# Foundation Model Distillation to Develop Robust and Efficient AI Models for MSI Testing in Endometrial Cancer

OWKIN  PartrAlt
OWKIN DIAGNOSTICS  MSD

September 2025

Alexandre Filiot[1], Elin Samuelsson[1], Mélanie Lubrano[1], John Klein[1], Aurore Lyon[1], Corentin Wiscart[1], Arnaud Fouillet[1], Lydwine van Praet[1], Aymeric Vinçotte[1], Catherine Genestie[2], Frederik Deman[3], Luis Cano Ayestas[1], Katharina von Loga[1]

1 Owkin, Paris, France, 2 Gustave Roussy, Department of Pathology, Paris, France, 3 Ziekenhuis aan de stroom (ZAS), Department of Pathology, Antwerp, Belgium

## Background

→ AI models for predicting microsatellite instability (MSI) from H&E slides require diverse, multi-center datasets. This is challenging in cancers with lower MSI prevalence, such as endometrial cancer (EC). While recent foundation models (FM) address data scarcity, they often fail to overcome biases in digital pathology and contribute to high computational costs, limiting clinical applicability.
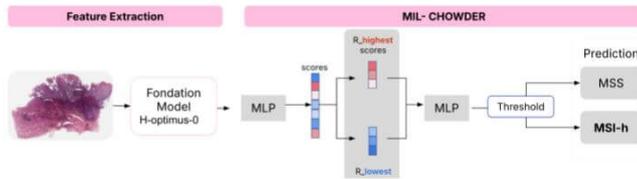
## Objectives

→ This study demonstrates that a distilled version of a large foundation model can offer an efficient, robust AI-assisted diagnostic tool for MSI testing in EC.

## Method

→ **Task**: Predict MSI status from digitized H&E slides in GC and EC
→ **Model**: Feature extraction with Foundation model + MIL (Chowder)



→**Foundation model comparison**: We compared the performance of a large foundation model ("H0") [1] with a 13x-smaller version ("H0-mini"). [2]
→ **Data**: multi-center cohorts

| Indication | Cohort | N. patients (MSI-h) | N WSI | Tissue type |
|---|---|---|---|---|
| EC | TCGA-UCEC | 478 (153) | 559 | Resections |
| EC | ZNA-GZA | 130 (41) | 216 | Resections + biopsies |
| EC | GR | 109 (41) | 109 | Resections |

## Discussion

→ Advanced distillation techniques yield an enhanced, efficient AI model for MSI detection in EC, improving detection capabilities while lowering computational costs. These findings highlight the potential for broader implementation of robust MSI testing in routine clinical practice, aiding streamline diagnosis and targeted therapy in this under-researched cancer population.

## References & acknowledgements

[1] Saillard, C., Jenatton, R., Llinares-López, F., Mariet, Z., Cahané, D., Durand, E., Vert, J.P.: H-optimus-0 (2024), https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0

[2] Filiot, A., Dop, N., Tchita, O., Riou, A., Dubois, R., Peeters, T., ... & Olivier, A. (2025). Distilling foundation models for robust and efficient models in digital pathology. arXiv preprint arXiv:2501.16239.

## Results

→ **Performance**: The H0-mini model achieved an AUROC of 0.769 (0.738, 0.760, 0.810) and specificity of 0.497 (0.451, 0.393, 0.648) at 0.9 sensitivity across three left-out cohorts, outperforming H0 by 0.027 (p=0.03) and 0.291 (p<0.001), respectively.

| | Training on TCGA-UCEC | | Training on GR and ZNA-GZA |
|---|---|---|---|
| Evaluation | GR | ZNA-GZA | TCGA-UCEC |
| H0 | AUC: 0.810 (0.733, 0.883) Sp: 0.493 (0.372, 0.609) | AUC: 0.760 (0.691, 0.827) Sp: 0.300 (0.231, 0.375) | AUC: 0.738 (0.696, 0.780) Sp: 0.363 (0.314, 0.412) |
| Distilled | AUC: **0.842** (0.767, 0.912) Sp: **0.648** (0.526, 0.754) | AUC: **0.768** (0.704, 0.832) Sp: **0.393** (0.316, 0.475) | AUC: **0.761** (0.721, 0.803) Sp: **0.451** (0.404, 0.501) |

**Table: MSI classification metrics on external cohorts for H0 and H0-mini**

→ **Robustness**: H0-mini achieved an intra-class correlation of 0.725 (95% CI: 0.651-0.793) between paired resections and biopsies, compared to 0.665 (95% CI: 0.573-0.750) for H0 (p<0.01).

| Bootstrap over 1000 repeats | Pearson R | CCC (Concordance Correlation Coefficient) | ICC (Intra-Class Correlation) |
|---|---|---|---|
| H0 | 0.668 (0.540, 0.775) | 0.614 (0.482, 0.726) | 0.665 (0.573, 0.750) |
| Distilled | **0.727 (0.625, 0.810)** | **0.718 (0.612, 0.802)** | **0.725 (0.651, 0.793)** |
| t-test p-value from paired bootstrap distributions | 0.043 | 0.003 | 0.009 |

**Table: Concordance between Rx and Bx scores for endo lock model depending on the FM**

→ **Computational cost:** The H0-mini model is 9.1 times faster than H0 on a GPU machine. On standard pathology workstation, processing a WSI takes 15.2 minutes with H0-mini, compared to 120.4 minutes for H0, highlighting the intractability of large FMs without dedicated infrastructure.

| Tiling time in minutes | H0 GPU | Distilled GPU | H0 CPU | Distilled CPU | Ratio GPU averaged over slide | Ratio CPU | Ratio Distilled CPU / H0 GPU |
|---|---|---|---|---|---|---|---|
| 10 TCGA slides (152k tiles in total) | 84 | 9 | – | – | 9.0 (0.8) | – | – |
| 2 TCGA slides (29.5k tiles in total) | – | – | 227 | 35 | – | 6.7 (0.4) | 1.9 (0.5) |

**Table: Tiling time in minutes on GPU and CPU**