RSNA 2018

Machine Learning for Identifying the Value of Digital Breast Tomosynthesis using Data from a Multicentre Retrospective Study

A. Alaa, **Fiona J. Gilbert,** Y. Huang and M. van der Schaar

November, 2018

Main Objectives

- We aim at using machine learning and data from the TOMMY trial to answer the following questions…

Which women benefit the most from DBT imaging?
Age and breast density groups

What are the accuracy gains achieved by DBT compared to 2D mammography for different types of lesions?

Background

- **Limitations of standard 2D mammography:** overlapping dense Fibroglandular tissue can decrease visibility of malignant abnormalities or simulate the appearance of an abnormality.

Negative impact on:

Sensitivity

False negatives

15–30% of cancers are not detected by standard screening.

Worse for women aged under **50 years** and in women with **dense** breasts

Specificity

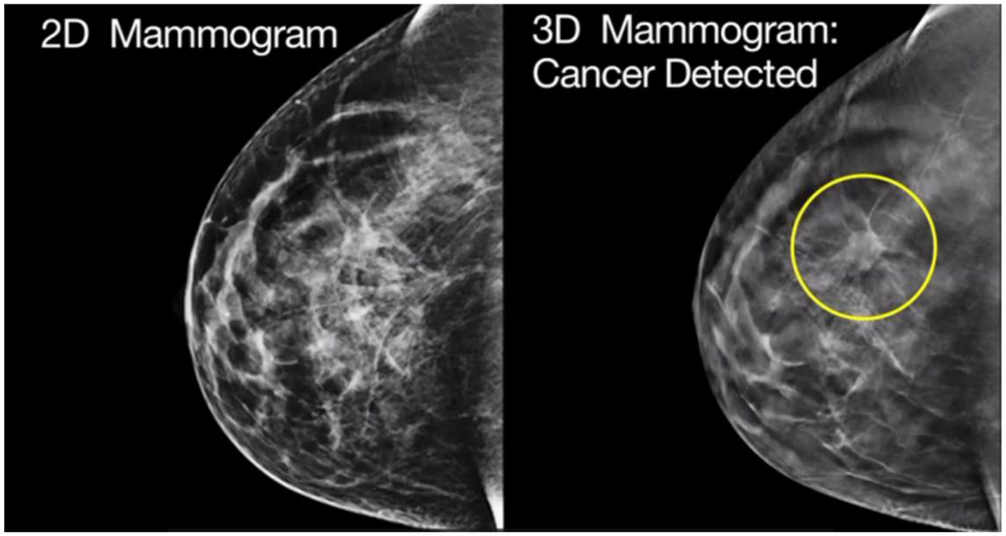
False positives

Unnecessary extra visit/imaging/biopsy

Digital Breast Tomosynthesis (DBT)

- Improve the conspicuity of mammography by reducing overlapping shadows from breast tissue that degrade the image quality in standard 2D projection imaging.

- Better differentiation between malignant and non-malignant features.



Digital Breast Tomosynthesis (DBT)

The Expectations from DBT are

Better Sensitivity

Small cancers obscured by normal Fibroglandular tissue in standard 2D imaging should be detected using DBT

Better Specificity

Identify features on 2D imaging as normal composite shadows and thereby decrease the number of false-positive recalls

Beneficial for women with dense breasts!

Our Analysis

- Using machine learning to <u>augment</u> radiological reports!

