An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

Alberto García-Galindo

agarciagali@unav.es

Universidad de Navarra | DATAI INSTITUTE OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

joint work with Marcos López-De-Castro and Rubén Armañanzas

Institute of Data Science and Artificial Intelligence (DATAI), Universidad de Navarra

12th Symposium on Conformal and Probabilistic Prediction with Applications

Limassol, Cyprus

An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

Neoadjuvant therapy: the gold-standard preoperative treatment for breast cancer patients but...

Neoadjuvant therapy: the gold-standard preoperative treatment for breast cancer patients but...

- Not all patients respond in the same way.

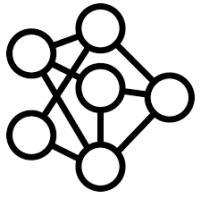
  Only 20 - 30 % achieve complete response.

Neoadjuvant therapy: the gold-standard preoperative treatment for breast cancer patients but...

- Not all patients respond in the same way.
  Only 20 - 30 % achieve complete response.
- Patients experiencing ineffective neoadjuvant therapy incur in toxicity and side effects without reaching the desired clinical benefit.

Neoadjuvant therapy: the gold-standard preoperative treatment for breast cancer patients but...

- Not all patients respond in the same way.
  Only 20 - 30 % achieve complete response.
- Patients experiencing ineffective neoadjuvant therapy incur in toxicity and side effects without reaching the desired clinical benefit.
- **So we need tools to predict how a patient will respond to neoadjuvant therapies.**



Machine learning to the rescue!

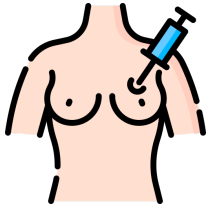
Response to neoadjuvant therapy has been approached through different data modalities

Response to neoadjuvant therapy has been approached through different data modalities

- Dynamic contrast-enhanced MRI features.
  - Provide a non-invase protocol.
  - No additional cost to those patients where MRI is part of their preoperative test.
  - Limited predictive power.

Response to neoadjuvant therapy has been approached through different data modalities

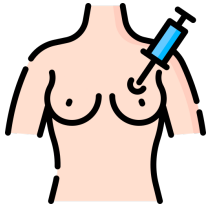
- Dynamic contrast-enhanced MRI features.
  - Provide a non-invase protocol.
  - No additional cost to those patients where MRI is part of their preoperative test.
  - Limited predictive power.



- Clinical and biomolecular predictors.
  - Gathered through invasive biopsy tests.
  - Allow a better understanding of biological processes.
  - Better predictive power.

Response to neoadjuvant therapy has been approached through different data modalities

- Dynamic contrast-enhanced MRI features.
  - Provide a non-invase protocol.
  - No additional cost to those patients where MRI is part of their preoperative test.
  - Limited predictive power.



- Clinical and biomolecular predictors.
  - Gathered through invasive biopsy tests.
  - Allow a better understanding of biological processes.
  - Better predictive power.

- **What if we could efficiently use each feature set?**

  Only in those cases in which the imaging features provide an uncertain prediction, should a biopsy be performed.

An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

Uncertainty-Aware Sequential Approach

Uncertainty-Aware Sequential Approach

Let assume a set of patient triplets

$$(x_1^{\text{MRI}}, x_1^{\text{BIO}}, y_1), \ldots, (x_n^{\text{MRI}}, x_n^{\text{BIO}}, y_n)$$

Uncertainty-Aware Sequential Approach

Let assume a set of patient triplets

$$(x_1^{\text{MRI}}, x_1^{\text{BIO}}, y_1), \ldots, (x_n^{\text{MRI}}, x_n^{\text{BIO}}, y_n)$$

- **Naive approach**

  Learn a predictive model using the whole set of features $f(x^{\text{MRI}}, x^{\text{BIO}})$.

