Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations (Duke-Breast-Cancer-MRI)

Reproduced from https://wiki.cancerimagingarchive.net

Summary

Redirection Notice

This page will redirect to https://www.cancerimagingarchive.net/collection/duke-breast-cancer-mri/ in about 5 seconds.

Breast MRI is a common image modality to assess the extent of disease in breast cancer patients. Recent studies show that MRI has a potential in prognosis of patients' short and long-term outcomes as well as predicting pathological and genomic features of the tumors. However, large, well annotated datasets are needed to make further progress in the field. We share such a dataset here.

In terms of design, the dataset is a single-institutional, retrospective collection of 922 biopsy-confirmed invasive breast cancer patients, over a decade, having the following data components:

1. Demographic, clinical, pathology, treatment, outcomes, and genomic data: Collected from a variety of sources including clinical notes, radiology report, and pathology reports and has served as a source for multiple published papers on radiogenomics, outcomes prediction, and other areas.
2. Pre-operative dynamic contrast enhanced (DCE)-MRI: Downloaded from PACS systems and de-identified for The Cancer Imaging Archive (TCIA) release. These include axial breast MRI images acquired by 1.5T or 3T scanners in the prone positions. Following MRI sequences are shared in DICOM format: a non-fat saturated T1-weighted sequence, a fat-saturated gradient echo T1-weighted pre-contrast sequence, and mostly three to four post-contrast sequences.
3. Locations of lesions in DCE-MRI: Annotations on the DCE-MRI images by radiologists.
4. Imaging features from DCE-MRI: A set of 529 computer-extracted imaging features by inhouse software. These features represent a variety of imaging characteristics including size, shape, texture, and enhancement of both the tumor and the surrounding tissue, which is combined of features commonly published in the literature, as well as the features developed in our lab.

If you use this dataset, please cite the following publication:

Publication Citation

Saha, A., Harowicz, M.R., Grimm, L.J., Kim, C.E., Ghate, S.V., Walsh, R. and Mazurowski, M.A., 2018. **A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features.** British journal of cancer, 119(4), pp.508-516. (*A free version of this paper is available here:* PMC6134102)

For more information see this site https://sites.duke.edu/mazurowski/resources/breast-cancer-mri-dataset/, the related publication, or contact TCIA Helpdesk at help@cancerimagingarchive.net. Please visit this discussion forum for any questions related to the data: https://www.reddit.com/r/DukeDCEMRIData/.

Acknowledgements

- Harmonization of the components of this dataset, including into standard DICOM representation, was supported in part by the NCI Imaging Data Commons consortium. NCI Imaging Data Commons consortium is supported by the contract number 19X037Q from Leidos Biomedical Research under Task Order HHSN26100071 from NCI.

Reproduced from https://wiki.cancerimagingarchive.net

Data Access

Data Access

Data Type	Download all or Query/Filter	License
Images (DICOM, 368.4 GB)	*Download* *Search* (Download requires the NBIA Data Retriever)	CC BY-NC 4.0
File Path mapping tables (XLSX, 49.6 MB)	*Download*	CC BY-NC 4.0
Clinical and Other Features (XLSX, 582 kB)	*Download*	CC BY-NC 4.0
Annotation Boxes (XLSX, 49 kB)	*Download*	CC BY-NC 4.0
Imaging features (XLSX, 6.44 MB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations (NRRD, 4 kB)	*Download* (Download requires IBM Aspera Faspex)	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations (DICOM, 23 kB)	*Download* (Download requires the NBIA Data Retriever)	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations README (.TXT, 3kB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentation Filepath Mapping (. CSV, 129KB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentation Test Ids (.CSV, 1KB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations Train Ids (.CSV, 1 kB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations Breast Radiologist Densities (.XLSX, 12KB)	*Download*	CC BY-NC 4.0
3D Breast and FGT MRI Segmentations (NRRD, 0.4GB)	*Download* (Download requires IBM Aspera Faspex)	CC BY-NC 4.0

Click the Versions tab for more info about data releases.

Reproduced from https://wiki.cancerimagingarchive.net

Additional Resources for this Dataset

The NCI Cancer Research Data Commons (CRDC) provides access to additional data and a cloud-based data science infrastructure that connects data sets with analytics tools to allow users to share, integrate, analyze, and visualize cancer research data.

- Imaging Data Commons (IDC) (Imaging Data)

Detailed Description

Detailed Description

	Radiology Image Statistics
Modalities	MR, SEG
Number of Participants	922
Number of Studies	922
Number of Series	5,161
Number of Images	773,888
Images Size (GB)	368.4

Users please note this caveat about DICOM tag **(0020,0052) Frame Of Reference UID**: This DICOM tag was lost during deidentification and curation, and has been replaced with a Dummy value **per study** that may not be reliable for image alignment.

