LAMIS-DMDB: A new full field digital mammography database for breast cancer AI-CAD researches

---

Database Overview:

Mammography databases play a pivotal role in advancing medical diagnostics and research related to breast cancer. These databases are indispensable tools for healthcare professionals and researchers, facilitating the diagnosis, monitoring, and treatment of breast cancer by providing comprehensive datasets. One such groundbreaking resource is the LAMISDMDB, a new mammography database designed to significantly enhance the detection and classification of breast cancer. Our new mammography database, LAMISDMDB, can make a breakthrough in detecting and classifying breast cancer. It is ready to use ML and DL algorithms to detect and classify different cancers within the breasts accurately. This database has a large size as compared to other public mammogram databases; this allows for more detailed analysis when it comes to detecting abnormalities or malignant tumors. Additionally, users can download images based on any combination of structures such as BIRADS (Breast Imaging Reporting And Data System), ACR (American College of Radiology), normal/benign/malignant classifications, etc., allowing for even greater accuracy when analyzing potential cases. In conclusion, LAMISDMDB aims to revolutionize how we diagnose breast cancer by providing accurate information. This makes it easier for medical professionals worldwide to ensure their patients get the best care possible no matter where they live, helping save countless lives a year that would otherwise succumb to diseases like these due to a lack of proper diagnosis time.

METADATA Overview

The metadata for our breast cancer diagnostic database is organized into four distinct Excel files, each meticulously tailored to correspond with the BIRADS (Breast Imaging Reporting And Data System) categories I, II, IV, and V. These files, formatted as xlsx, serve as comprehensive repositories of crucial information for medical analysis. Within each file, essential patient-specific details such as patient ID, laterality (indicating whether the examination pertains to the left or right breast), view (e.g., MLO - Mediolateral Oblique, CC - Cranio-caudal), age, examination date, type of abnormality, and density category are systematically documented. This strategic decomposition based on BIRADS categories not only facilitates a nuanced understanding of breast health but also enables healthcare professionals and researchers to discern patterns and trends associated with specific diagnostic classifications. The meticulous inclusion of such diverse information empowers practitioners to conduct detailed analyses, contributing to the advancement of breast cancer diagnostics and personalized patient care.

LAMISDMDB Dataset:

The LAMISDMDB dataset is an amalgamation of medical images and accompanying metadata (annotations).

Images Organization:

Two organizational structures have been implemented for the categorization of medical images within the dataset. The initial structure classifies images based on mammogram classes, specifically normal, benign, and malignant designations. In contrast, the second structure adopts a hierarchical arrangement, with the primary tier delineated by BIRADS categories. Subsequent levels encompass classifications corresponding to specific abnormalities, culminating in a stratification based on density assessments. These frameworks serve as systematic and comprehensive methodologies for the organization and retrieval of medical images, facilitating distinct analytical perspectives for research and clinical investigations.

Structure I:

The image files are organized into three distinct directories, each dedicated to a specific category: normal, benign, and malignant mammogram images. This meticulous categorization serves to systematically structure the dataset, enabling precise access and analysis for academic and research purposes. The inclusion of metadata enhances the dataset's academic value, providing detailed annotations that contribute to a deeper understanding of the medical images within each category. This dataset, characterized by its well-organized structure and rich annotations, serves as a valuable resource for formal research investigations and academic endeavors in the fields of mammography and breast cancer diagnostics.

Structure II: Hierarchical Categorization

The LAMISDMDB dataset is thoughtfully organized to facilitate efficient navigation and analysis. The hierarchical structure is designed to cater to abnormalities (normal, benign, malign), BIRADS classifications, specific abnormalities (mass, calcifications, etc.), and ACR assessments, ensuring a comprehensive and accessible resource for researchers and healthcare professionals. Below is a detailed description of the dataset structure:
The dataset is hierarchically structured into three main directories:

Normal Directory:

> Directly connected to BI-RADS 1, leading to four ACR categories (ACR 1 to ACR 4).

Benign Directory:

Connected to BI-RADS 2, with branches for calcifications, masses, etc., leading to ACR categories.

