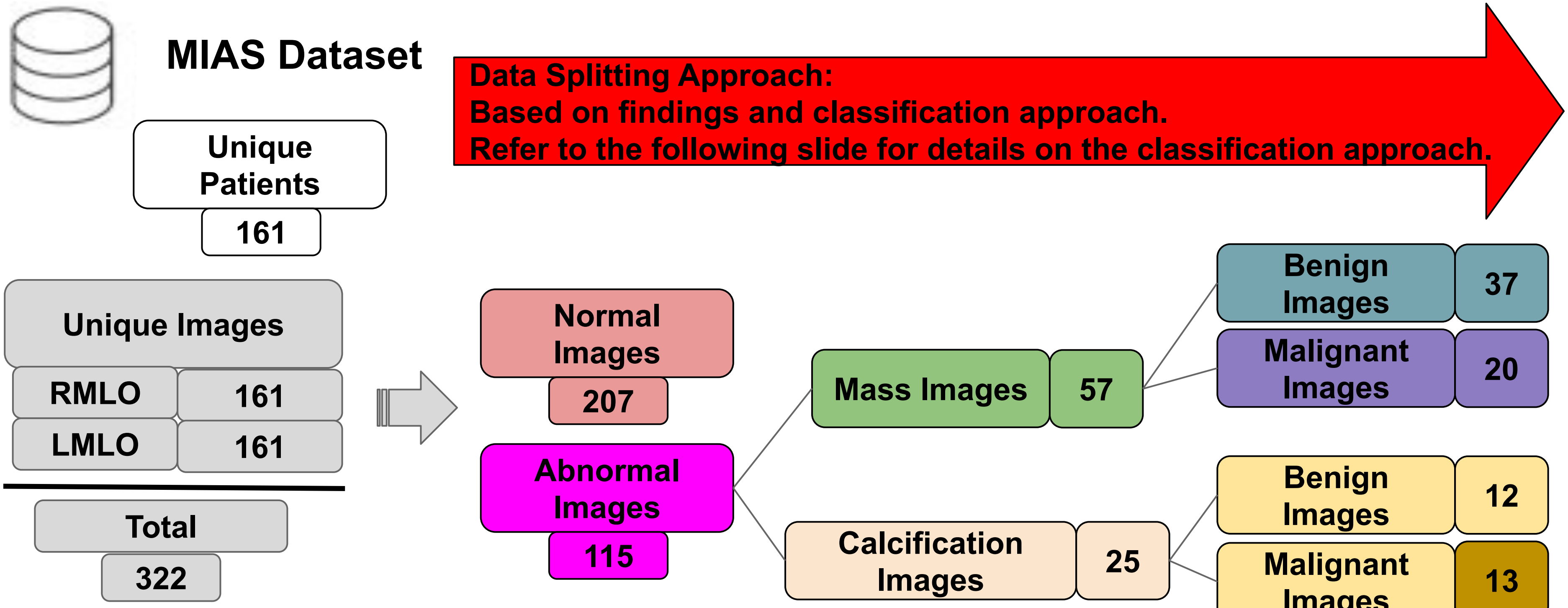
- Each finding includes breast density, abnormality type, severity, x and y coordinates, and abnormality radius.
- Abnormalities classified into seven categories, including calcification, well-defined masses, spiculated masses, ill-defined masses, architectural distortion (18 images), asymmetry (14 images), and normal.

View Orientation:

- Images captured in the mediolateral oblique (MLO) view.
- Each patient has both left MLO (LMLO) and right MLO (RMLO) views.

