PS3-06-10: Pan-cancer AI foundation models yield accurate biomarker and survival predictions in breast cancer

Joseph Cappadona[1], Ken Zeng[1], Jungkyu Park[1], Bartosz Machura[1], Jan Witowski[1], Krzysztof J. Geras[1]
[1]Ataraxis AI

Introduction

The application of machine learning methods to oncology has historically been challenging due to the high resolution of medical imaging modalities and scarcity of downstream supervised data. Over the past years, however, foundation models have enabled the field to bypass these constraints by leveraging self-supervised learning on large quantities of unsupervised imaging data. These foundation models learn effective latent representations for the various morphologies seen throughout the training data. The representations produced can then be used downstream for supervised machine learning tasks with minimal additional training.

Methods

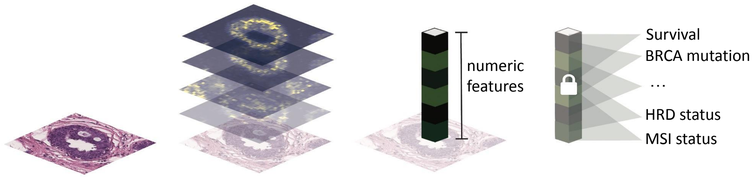
A giant vision transformer (ViT-g; 1.1B parameters) digital pathology foundation model was trained using self-supervised learning method DINOv2 on 250M patches extracted from 260k whole slide images (WSIs). We evaluated the model on 24 biomarker classification and survival analysis tasks across a total of 17 cohorts covering 12 cancer types. For all evaluations, we kept the foundation model frozen and trained regressors or classifiers on top of mean-pooled embeddings obtained by passing hematoxylin and eosin (H&E)-stained slide patches through the model. No clinical data was used for any evaluation. For all tasks, we perform 10 seeds of 5-fold cross-validation and report AUROC for classification tasks and C-index for survival tasks. For classification tasks, we used k-NN (k=20), and for survival tasks, we used the Cox Proportional Hazards model.

Conclusions

Pathology foundation models are applicable to a wide variety of tasks across a range of cancer subtypes, even in spite of sparse data and naive downstream model architectures. With the right data, similar methodology could be used to train predictors of recurrence risk, metastasis risk, treatment benefit, and more.

Notably, we see large improvements in performance over Kestrel, a ViT-Large (303M parameters), suggesting the limits of scaling these models has not yet been reached.

Foundation model extracted features



1. Image patch from a digitized H&E slide.
2. Foundation models extract morphological features from the image.
3. The model compresses these features into a single vector.
4. Embeddings are extracted from each foundation model (frozen) for 17 cohorts and evaluated on 22 tasks.

Figure 1: Each foundation model extracts a patch-level embedding for all patches in each slide. Then, all embeddings from a slide are mean-pooled, and the mean-pooled embeddings from all slides in each cohort are used to evaluate the performance of the foundation model on survival or classification tasks.

Falcon training methodology

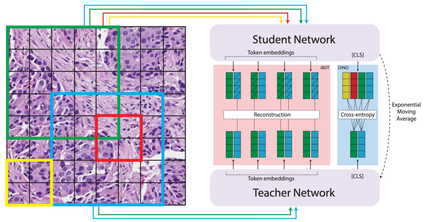


Figure 2: We train Falcon using state-of-the-art self-supervised learning method DINOv2. DINOv2 combines a contrastive loss over the ViT's CLS token (DINO) with a reconstruction loss over the patch tokens (iBOT). The DINO loss is computed over global crops (blue and green) and local crops (red and yellow), while the iBOT loss is computed only over global crops.

CRC-No-Norm	colon	histology (general)
WILDS-C17	lymph node	contains tumor
UniToPatho	colon	histology (adenomas)
WSSS4LUAD	lung	histology (general)
PatchCamelyon	lymph node	metastases
Camel	colon	histology (adenomas)
IDC	breast	histology (IDC)
Gleason	prostate	Gleason score
TCGA-TILs	various (23)	TILs status
Chao-Yang	colon	histology (adenomas)
Osteosarcoma	bone	necrosis
BRACS	breast	histology (lesions)
Kather-MSI-Subtyping	colon/stomach	organ
MHIST	colon	histology (polyps)
CoNIC	colon	cell types
Cytomorphology	bone marrow	cell types
TCGA-Subtyping	various (23)	organ
GlySAC	stomach	cell nuclei
Kather-MSI-CRC	colon	MSI status
BreastPathQ	breast	tumor cellularity
BreakHis	breast	histology (lesions)
Kather-MSI-STAD	stomach	MSI status
IDC Grading	breast	grade of IDC
BCSS	breast	histology (general)