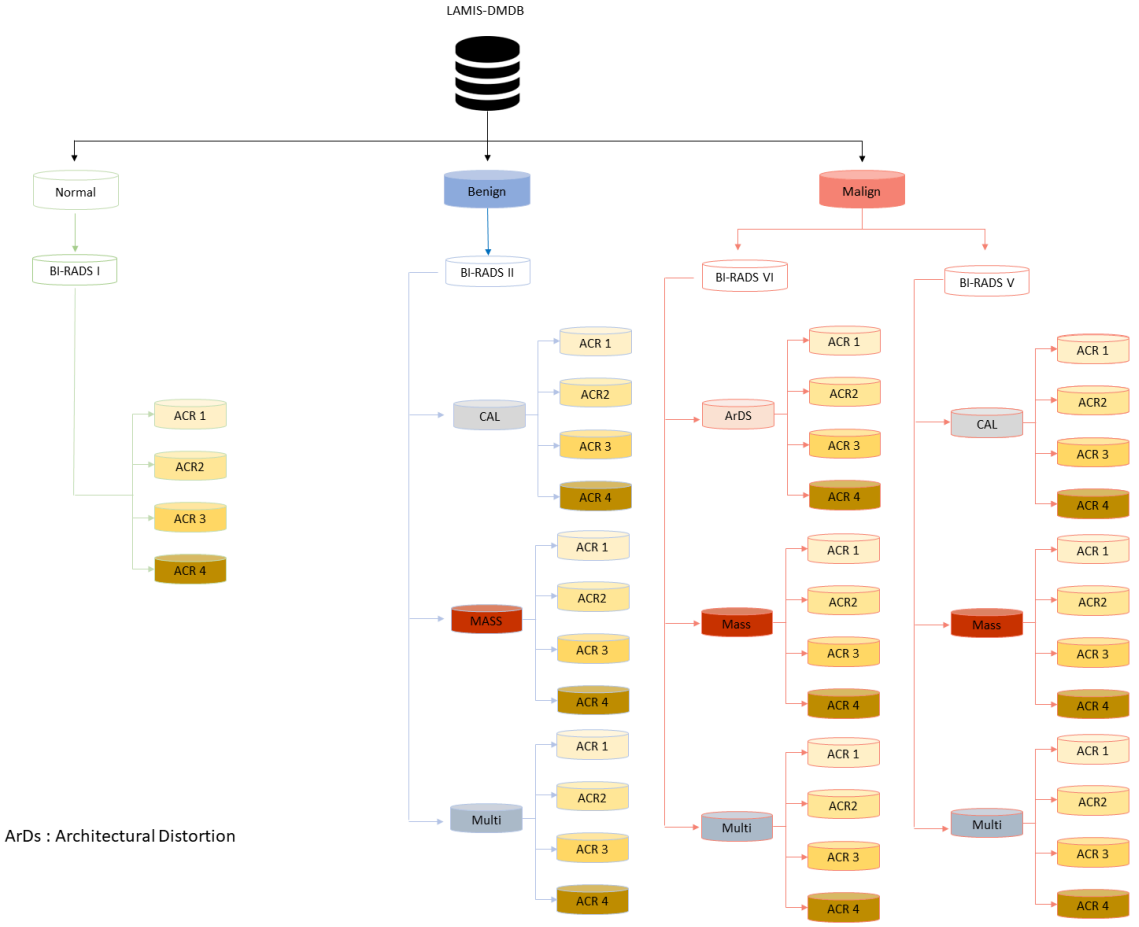
Malign Directory:

Comprising two subcategories:

BI-RADS 4: Connects to the Specific Abnormalities categories (mass, architectural distortion, etc.), with sub-branches for ACR as in the Benign section.

BI-RADS 5: Similar to BI-RADS 4, with sub-branches for Masses and calcification, etc., leading to ACR categories.

The figure below illustrates the hierarchical structure:



This structured dataset encourages a systematic approach to access and analyze medical images based on BIRADS classifications, specific abnormalities, and detailed ACR assessments. Researchers and healthcare professionals can leverage this organization for focused investigations and comprehensive studies in the realm of breast cancer diagnostics.

Metadata Files Organization:

LAMISDMDB_BIRADS I.xlsx : The initial file within our LAMISDMDB metadata is dedicated to mammogram images categorized under BIRADS I. This file is thoughtfully structured into four distinct Excel sheets, each corresponding to a different ACR. Each sheet within this file captures essential metadata, including patient ID, laterality, view, age, examination date, type of abnormality, and density category.

LAMISDMDB_BIRADS II.xlsx : The second file of our LAMISDMDB metadata focuses on mammogram images classified under BIRADS II. This file is intricately organized into three separate Excel sheets, with each sheet aligning with specific abnormality types such as mass, calcifications, or a combination of both. Within each sheet, critical metadata is documented, encompassing patient ID, laterality, view, age, examination date, type of abnormality, density category, and additional details like the number of abnormalities present in the image.

LAMISDMDB_BIRADS IV.xlsx : The third file of our LAMISDMDB metadata is dedicated to mammogram images classified under BIRADS IV. This file is meticulously structured into three distinct Excel sheets, each tailored to specific abnormality types, including mass, architectural distortion, or a combination of both. Within each sheet, essential metadata is comprehensively documented, encompassing patient ID, laterality, view, age, examination date, type of abnormality, density category, and crucial additional details such as the number of abnormalities present in the image.

LAMISDMDB_BIRADS V.xlsx : The fourth file of our LAMISDMDB metadata centers on the metadata of mammogram images categorized under BIRADS V. This file is meticulously structured into three distinct Excel sheets, each tailored to specific abnormality types, including mass, calcifications, or a combination of both. Within each sheet, crucial metadata is thoroughly documented, encompassing patient ID, laterality, view, age, examination date, type of abnormality, density category, and key additional details such as the number of abnormalities present in the image.

Access to images:

Accessing and creating filenames based on the metadata in our database is a streamlined process designed for efficiency and clarity. To access a specific file, users can leverage the structured metadata, which includes essential information such as Patient Code, Patient ID, Laterality, View, and Examination Date. For instance, if a user is seeking a particular mammogram image, they can utilize the provided metadata to formulate the filename in the prescribed structure:

"**dataMG-PROC**_PatientCode_ PatientID_ Laterality_VIEW_ExamDate**.dcm**"

Taking an illustrative example, a filename could appear as follows:

"**dataMG-PROC**_PQ25_574_R_MLO_29_December_2018**.dcm**"

"

This filename accurately reflects the metadata details, with **"PQ25"** denoting the Patient Code, **"574"** representing the Patient ID, **"R"** indicating the Laterality (Right), **"MLO"** specifying the view as Mediolateral Oblique, and **"29 December 2018"** serving as the Examination Date. This systematic approach to naming files ensures that users can easily identify and access the specific mammogram image they are seeking, streamlining the retrieval process and contributing to efficient data management.

Please cite this article as: Otmani Imane, Amroune Mohamed, Rahmani Foued Lazhar, Soltani Hama, Benkhelifa Elhadj, Aura Conci, LAMIS-DMDB: A new full field digital mammography database for breast cancer AI-CAD researches, Biomedical Signal Processing and Control,Volume 90, 2024, https://doi.org/10.1016/j.bspc.2023.105823.