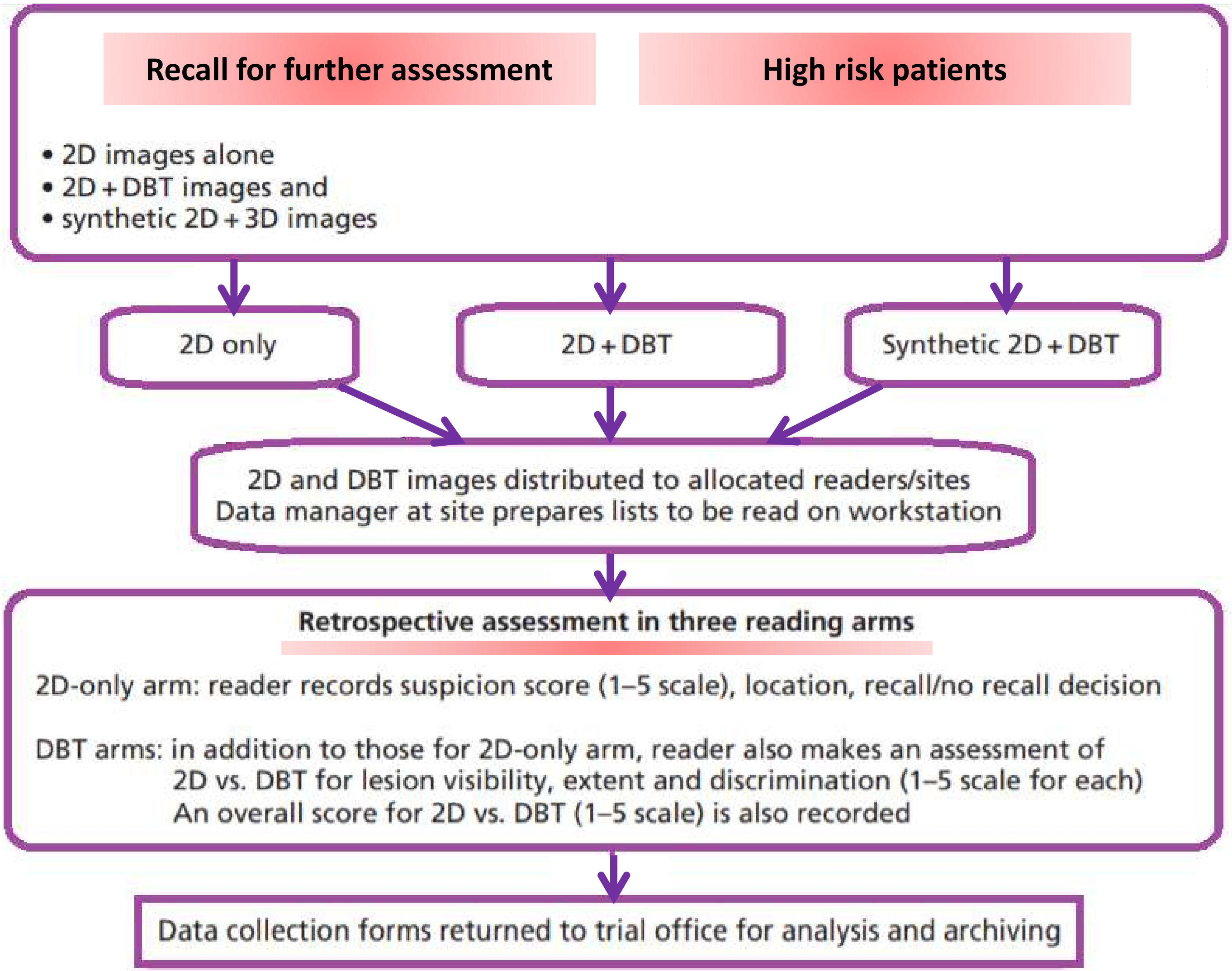
Age and breast density

AutoPrognosis

Prediction of malignancy

Radiological report

The TOMMY Study Design: Reading Study



Radiologist reports

DBT-related	2D-related
MLO sign	MLO sign
MLO suspicion	MLO suspicion
MLO conspicuity	MLO conspicuity
CC sign	CC sign
CC suspicion	CC suspicion
CC conspicuity	CC conspicuity

