CBIS-DDSM: A CURATED MAMMOGRAPHY DATA SET FOR USE IN COMPUTER-AIDED DETECTION AND DIAGNOSIS RESEARCH

Daniel L. Rubin, MD, MS

Department of Biomedical Data Science, Radiology, and Medicine (Biomedical Informatics)
Stanford University

Acknowledgements

- Rebecca Lee, PhD
- Berkman Sahiner, PhD
- Justin Kirby, John Freymann, TCIA team
- Funding support

NCI QIN grants U01CA142555,U01CA190214, U01CA187947

Motivation

- Many different **AI/ML applications for mammography** are being developed
  - Detection of suspicious lesions (CADe)
  - Diagnosis of cancer (CADx)
- Algorithm performance evaluated on **different datasets**
  - Private data sets
  - Unspecified subsets of public databases
  - Variable dataset sizes
- Not possible to directly **compare the performance of methods or to replicate prior results**

Some CADe systems reported in the literature

Performance statistics of selected CADe methods for the detection of abnormalities

Authors	Size of Data set (Cases)	Public or Private Data	Accuracy	Sensitivity	False Positives Per Image
Karssemeijer and te Brake[13]	50	Public (MIAS*)	NA	90%	1
Mudigonda *et al.*[14]	56	Public (MIAS*)	NA	81%	2.2
Liu *et al.*[15]	38	Public (MIAS*)	NA	90%	1
Li *et al.*[16]	94	Private	NA	91%	3.21
Baum *et al.*[17]	63	Private	NA	89%	0.61
Kim *et al.*[18]	83	Private	NA	96%	0.2
Yang *et al.*[19]	203	Private	96.1%	95–98%	1.8
The *et al.*[20]	123	Private	NA	94%	2.3 per case
Sadaf *et al.*[21]	127	Private	NA	91%	NA
Chu *et al.*[22]	230	Public (DDSM†)	NA	98.5%	0.84

Scientific data, *4*, 170177. doi:10.1038/sdata.2017.177

Some CADx systems reported in the literature

Performance statistics of selected CADx methods for the classification of masses

Authors	Size of Data set (Cases)	Public or Private Data	Classification Accuracy	Az*
Brzakovic *et al.*[23]	25	Private	85%	NA
Huo *et al.*[24]	65	Private	NA	0.94
Rangayyan *et al.*[25]	54	Public (MIAS†) and Private	91%	NA
Mudigonda *et al.*[26]	39	Public (MIAS†)	82.1%	0.85
Sahiner *et al.*[27]	102	Private	NA	0.91
Timp *et al.*[28]	465	Private	NA	0.77
Ganesan *et al.*[29]	282	Private	88.8%	NA
Görgel *et al.*[30]	78, 65	Private, Public (MIAS†)	91.4%, 90.1%	NA
Qiu *et al.*[31]	560	Private	77.14%	0.81
Choi *et al.*[32]	600	Public (DDSM†)	NA	0.88

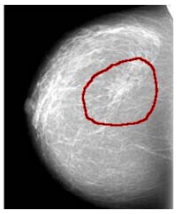
Scientific data, *4*, 170177. doi:10.1038/sdata.2017.177

Standard image datasets are being developed and popular

- ImageNet (14 million images from 27 categories)
- Mixed National Institute of Standards and Technology (MNIST) database (hand-written digits)
- ChestX-ray8 and OpenI (Chest X-rays)
- DeepLesion (CT)
- Digital Database for Screening Mammography (DDSM)

DDSM is limited for evaluating AI algorithms

- Images saved in non-standard format compressed files, difficult to use
- Image metadata are fragmented, unwieldy to access/use
- ROIs are very coarse, limited value for lesion localization
- Cases not curated into training/testing subsets



CBIS-DDSM (Curated Breast Imaging Subset of DDSM)

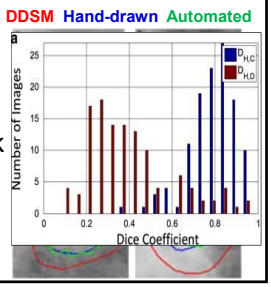
- An updated, standardized subset of DDSM
- Curated
  - Images in standard format (DICOM)
  - Unified metadata in single files
  - Improved ROIs
  - Training/testing subsets
- Can serve as a common dataset for comparing performance of AI/ML algorithms (CADe, CADx; NB: not segmentation)

CBIS-DDSM: Images

- DDSM scanned film mammography studies
  - Source: MGH, Wake Forest Univ, Sacred Heart Hospital, WUSTL
  - Image labels: **Normal**, **benign**, and **malignant** (latter verified by pathology)
- Decompressed and converted to DICOM
- Case selection and curation by expert mammographer
  - Removal of questionable mass cases and cases containing PHI
- Convenience images: focused crops of abnormalities based on ROI bounding box

CBIS-DDSM: ROIs

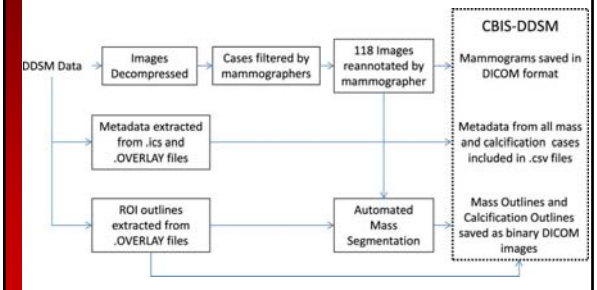
- ROIs for masses
  - Mass ROIs refined by automated segmentation
  - Mammographer verification in 188 cases (Dice 0.8 ± 0.1)
- ROIs are binary mask images delineating the abnormality



