Performance Measures

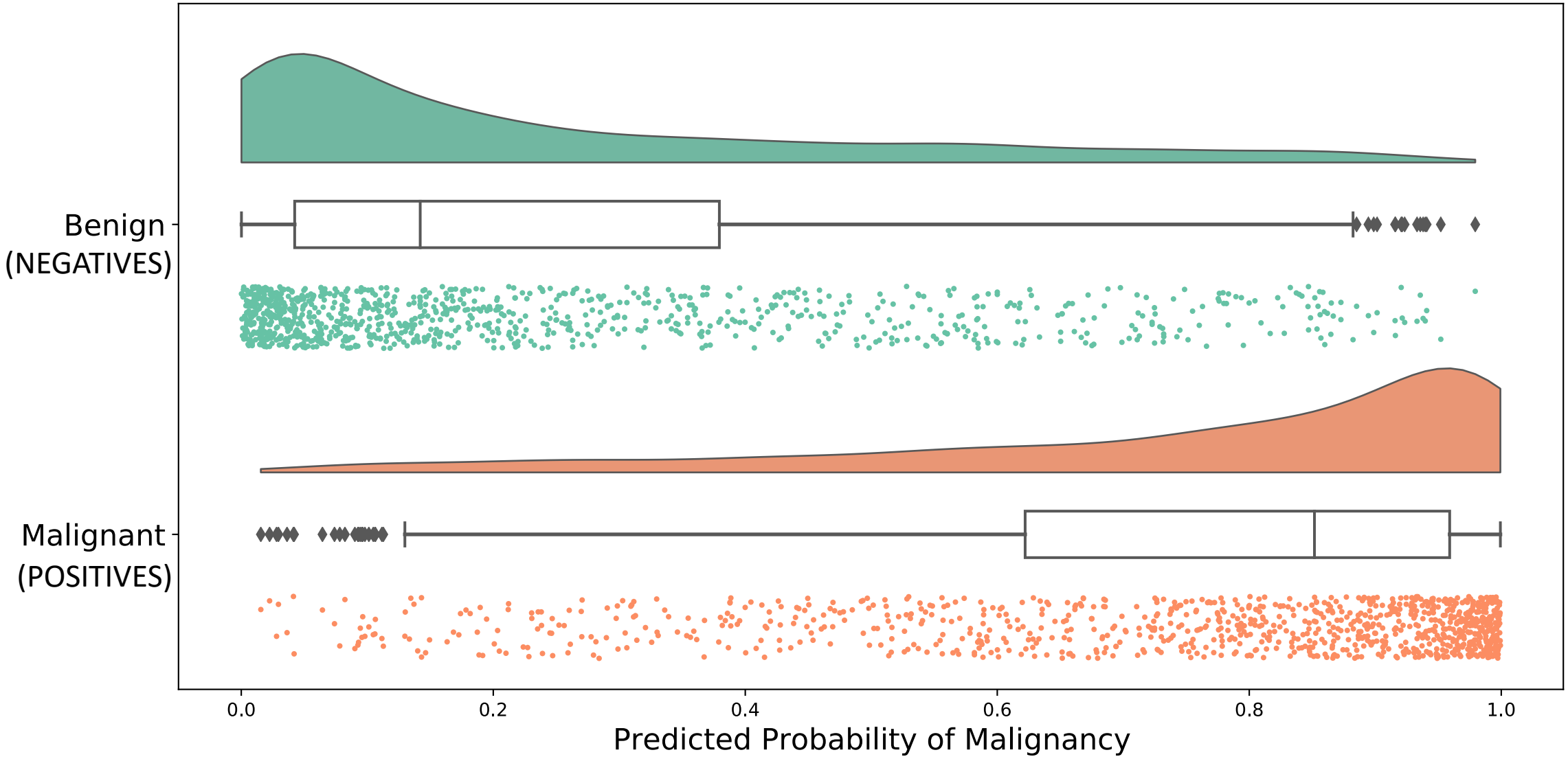
MMCi Block 2

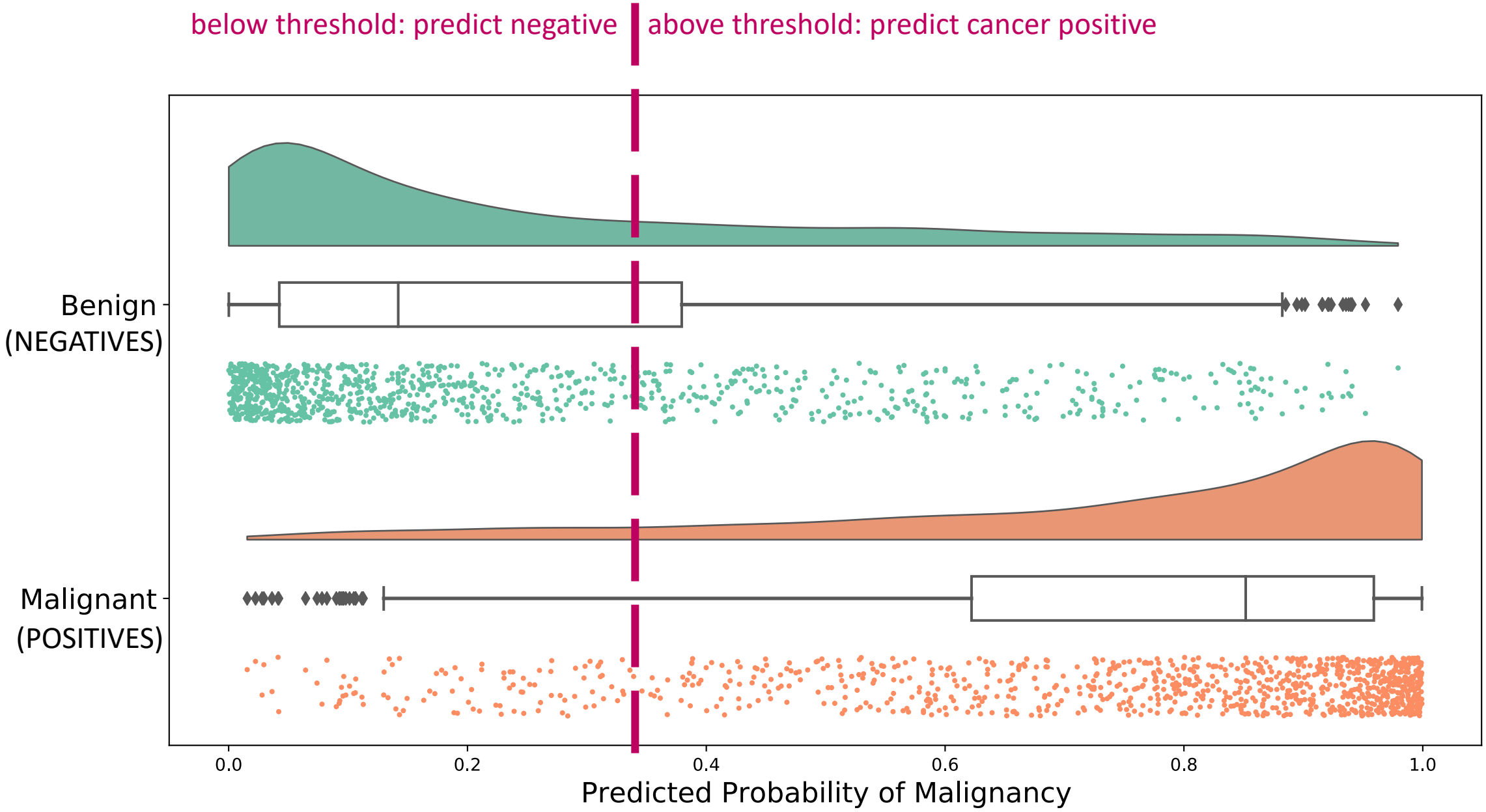
Matthew Engelhard

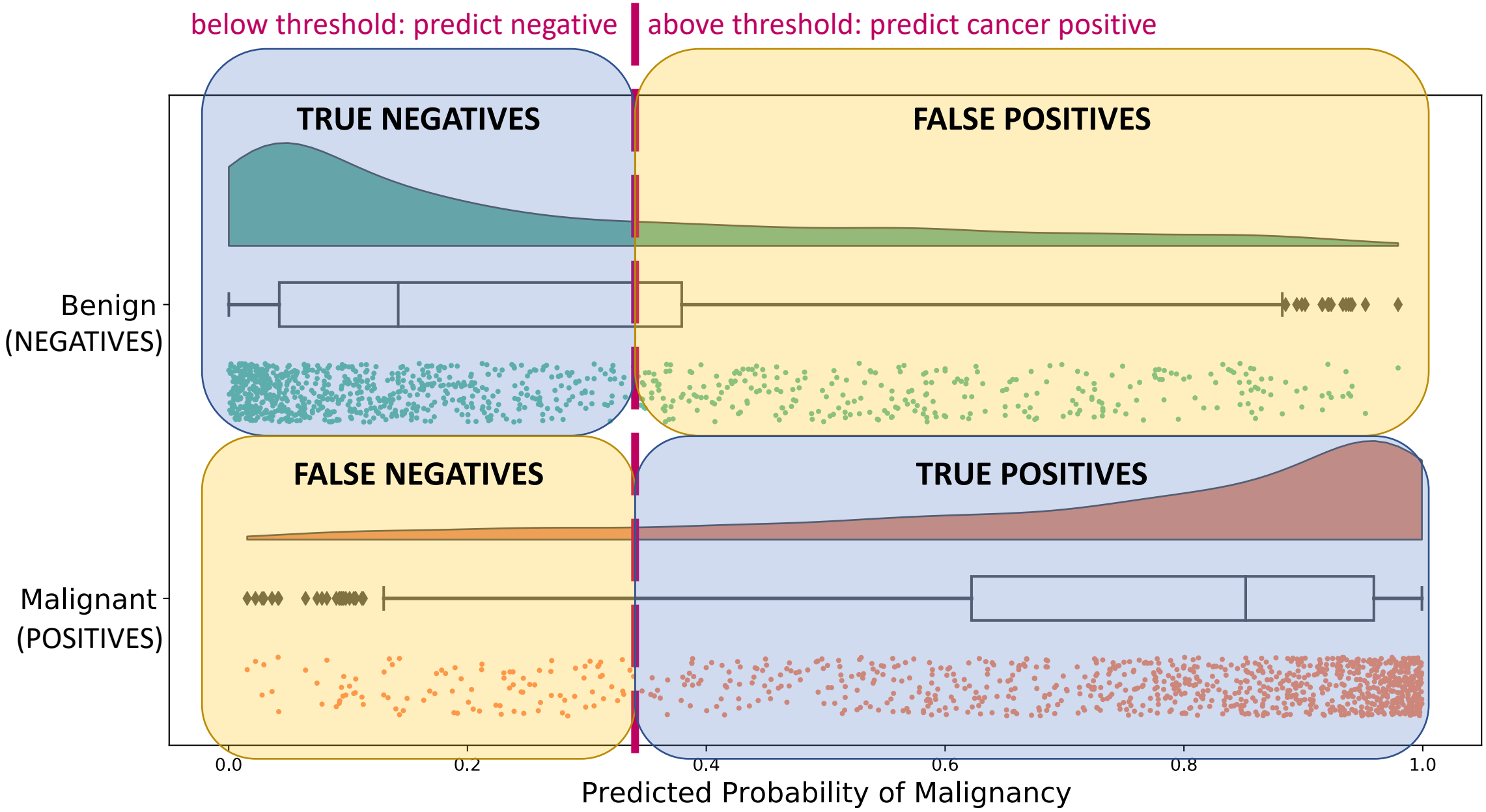
Goals

- Understand common performance measures for binary classification
- Recognize that which measure(s) are most appropriate depends on the application
- Run through a few different clinical scenarios
- Touch on metrics for problems other than binary classification