Prognostic Modelling with
Age & breast density & Volpara measures

Prognostic Model	AUC-ROC	Prognostic Model	AUC-ROC
Logistic Regression	0.679 ± 0.033	**Multinomial Naïve Bayes**	0.541 ± 0.040
SGD Perceptron	0.512 ± 0.049	**AdaBoost**	0.666 ± 0.030
KNN	0.539 ± 0.012	**Bagging**	0.585 ± 0.020
Decision Tree	0.518 ± 0.012	**Gradient Boosting**	0.674 ± 0.029
Linear SVM	0.556 ± 0.021	**XGBoost**	0.674 ± 0.028
Gauss. Naïve Bayes	0.660 ± 0.024	**MLP**	0.482 ± 0.033
Bern. Naïve Bayes	0.499 ± 0.000	**Random Forest**	0.596 ± 0.023
LDA	0.663 ± 0.023	**AutoPrognosis**	**0.685 ± 0.025**

Prognostic Modelling with
Age & breast density & Volpara measures

Prognostic Model	AUC-PR	Prognostic Model	AUC-PR
Logistic Regression	0.271 ± 0.032	**Multinomial Naïve Bayes**	0.179 ± 0.024
SGD Perceptron	0.164 ± 0.021	**AdaBoost**	0.255 ± 0.030
KNN	0.167 ± 0.029	**Bagging**	0.201 ± 0.009
Decision Tree	0.162 ± 0.003	**Gradient Boosting**	0.256 ± 0.025
Linear SVM	0.182 ± 0.008	**XGBoost**	0.261 ± 0.024
Gauss. Naïve Bayes	0.266 ± 0.024	**MLP**	0.138 ± 0.013
Bern. Naïve Bayes	0.154 ± 0.000	**Random Forest**	0.201 ± 0.011
LDA	0.271 ± 0.032	**AutoPrognosis**	**0.264 ± 0.028**

Prognostic Modelling with Radiologist report extracted features

- **Radiological assessments are used as the predictors.**

DBT-related	2D-related
MLO sign	MLO sign
MLO suspicion	MLO suspicion
MLO conspicuity	MLO conspicuity
CC sign	CC sign
CC suspicion	CC suspicion
CC conspicuity	CC conspicuity

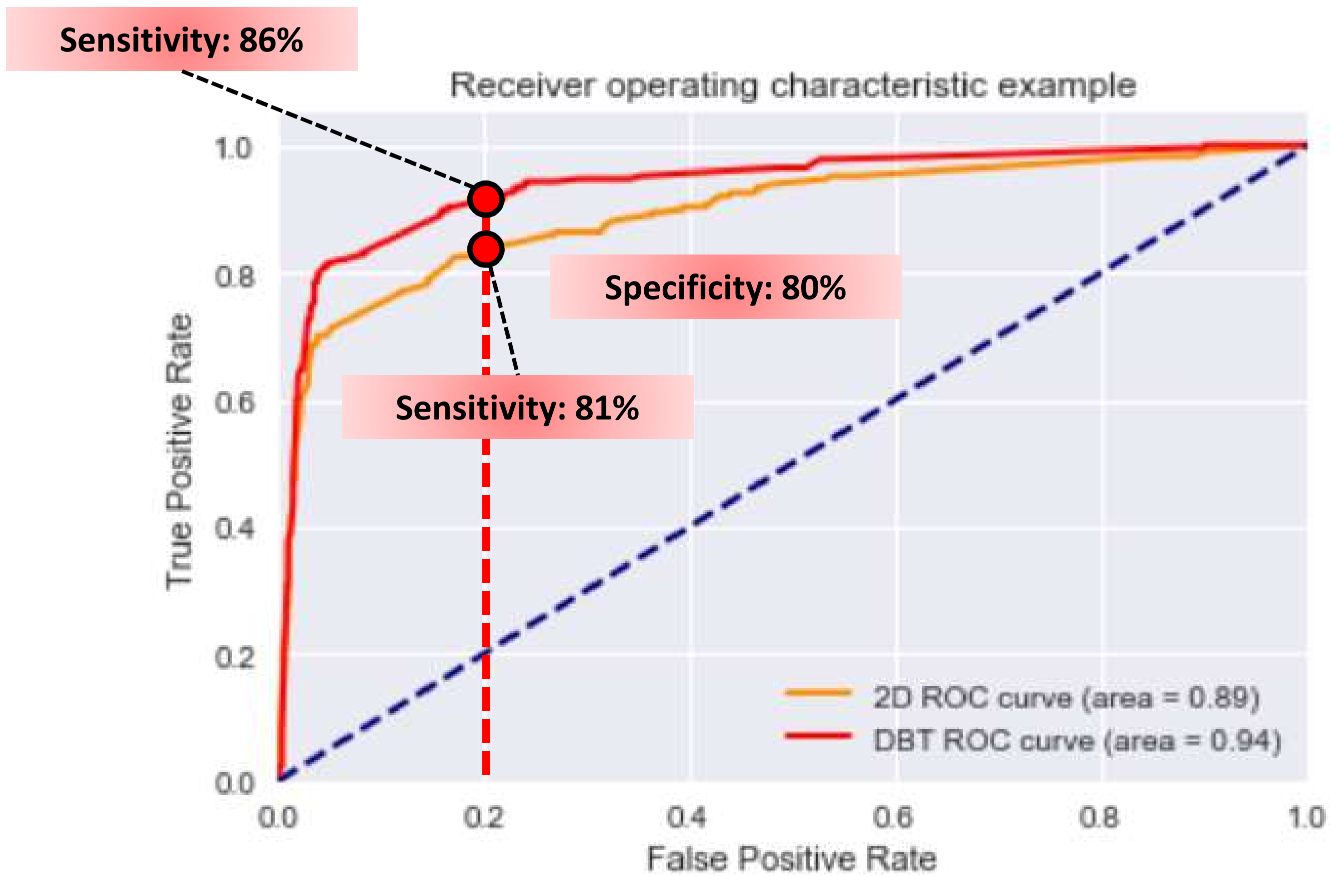
Prognostic Modelling with 2D radiologist reports

