10P — AI foundation model for breast cancer characterization and outcome prediction

Antonis Valachis[1], Jan Witowski[2], Joseph Cappadona[2], Ken Zeng[2], Krzysztof J. Geras[1]
[1]Department of Oncology, Örebro University Hospital, Örebro, Sweden; [2]Ataraxis AI, New York, NY, USA

2025 ESMO BREAST CANCER Annual Congress
MUNICH GERMANY
14-17 MAY 2025

Introduction

The advent of Foundation Models (FMs) in artificial intelligence presents a transformative opportunity for oncology. These large-scale models, trained via self-supervised learning on large repositories of unlabeled data, can learn to recognize complex patterns within images and can be easily adapted to various tasks with minimal additional training. FMs excel at tasks with relatively few training data, which is especially important in oncology where obtaining large and diverse labeled datasets is challenging, time-consuming, and expensive.

To demonstrate the potential of FMs in oncology, we train a pan-cancer digital pathology FM and evaluate it on a diverse suite of clinically relevant breast cancer tasks. The presented FM, *Kestrel*, was pretrained using self-supervised learning method DINOv2 on 133 million pathology image patches from 10,000 H&E-stained whole slide images.

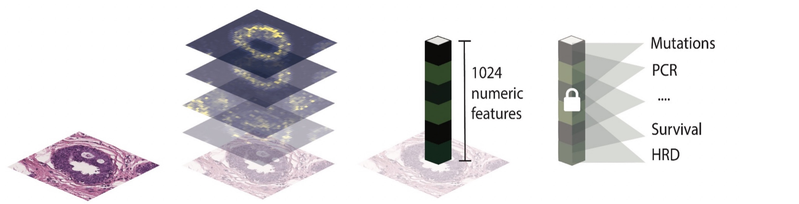
We apply Kestrel to over 30 tasks, including prediction of breast cancer recurrence and survival with a novel AI test called Ataraxis Breast. We also apply the FM to prediction of neoadjuvant treatment response, as well as identification of tissue, tumor, and patient characteristics from a single H&E-stained slide. Kestrel was locked for all evaluations. Generated morphological features were used to train and evaluate downstream models for survival and other tasks.

Conclusions

In this study, we have demonstrated that our AI foundation model can be used to predict recurrence, survival, and infer a wide range of breast cancer characteristics from a single H&E-stained slide, including histological subtypes and clinically relevant gene alterations.
The unique strengths of our approach—leveraging small amounts of supervised data along with large-scale unsupervised pretraining—create significant opportunities for advancements in pathology research and clinical deployment.

AI foundation model and prognostic test development

Foundation model extracted features



1. Image patch from a digitized H&E slide.
2. Kestrel extracts morphological features from the image.
3. The model compresses these features into a single embedding.
4. Kestrel is locked and used to extract embeddings for all prediction tasks.

Figure 1: Our AI foundation model, Kestrel, extracts morphological features from each image and combines them into a single feature embedding.

Integrating foundation model morphological features with clinical data into a multi-modal prognostic score

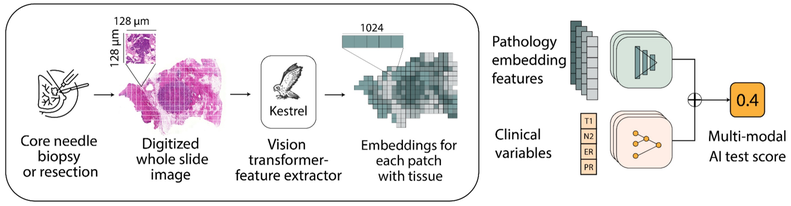


Figure 2: Morphological features extracted by Kestrel are integrated with standard clinical variables, such as tumor size, nodal status, ER, PR, HER2 status.

The multi-modal (Kestrel + clinical data) AI test score, *Ataraxis Breast*, was trained to predict breast cancer recurrence using 4,659 patients across 10 cohorts from six countries.

