Name: EMBED
Information: Stands for EMory BrEast imaging Dataset. This dataset contains 3,383,659 screening and diagnostic mammogram images from 115,910 patients. Among these, 20% of the total 2D and C-view dataset is available for research use.
Can be used for: Breast_Cancer_Detection, Breast_Cancer_Risk_Prediction, Mammographic_Report_Generation, Breast_Cancer_Type_Classification, Breast_Tumor_Classification, Tumor_Localization, Breast_Density_Estimation, Synthetic_Data_Generation
Data Availability: Available upon signing agreement.
Data Link: Available at https://aws.amazon.com/marketplace/pp/prodview-unw4li5rkivs2#overview
Associated Article: The EMory BrEast imaging Dataset (EMBED): A RaciallyDiverse, Granular Dataset of 3.4 Million Screening andDiagnostic Mammographic Images
Data/Article published on: 2023
Article available at: https://pubs.rsna.org/doi/10.1148/ryai.220047
Types of data in dataset: Mammogram_Images
Types of files in dataset: dcm
Data collected from: USA

Name: DREAM
Information: Stands for Dialogue on Reverse Engineering Assessment and Methods. This dataset contains 144,231 screening mammograms from 85,580 women from the United States and 166,578 screening mammograms from 68,008 women from Sweden.
Can be used for: Breast_Cancer_Detection, Breast_Tumor_Classification
Data Availability: Not Accessible
Data Link: Only metadata available at https://www.synapse.org/Synapse:syn1918762
Associated Article: Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms
Data/Article published on: 2020
Article available at: https://jamanetwork.com/journals/jamanetworkopen/article-abstract/2761795
Types of data in dataset: Mammogram_Images
Types of files in dataset: csv
Data collected from: USA, Sweden

Name: DDSM
Information: Stands for Digital Database for Screening Mammography. The dataset contains 2620 scanned film mammograms (images), 339 images with abnormalities, 373 normal cases, 223 cancer cases.
Can be used for: Breast_Cancer_Detection, Lesion_Localization
Data Availability: Accessible
Data Link: Available at http://www.eng.usf.edu/cvprg/Mammography/Database.html and a mini version is available at https://www.kaggle.com/datasets/skooch/ddsm-mammography
Associated Article: Current Status of the Digital Database for Screening Mammography
Data/Article published on: 1998
Article available at: https://link.springer.com/chapter/10.1007/978-94-011-5318-8_75
Types of data in dataset: Mammogram_Images, Metadata
Types of files in dataset: Ics, Overlay, Ljpeg
Data collected from: USA

Name: CBIS-DDSM

Information: Stands for Curated Breast Imaging Sub-set of the Digital Database for Screening Mammography. It's an updated version of DDSM. This dataset contains 2620 scanned film mammograms converted into dicom images making a total of 10,239 images.
Derived from: DDSM
Can be used for: Breast_Cancer_Detection, Breast_Tumor_Classification, Image_Segmentation, Mammographic_Report_Generation, Multi_Modal_Learning, Synthetic_Data_Generation
Data Availability: Accessible
Data Link: Available at https://www.cancerimagingarchive.net/collection/cbis-ddsm/
Associated Article: A curated mammography data set for use in computer-aided detection and diagnosis research
Data/Article published on: 2017
Article available at: https://www.nature.com/articles/sdata2017177
Types of data in dataset: Mammogram_Images, Radiomic_Features
Types of files in dataset: dcm, csv
Data collected from: USA

Name: RBIS-DDSM
Information: Stands for Re-curated Breast Imaging Subset of the Digital Database for Screening Mammography. It's an updated, extended, and standardized version of CBIS-DDSM. It contains 849 mammogram images.
Derived from: CBIS-DDSM
Can be used for: Breast_Cancer_Detection, Breast_Tumor_Classification, Image_Segmentation, Mammographic_Report_Generation, Multi_Modal_Learning, Synthetic_Data_Generation
Data Availability: Accessible
Data Link: Available at
https://ieee-dataport.org/documents/re-curated-breast-imaging-subset-ddsm-dataset-rbis-ddsm
Associated Article: No associated paper
Types of data in dataset: Mammogram_Images, Radiomic_Features
Types of files in dataset: dcm, csv
Data collected from: USA

