



# Predicting Breast Cancer with Deep Learning

Journal Club

Brian Pollack

8/6/19

Paper: <https://pubs.rsna.org/doi/pdf/10.1148/radiol.2019182622>

# Study Overview



- ★ **Eval ML methods for early breast cancer detection using images and EHR.**

- Predict biopsy malignancy and classify normal vs abnormal

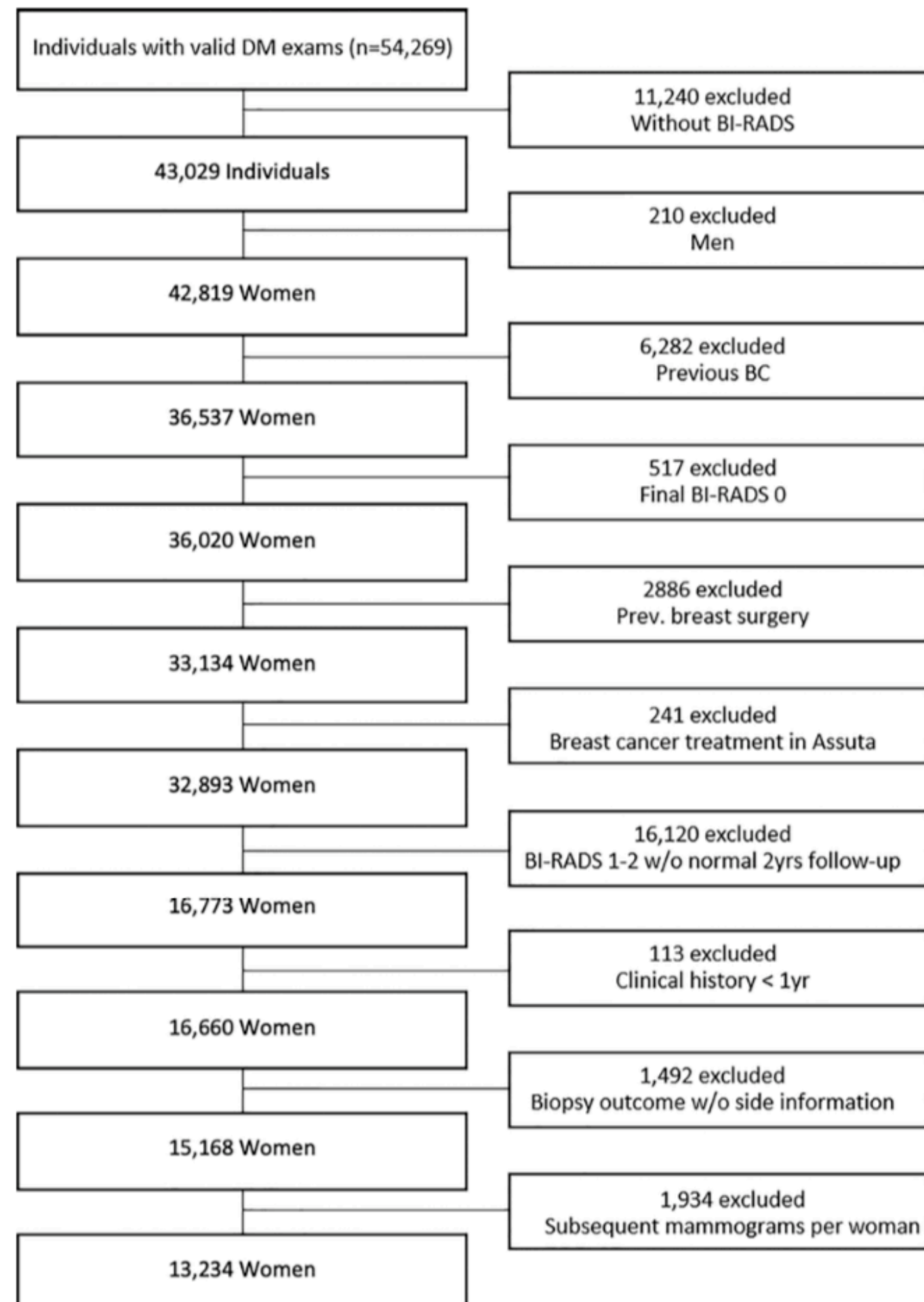
- ★ **Data:**

- 13k patients (52k images)
- Years: 2013-2017
- All had existing EHRs, all have 1 year follow-up
- Trained on 9.6k mammograms

