CUSTOMIZATION OF AI-BASED BREAST CANCER RISK MODELS FOR VALIDATION IN AANHPI WOMEN

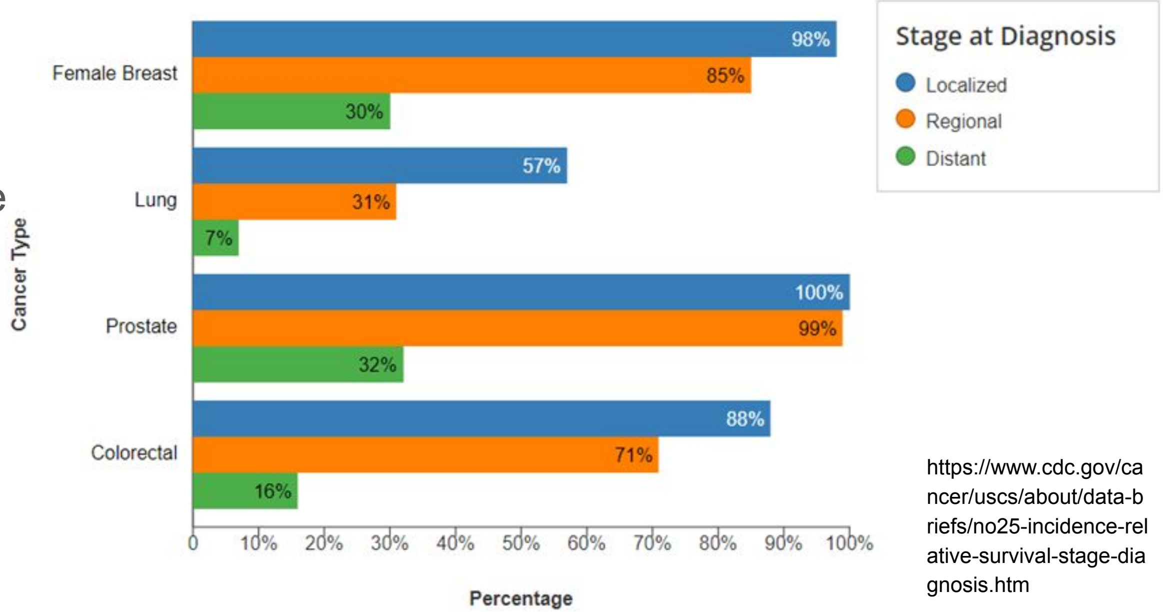
MARC
at UH MĀNOA
Maximizing Access to Research Careers

Cade Kane
SURE Symposium 2024

Breast Cancer Risk Assessment

- Ultimate goal: catch breast cancer earlier, when there’s a higher chance of survival for the patient.
- 5-year or lifetime risk assessments