Reproduced from https://wiki.cancerimagingarchive.net

POPULATION

The breast MRI dataset contains 922 patients gathered in Duke Hospital from 1 January, 2000 to 23 March, 2014 with invasive breast cancer and available pre-operative MRI at Duke Hospital. Detailed inclusion/exclusion criteria are described in the publication listed in the HOW TO REFERENCE THIS DATA section.

Reproduced from https://wiki.cancerimagingarchive.net

IMAGES

The images (DICOM) are located here https://doi.org/10.7937/TCIA.e3sv-re93. Under “Data Access”, use “Data Type” item “Images” to download. The respective paths to the image slices are indicated in Breast-Cancer-MRI-filepath_filename-mapping.xlsx.

Reproduced from https://wiki.cancerimagingarchive.net

IMAGE ANNOTATIONS

The Image Annotations are located here https://doi.org/10.7937/TCIA.e3sv-re93. Under “Data Access”, use “Data Type” item “Annotation Boxes” to download the Annotation_Boxes.xlsx spreadsheet. Apart from the header, each row corresponds to a unique patient denoted by the “Patient ID” in the first column. The next four columns describe the annotation box as follows: the first two values are the start and end row of the box, the third and the fourth values are the start and end column of the box, the fifth and sixth values are the start and end slice number of the box. If the original box coordinates were not integers, they were rounded for this spreadsheet.

Guide to obtain the overlay of Annotation Boxes on the images: Please use the NBIA Data Retriever with the corresponding manifest file to download this imaging data. Please select the fat saturated sequences only. For each sequence, read the filenames in ascending/increasing order. The slice number of the annotation box corresponds to the characters after the hyphen (-) symbol in the filename, excluding the extension. For e.g., the filename ‘1-050.dcm’ corresponds to slice 50.

For a given patient, note that the sequences may not always begin with a file (slice) having the lowest height (check Z coordinate of ImagePositionPatient DICOM tag to confirm that). It can begin with a slice with the highest height. In such a scenario, reading the filenames in ascending/increasing order will result in descending order of height with the increase of slice number.

If you chose to use a software that returns the volume directly from a sequence considering the slice with the lowest height as the first slice, please check if reading the filenames in ascending/increasing order results in ascending order of height. If not, you will need to change the slice order in the following manner to obtain an overlay. If a total of M slices is present in the sequence, Start Slice is A, and End Slice is B as per ‘Annotation_Box.xlsx’; change Start Slice to (M – B + 1) and End Slice to (M – A + 1) to obtain the overlay.

Other details pertaining to annotation: The boxes were drawn by 8 radiologists through inhouse graphical user interface developed in MATLAB. The MRI sequences that were involved in annotation were: (a) pre-contrast, (b) first post-contrast, and (c) subtracted (obtained by subtracting the pre-contrast from the first post-contrast).

The data was annotated in two parts with some differences in the procedures followed. The first procedure resulted in annotation of a subset of 271 patients, and the second one resulted in the annotation of the remaining 651 patients.

For 271 of the patients, a panel of 6 fellowship-trained radiologists was formed. One of 6 radiologists annotated a study randomly assigned to him/her. Each radiologist was responsible for annotating a subset only. The radiologists used a graphical user interface to draw a three-dimensional box around any areas of mass and non–massenhancement for up to five lesions. If multiple lesions were annotated, the biopsied tumor was selected after further review of relevant radiology and pathology reports. If there were multiple biopsies, the largest biopsied tumor was selected for feature extraction.

For the remaining 651 patients, a panel of 4 fellowship-trained radiologists was formed and a slight modification in the annotation procedure was made. The radiologist was provided with location(s) of the biopsies and were told to annotate the largest biopsied lesion. One of 4 radiologists annotated each study randomly assigned to him/her. In contrast to the annotations in the first phase, radiologists had access to the PACS system, should they need it.

Reproduced from https://wiki.cancerimagingarchive.net

DEMOGRAPHIC, CLINICAL, PATHOLOGY, GENOMIC, TREATMENT, OUTCOMES, AND OTHER DATA

This data (tabular) is located here https://doi.org/10.7937/TCIA.e3sv-re93. Under “Data Access”, use the “Data Type” item “Clinical and Other Features” to download. Following are specific elements of this file:

Demographics. These data were obtained from the Oncology clinic note in the electronic medical record.