Uncertainty-Aware Sequential Approach

Let assume a set of patient triplets

$$(x_1^{\text{MRI}}, x_1^{\text{BIO}}, y_1), \ldots, (x_n^{\text{MRI}}, x_n^{\text{BIO}}, y_n)$$

- **Naive approach**

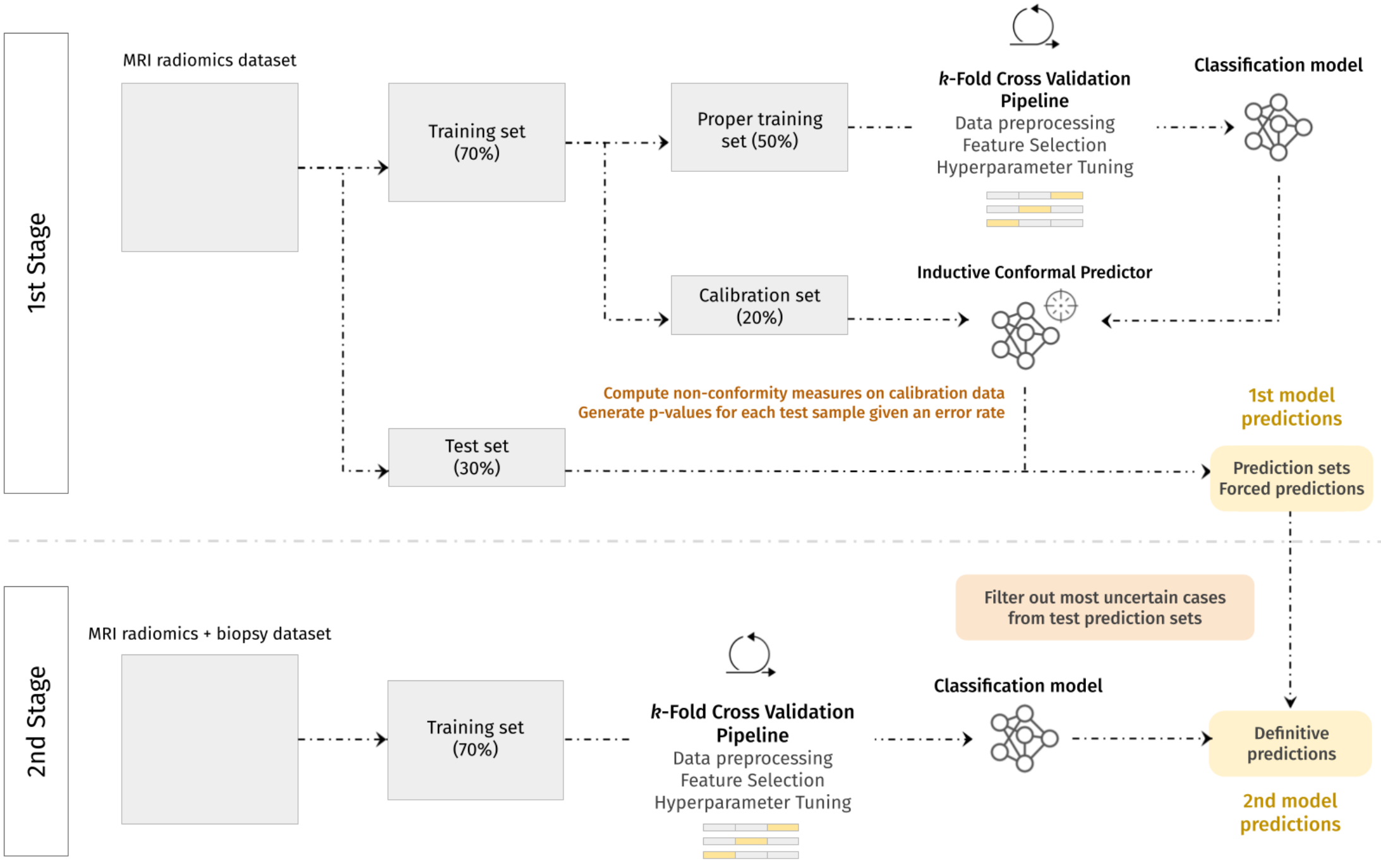
  Learn a predictive model using the whole set of features $f(x^{\text{MRI}}, x^{\text{BIO}})$.