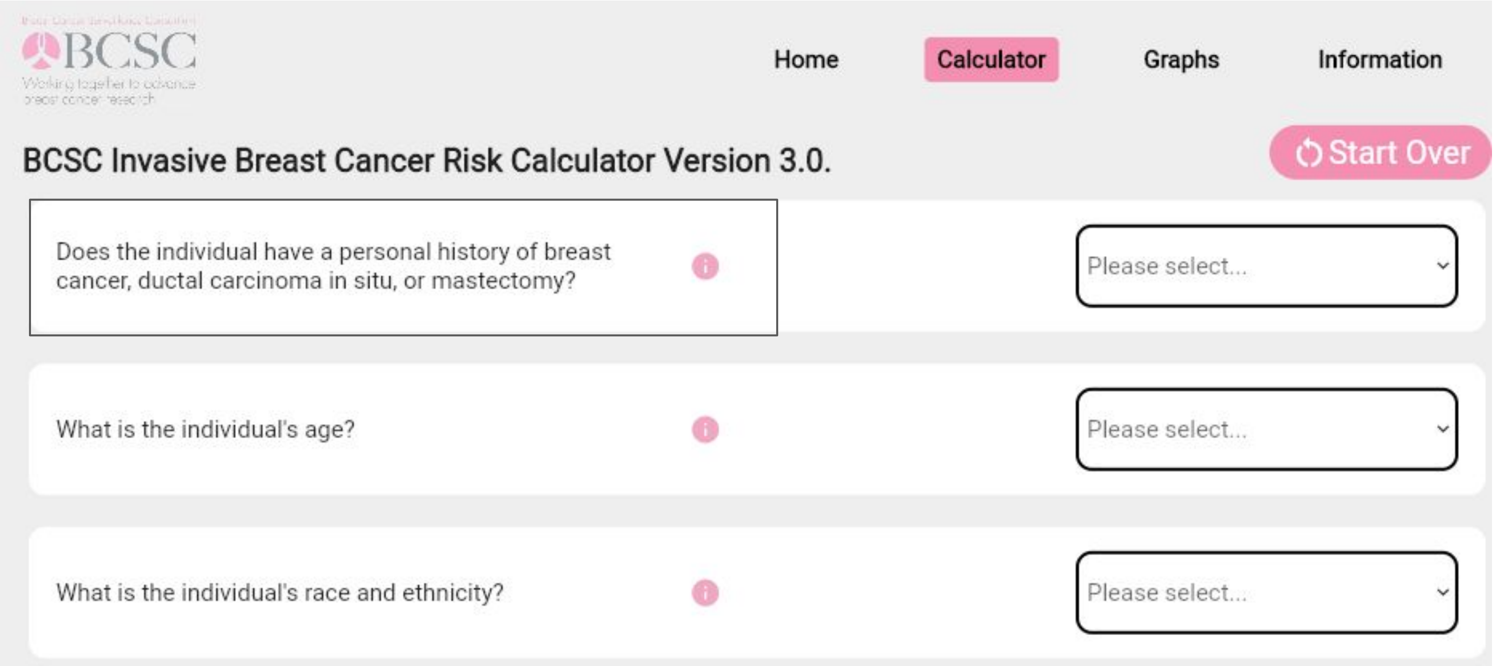
5-year relative survival estimates the percentage of cancer patients who will have **not** died from their cancer **5 years** after diagnosis



https://www.cdc.gov/cancer/uscs/about/data-briefs/no25-incidence-relative-survival-stage-diagnosis.htm

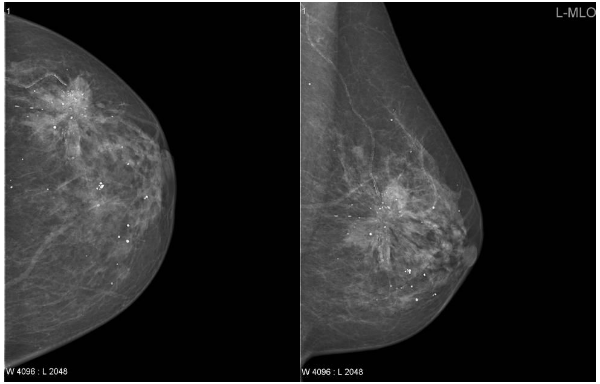
Clinical Breast Cancer Risk

- Demographics and clinical risk factors only
- Can be extended polygenic risk scores



Mammography

- Mammography is the recommended primary breast cancer screening modality for women in the U.S.
- Mammography consists of a low-dose x-ray of the breast tissue
- Mammography has four views: R-CC, R-MLO, L-CC, L-MLO