The external validation sets for Ataraxis Breast included five cohorts (N=3,502):

- Providence (n=1733) and TCGA (n=911) cohorts comprised all invasive breast cancer subtypes
- UChicago (n=421), Basel (n=269), and Karmanos (n=168) cohort included only HR+ HER2- patients previously tested with Oncotype DX as part of the standard diagnostic workflow.

Results

Multi-modal AI test was prognostic of recurrence and survival

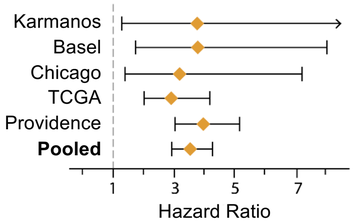


Figure 3: Ataraxis Breast, the multi-modal AI test, was found to be prognostic in all five patient cohorts used as external evaluation sets. Pooled results are computed using random effects model. Hazard ratios are computed for every 0.2 increase in the AI test score.

Breast cancer characterization directly from AI morphological features

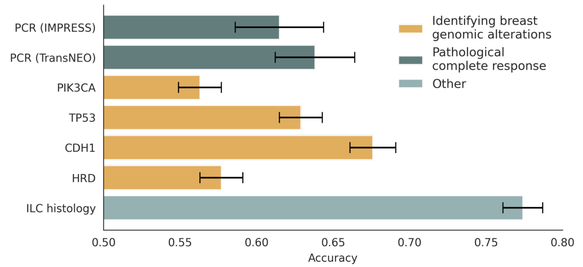


Figure 4: The AI foundation model was applied to prediction of response to neoadjuvant therapy (NAT) and identification of genomic alterations from a single H&E slide.

In predicting NAT response, Kestrel had an out-of-distribution AUC of 0.64 for Cambridge University dataset and 0.63 for the Ohio State dataset.

Kestrel also identified genomic alterations in PIK3CA, TP53, CDH1 genes, as well as homologous recombination deficiency (HRD).

Potential clinical use cases

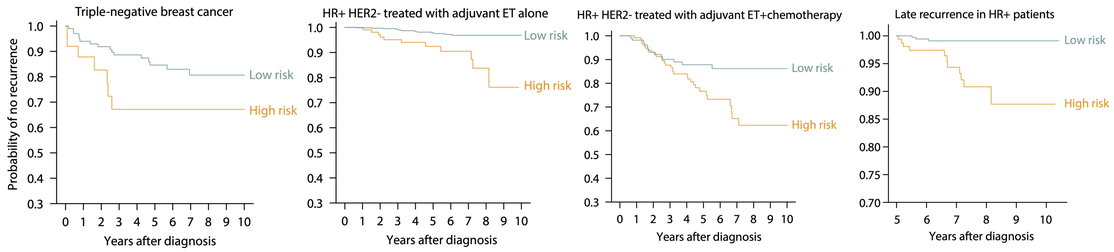


Figure 5: Risk stratification reveals opportunities for personalized treatment across breast cancer subtypes. In TNBC, patients with very low AI risk score may be candidates for KEYNOTE-522 de-escalation. In HR+ HER2-, patients with high AI risk score recur despite ET±CTx, highlighting need for further escalation with e.g., CDK4/6 inhibitors. Finally, among HR+ patients post-5 years of ET, risk stratification identifies candidates for extended endocrine therapy.

Citations

Witowski J. et al. "Multi-modal AI for comprehensive breast cancer prognostication." arXiv preprint arXiv:2410.21256 (2024).

Cappadona J, et al. "Squeezing performance from pathology foundation models with chained hyperparameter searches," in *NeurIPS 2024 Workshop: Self-Supervised Learning - Theory and Practice*, 2024.

Disclosures

A. Valachis has no conflicts of interest to declare.
Study funded by Ataraxis AI, Inc.

Correspondence: Antonios.Valachis@oru.se, jan.witowski@ataraxis.ai