1 Supplementary Materials

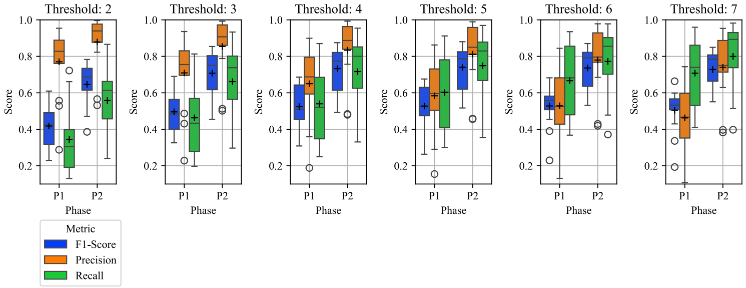


Fig. S1. Comparison of F1-score, precision, and recall for individual raters against the multi-rater consensus. The analysis considers varying thresholds for the number of raters required to agree on a MF in forming a consensus. Higher inter-rater agreement is found for PHH3-assisted annotation (P2) compared to HE-based annotation (P1) for different consensus thresholds.

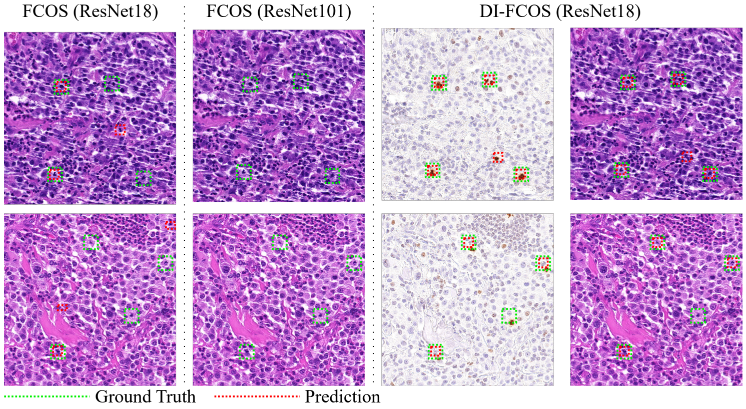


Fig. S2. Qualitative detection examples of the FCOS and DI-FCOS models on the PHH3-assisted labels of the sutdy dataset: Despite imperfect alignment of HE- and PHH3-stained patches (lower row), the DI-FCOS model successfully detected MFs.