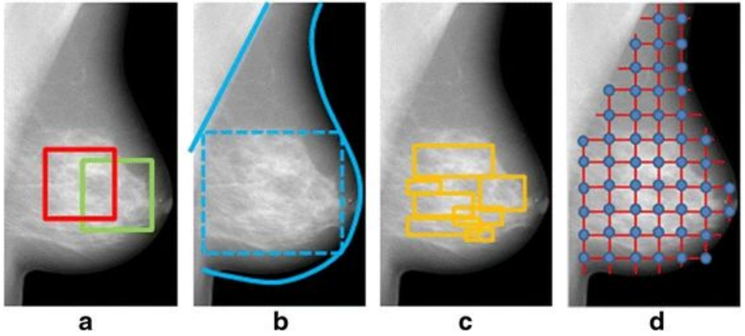


L-CC L-MLO

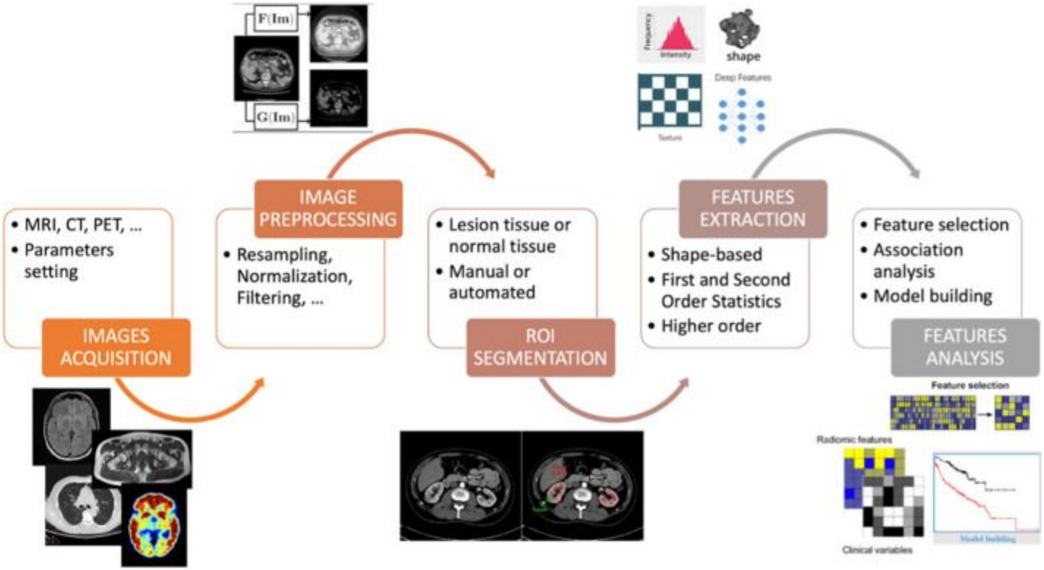
(Radswiki, 2011)

Breast Cancer Risk with Radiomics/AI

- BC risk can be found from radiomics through calculation of traditional computer vision features or AI-derived features from breast imaging.



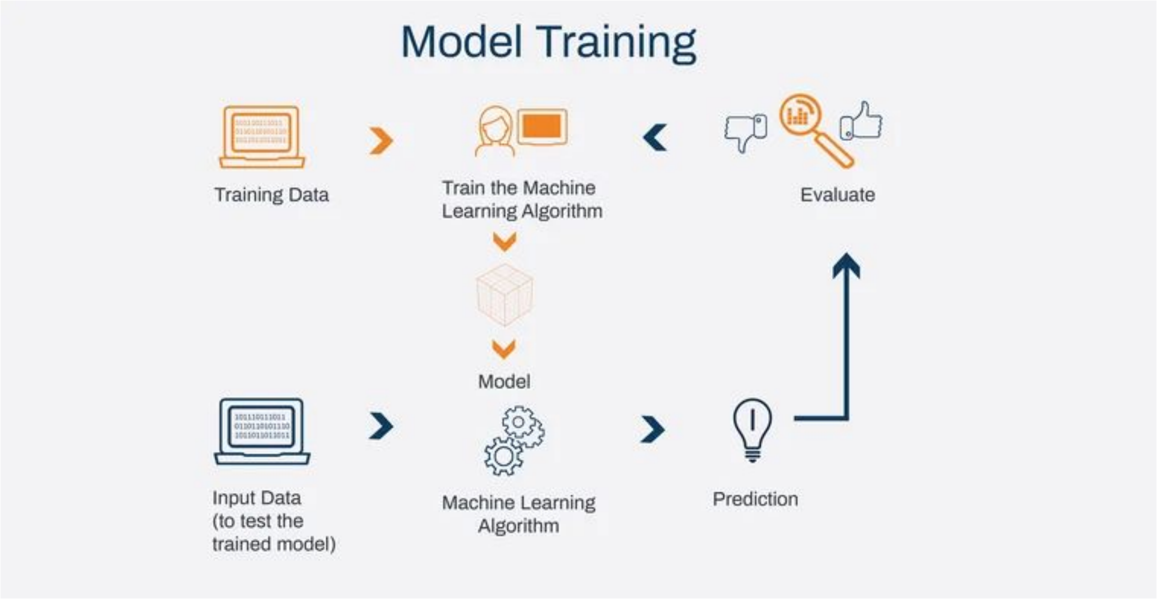
(Pertuz et al., 2019)



(Scapicchio et al., 2021)

Machine Learning and AI in Breast Cancer Risk Prediction

- Pattern recognition → *Learning patterns and features*
- A trained model can be given NEW data to evaluate its performance



(Chaitanya, n.d.)