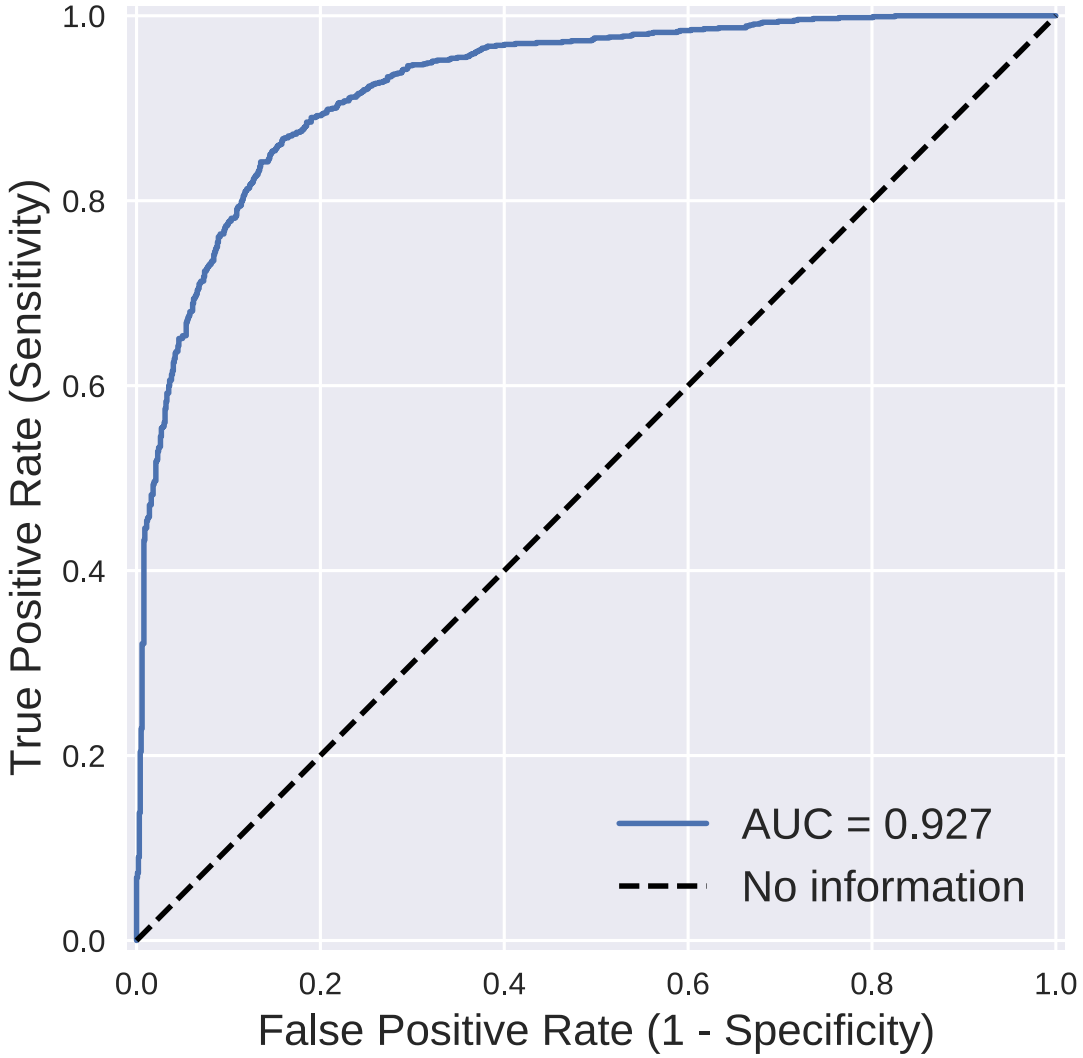
Let’s go back to cancer prediction

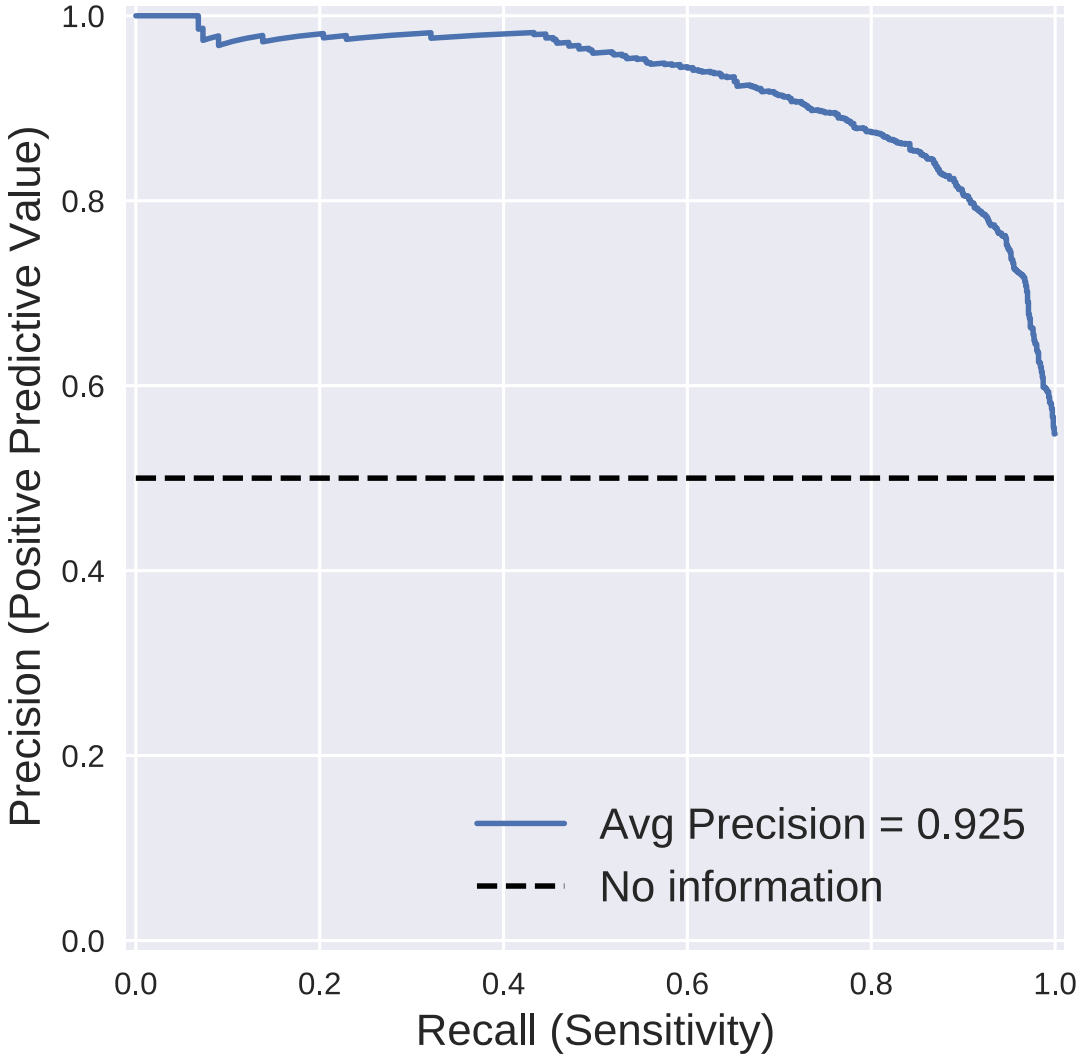



below threshold: predict negative
above threshold: predict cancer positive
Benign
(NEGATIVES)
Malignant
(POSITIVES)
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Probability of Malignancy


below threshold: predict negative
above threshold: predict cancer positive
TRUE NEGATIVES
FALSE POSITIVES
FALSE NEGATIVES
TRUE POSITIVES
Benign
(NEGATIVES)
Malignant
(POSITIVES)
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Probability of Malignancy