- Age in days at diagnosis
- Menopausal status at diagnosis (based on clinical notes in the electronic medical record)
- Race/ethnicity (White, Black, Asian, Native American, Hispanic, Multiethnic, Hawaiian, American Indian)
- Metastatic disease at presentation (no,yes)

Tumor Characteristics. These data were obtained from the pathology biopsy report.

- Estrogen receptor status (negative, positive)
- Progesterone receptor status (negative, positive)
- Human epidermal growth factor 2 receptor status (negative, positive)
- Molecular subtype (luminal-like, ER/PR positive and HER2 positive, HER2, triple negative)
- Oncotype score
- TNM staging (based on combined pathologic and clinical staging)
- Tumor grade (tubule, nuclear, and mitotic)
- Nottingham grade (low, intermediate, high)
- Histologic type (ductal carcinoma in situ, invasive ductal carcinoma, invasive lobular carcinoma, metaplastic, lobular carcinoma in situ, tubular, mixed type, micropapillary, colloid)
- Tumor location (left,right)
- Tumor position (clock face position, i.e. L 12 means left breast 12 o’clock)
- Bilateral breast cancer (yes,no), if bilateral breast cancer different receptor status (yes,no)
- Side annotated on the imaging (left,right)
- For other side if bilateral: side of cancer, oncotype score, nottingham grade, ER status, PR status, HER2 status, molecular subtype)

MRI Findings. These data were obtained from the radiologist MRI report.

- Multicentric/multifocal (no,yes)
- Contralateral breast involvement (no,yes)
- Lymphadenopathy or suspicious lymph nodes (no,yes)
- Skin/nipple involvement (no,yes)
- Pectoral muscle/chest involvement (no,yes)

Surgery. These data were obtained from the Oncology clinic note in the electronic medical record.

- Surgery status (no,yes)

Reproduced from https://wiki.cancerimagingarchive.net

- Days to surgery from diagnosis
- Definitive surgery type (breast conservation therapy, mastectomy)

Radiation Therapy. These data were obtained from the Oncology clinic note in the electronic medical record.

- Neoadjuvant radiation (no,yes)
- Adjuvant radiation (no,yes)

Tumor response. These data were obtained from the Oncology clinic note in the electronic medical record. Please note this data was obtained from the initial evaluation of the electronic medical record. Columns further to the right in the spreadsheet, labeled Pathological Response to Neo-Adjuvant Therapy and Near-Complete Response were obtained on second review of the electronic medical record with a few updates made to the data.

- Clinical response (obtained from radiologist imaging report)
- Pathologic response to neoadjuvant therapy (complete response, not complete response, DCIS only remaining, LCIS only remaining, treatment response assessment unavailable, not applicable)

Recurrence. These data were obtained from the Oncology clinic note in the electronic medical record.

- No, yes
- If yes: days to local recurrence and/or days to distant recurrence from date of diagnosis

Follow-up. These data were obtained from all clinical notes in the electronic medical record.

- Days to death from diagnosis
- Days to last local recurrence free assessment (based on clinical notes in the electronic medical record)
- Days to last distant recurrence free assessment (based on clinical notes in the electronic medical record)
- Days to last contact in electronic medical record (last time patient known to be alive, unless age of death is reported)

Mammography Characteristics. These data were obtained from the radiologist preoperative mammogram report.

- Age at mammogram
- Breast density (heterogeneous, scattered, minimal, moderate, extremely, predominantly fatty)
- Lesion shape (oval, irregular, lobular, reniform, stellate)
- Lesion margin (obscured, spiculated, indistinct/ill-defined, circumscribed)
- Architectural distortion (no,yes)
- Lesion density
- Calcifications (yes, pleomorphic, heterogeneous, microcalcification, linear, clustered, amorphous, branching)
- Lesion size (cm)

Reproduced from https://wiki.cancerimagingarchive.net

Ultrasound (US) features. These data were obtained from the radiologist preoperative ultrasound report.

- Lesion shape (oval, irregular, lobular)
- Lesion margin (obscured, ill-defined, spiculated, indistinct, circumscribed, microlobulated, angular, irregular)
- Lesion size (cm)
- Lesion echogenicity (hypoechoic, hyperechoic, isoechoic, anechoic, irregular, mixed, boundary)
- Solid
- Posterior acoustic shadowing

Therapy data. Please note this data was obtained on second review of the electronic medical record with a few updates made to the data.