Breast Cancer Risk by Ethnicity/Race

- Due to high proportion of women of Asian descent in HI we can **disaggregate** the “**AANHPI**” ethnicity group into **specific subgroups**: i.e., Japanese, Chinese, etc.
- In HI we have populations which aren’t seen in statistically significant numbers on the continental US (NHPI), allowing us to do **subgroup analysis**.

BCSC 5-year risk	White N=91,308	Filipina N=6,551	Chinese N=24,051	Japanese N=2,485
	Invasive cancers %			
<1.67%	54	90	93	89
≥1.67%	46	10	7	11

UCSF breast cancer surveillance consortium risk model

Overall Problem Statement

Examine clinical and radiomic risk factors (family history of breast cancer, BMI, BI-RADS breast density) and their association with breast cancer risk in women undergoing breast cancer screening by **Asian and Native Hawaiian/Pacific Islander** ancestry <u>compared</u> to non-Hispanic **White**, non-Hispanic **Black**, and **Hispanic** women.

Hawaii & Pacific Islands Mammography Registry (HIPIMR)

- Collects breast imaging and breast health information (2009-present)

- Linked to the Hawaiʻi Tumor Registry to identify cases

- Consist of imaging, metadata, clinical variables, patient characteristics, and biopsy-confirmed cancer status

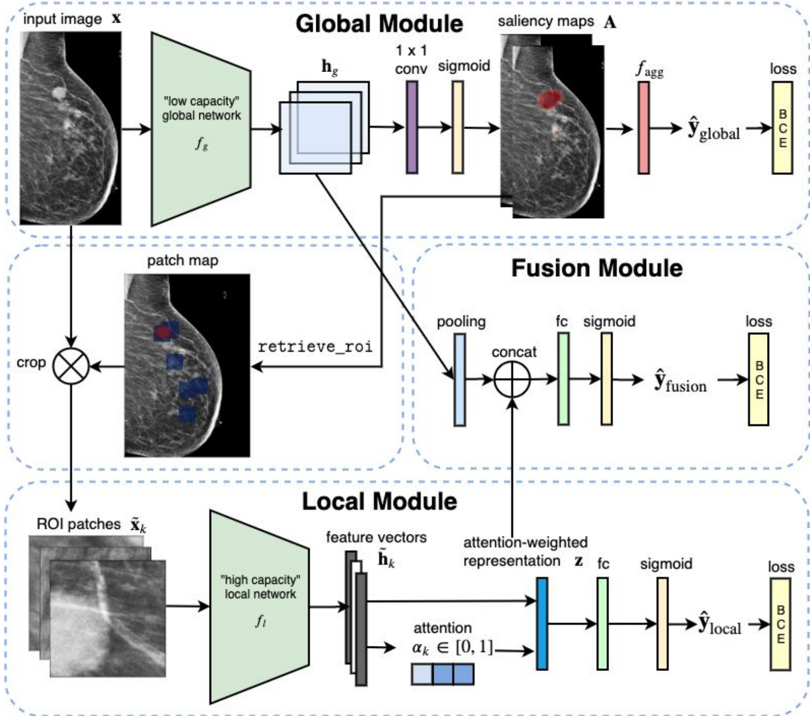
Model	***Required Views***	***Processed or Raw***	***File Format***
Transpara	R/L MLO/CC	Processed	DICOM or JPEG
iCAD (ProFound AI)	R/L MLO/CC	Processed (Dimensions)	DICOM
Malkov Features	R/L CC	Processed	PNG
OpenBreast	R/L MLO/CC	Processed	DICOM
CaPTk	Any	Processed or Raw	DICOM
LIBRA	Any	Processed or Raw	DICOM
Geras Density	R/L MLO/CC	Processed	PNG
Xun	R/L MLO/CC	Raw (Proc2Pres)	DICOM
Mirai	R/L MLO/CC	Processed	DICOM or PNG
Geras Classifier (GMIC)	R/L MLO/CC	Processed	PNG
RADIFUSION	R/L MLO/CC	Processed	DICOM