Current Research Focus - **Proposed Method (2/5)**

1. Data Acquisition and Splitting **(cond't)**



Current Research Focus - **Proposed Method (⅗)**

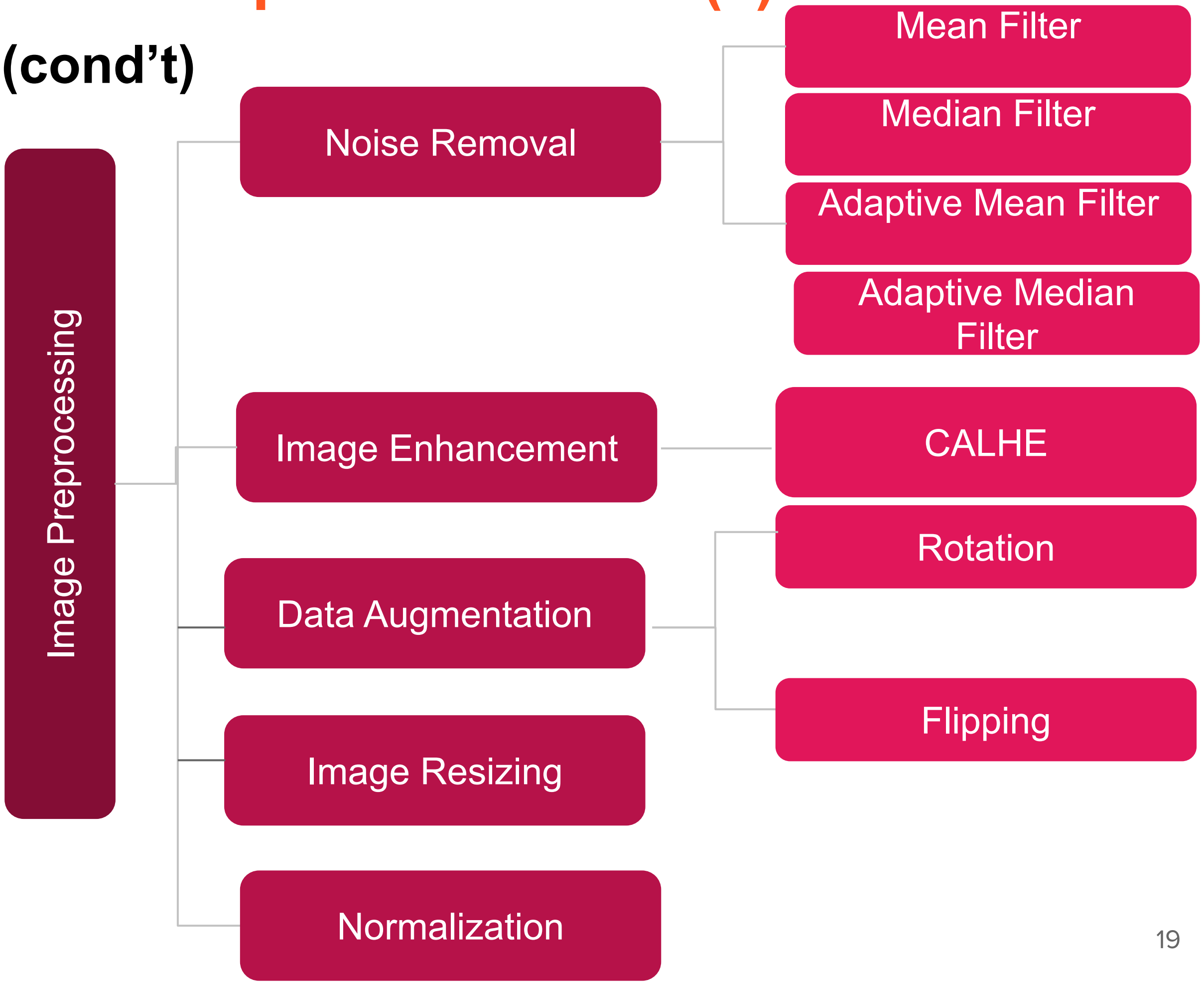
2. Image Preprocessing

Drawbacks of MIAS Dataset	**Mitigation Strategies**
Limited Dataset Size (322 single-slice digital mammograms)	• Apply data augmentation techniques (random transformations). • Mitigate risk of overfitting in training DL models.
Imbalanced Dataset (From total 322 images 207 cases are normal and 115 cases are abnormal)	• Implement One-Class Classification approach. • Address bias concerns by enhancing precision for limited classes.
Noise and Extraneous Data (i.e. tape artifact, high intensity rectangular label)	• Utilize denoising filters (e.g., mean filter, adaptive mean filter). • Remove unwanted information from images.
Limited Contrast	• Apply contrast enhancement techniques (contrast limited adaptive histogram equalization (CLAHE)).
Further Pre-processing Operations	• Resize and normalize images for compatibility with different pre-trained CNNs. • Include various operations during preprocessing (e.g., image resizing, normalization).