Prognostic Model	AUC-PR	Prognostic Model	AUC-PR
Logistic Regression	0.742 ± 0.057	**Multinomial Naïve Bayes**	0.296 ± 0.035
SGD Perceptron	0.719 ± 0.067	**AdaBoost**	0.765 ± 0.061
KNN	0.689 ± 0.060	**Bagging**	0.735 ± 0.051
Decision Tree	0.675 ± 0.053	**Gradient Boosting**	0.772 ± 0.057
Linear SVM	0.729 ± 0.058	**XGBoost**	0.775 ± 0.057
Gauss. Naïve Bayes	0.716 ± 0.059	**MLP**	0.760 ± 0.065
Bern. Naïve Bayes	0.286 ± 0.021	**Random Forest**	0.735 ± 0.063
LDA	0.743 ± 0.058	**AutoPrognosis**	**0.776 ± 0.051**

Prognostic Modelling with DBT radiologist reports

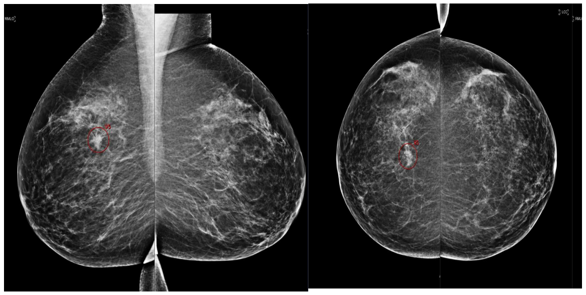
Prognostic Model	AUC-PR	Prognostic Model	AUC-PR
Logistic Regression	0.797 ± 0.043	Multinomial Naïve Bayes	0.347 ± 0.058
SGD Perceptron	0.797 ± 0.043	AdaBoost	0.808 ± 0.052
KNN	0.723 ± 0.057	Bagging	0.793 ± 0.047
Decision Tree	0.745 ± 0.042	Gradient Boosting	0.804 ± 0.045
Linear SVM	0.780 ± 0.030	XGBoost	812 ± 0.047
Gauss. Naïve Bayes	0.774 ± 0.046	MLP	0.812 ± 0.052
Bern. Naïve Bayes	0.304 ± 0.020	Random Forest	0.797 ± 0.048
LDA	0.799 ± 0.043	**AutoPrognosis**	**0.812 ± 0.042**