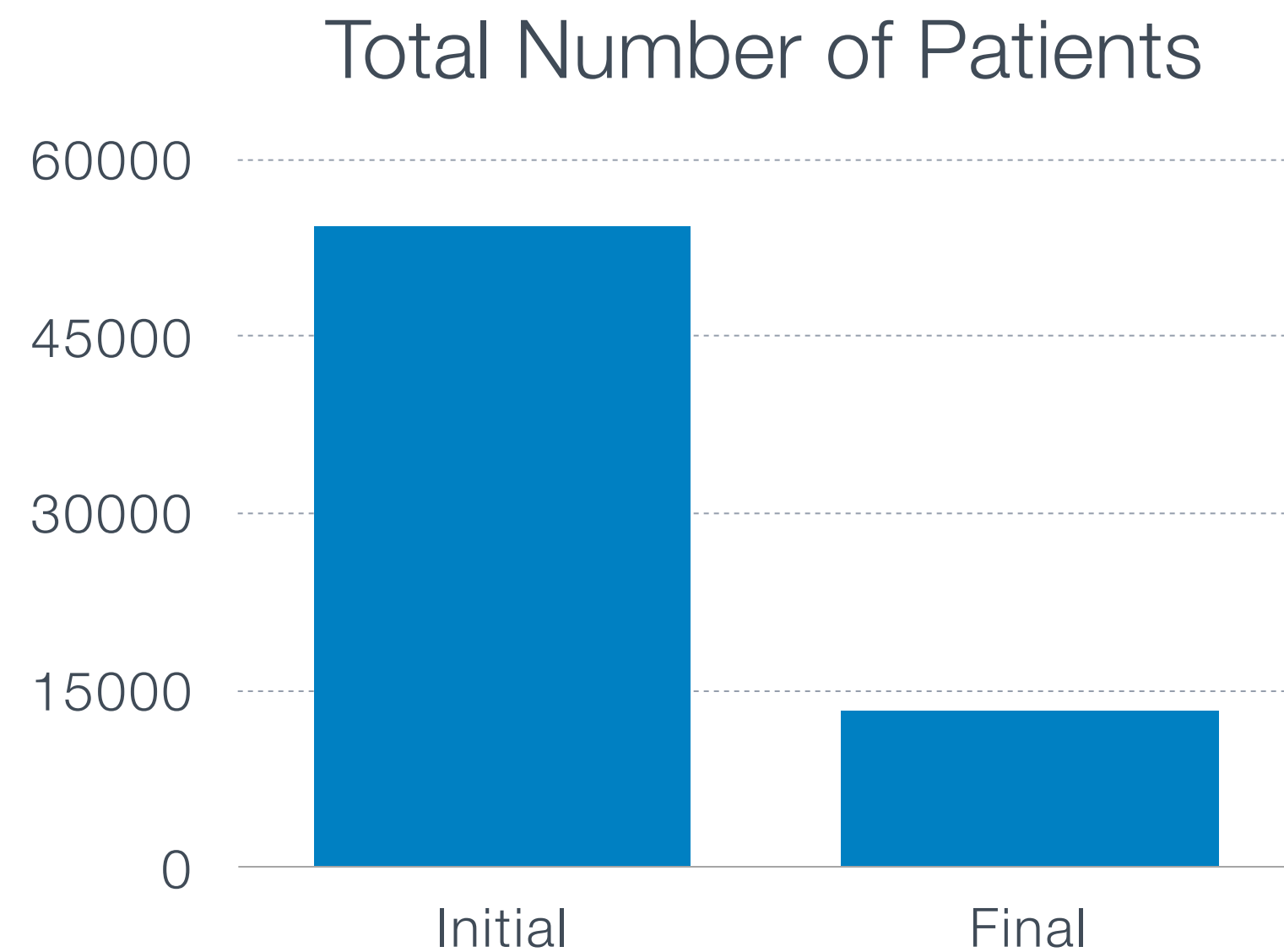
- ★ **Results:**

- Val: 1k patients, Test: 2.5k Patients
- Algo caught 34/71 false negatives (bad doctor predictions)
- Does about as well as an average radiologist

# Data Selection, Grouping



**Figure 1:** Flowchart of study inclusion and exclusion on the basis of the Strengthening Reporting of Observational Studies in Epidemiology (known as STROBE). BC = breast cancer, BI-RADS = Breast Imaging Reporting and Data System, DM = digital mammography, w/o = without.