Annotations: Parsed semantic features

- **Patient age**
- **BI-RADS descriptors (**mass shape, mass margin, calcification type, calcification distribution, and breast density)
- **BI-RADS final assessment category** (0 to 5)
- Rating of the **subtlety of abnormality** (1 to 5)
- **Type of abnormality** (mass or calc)
- Annotations formatted in CSV format similar to modern computer vision data sets

Summary: Preparation of CBIS-DDSM

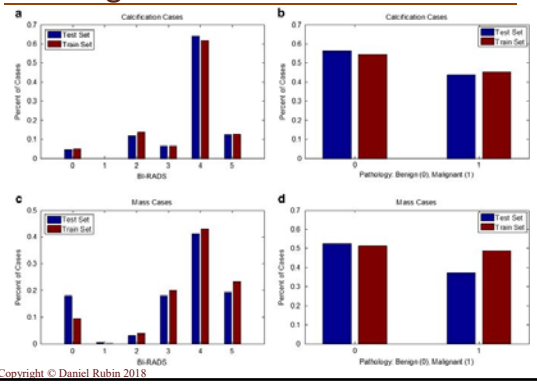


Training/testing splits of cases

- 1,644 cases: 753 calc cases, 891 mass cases
- Separate sets of cases for training (80%) and testing (20%)
  - Enables all researchers to use same cases for these tasks
- Separate splits for mass and calc cases
  - Contain full range of label values (BI-RADS codes)
  - Equal distribution in training and testing cases

	Benign Cases	Malignant Cases
Calcification Training Set	329 cases (552 abnormalities)	273 cases (304 abnormalities)
Calcification Test Set	85 cases (112 abnormalities)	66 cases (77 abnormalities)
Mass Training Set	355 cases (387 abnormalities)	336 cases (361 abnormalities)
Mass Test Set	117 cases (135 abnormalities)	83 cases (87 abnormalities)

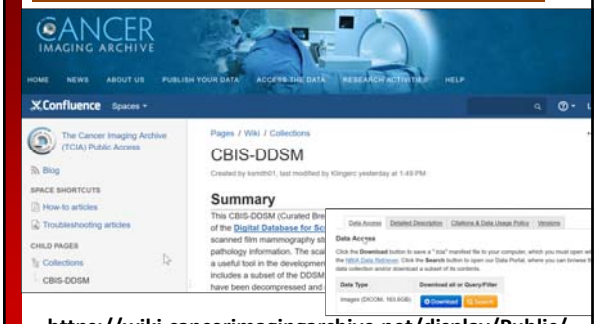
Distribution of case labels for training and test sets



Limitations of CBIS-DDSM

- **Limited dataset size** (1,644 cases reasonable for quantitative imaging; somewhat small for deep learning)
- **Film-screen images**; FFDM and tomosynthesis are modern techniques
  - Could produce similar dataset if such images become publicly available
- **Segmentations** of lesions are improved over original DDSM, but not all hand-drawn

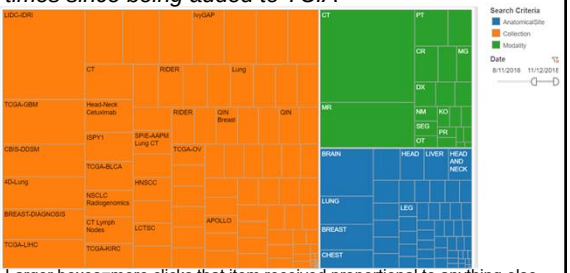
Obtaining CBIS-DDSM



https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM

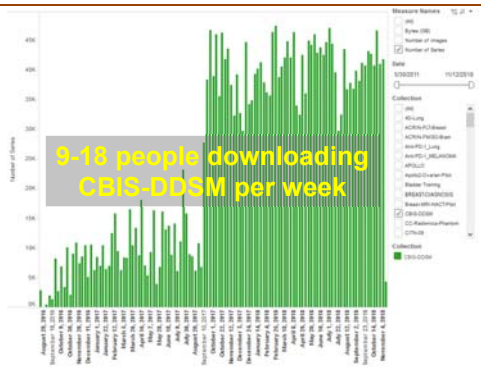
CBIS-DDSM is a popular dataset…

CBIS-DDSM is the 3rd most popular dataset in the time it has been in TCIA; selected in search 22,557 times since being added to TCIA



Larger boxes=more clicks that item received proportional to anything else in TCIA search interface.

Download history since 2016



Summary

- CBIS-DDSM is a curated set of benign/malignant mammography cases
  - Annotations of mass lesions and calcifications
  - Provides data formats suitable for AI/ML development and testing
- Potentially useful as a common dataset for evaluating and comparing CADe/CADx methods
- Preferably similar dataset will be built using FFDM/tomosynthesis images if public data becomes available

Thank you.

Contact info:
dlrubin@stanford.edu