- **Our proposal**

  Learn an **inductive conformal predictor** on top of a non-invasive MRI predictive model.

  If the model is certain enough for an specific patient, compute a prediction using the non-invasive model $f(x^{\text{MRI}})$

  If not, compute a prediction with a biopsy-enriched invasive model $f(x^{\text{MRI}}, x^{\text{BIO}})$.


1st Stage
MRI radiomics dataset
Training set (70%)
Proper training set (50%)
k-Fold Cross Validation Pipeline
Data preprocessing
Feature Selection
Hyperparameter Tuning
Classification model
Calibration set (20%)
Inductive Conformal Predictor
Compute non-conformity measures on calibration data
Generate p-values for each test sample given an error rate
Test set (30%)
1st model predictions
Prediction sets
Forced predictions
2nd Stage
MRI radiomics + biopsy dataset
Training set (70%)
k-Fold Cross Validation Pipeline
Data preprocessing
Feature Selection
Hyperparameter Tuning
Classification model
Filter out most uncertain cases from test prediction sets
Definitive predictions
2nd model predictions

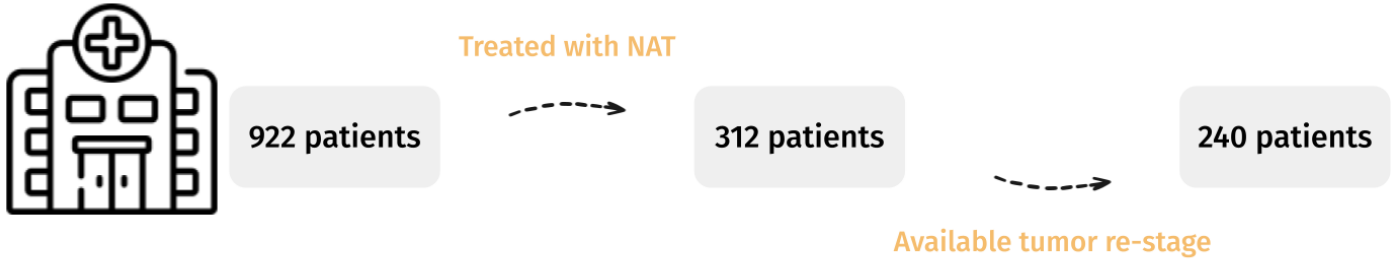
An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

Experimental analysis

Experimental analysis

Duke Breast Cancer MRI dataset

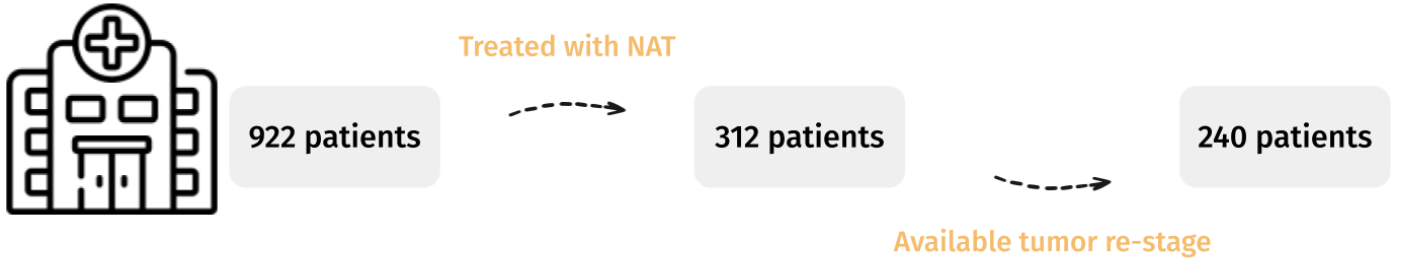
A fully anotated and anonymazed collection of 922 breast cancer patients admitted at Duke University Hospital between January 1st, 2000 and March 23rd, 2014.