An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization

Yiqiu Shen[a], Nan Wu[a], Jason Phang[a], Jungkyu Park[a], Kangning Liu[a], Sudarshini Tyagi[d], Laura Heacock[b,e], S. Gene Kim[b,c,e], Linda Moy[b,c,e], Kyunghyun Cho[a,d,f], Krzysztof J. Geras[b,c,a]

- GMIC: globally-aware multiple instance classifier

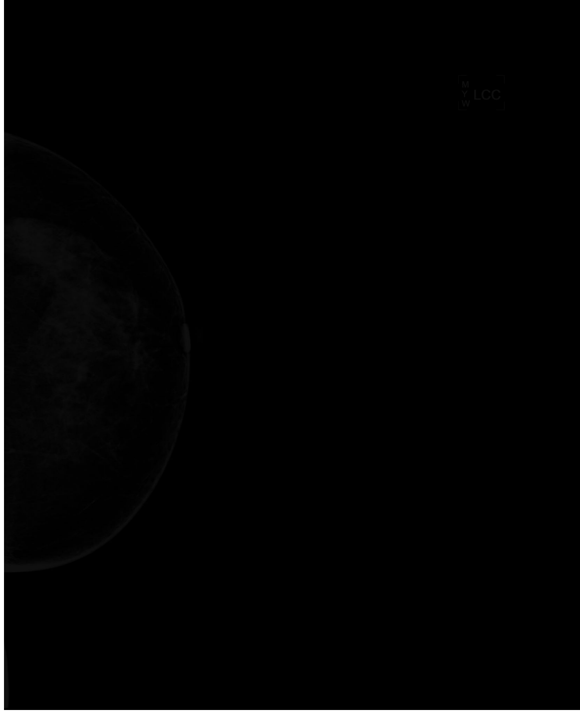
- Input: mammogram PNG + exam list
- Output: cropped mammograms, probability of malignant and benign



GMIC Pipeline Alterations

- Preprocessing:
  - DICOM conversion to PNG format
    - Bit difference 12 bit vs. 16 bit
  - CSV conversion to exam list

- Wrapper script
  - Base directory
  - DICOM directory
  - Path to CSV



Mammogram

CSV Input Data Example

Image_file	ANALYSIS_ID	CANCER_STATUS	ViewPosition	ImageLaterality	StudyInstanceUID	Visit
1.2.840.114191.6.!	100685956	0	CC	R	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100685956	0	CC	L	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100685956	0	MLO	L	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100685956	0	MLO	R	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100689996	1	CC	R	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100689996	1	MLO	L	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100689996	1	CC	L	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100689996	1	MLO	R	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100689996	1	MLO	R	1.2.840.114191.6.9(	1
1.2.840.114191.6.!	100689996	1	MLO	L	1.2.840.114191.6.9(	1
1.2.840.114191.6.!	100689996	1	CC	L	1.2.840.114191.6.9(	1
1.2.840.114191.6.!	100689996	1	CC	R	1.2.840.114191.6.9(	1
1.2.840.114191.6.!	100985944	0	CC	R	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100985944	0	CC	L	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100985944	0	MLO	L	1.2.840.114191.6.9(	0
1.2.840.114191.6.!	100985944	0	MLO	R	1.2.840.114191.6.9(	0

GMIC Pipeline Results Example

image_index	benign_pred	malignant_pred	benign_label	malignant_label
0_L-CC	0.1356	0.0081	0	0
0_R-CC	0.8928	0.3259	1	0
0_L-MLO	0.2368	0.0335	0	0
0_R-MLO	0.9509	0.1812	1	0
1_L-CC	0.0508	0.0144	0	0
1_R-CC	0.5679	0.9965	0	1
1_L-MLO	0.0545	0.0154	0	0
1_R-MLO	0.4722	0.7178	0	1
2_L-CC	0.0746	0.016	0	0
2_R-CC	0.3551	0.0884	1	0
2_L-MLO	0.0953	0.0086	0	0
2_R-MLO	0.2267	0.0395	1	0
3_L-CC	0.6162	0.8992	0	1
3_R-CC	0.2945	0.2116	0	0
3_L-MLO	0.6677	0.722	0	1