Name: OMI-DB
Information: Stands for Oncology Medical Image DataBase. This dataset contains 34,104 2D mammography images from 2,623 individuals.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: The oncology medical image database (OMI-DB)
Data/Article published on: 2014
Article available at:
https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9039/1/The-oncology-medical-image-database-OMI-DB/10.1117/12.2041674.short
Types of data in dataset: Mammogram_Images
Types of files in dataset: dcm
Data collected from: UK

Name: OPTIMAM
Information: The dataset is an extended version of OMI-DB. This dataset contains 3,072,878 images from 172,282 women. The data has 4 classes namely "Normal", "Interval Cancer", "Benign", and

"Malignant". Of the 7661 screen-detected cancers from episodes with digital images, 5097 have been marked (67% of cancers).
Derived from: OMI-DB
Can be used for: Breast_Cancer_Detection, Breast_Cancer_Risk_Prediction, Breast_Cancer_Type_Classification, Synthetic_Data_Generation, Tumor_Localization
Data Availability: Available upon signing agreement.
Data Link: Available at https://medphys.royalsurrey.nhs.uk/omidb/
Associated Article: OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data
Data/Article published on: 2020
Article available at: https://pubs.rsna.org/doi/full/10.1148/ryai.2020200103
Types of data in dataset: Mammogram_Images, Clinical_Data
Types of files in dataset: dcm, csv
Data collected from: UK

Name: MIAS
Information: Stands for Mammographic Imaging Analysis Society. This database contains 332 images.
Can be used for: Breast_Cancer_Detection, Lesion_Localization, Breast_Cancer_Type_Classification
Data Availability: Accessible
Data Link: Available at https://www.repository.cam.ac.uk/items/b6a97f0c-3b9b-40ad-8f18-3d121eef1459
Associated Article: The mammographic images analysis society digital mammogram database
Data/Article published on: 1994
Article available at: https://cir.nii.ac.jp/crid/1572261551129161728
Types of data in dataset: Mammogram_Images, Metadata
Types of files in dataset: pgm, tar, tz, txt
Data collected from: UK

Name: VinDr-Mammo
Information: The dataset contains 20000 images from 5000 exams (4-view), 241 of these findings are marked BIRADS 5, 995 of the findings are marked BIRADS 4, and 2254 of the findings are marked BIRADS 3.
Can be used for: Breast_Cancer_Detection, Breast_Cancer_Risk_Prediction, Mammographic_Report_Generation, Breast_Cancer_Type_Classification, Breast_Tumor_Classification, Tumor_Localization, Breast_Density_Estimation, Synthetic_Data_Generation, Anomaly_Detection, Multi_Modal_Learning
Data Availability: Available upon signing agreement but an Accessible version is available.
Data Link: Original version available at https://www.physionet.org/content/vindr-mammo/1.0.0/ and the accessible version is available at https://www.kaggle.com/datasets/shantanughosh/vindr-mammogram-dataset-dicom-to-png
Associated Article: VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography
Data/Article published on: 2022
Article available at: https://www.physionet.org/content/vindr-mammo/1.0.0/
Types of data in dataset: Mammogram_Images, Metadata
Types of files in dataset: dcm, png, csv
Data collected from: Vietnam

Name: RSNA

Information: Stands for Royal Society of North America. The dataset contains 54713 image files (54706 in the training set), 1158 cancer images, 53548 non_cancer images, 11427 patients negative, 486 patients positive.
Can be used for: Breast_Cancer_Detection, Class_Balancing
Data Availability: Accessible
Data Link: Available at https://www.kaggle.com/competitions/rsna-breast-cancer-detection/data
Associated Article: No associated paper
Data/Article published on: 2022
Types of data in dataset: Mammogram_Images, Metadata
Types of files in dataset: dcm
Data collected from: USA, Australia