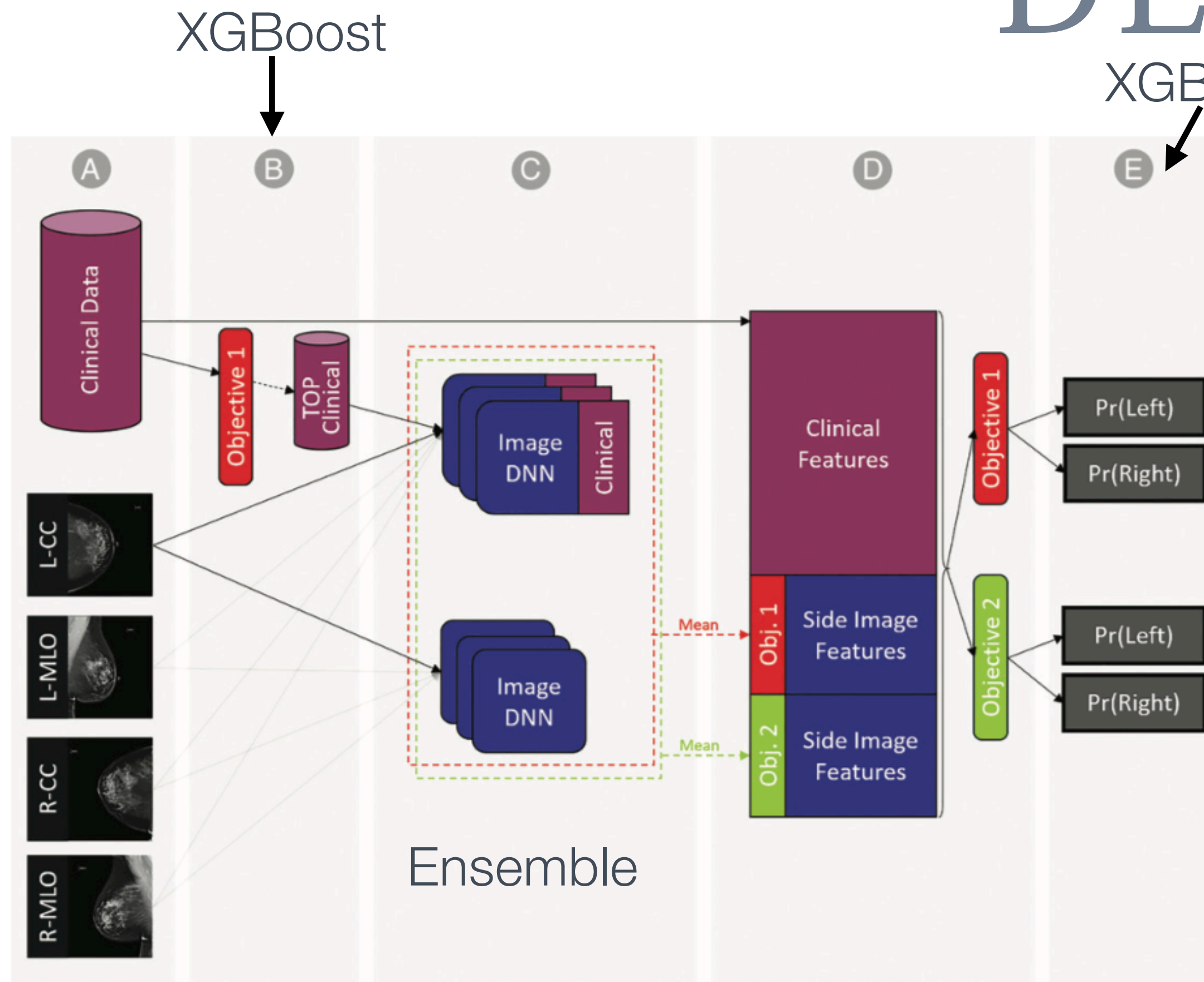


**Table E1: Training, validation and test sizes by cohort**

Cohort	# Samples		
	Train	Validation	Test
General Cohort	9611	1062	2548
Excluding US-only suspicious findings	8272	940	2202
First Examination	3043	348	839
First Examination + Excluding US-only suspicious findings	2523	296	682