GMIC Pipeline Results Example

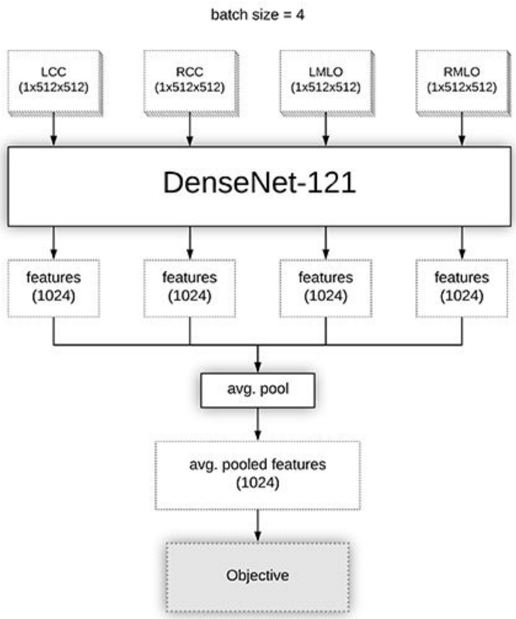
image_index	benign_pred	malignant_pred	benign_label	malignant_label
0_L-CC	0.1356	0.0081	0	0
0_R-CC	0.8928	0.3259	1	0
0_L-MLO	0.2368	0.0335	0	0
0_R-MLO	0.9509	0.1812	1	0
1_L-CC	0.0508	0.0144	0	0
1_R-CC	0.5679	0.9965	0	1
1_L-MLO	0.0545	0.0154	0	0
1_R-MLO	0.4722	0.7178	0	1
2_L-CC	0.0746	0.016	0	0
2_R-CC	0.3551	0.0884	1	0
2_L-MLO	0.0953	0.0086	0	0
2_R-MLO	0.2267	0.0395	1	0
3_L-CC	0.6162	0.8992	0	1
3_R-CC	0.2945	0.2116	0	0
3_L-MLO	0.6677	0.722	0	1
3_R-MLO	0.0722	0.0303	0	0

Deep Learning Predicts Interval and Screening-detected Cancer from Screening Mammograms: A Case-Case-Control Study in 6369 Women

Xun Zhu, Thomas K. Wolfgruber, Lambert Leong, Matthew Jensen, Christopher Scott, Stacey Winham, Peter Sadowski, Celine Vachon, Karla Kerlikowske, John A. Shepherd

- Interval-based breast cancer
- Clinical factors
- Engineering differences (Xun uses Proc2Pres)

A



Xun Pipeline Alterations

- Preprocessing
  - Adjusting CSV column names and paths
  - Add acquisition_date and acquisition_time to patient metadata’
  - Use patient_id_acq_time as the unique identifier

- Convolutional Neural Network (CNN) Model
  - Use a MyDenseNet object that Arianna wrote instead of Xun’s code

