ESMO Congress 2022, 9-13 September, Paris

Generalization of a deep learning model for HER2 status prediction on H&E-stained whole slide images from 3 neoadjuvant clinical studies

Miriam Hägele, Klaus-Robert Müller, Carsten Denkert, Andreas Schneeweiss, Bruno Sinn, Michael Untch, Marion van Mackelenbergh, Christian Jackisch, Valentina Nekljudova, Thomas Karn, Maximilian Alber, Frederik Marmé, Christian Schem, Elmar Stickeler, Peter A. Fasching, Volkmar Müller, Karsten Weber, Bianca Lederer, Sibylle Loibl, Frederick Klauschen

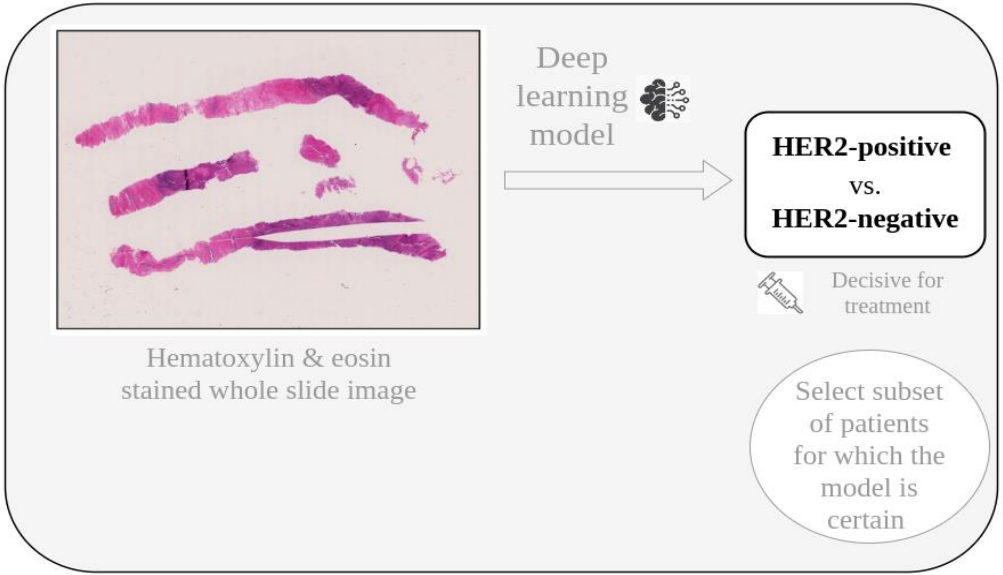
Disclosure

Miriam Hägele

Part-time employment at Aignostics GmbH.

Background and Aim

- Predict HER2 status in breast cancer from routine diagnostic histological slides via deep learning algorithms
- Develop model to generalize across clinical studies



Materials and Methods

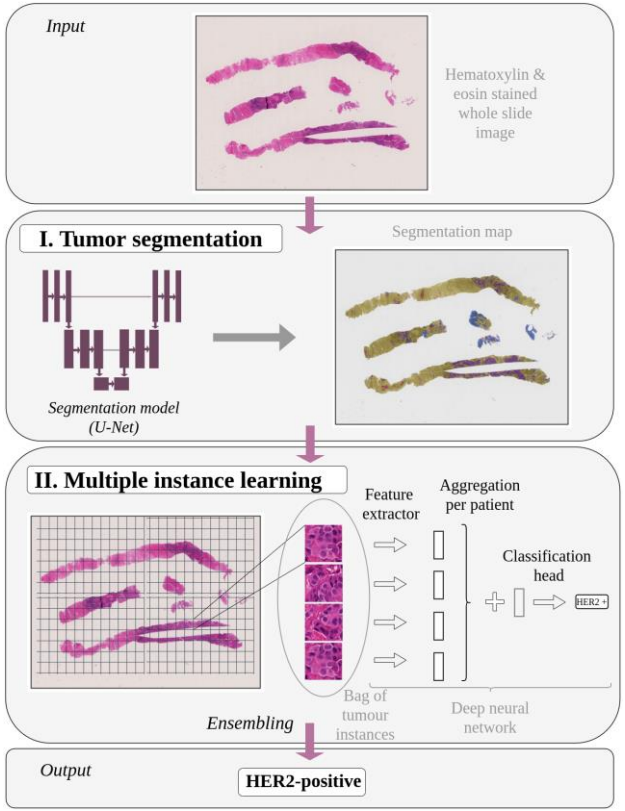
Data

- Three neoadjuvant clinical studies[1,2,3]