- Chemotherapy (neoadjuvant, adjuvant)
- Endocrine Therapy (neoadjuvant, adjuvant, known ovarian status, number of ovaries in situ, therapeutic or prophylactic oophorectomy as part of endocrine therapy)
- Anti-Her2/Neu Therapy (neoadjuvant, adjuvant)
- Neo-Adjuvant Therapy (received neoadjuvant or not)
- Pathologic Response to Neo-Adjuvant Therapy (pathologic stage (T) following neoadjuvant therapy, pathologic stage (N) following neoadjuvant therapy, pathologic stage (M) following neoadjuvant therapy)
- Near-Complete Response (overall near-complete response (stricter definition), overall near-complete response (looser definition), near-complete response (graded measure))

MRI technical information. These fields were collected from the pre-contrast sequence and, if they were not available in the pre-contrast sequence, they were collected from the post-contrast sequences. To calculate the ‘Days to MRI from Diagnosis’, the date of diagnosis obtained in the clinical report was subtracted from the date of MRI acquisition.

- Days to MRI from diagnosis
- Manufacturer
- Manufacturer model name
- Scan option
- Field strength (Tesla)
- Patient position during MRI
- Image position of patient
- Contrast bolus volume (mL)
- Repetition time
- Echo time
- Acquisition matrix
- Slice thickness
- Rows
- Columns
- Reconstruction diameter
- Flip angle
- Field of view (cm)

Reproduced from https://wiki.cancerimagingarchive.net

IMAGE FEATURES

This data (tabular) is located here https://doi.org/10.7937/TCIA.e3sv-re93. Under “Data Access”, use the “Data Type” item “Imaging features” to download.

Feature Source: We extracted 529 imaging features from the tumor and automatically segmented FGT (fibroglandular tissue). FGT was segmented using the (i) T1- fat saturated sequence and (ii) T1-non fat saturated first post-contrast sequence both pre-processed with N4-ITK (https://pubmed.ncbi.nlm.nih.gov/20378467/). While extracting features from FGT, tumor pixels were subtracted from it.

Feature Groups: We extracted features pertaining to (a) the volume of breast and FGT, (b) size and morphology of the tumor, (c) enhancement of the FGT, (d) enhancement of the tumor, (e) combined enhancement of the tumor and FGT, (f) texture of FGT enhancement, (g) texture of tumor enhancement, (f) spatial heterogeneity related to tumor enhancement, (h) variations in FGT enhancement, and (i) variations in tumor enhancement.

Feature Names: The features used in this file are enlisted in the supplemental material of the paper [1] (a version is available in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6176866/). Note that, there might be minor changes in the feature names. For e.g., in the feature ‘Max_Enhancement_from_char_curv’ from the TCIA file is the same as the feature ‘Max_Enhancement_from_char_curve_tumor’ found in the supplemental material.

Analyses of the features: We conducted a series of analyses with the features extracted as follows:

1. Assessment of variability of the features due to changes in MRI protocol:

We conducted an analysis to determine how the features were affected by changes in (i) manufacturer of the MRI scanner (ii) magnetic field strength of the MRI scanner, and (iii) slice thickness of the sequence using of subset of 272 patients. Our findings are reported in the paper [1] (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6176866/).

1. Assessment of inter-reader (inter-radiologist) stability of the features:

To perform this, we obtained a sub-sample of 50 patients of the whole dataset and four fellowship trained breast imagers annotated the tumors present in those. Then corresponding to these annotations, 4 sets of 529 features were extracted from each of the 50 patients. Intra-class correlation coefficient was used as the measure of inter-reader stability. Our findings are reported in the paper [2] (https://pubmed.ncbi.nlm.nih.gov/29663411/).

1. Validation of the features in independent test for predicting tumor subtype, receptor status, and Ki-67 status:

All of these three validations were reported in the paper [3] (https://pubmed.ncbi.nlm.nih.gov/30033447/).

Reproduced from https://wiki.cancerimagingarchive.net

1. Validation of the features in independent test for predicting Oncotype DX (ODX) recurrence score levels:

We validated the features for discriminating different levels of ODX scores using the subset 261 of patients having pertinent data. Our findings are reported in the paper [4] (https://pubmed.ncbi.nlm.nih.gov/29427210/).

1. Validation of the features in independent test for predicting complete response (pCR) to neoadjuvant therapy (NAT):

We validated the features for discriminating pCR using the subset of 288 patients having pertinent data. Evaluation was carried out in three relevant groups of patients- (i) all with NAT, (ii) triple-negative or human epidermal growth factor receptor 2-positive (TN/HER2+) patients who had NAT, and (ii) with neoadjuvant chemotherapy. Our findings can be found in the paper [5] (https://pubmed.ncbi.nlm.nih.gov/30328048/).

1. Validation of features associated with distant recurrence-free survival:

We validated the association of the features with distant recurrence-free survival over median follow-up period of almost 4 years. This study was carried out using a subset of 892 patients having pertinent follow-up data. Our findings are reported in the paper [6] (https://pubmed.ncbi.nlm.nih.gov/30672045/).