2D vs. DBT Diagnostic Accuracy



MLO vs. CC Diagnostic Accuracy

- **Which of the two views (CC and MLO) is more predictive?**



2D-CC	0.888 ± 0.025
2D-MLO	0.892 ± 0.023
2D-CC+MLO	0.915 ± 0.018

MLO is slightly more predictive

Both CC and MLO are complementary

DBT-CC	0.918 ± 0.014
DBT-MLO	0.932 ± 0.009
DBT-CC+MLO	0.942 ± 0.009

MLO is slightly more predictive

Both CC and MLO are complementary

How predictive are single radiological assessments?

- **Suspicion scores are the most predictive.**

Assessment Score	AUC-PR	Assessment Score	AUC-PR
2D-CC Sign	0.2375	DBT-CC Sign	0.2858
2D-CC Suspicion	0.5757	DBT-CC Suspicion	0.6409
2D-CC Conspicuity	0.2380	DBT-CC Conspicuity	0.2574
2D-MLO Sign	0.2405	DBT-MLO Sign	0.2900
2D-MLO Suspicion	0.5707	DBT-MLO Suspicion	0.6407
2D-MLO Conspicuity	0.2396	DBT-MLO Conspicuity	0.2564

How well can machine learning augment radiological assessment?

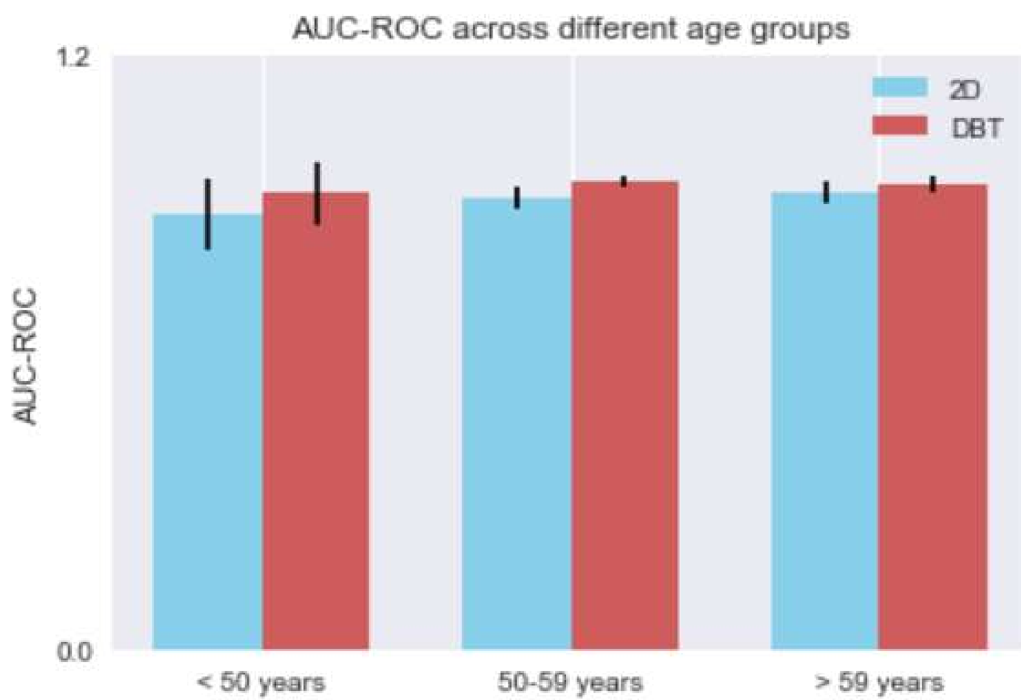
- Machine learning constructs a scoring rule that combines the different assessments in a radiological report

Assessment Score	AUC-ROC
Most Accurate single 2D Radiological Assessment	0.8669
Most Accurate single DBT Radiological Assessment	0.9119
Machine Learning Accuracy	0.9435

Machine learning applied to all assessments for MLO and CC images

Which Women Benefit the most from DBT Imaging?