# DL Model



**Figure 2:** Outline of machine- and deep-learning model. *A*, Input is standard four-view mammograms and clinical data (1343 features). *B*, Selection of top clinical features for deep neural network by using the XGBoost algorithm. *C*, Ensemble of deep neural network (DNN) algorithms trained separately for each prediction objective (objective 1, biopsy positive for cancer; objective 2, normal examination identification) per image either with or without the subset of clinical features. *D*, A woman's side is represented by a feature set composed of the imaging features obtained from both views of the breast (craniocaudal [CC] and mediolateral oblique [MLO]) for both prediction objectives (Obj) and joined with the entire set of clinical features. *E*, Predicted probability of outcome (Pr) per objective and per side is obtained by training an XGBoost algorithm on the vector formed in *D*. L = left, R = right.

## Inception Net

**Table E2: Description of the IM-DNN architecture per image**

Input	layer	#in ch	#out ch	kernel	stride	output x-y
*	Input image					2048 × 1024
	conv2d	1	32	3 × 3	2 × 2	
	conv2d	32	32	3 × 3	1 × 1	
	conv2d	32	64	3 × 3	1 × 1	
	maxpool	64	64	2 × 2	2 × 2	
	conv2d	64	80	3 × 3	1 × 1	
	conv2d	80	192	3 × 3	1 × 1	
	maxpool	192	192	2 × 2	2 × 2	
	mixed5b†	192	320			
	10xBlock35†	320	320			
	mixed6a†	320	1088			
	3xBlock17†	1088	1088			126 × 64
	GMP	1088	1088			1 × 1
*	Clinical input		30			
dense_1,GMP	dense_1	30	256			
	concat0		1344			
	dense_2	1344	256			
	dense_3	256	3			
	softmax	3	3			

“To view the areas on the image that were suspicious for cancer, we ... identified the smallest continuous areas in the input image that contributed the most to the malignancy prediction by the ML-DL model.”

# More Model Details



- ★ Images cropped to 4096x2048, resized to 2048x1024.
- ★ Simple pixel normalization (-1,1).
- ★ InceptionResnetV2<sup>54</sup>
- ★ Final image feature vectors pooled to single vector of 1088
- ★ BatchNorm, random affine, color transformation (?)
- ★ 8 Nvidia 1080 Tis (or Titan Xs)

# Population Normalization



- ★ **Training cohort is not a representative sample of local population:**
  - Training are typically more cancerous than general pop due to selection biases.
  - Bootstrapping data to counteract category imbalances:
    - Normal, Benign, Malignant
  - Eventually matched numbers from the Breast Cancer Surveillance Consortium