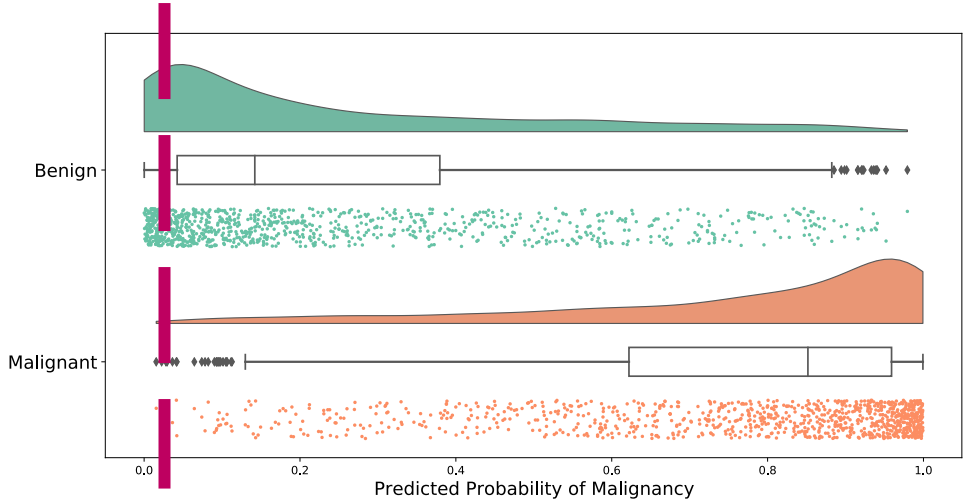
ROC versus PR curve: two different tradeoffs

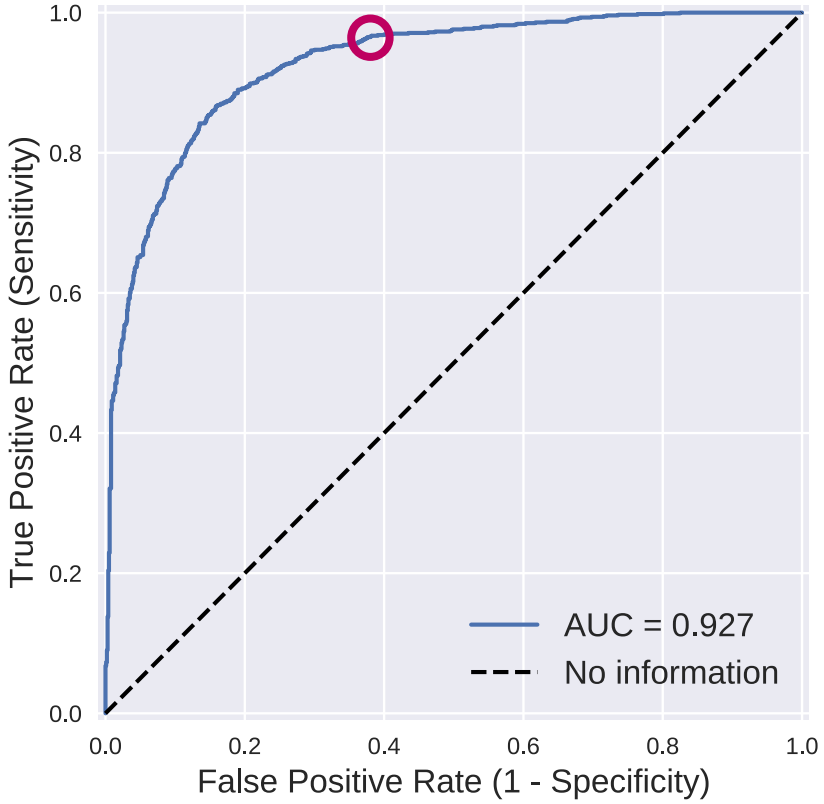


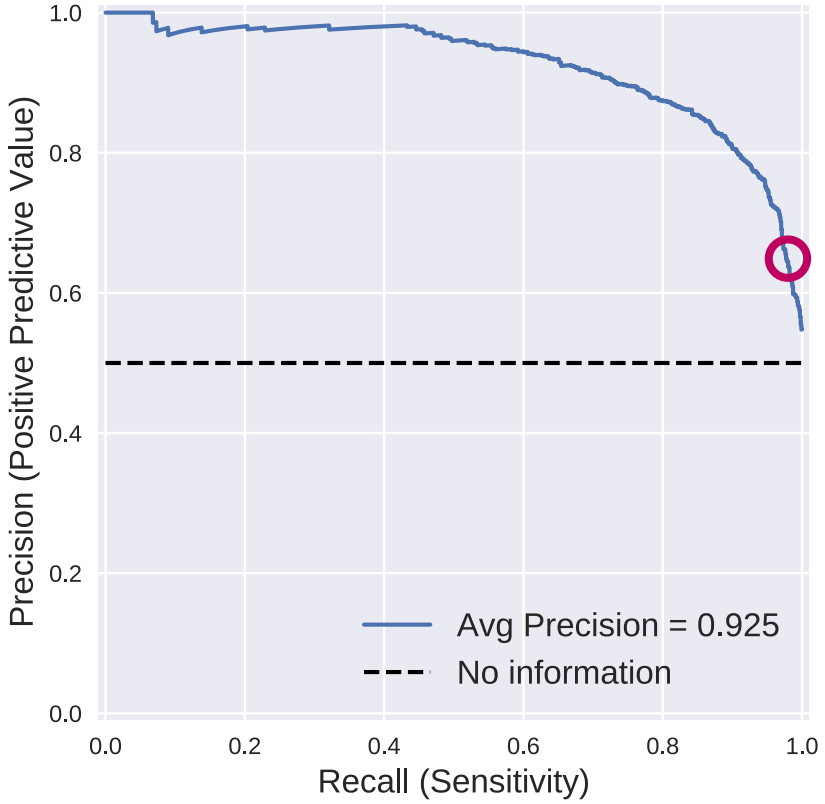


Operating Point:

high sensitivity

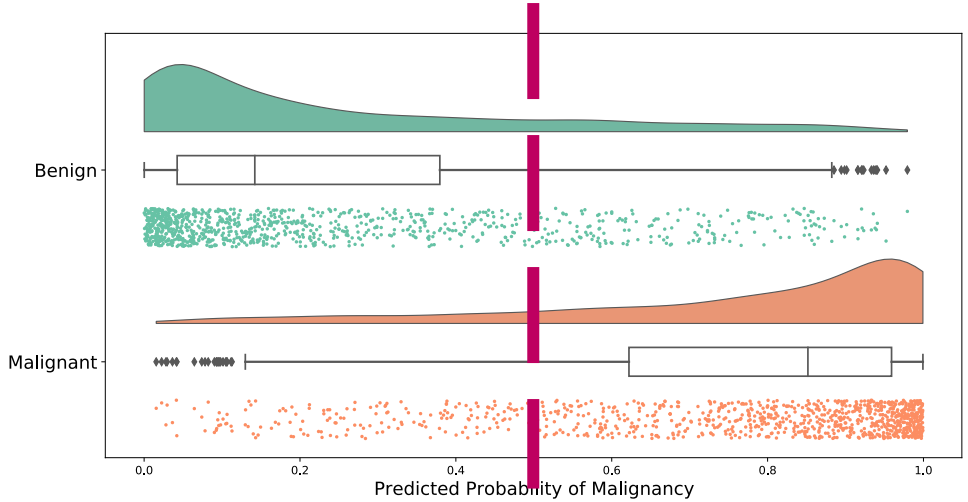


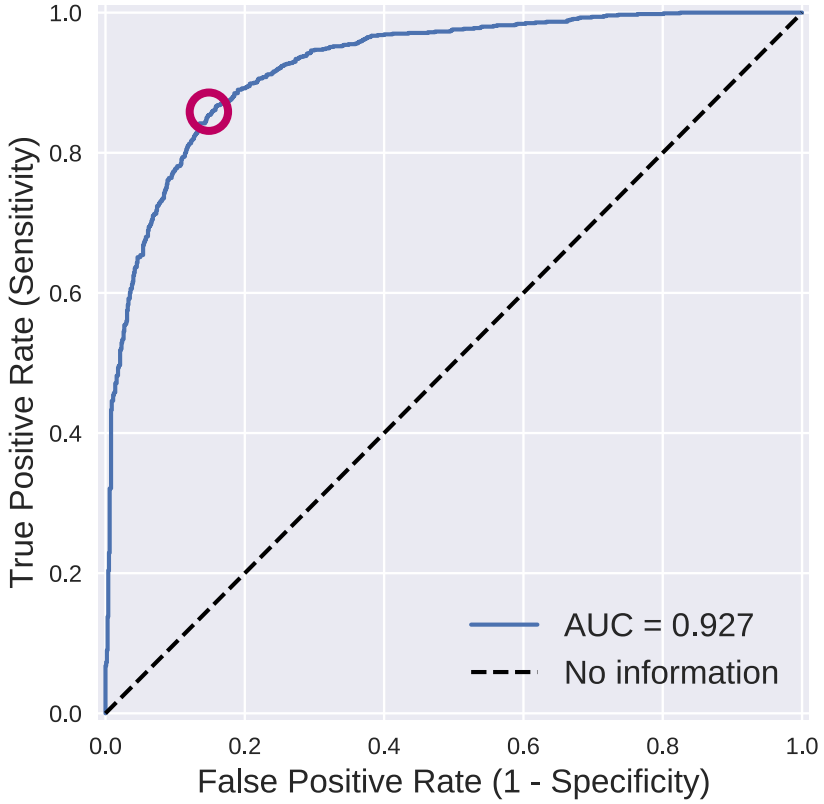


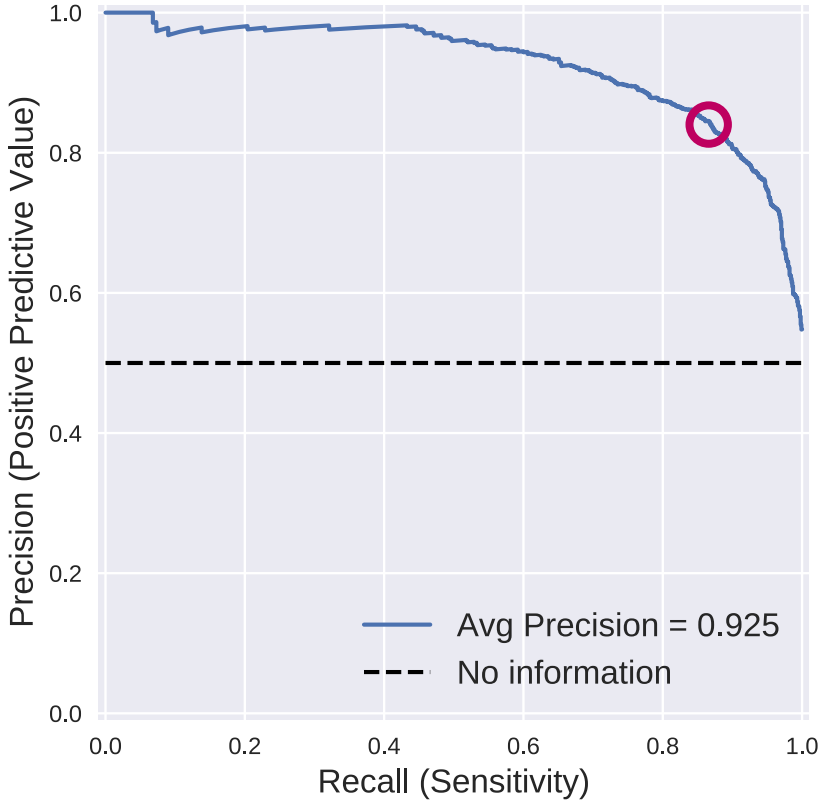


Operating Point:

balanced

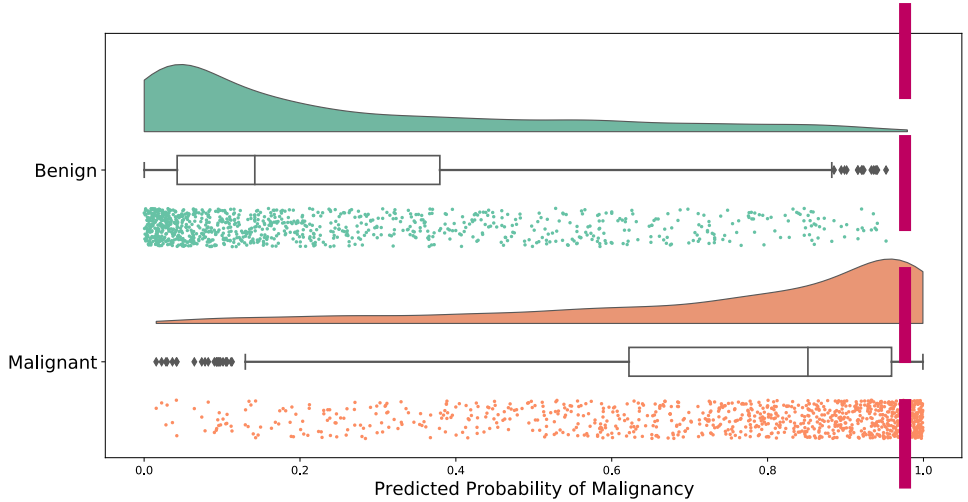


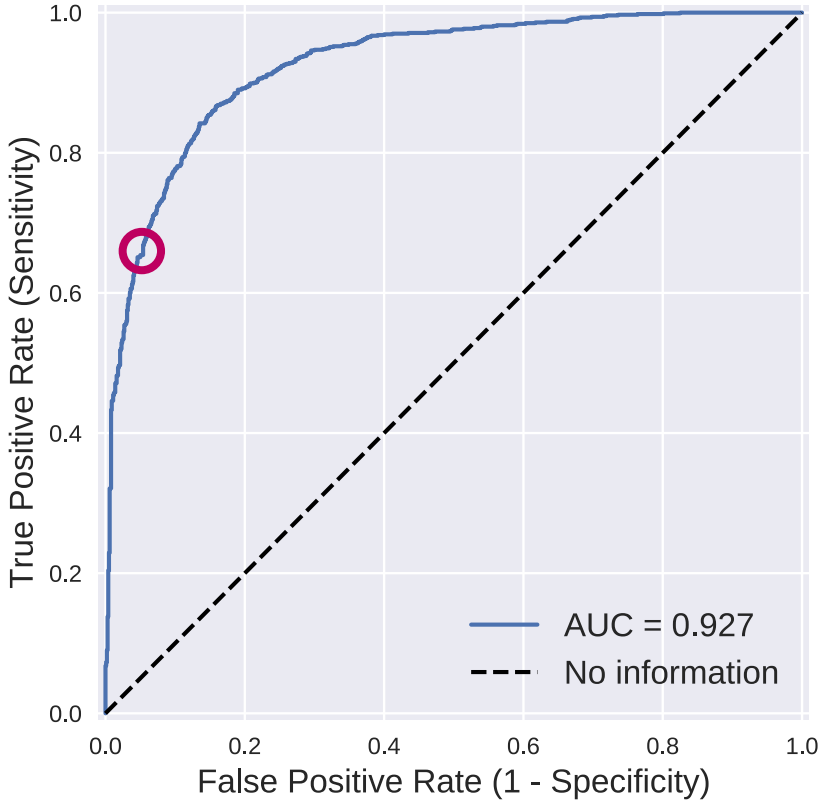


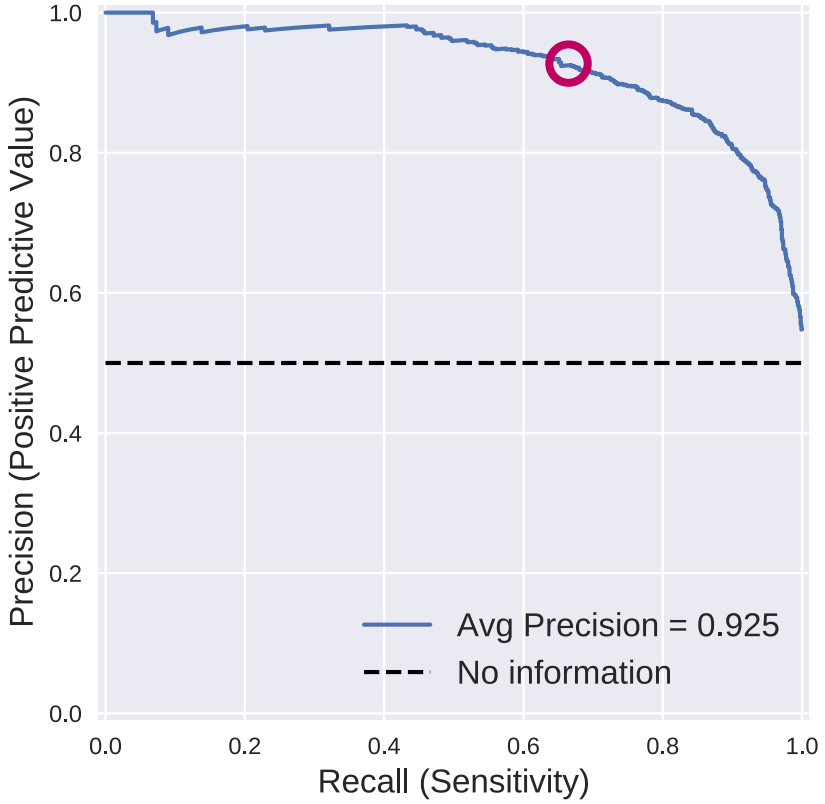


Operating Point:

high specificity







Healthcare Scenarios

1. A computer vision model that detects carcinoma

Healthcare Scenarios

1. A computer vision model that detects carcinoma

2. An EHR-based model that surveils autism risk

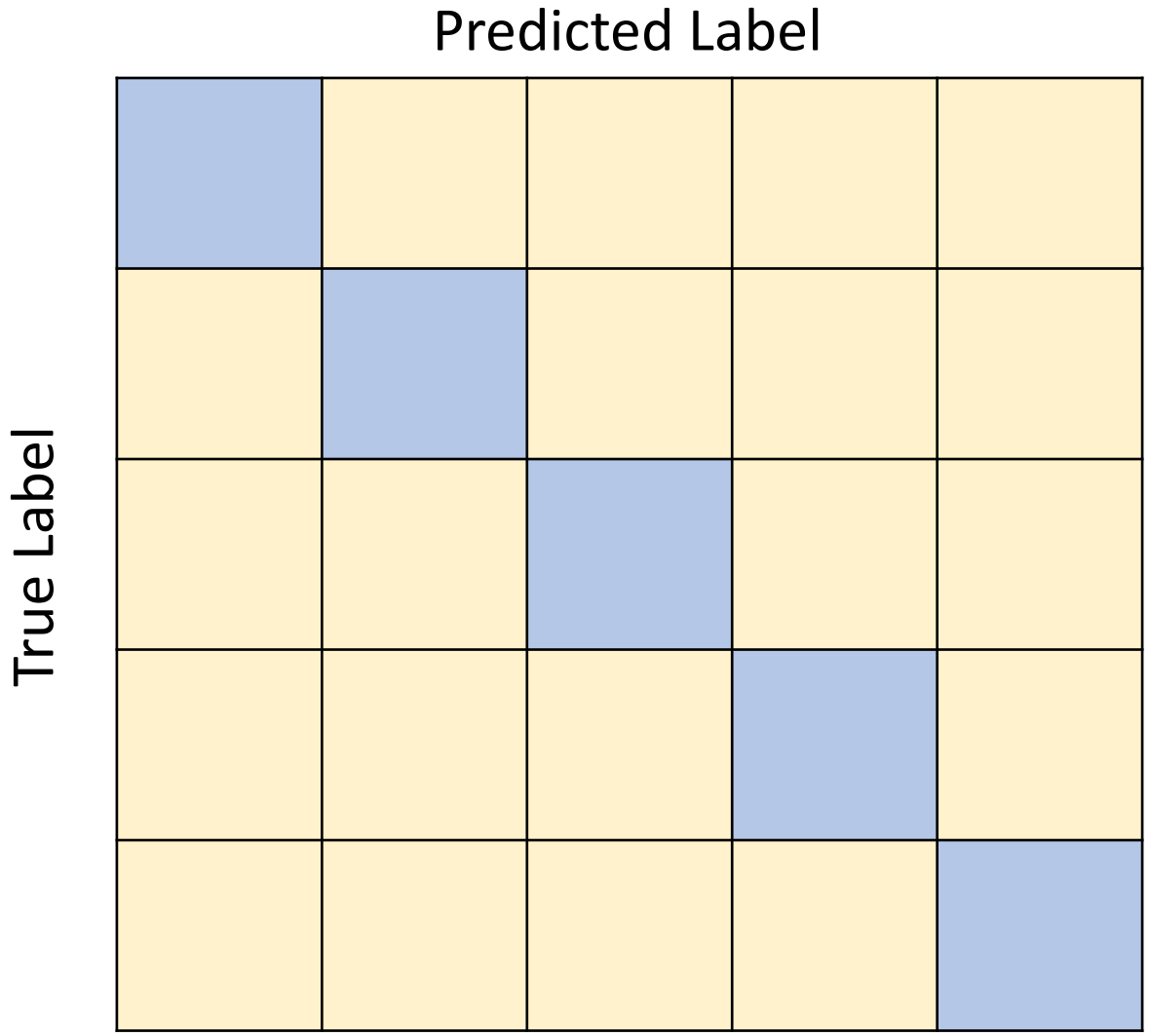
Healthcare Scenarios

1. A computer vision model that detects carcinoma
2. An EHR-based model that surveils autism risk
3. An algorithm that detects COVID in Apple watch users