Experimental analysis

Duke Breast Cancer MRI dataset

A fully anotated and anonymazed collection of 922 breast cancer patients admitted at Duke University Hospital between January 1st, 2000 and March 23rd, 2014.

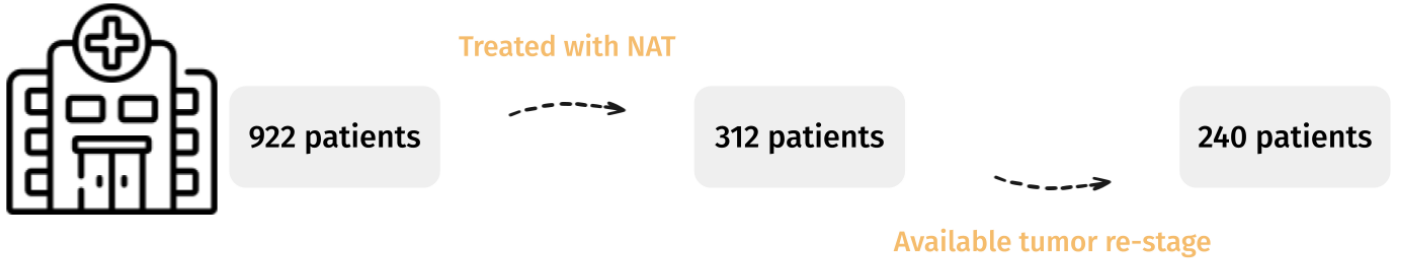


Response to NAT	Cancer re-stage	Sample size
Pathological complete response	$\varnothing$	71 (29.6 %)
Early stage	IA or IIA	104 (44.3 %)
Locally advanced or metastasis stage	From IIB to IV	65 (27.1 %)

Experimental analysis

Duke Breast Cancer MRI dataset

A fully anotated and anonymazed collection of 922 breast cancer patients admitted at Duke University Hospital between January 1st, 2000 and March 23rd, 2014.



Response to NAT	Cancer re-stage	Sample size
Pathological complete response	$\varnothing$	71 (29.6 %)
Early stage	IA or IIA	104 (44.3 %)
Locally advanced or metastasis stage	From IIB to IV	65 (27.1 %)

- **521 MRI numerical features** describing tumor and fibroglandular tissue characteristics.
- **12 clinical features** describing tumor biology from biopsy.

Experimental analysis

Experimental analysis

Classifiers

We tested three diferents algorithms: **logistic regression**, **random forest** and **xgboost**.

Experimental analysis

Classifiers

We tested three diferents algorithms: **logistic regression, random forest** and **xgboost.**

Non-conformity functions

- **Inverse Probability Error :** $\Delta(y, f(x)) = 1 - \hat{\mathbb{P}}\left(y_i | x\right)$
- **Margin Error :** $\Delta(y, f(x)) = 0.5 - \frac{\hat{\mathbb{P}}\left(y_i | x\right) - \max_{y != y_i} \hat{\mathbb{P}}\left(y | x\right)}{2}$

Experimental analysis

Classifiers

We tested three diferents algorithms: **logistic regression**, **random forest** and **xgboost**.

Non-conformity functions

- **Inverse Probability Error :** $\Delta(y, f(x)) = 1 - \hat{\mathbb{P}}\left(y_i | x\right)$
- **Margin Error :** $\Delta(y, f(x)) = 0.5 - \frac{\hat{\mathbb{P}}\left(y_i | x\right) - \max_{y != y_i} \hat{\mathbb{P}}\left(y | x\right)}{2}$

Filtering strategy

Patients with $|\Gamma(x_i)| = 1$ will be retained within the 1st stage.