Which women benefit the most from DBT imaging?
Stratify by age groups

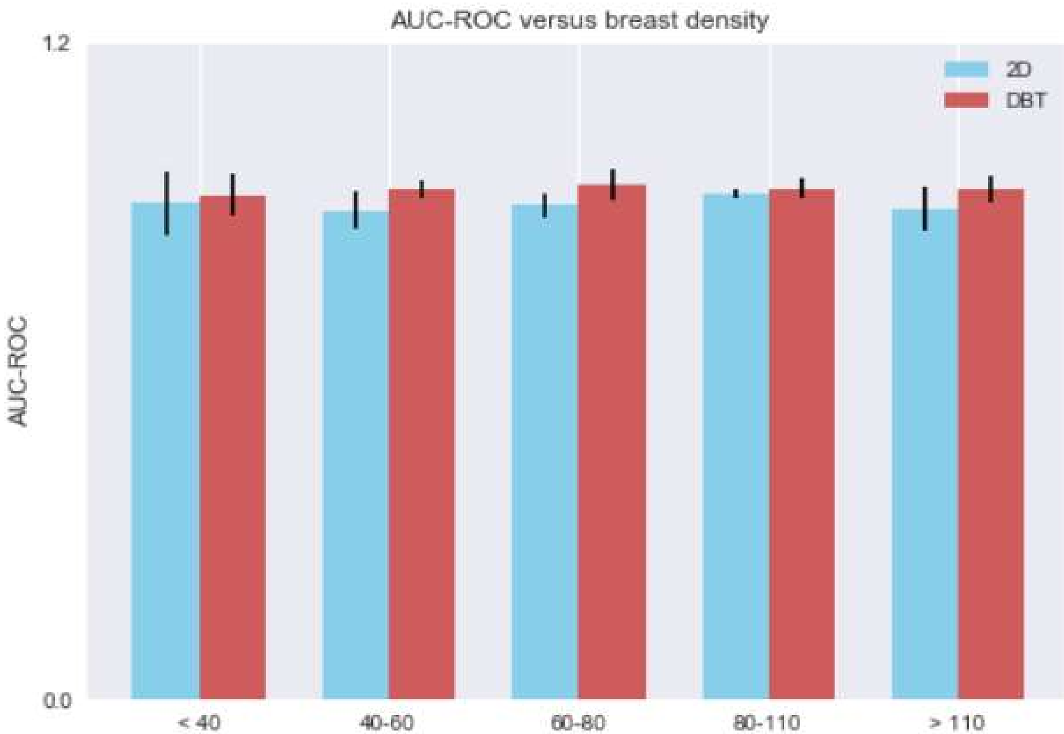


Fibrograndular Tissue Volume	89.5%	80.6 %	68.9 %

Younger age groups benefit more from DBT

Which Women Benefit the most from DBT Imaging?

Which women benefit the most from DBT imaging?
Breast density groups



Fibrograndular Tissue Volume

Group with density 40-80 and > 110 benefit most from DBT

Accuracy of DBT Imaging for different types of tumours

What is the accuracy of DBT for different types of tumours?

Three types of Malignant Tumours

Invasive ductal	
Invasive lobular	
Other + DCIS	• Medullary • Mucinous • Tubular • Intracystic papillary • Lymphoma • Ductal In situ

Accuracy of DBT Imaging for different types of tumours

What is the accuracy of DBT for different types of tumours?

Three types of Malignant Tumors

	Invasive ductal	Invasive lobular	Other invasive and DCIS
2D	0.919 ± 0.021	0.873 ± 0.014	0.875 ± 0.035
DBT	0.943 ± 0.011	0.905 ± 0.015	0.894 ± 0.029

Conclusions

- Machine learning combining different radiological assessments in a radiologist's report together with age and breast density improve predictions of malignancy.
- Suspicious assessments from MLO view are dominant predictive features in radiological reports.
- DBT improves accuracy (compared to 2D) for younger patients.
- Accuracy improvement with DBT is uniform over all types of tumours.