Name: CMMD
Information: Stands for Chinese MaMmography Database. This dataset contains 3,728 mammographies from 1,775 patients. Patients' molecular subtypes were included for 749 of these patients (1,498 mammographies). There are a total of 5202 images.
Can be used for: Mammographic_Report_Generation, Breast_Cancer_Detection, Image_Segmentation, Multi_Modal_Learning, Breast_Cancer_Type_Classification, Synthetic_Data_Generation
Data Availability: Accessible
Data Link: Available at https://www.cancerimagingarchive.net/collection/cmmd/#citations
Associated Article: The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast
Data/Article published on: 2021
Article available at: https://onlinelibrary.wiley.com/doi/full/10.1155/2019/2717454
Types of data in dataset: Mammogram_Images, Clinical_Data
Types of files in dataset: dcm, xlsx
Data collected from: China

Name: Inbreast
Information: The dataset contains 115 cases (410 images), 90 women with 4 images, 25 women with 2 images (Masectomy).
Can be used for: Breast_Cancer_Detection, Lesion_Localization, Image_Segmentation, Breast_Density_Estimation, Synthetic_Data_Generation
Data Availability: Accessible
Data Link: Available at https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset
Associated Article: INbreast: toward a full-field digital mammographic database
Data/Article published on: 2012
Article available at:
https://www.sciencedirect.com/science/article/pii/S107663321100451X?casa_token=scHMLQAy4ZEAAA
AA:i2kUWh2XkGKwQTmJYNK6TGNZRg--dgywtj1WWtuB2-HKA-aL8lv4r6ct7FGTZchvKZvydD4ttw
Types of data in dataset: Mammogram_Images, Metadata
Types of files in dataset: dcm
Data collected from: Portugal

Name: BCS-DBT
Information: Contains Digital Breast Tomosynthesis (3D) Images. 22,032 reconstructed DBT volumes belonging to 5,610 studies from 5,060 patients. 5,129 normal studies, 280 studies where additional imaging was needed, 112 benign biopsied studies, 89 studies with cancer.

Can be used for: Breast_Cancer_Detection, Breast_Density_Estimation, Breast_Cancer_Type_Classification, Tumor_Localization, Multi_Modal_Learning, Synthetic_Data_Generation
Data Availability: Accessible
Data Link: Available at https://www.cancerimagingarchive.net/collection/breast-cancer-screening-dbt/
Associated Article: Detection of masses and architectural distortions in digital breast tomosynthesis: a publicly available dataset of 5,060 patients and a deep learning model
Data/Article published on: 2020
Article available at: https://arxiv.org/abs/2011.07995
Types of data in dataset: Mammogram_Images, Radiomic_Features, Metadata, Clinical_Data, Annotation_Box
Types of files in dataset: dcm, xlsx
Data collected from: USA

Name: KAU-BCMD
Information: Stands for King Abdulaziz University Breast Cancer Mammogram Dataset. First significant mammogram dataset from Saudi Arabia. 1416 cases each with 2 views of both breasts making a total of 5664 images. Also contains 205 ultrasound cases corresponding to a part of the mammogram cases, with 405 images as a total.
Can be used for: BIRADS_Category_Classification, Breast_Cancer_Detection, Tumor_Localization, Breast_Tumor_Classification
Data Availability: Accessible
Data Link: Available at
https://www.kaggle.com/datasets/asmaasaad/king-abdulaziz-university-mammogram-dataset
Associated Article: King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD)
Data/Article published on: 2021
Article available at: https://www.mdpi.com/2306-5729/6/11/111
Types of data in dataset: Mammogram_Images, Metadata
Types of files in dataset: jpg, xlsx
Data collected from: Saudi Arabia