Current Research Focus - **Proposed Method (3/5)**

2. Image Preprocessing **(cond't)**

To improve dataset quality and applicability, we employ preprocessing techniques such as **noise removal, contrast enhancement, data augmentation, resizing**, and **normalization** for each MIAS dataset mammogram, as shown in the figure below.



Current Research Focus - **Proposed Method (4/5)**

3. Image Data Classification

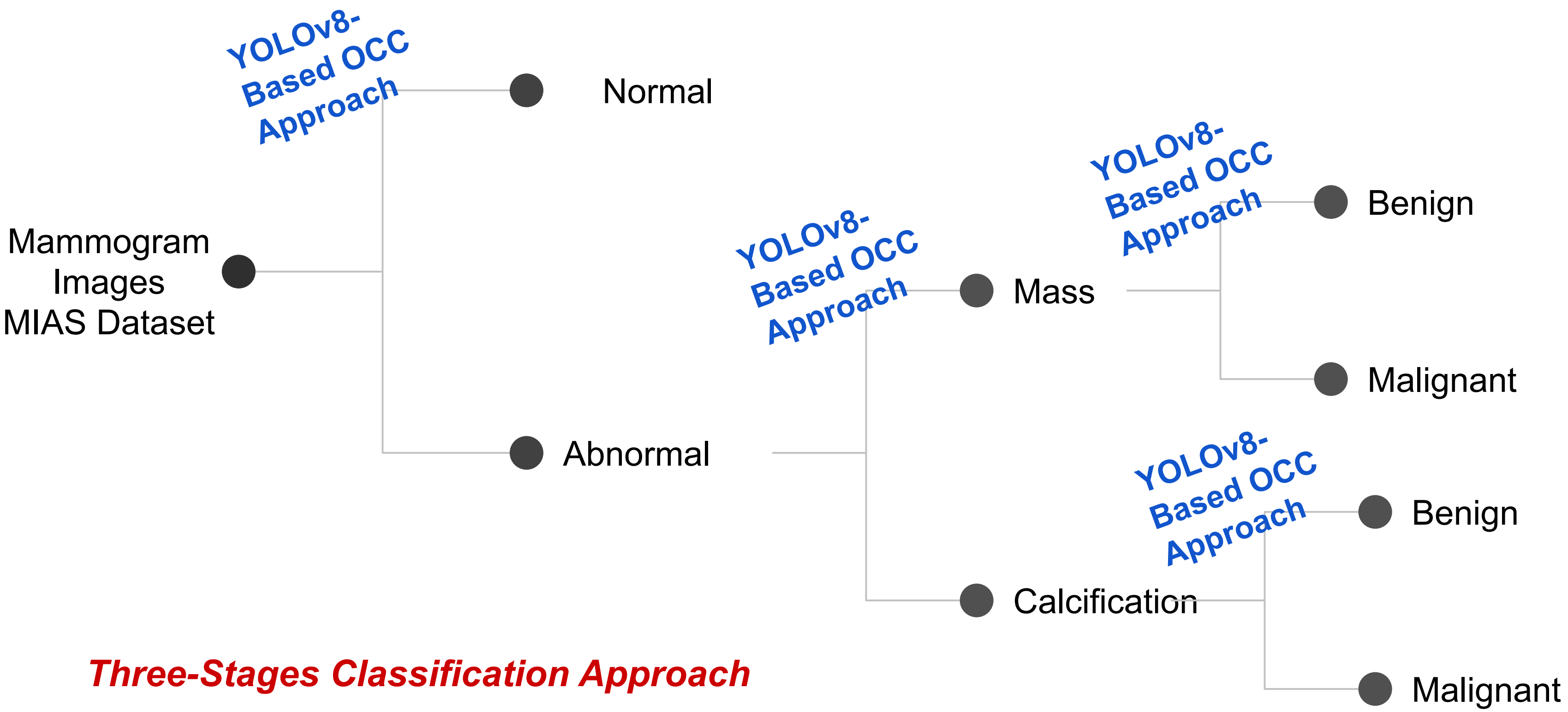
- **Establishing Pipeline:** Develop an advanced pipeline for breast mammogram detection and classification using the YOLOv8 model - based on One-Class Classification (OCC) approach.