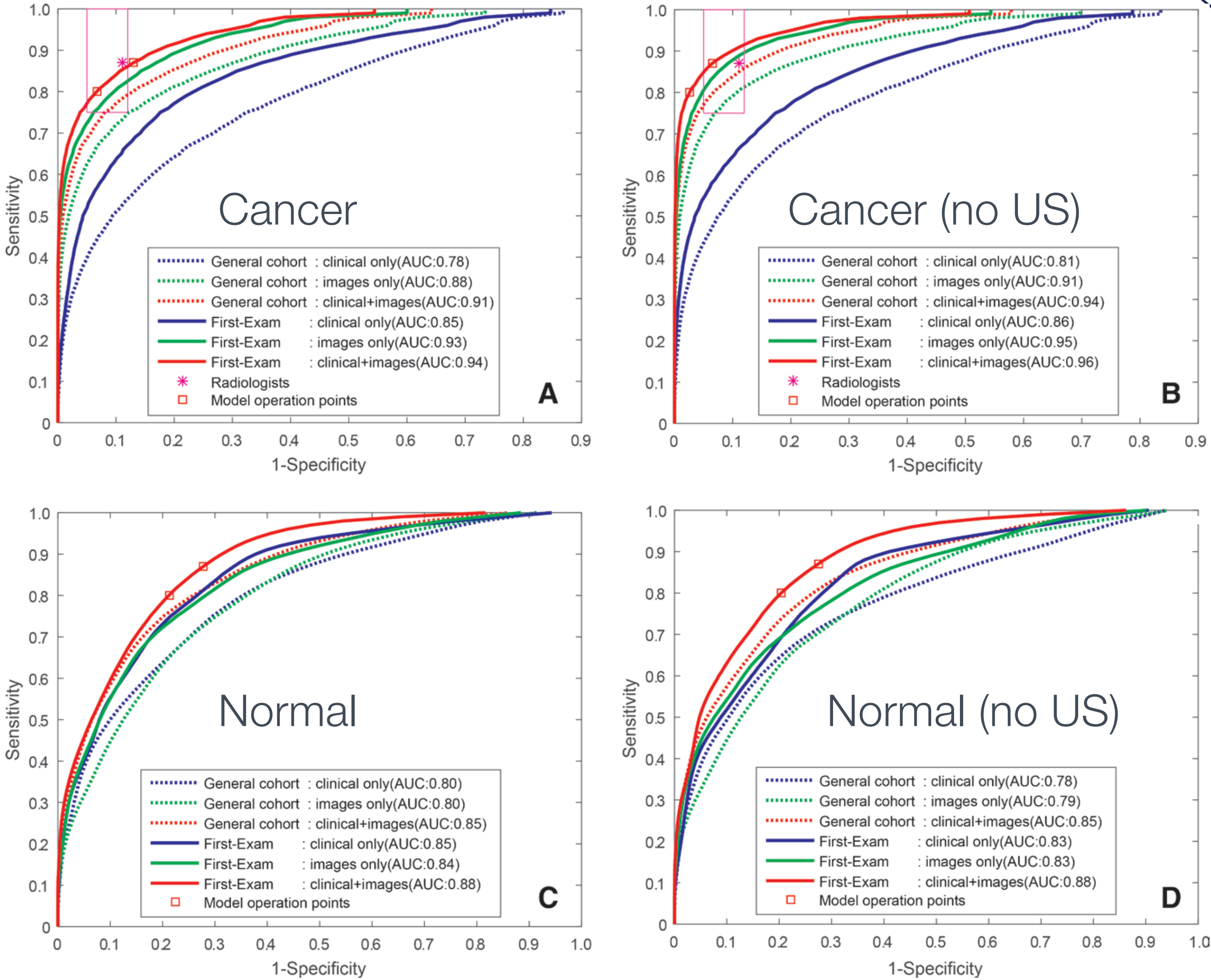


Table 3: Results of the Prediction Objectives Compared with Deep-Learning Models on the Breast Level			
Prediction Objective	AUC*	Specificity with Sensitivity of 87%†	Specificity with Sensitivity of 80%†
Objective 1, prediction of malignancy			
General cohort			
All features	0.91 (0.89, 0.93)	3139/4061 (77.3)[69.2, 85.4]	3537/4061 (87.1) [81.5, 92.7]
DM images only	0.88 (0.86, 0.90)	2835/4061 (69.8) [59.3, 80.3]	3314/4061 (81.6) [74.5, 88.7]
Clinical only	0.78 (0.75, 0.81)	1888/4061 (46.5) [38.5, 54.5]	2392/4061 (58.9) [50.9, 66.9]
Excluding US-only suspicious findings subcohort			
All features	0.94 (0.93, 0.95)	3196/3691 (86.6) [80.7, 92.5]	3444/3691 (93.3) [89.3, 97.3]
DM images only	0.91 (0.89, 0.93)	2942/3691 (79.7) [78.7, 80.8]	3296/3691 (89.3) [85.9, 92.7]
Clinical only	0.81 (0.78, 0.84)	1923/3691 (52.1) [45.4, 58.8]	2399/3691 (65) [55.4, 74.6]
First examination subcohort			
All features	0.94 (0.93, 0.95)	1055/1213 (87) [82.7, 91.3]	1132/1213 (93.3) [90.0, 96.6]
DM images only	0.93 (0.91, 0.95)	1008/1213 (83.1) [76.8, 89.4]	1095/1213 (90.3) [87.4, 93.2]
Clinical only	0.85 (0.82, 0.88)	787/1213 (64.9) [53.2, 76.6]	929/1213 (76.6) [70.1, 83.1]
First examination and excluding findings suspicious for cancer at US only subcohort			
All features	0.96 (0.95, 0.97)	979/1047 (93.5) [90.4, 96.6]	1020/1047 (97.4) [95.7, 99.1]
DM images only	0.95 (0.94, 0.96)	951/1047 (90.8) [87.3, 94.3]	997/1047 (95.2) [93.6, 96.8]
Clinical only	0.86 (0.84, 0.88)	689/1047 (65.8) [57.5, 74.1]	805/1047 (76.9) [70.5, 83.3]
Breast Cancer Surveillance Consortium radiologists		89	
DREAM model		0.87	81‡
Objective 2, identification of normal DM examinations			
General cohort			
All features	0.85 (0.84, 0.86)	648/1016 (63.8) [60.8, 66.8]	754/1016 (74.2) [71.1, 77.3]
DM images only	0.80 (0.79, 0.81)	555/1016 (54.6) [51.5, 57.7]	650/1016 (64) [61.6, 66.4]
Clinical only	0.80 (0.79, 0.81)	533/1016 (52.5) [48.5, 56.5]	656/1016 (64.6) [62.6, 66.6]
Subcohort excluding findings suspicious for cancer found at US only			
All features	0.85 (0.84, 0.86)	374/597 (62.6) [58.2, 67.0]	439/597 (73.5) [71.1, 75.9]