Two-step machine learning approach

1. Extract tumor regions: *Segmentation model*
2. Weakly supervised approach: *Multiple instance learning model*

- Selective prediction evaluation



GeparSixto: Von Minckwitz et al. Lancet Oncol (2014).
GeparSepto: Untch et al. Lancet Oncol (2016).
GeparOcto: Schneeweiss et al. Eur J Cancer (2019).

Results

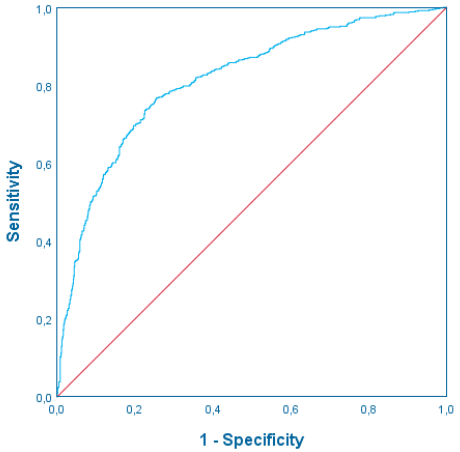
- Validated on 1567 independent patients
- Tested generalization to independent clinical study GeparOcto[3]

	Training	Validation
GeparSixto[1]	205	156
GeparSepto[2]	639	496
GeparOcto[3]	-	915

	ROC AUC (CI 95%)	**Balanced accuracy***
Test cohorts (GeparSixto, GeparSepto, GeparOcto)	81.2 (79.0-83.4)	73.1%
Independent study cohort (GeparOcto)	79.9 (76.9-82.9)	70.4%

Balanced accuracy = (Sensitivity + Specificity) / 2

ROC curve *(overall)*

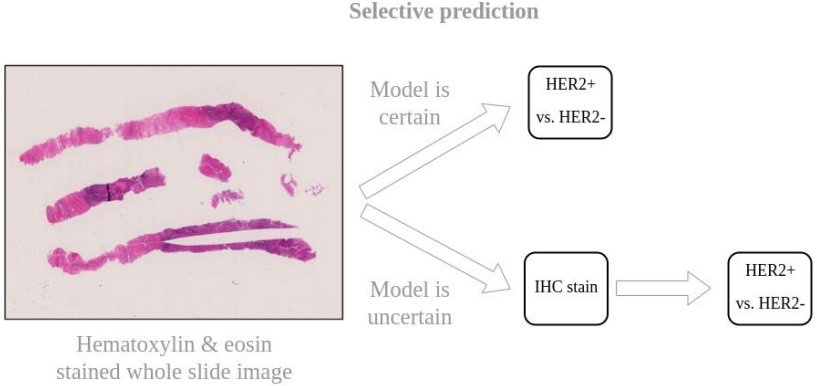


GeparSixto: Von Minckwitz et al. Lancet Oncol (2014).
GeparSepto: Untch et al. Lancet Oncol (2016).
GeparOcto: Schneeweiss et al. Eur J Cancer (2019).

Selective Prediction

- *Concept:* Automatically select cases for which the model is very certain, i.e. define a subset with high predictive performance
- Trade-off between predictive performance and coverage

	Total	Selection with guaranteed risk[1]	Ensemble tail
Coverage	100%	32%	13%
Balanced accuracy	0.73	0.81	0.85



[1] Geifman et al. (2017).

Conclusion

- Our deep learning model predicts HER2 status with state-of-the-art performance
- We validate the performance of our deep learning model on an independent clinical study
- Substantial performance increases can be achieved for subsets of patients based on the model's confidence via selective prediction

Acknowledgments

- Thanks to all Co-Authors.
- Thanks to all cooperating partners: German Breast Group, Charité, TU Berlin and BIFOLD.

This work was funded by the German Ministry for Education and Research as BIFOLD - Berlin Institute for the Foundations of Learning and Data (ref. 01IS18025A and ref. 01IS18037A) and as BBDC (KS 89715122).