Citations & Data Usage Policy

Citations & Data Usage Policy

Users must abide by the TCIA Data Usage Policy and Restrictions. Attribution should include references to the following citations:

Data Citation

Saha, A., Harowicz, M. R., Grimm, L. J., Weng, J., Cain, E. H., Kim, C. E., Ghate, S. V., Walsh, R., & Mazurowski, M. A. (2021). **Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations [Data set]**. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.e3sv-re93

Publication Citation

Saha, A., Harowicz, M. R., Grimm, L. J., Kim, C. E., Ghate, S. V., Walsh, R., & Mazurowski, M. A. (2018). **A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features.** British journal of cancer, 119(4), 508-516. DOI: https://doi.org/10.1038/s41416-018-0185-8 , PMC6134102

TCIA Citation

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**. Journal of Digital Imaging, 26(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7

Reproduced from https://wiki.cancerimagingarchive.net

Other Publications Using This Data

TCIA maintains a list of publications which leverage TCIA data. If you have a manuscript you'd like to add please contact TCIA's Helpdesk.

Versions

Version 3 (Current): Updated 2022/08/31

Data Type	Download all or Query/Filter	License
Images (DICOM, 368.4 GB)	*Download* *Search* (Download requires the NBIA Data Retriever)	CC BY-NC 4.0
File Path mapping tables (XLSX, 49.6 MB)	*Download*	CC BY-NC 4.0
Clinical and Other Features (XLSX, 582 kB)	*Download*	CC BY-NC 4.0
Annotation Boxes (XLSX, 49 kB)	*Download*	CC BY-NC 4.0
Imaging features (XLSX, 6.44 MB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations (NRRD, 4 kB)	*Download* (Download requires IBM Aspera Faspex)	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations (DICOM)	*Download* (Download requires the NBIA Data Retriever)	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations README (.TXT)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentation Filepath Mapping (.CSV, 129KB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentation Test Ids (.CSV, 1KB)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations Train Ids (.CSV)	*Download*	CC BY-NC 4.0
2D Breast and FGT MRI Segmentations Breast Radiologist Densities (.XLSX, 12KB)	*Download*	CC BY-NC 4.0
3D Breast and FGT MRI Segmentations (NRRD, 0.4GB)	*Download*	CC BY-NC 4.0

	(Download requires IBM Aspera Faspex)	

Added 3D Breast and FGT MRI Segmentation Supplemental Data.

Version 2: Updated 2022/06/13

Data Type	Download all or Query/Filter	License
Images (DICOM, 368.4 GB)	*Download* *Search* (Download requires the NBIA Data Retriever)	CC BY-NC 4.0
File Path mapping tables (XLSX, 49.6 MB)	*Download*	CC BY-NC 4.0
Clinical and Other Features (XLSX, 582 kB)	*Download*	CC BY-NC 4.0
Annotation Boxes (XLSX, 49 kB)	*Download*	CC BY-NC 4.0
Imaging features (XLSX, 6.44 MB)	*Download*	CC BY-NC 4.0
Supplemental Segmentation (TXT, XLSX)	• 2D Breast and FGT MRI Segmentation Supplemental Data in nrrd format *Download* (Download requires IBM Aspera Faspex) • 2D Breast and FGT MRI Segmentation Supplemental Data in DICOM format *Download* (Download requires the NBIA Data Retriever)	CC BY-NC 4.0
Supplemental Segmentation (README.txt)	*Download*	CC BY-NC 4.0
Segmentation_filepath_mapping.csv	*Download*	CC BY-NC 4.0
Supplemental Segmentation (Test_ids.csv)	*Download*	CC BY-NC 4.0
Supplemental Segmentation (train.ids.csv)	*Download*	CC BY-NC 4.0
Supplemental Segmentation (Breast_Radiologist_Density.xlsx)	*Download*	CC BY-NC 4.0

Added 2D Breast and FGT MRI Segmentation Supplemental Data.

Reproduced from https://wiki.cancerimagingarchive.net

Version 1: Updated 2021/04/06

Data Type	Download all or Query/Filter
Images (DICOM, 368.4 GB)	*Download* (Download requires the NBIA Data Retriever)
File Path mapping tables (XLSX, 49.6 MB)	*Download*
Clinical and Other Features (XLSX, 582 kB)	*Download*
Annotation Boxes (XLSX, 49 kB)	*Download*
Imaging features (XLSX, 6.44 MB)	*Download*

Added new subjects.

Reproduced from https://wiki.cancerimagingarchive.net