DM images only	0.79 (0.78, 0.80)	304/597 (51) [47.9, 54.1]	367/597 (61.4) [58.7, 64.1]
Clinical only	0.78 (0.77, 0.79)	252/597 (42.2) [36.4, 48.0]	346/597 (58) [55.0, 61.0]
First examination subcohort			
All features	0.88 (0.87, 0.89)	341/472 (72.3) [70.6, 74.0]	372/472 (78.8) [77.5, 79.9]
DM images only	0.84 (0.83, 0.85)	297/472 (62.9) [60.4, 65.4]	339/472 (71.8) [69.3, 74.3]
Clinical only	0.85 (0.84, 0.86)	313/472 (66.4) [64.5, 68.3]	346/472 (73.4) [71.1, 75.7]
First examination and subcohort excluding findings suspicious for cancer found at US only			
All features	0.88 (0.87, 0.89)	210/289 (72.6) [69.4, 75.8]	230/289 (79.6) [77.1, 82.1]
DM images only	0.83 (0.82, 0.84)	163/289 (56.5) [52.0, 61.0]	196/289 (67.8) [65.0, 70.6]
Clinical only	0.83 (0.82, 0.84)	188/289 (65.2) [64.1, 66.3]	207/289 (71.7) [69.7, 73.7]
Normal identification with deep learning§		0.61	20‡

Note.—Unless otherwise indicated, data are numerator/denominator. Individual level results are reported in Table E8 (online). AUC = area under the receiver operating characteristic curve, The excluding US-only suspicious findings subcohort excluded examinations in which the digital mammography final BI-RADS assessment was 1 or 2 and the US final BI-RADS assessment was 3 or higher. DREAM = Dialogue for Reverse Engineering Assessments and Methods, DM = digital mammography.  
\* Data in parentheses are 95% confidence intervals.  
† Data in parentheses are percentages; data in brackets are 95% confidence intervals.  
‡ Results were previously obtained by Geras et al (32).

# Results

Predicting Breast Cancer by Using Deep Learning on Electronic Health Records and Mammograms



# Discussion



- ★ **Performs about as well as radiologists.**
- ★ **Caught some false-negatives.**
- ★ **Does not use:**
  - Context (reference images)
  - Transfer learning (imagenet or other)
  - Self-supervised methods
  - Unlabeled data
- ★ **Addition of clinical data improves performance.**
- ★ **Does not use US images**
- ★ **Marketed as a double-reading enabler, lots of attention called to the correct diagnosis of false-negatives.**