Name: DMID
Information: This dataset contains 510 images. Among them, 274 are abnormal images, including benign and malignant cases.
Can be used for: Breast_Cancer_Detection, BIRADS_Category_Classification, Mammographic_Report_Generation
Data Availability: Accessible
Data Link: Available at
https://figshare.com/articles/dataset/_b_Digital_mammography_Dataset_for_Breast_Cancer_Diagnosis_Research_DMID_b_DMID_rar/24522883
Associated Article: Digital mammography dataset for breast cancer diagnosis research (DMID) with breast mass segmentation analysis
Data/Article published on: 2024
Article available at: https://link.springer.com/article/10.1007/s13534-023-00339-y
Types of data in dataset: Mammogram_Images, Metadata, Radiology Reports
Types of files in dataset: dcm, xlsx, txt
Data collected from: India

Name: BancoWeb
Information: This dataset Contains 1400 images from 320 cases.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [Online Mammographic Images Database for Development and Comparison of CAD Schemes](#)
Data/Article published on: 2010
Article available at: https://link.springer.com/article/10.1007/s10278-010-9297-2
Types of data in dataset: Mammogram_Images
Types of files in dataset: tiff
Data collected from: Brazil

Name: Miracle DB
Information: This dataset contains 204 mammography images collected from 196 patients.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [A web-accessible mammographic image database dedicated to combined training and evaluation of radiologists and machines](#)
Data/Article published on: 2009
Article available at: https://ieeexplore.ieee.org/abstract/document/5394465
Types of data in dataset: Mammogram_Images
Types of files in dataset: tiff
Data collected from: Greece

Name: Nijmegen
Information: This dataset contains 40 images from 21 cases.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [A computerized system for detection and segmentation of clustered microcalcifications](#)
Data/Article published on: 1996
Article available at: http://dspace.fcu.edu.tw/handle/2377/2738
Types of data in dataset: Mammogram_Images
Types of files in dataset: tiff
Data collected from: Taiwan

Name: Malaga
Information: This dataset contains 320 images from 89 cases.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [Automatic mass segmentation in mammographic images](#)
Data/Article published on: 2008
Article available at: https://dugi-doc.udg.edu/handle/10256/4909
Types of data in dataset: Mammogram_Images
Types of files in dataset: dcm
Data collected from: Spain

Name: LLNL
Information: This dataset contains 198 images from 50 cases.

Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [Toward a standard reference database for computer-aided mammography](#)
Data/Article published on: 2008
Article available at:
https://www.spiedigitallibrary.org/conference-proceedings-of-spie/6915/1/Toward-a-standard-reference-database-for-computer-aided-mammography/10.1117/12.770325.short
Types of data in dataset: Mammogram_Images
Types of files in dataset: Ics
Data collected from: USA

Name: Magic-5
Information: This dataset contains 3369 images from 967 cases.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [MAGIC-5: an Italian mammographic database of digitised images for research](#)
Data/Article published on: 2002
Article available at:
https://www.researchgate.net/profile/Roberto-Bellotti-3/publication/5320442_MAGIC-5_An_Italian_mammographic_database_of_digitised_images_for_research/links/09e4150a2b4d4393bd000000/MAGIC-5-An-Italian-mammographic-database-of-digitised-images-for-research.pdf
Types of data in dataset: Mammogram_Images
Types of files in dataset: dcm
Data collected from: Italy

Name: Mammogrid
Information: No information found.
Can be used for: Breast_Cancer_Detection
Data Availability: Not Accessible
Associated Article: [MammoGrid: a prototype distributed mammographic database for Europe](#)
Data/Article published on: 2007
Article available at:
https://www.sciencedirect.com/science/article/pii/S0009926007001651?casa_token=-E26snreQ0cAAAAA:-R0727xsBGX71Pb3SovXb13mLue-io3wqbE6FcBdRhQsNxCg1z0_5yDTD_Qgl-upkf1EQNpsNA
Types of data in dataset: Mammogram_Images
Types of files in dataset: dcm
Data collected from: UK, Italy, Switzerland