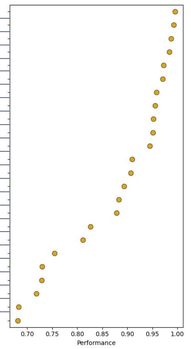


Figure 3: Hyperparameters for Falcon were tuned by optimizing performance on a suite of 24 patch-level histopathology evaluation tasks. Each task tests the model's ability to identify different properties of cancerous tissue across a variety of cancer subtypes, for example, classifying different histologies, identifying the presence of tumor-infiltrating lymphocytes, or predicting biomarker statuses. All tasks are classification tasks, except for BreatPathQ and CoNIC which are regression tasks. Visualized are the performances of the final model. Improvements in performance on these tasks have been found to correlate with improvements on slide-level tasks.

Results and comparison with public models

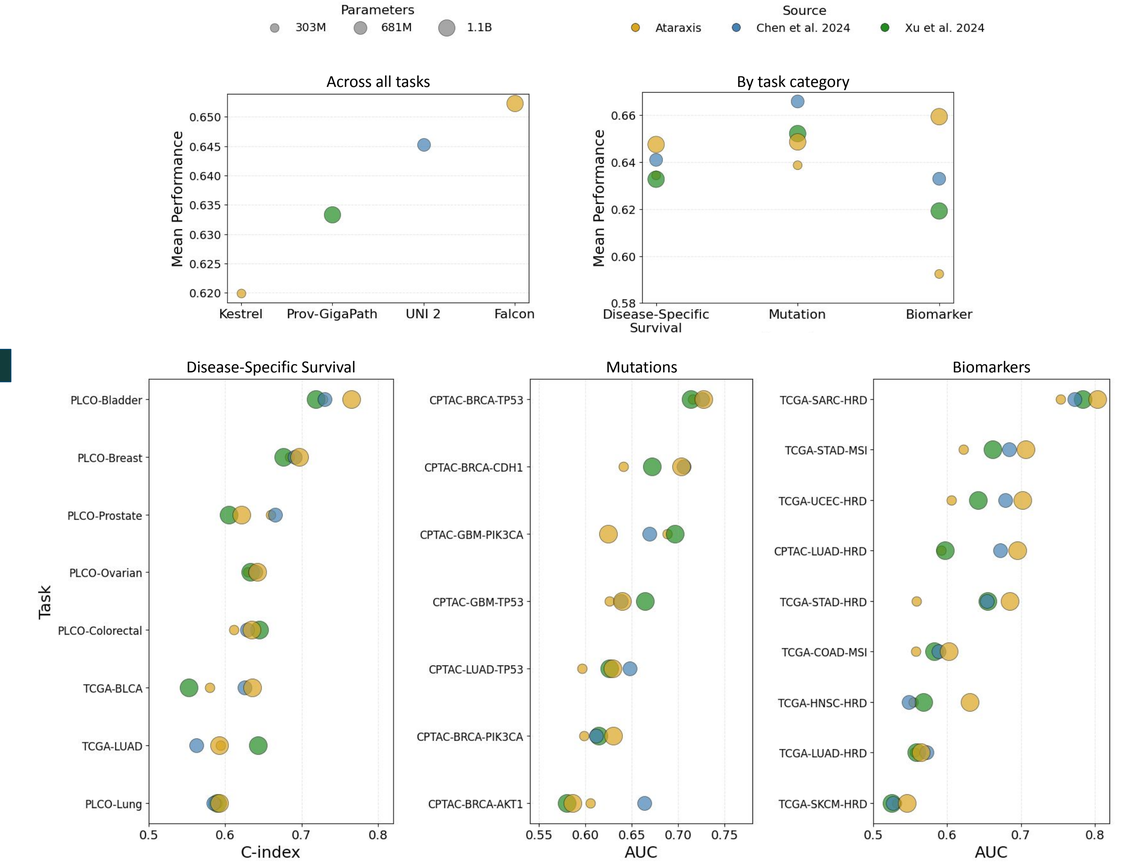


Figure 4: (top left) Model performance aggregated across all 22 tasks. (top right) Model performance broken down by task category. (bottom left) Model performance on disease-specific survival tasks. (bottom middle) Model performance on mutation classification tasks. (bottom right) Model performance on biomarker classification tasks.

This presentation is the intellectual property of the author/presenter. Contact them at Ataraxis AI for permission to reprint and/or distribute.