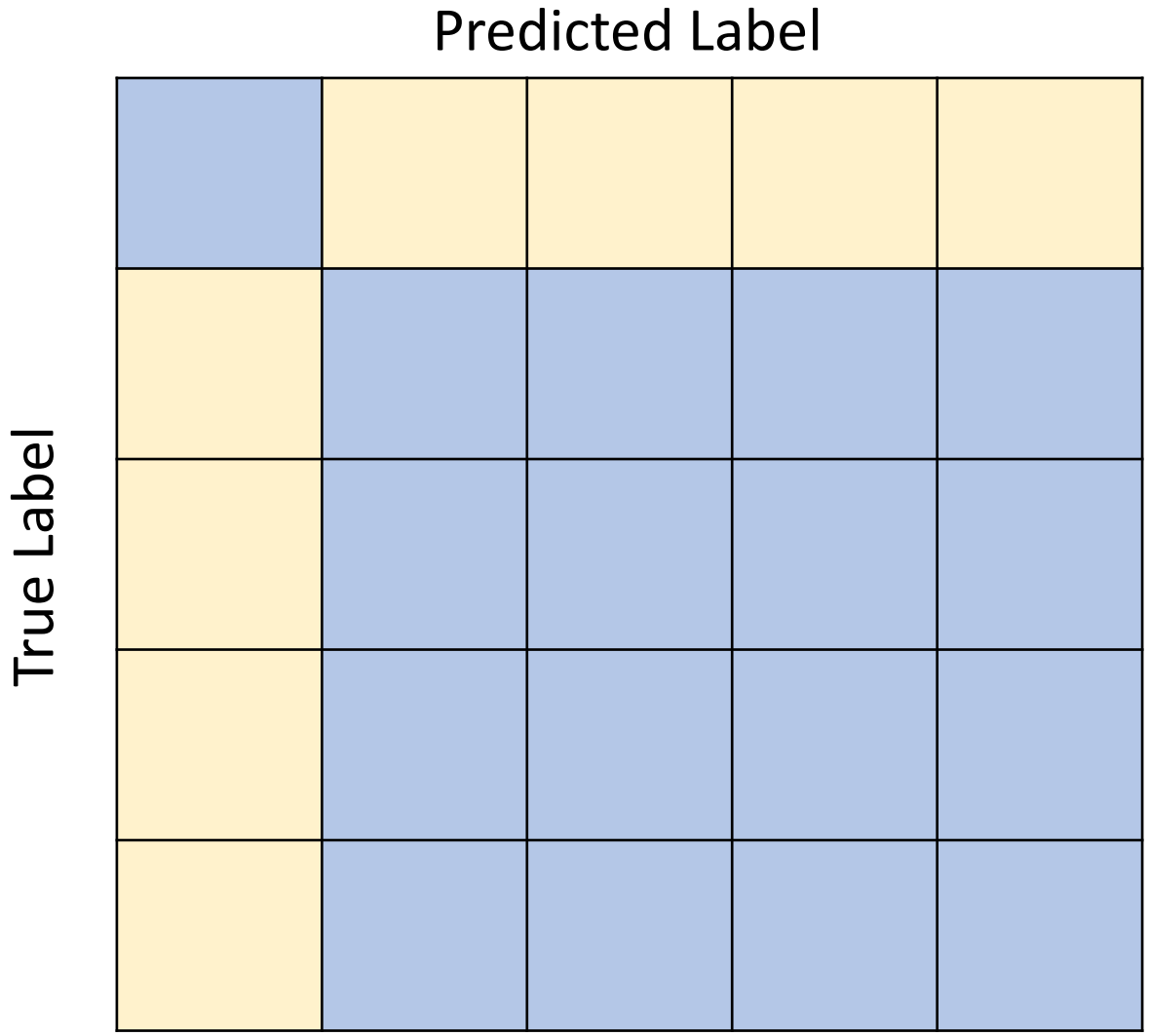
Healthcare Scenarios

1. A computer vision model that detects carcinoma
2. An EHR-based model that surveils autism risk
3. An algorithm that detects COVID in Apple watch users
4. An NLP model that identifies urgent text messages received through a maternal health platform with 2 million users

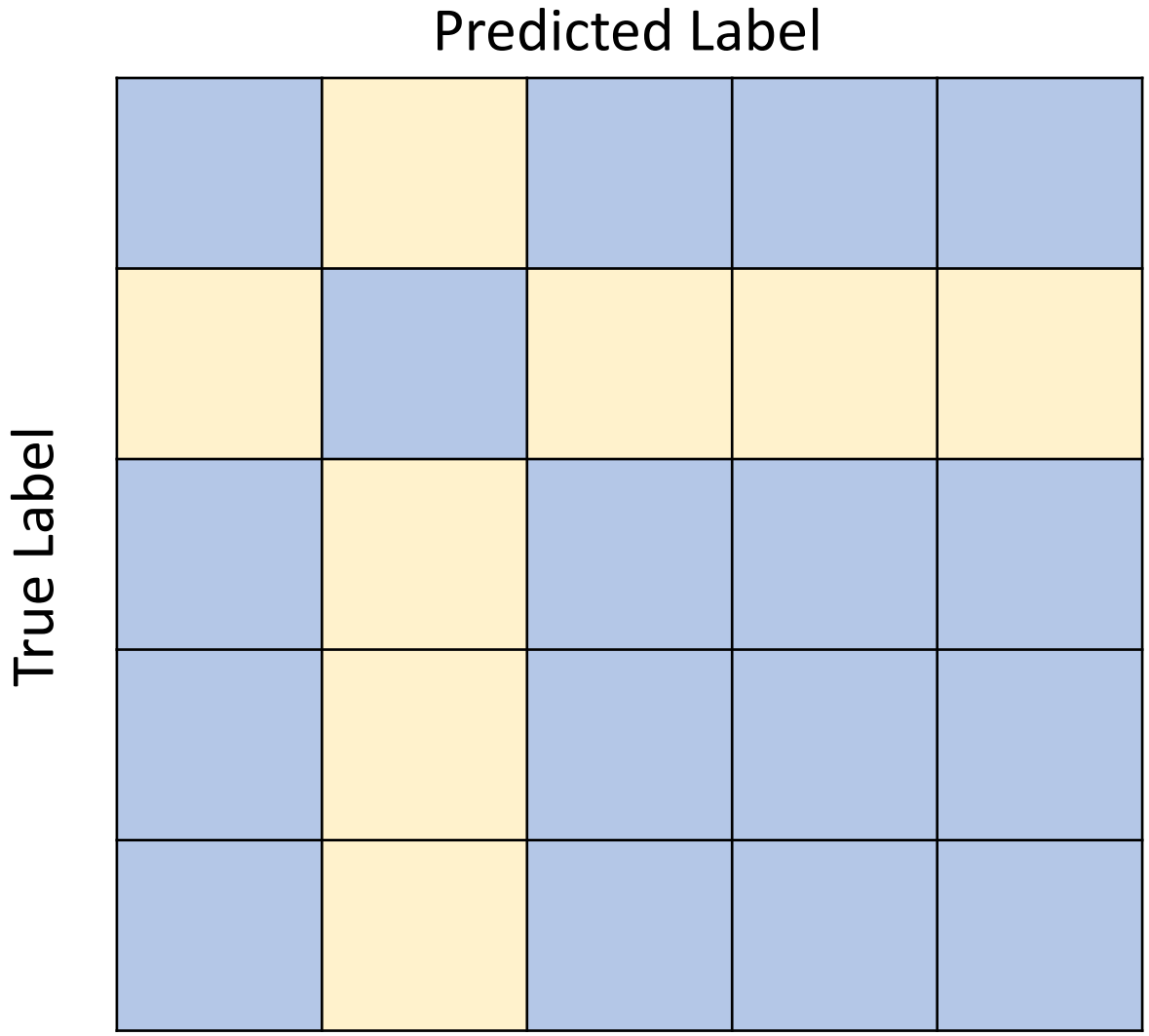
Multi-class problems: “Confusion Matrix”



Multi-class problems: Binary for Label 1



Multi-class problems: Binary for Label 2



There are many more, of course, but classification metrics go a long way.

- Regression
  - Mean squared error (MSE)
  - Mean absolute error (MAE)
  - R^2
- Survival Analysis (i.e. failure time)
  - Concordance index
  - MSE, MAE
  - Brier Score
  - AUC_t