- **Three Key Stages Implementation:**
    - ***Stage I - Abnormality Detection:*** Employ YOLOv8 to accurately identify abnormal locations within the breast, effectively categorizing images as either normal or abnormal.
    - ***Stage II - Masses vs. Microcalcifications:*** Utilize YOLOv8 for the second iteration to differentiate between masses and microcalcifications, essential in the detailed analysis of mammograms.
    - ***Stage III - Benign vs. Malignant Classification:*** Utilize YOLOv8's capabilities to classify identified abnormalities into benign or malignant cases, a critical step in breast cancer diagnosis.

++ See the Figure in the next slide for a graphical representation of the Three-stages classification approach.

Current Research Focus - **Proposed Method (4/5)**

3. Image Data Classification **(cond't)**



Three-Stages Classification Approach

Current Research Focus - **Proposed Method (4/5)**

3. Image Data Classification **(cond't)**

What is the reason into choose YOLOv8 model for this task? *The selection of the YOLOv8 model for our breast cancer diagnosis study using mammograms is based on several key considerations that make it suitable for this task:*

- **Efficiency and Speed:** YOLOv8 is renowned for its efficiency and speed in object detection and classification tasks. As our aim is to provide an efficient model for breast cancer diagnosis, YOLOv8's real-time capabilities align with our goal of swift and accurate diagnosis.
- **Accuracy:** YOLOv8 has demonstrated exceptional accuracy in object detection, making it a strong candidate for identifying abnormalities in mammograms, which is crucial for breast cancer diagnosis.
- **Versatility:** YOLOv8's architecture is highly versatile and can handle various object detection tasks. In the context of breast cancer diagnosis, where diverse abnormalities and conditions can be present in mammograms, this versatility is advantageous.
- **Transferability:** Our focus on achieving high transferability across different datasets is pivotal. YOLOv8's ability to adapt to different data distributions and generalize well across various datasets ensures that our model can maintain its performance consistency, regardless of the dataset it is applied to.
- **Community Support and Research:** YOLOv8 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 community of researchers who continually work on model improvements and adaptations.

Current Research Focus - **Proposed Method (5/5)**

4. Comprehensive Performance Evaluation

For a comprehensive evaluation of our Proposed Model. we will utilize a total of ten essential metrics: accuracy ***(Acc)***, sensitivity ***(Sn)***, specificity ***(Sp)***, precision ***(Pr)***, 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)***.

- **TN** shows the number of negative examples classified accurately.
- Similarly, **TP** indicates the number of positive examples classified accurately.

For Detection Task

- $\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{APi}$
- $\text{IOU score} = \text{Area of Intersection}/\text{Area of Union}$

For classification Task

- FP is the number of actual negative examples classified as positive.
- FN value is the number of actual positive examples classified as negative.
- $\text{Acc} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$
- $\text{Pr} = \text{TP}/(\text{TP} + \text{FP})$
- $\text{Sn} = \text{TP}/(\text{TP} + \text{FN})$
- $\text{Sp} = \text{TN}/(\text{TN} + \text{FP})$
- $\text{F1} - \text{Score} = 2 \times (\text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision})$
- $\text{ROC} - \text{AUC} = \text{FP}/(\text{FP} + \text{TN})$

Current Research Focus - **Conclusion & Future Work**

Our aim is to develop an advanced end-to-end learning model for breast cancer diagnosis through mammogram analysis, focusing on reliability and transferability.

- **Key Milestones Achieved:**
  - Image modality and dataset selection.
  - Adoption of widely used image preprocessing techniques.
  - Identification of the deep learning model and classification approach.
  - Careful selection of evaluation metrics.
- **Progress Highlights:**
  - Successful development of an initial YOLOv8-based algorithm.
  - Proficiency in detecting breast lesions and distinguishing between normal and abnormal cases.
- **Ongoing experiments with promising preliminary findings.**
  - Anticipation of groundbreaking results reshaping breast cancer diagnosis.
- **Future Work:**
  - Proposal for a multimodal fusion architecture integrating breast images and non-image data.
  - Decision-making pipeline for hierarchical breast cancer classification.
  - Aggregation of predictions from DSVDD and OCCNN models using a meta-model.
  - Exploration of image and non-image feature fusion for enhanced breast cancer prediction.

THANK YOU FOR LISTENING

Contact
PhD Student. Marwa Ben Ammar

Higher Institute of Medical Technologies of Tunis (ISTMT), University of Tunis El Manar
Research Laboratory in Biophysics and Medical Technologies (LRBTM)

+21658632141
marwa.ammar@istmt.utm.tn