In order to produce prediction sets, we test two error rates: $\epsilon \in \{0.1, 0.2\}$

Experimental analysis

Classifiers

We tested three diferents algorithms: **logistic regression, random forest** and **xgboost.**

Non-conformity functions

- **Inverse Probability Error :** $\Delta(y, f(x)) = 1 - \hat{\mathbb{P}}\left(y_i | x\right)$
- **Margin Error :** $\Delta(y, f(x)) = 0.5 - \frac{\hat{\mathbb{P}}\left(y_i | x\right) - \max_{y != y_i} \hat{\mathbb{P}}\left(y | x\right)}{2}$

Filtering strategy

Patients with $|\Gamma(x_i)| = 1$ will be retained within the 1st stage.

In order to produce prediction sets, we test two error rates: $\epsilon \in \{0.1, 0.2\}$

Meaningful metrics

- **1st stage (conformal predictor)**

  Single rate (patients assessed through 1st model) : $\frac{1}{N} \sum_i^N (|\Gamma(x_i)| = 1)$
- **Entire pipeline**

  F1 macro (unweighted per-class F1) : $\frac{1}{3} \left(\mathrm{F1}_{\mathrm{CR}} + \mathrm{F1}_{\mathrm{ES}} + \mathrm{F1}_{\mathrm{LA}}\right)$

  True Positive Rate (TPR) for each class.

Experimental analysis

Some key insights...

Experimental analysis

Some key insights...

- F1-macro using a RF model $(\epsilon = 0.1, \text{marginal error})$.

RF_{MRI}	$\text{RF}_{\text{MRI+BIO}}$	Our approach	Single rate
0.408	0.525	0.513	12.6 $\%$

Experimental analysis

Some key insights...

- F1-macro using a RF model $(\epsilon = 0.1, \text{marginal error})$.

RF_{MRI}	$\text{RF}_{\text{MRI+BIO}}$	Our approach	Single rate
0.408	0.525	0.513	12.6 $\%$

- TPR for locally advanced patients using a RF model $(\epsilon = 0.2, \text{marginal error})$.

RF_{MRI}	$\text{RF}_{\text{MRI+BIO}}$	Our approach	Single rate
0.538	0.602	0.618	27.6 $\%$

Even outperforming the model trained with the whole set of features!

Experimental analysis

Some key insights...

- F1-macro using a RF model $(\epsilon = 0.1, \text{marginal error})$.

RF_{MRI}	$\text{RF}_{\text{MRI+BIO}}$	Our approach	Single rate
0.408	0.525	0.513	12.6 %

- TPR for locally advanced patients using a RF model $(\epsilon = 0.2, \text{marginal error})$.

RF_{MRI}	$\text{RF}_{\text{MRI+BIO}}$	Our approach	Single rate
0.538	0.602	0.618	27.6 %

Even outperforming the model trained with the whole set of features!

- TPR for early stage patients using a xgboost model $(\epsilon = 0.2, \text{inverse probability error})$

XGB_{MRI}	$\text{XGB}_{\text{MRI+BIO}}$	Our approach	Single rate
0.461	0.672	0.659	13.7 %

An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

Conclusions

- Machine learning has the potential to assess how a breast cancer patient will respond to neoadjuvant therapies.
- Our conformal prediction-based approach helps identify patients whose prognosis is uncertain using non-invasive protocols.
- These patients are refered to a second assessment with invasive test, providing a more accurate prediction.
- Patients retained within the non-invasive model avoid unnecessary biopsies.

Future work

- Additional non-conformity measures (e.g., ordinal prediction sets).
- Other clinical applications in cost-variable problems.
- Limited data regime $\rightarrow$ cross-conformal prediction.

An Uncertainty-Aware Sequential Approach for Predicting Response to Neoadjuvant Therapy in Breast Cancer

Acknowledgments

Gobierno de Navarra
Nafarroako Gobernua
AGENDA 2030

European Commission
Horizon 2020
European Union funding
for Research & Innovation

ERA PerMed