CSV Input Example for Xun

path_to_dicom	patient_id	CANCER_STATUS	view_position	image_laterality	dicom_id	Visit	acquisition_date	acquisition_time
/mnt/srl-oahu-1/disk/a	100685956	0	CC	R	1.2.840.114191	0	20200612	152301
/mnt/srl-oahu-1/disk/a	100685956	0	CC	L	1.2.840.114191	0	20200612	152301
/mnt/srl-oahu-1/disk/a	100685956	0	MLO	L	1.2.840.114191	0	20200612	152301
/mnt/srl-oahu-1/disk/a	100685956	0	MLO	R	1.2.840.114191	0	20200612	152301
/mnt/srl-oahu-1/disk/a	100689996	1	CC	R	1.2.840.114191	0	20210713	105623
/mnt/srl-oahu-1/disk/a	100689996	1	MLO	L	1.2.840.114191	0	20210713	105623
/mnt/srl-oahu-1/disk/a	100689996	1	CC	L	1.2.840.114191	0	20210713	105623
/mnt/srl-oahu-1/disk/a	100689996	1	MLO	R	1.2.840.114191	0	20210713	105623
/mnt/srl-oahu-1/disk/a	100689996	1	MLO	R	1.2.840.114191	1	20221025	90505
/mnt/srl-oahu-1/disk/a	100689996	1	MLO	L	1.2.840.114191	1	20221025	90505
/mnt/srl-oahu-1/disk/a	100689996	1	CC	L	1.2.840.114191	1	20221025	90505
/mnt/srl-oahu-1/disk/a	100689996	1	CC	R	1.2.840.114191	1	20221025	90505
/mnt/srl-oahu-1/disk/a	100985944	0	CC	R	1.2.840.114191	0	20170203	131502
/mnt/srl-oahu-1/disk/a	100985944	0	CC	L	1.2.840.114191	0	20170203	131502
/mnt/srl-oahu-1/disk/a	100985944	0	MLO	L	1.2.840.114191	0	20170203	131502
/mnt/srl-oahu-1/disk/a	100985944	0	MLO	R	1.2.840.114191	0	20170203	131502

Xun Output Example

patient_id_acq_time	label	group	LCC	RCC	LMLO	RMLO	n_dicoms	prob_cancer
100685956_152301	0	test	1.2.840.	1.2.840.1	1.2.840.11	1.2.840.114	4	0.19794065
100689996_105623	1	test	1.2.840.	1.2.840.1	1.2.840.11	1.2.840.114	4	0.11494599
100689996_90505	1	test	1.2.840.	1.2.840.1	1.2.840.11	1.2.840.114	4	0.14777811
100985944_131502	0	test	1.2.840.	1.2.840.1	1.2.840.11	1.2.840.114	4	0.14504619

Acknowledgements

Maximizing Access to Research Careers (MARC) grant 1T34GM141986

- Dr. Peter Sadowski
- Arianna Bunnell

MARC
at UH MĀNOA
Maximizing Access to Research Careers

References

Shen, Y., Wu, N., Phang, J., Park, J., Liu, K., Tyagi, S., Heacock, L., Kim, S. G., Moy, L., Cho, K., & Geras, K. J. (2021). An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. *Medical Image Analysis*, *68*, 101908. https://doi.org/10.1016/j.media.2020.101908

Zhu, X., Wolfgruber, T. K., Leong, L., Jensen, M., Scott, C., Winham, S., Sadowski, P., Vachon, C., Kerlikowske, K., & Shepherd, J. A. (2021). Deep Learning Predicts Interval and Screening-detected Cancer from Screening Mammograms: A Case-Case-Control Study in 6369 Women. *Radiology*, *301*(3), 550–558. https://doi.org/10.1148/radiol.2021203758

Images Cited

Chaitanya, N. (n.d.). *Critical steps to training and evaluating AI and ML models*. Retrieved July 31, 2024, from https://www.columbusglobal.com/en/blog/critical-steps-to-training-and-evaluating-ai-and-ml-models

Pacifici S, Murphy A, Hacking C, et al. Mediolateral oblique view. Reference article, Radiopaedia.org (Accessed on 30 Jul 2024) https://doi.org/10.53347/rID-15338

Pertuz, S., Torres, G. F., Tamimi, R., & Kamarainen, J. (2019). Open Framework for Mammography-based Breast Cancer Risk Assessment. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 1–4. https://doi.org/10.1109/BHI.2019.8834599

Scapicchio, C., Gabelloni, M., Barucci, A., Cioni, D., Saba, L., & Neri, E. (2021). A deep look into radiomics. *La Radiologia Medica*, *126*(10), 1296–1311. https://doi.org/10.1007/s11547-021-01389-x

Mahalo!
Questions?

BREAST DENSITY CLASSIFICATION WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

Nan Wu[1], *Krzysztof J. Geras*[1], *Yiqiu Shen*[1], *Jingyi Su*[1], *S. Gene Kim*[3,4], *Eric Kim*[3], *Stacey Wolfson*[3], *Linda Moy*[3,4] & *Kyunghyun Cho*[1,2]