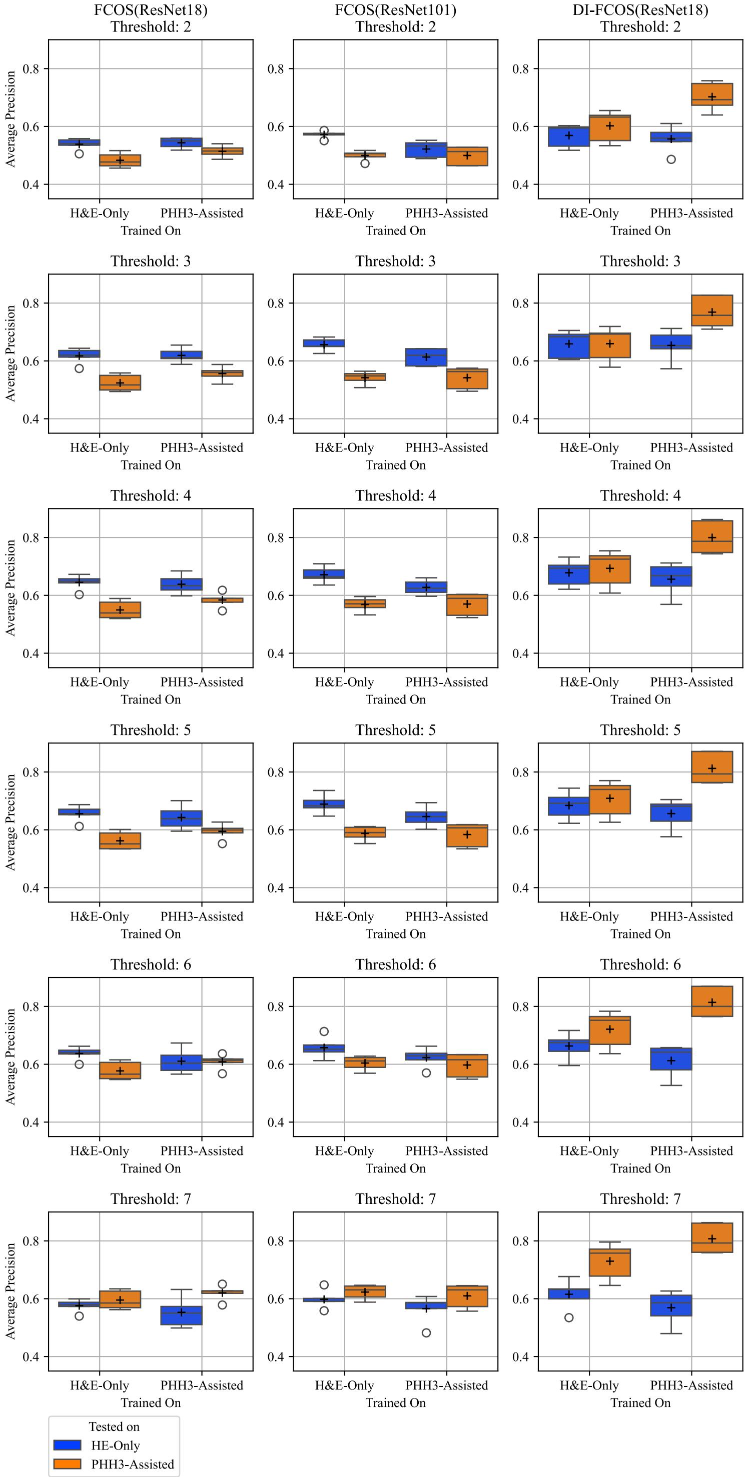
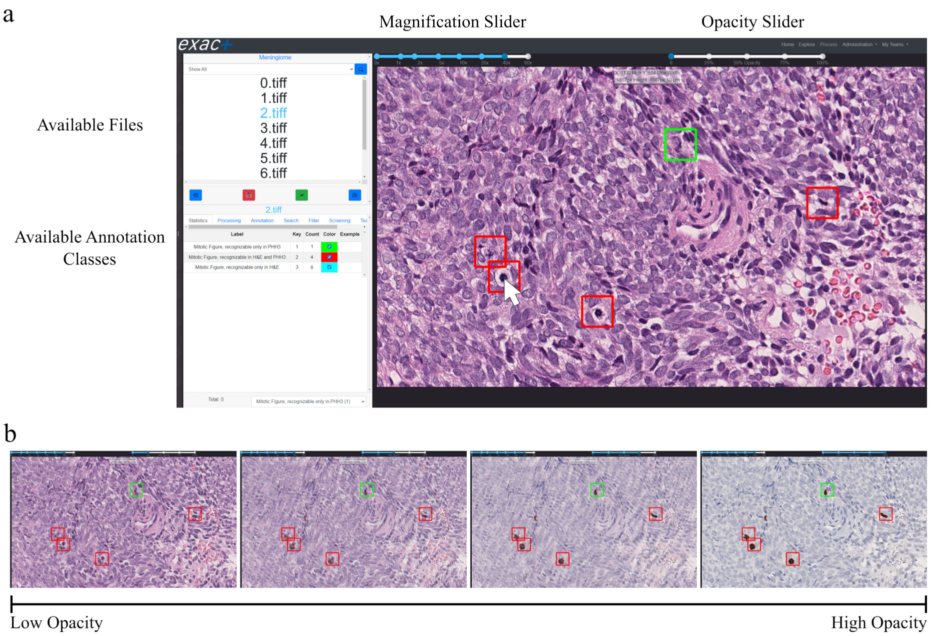


Fig. S3. Results of five-fold cross-validation of the FCOS and DI-FCOS models on the study dataset, as a function of the number of raters necessary to agree upon an object to be counted as a ground truth (GT) MF from two (top) to seven (bottom). For higher thresholds, the performance of the detector drops, as the number of GT MFs drops due to a high omission rate in the H&E annotations, while the detector performance is not altered significantly when using the PHH3-assisted annotation, as the omission rate is strongly reduced and rater agreement increased.



0.tiff
1.tiff
2.tiff
3.tiff
4.tiff
5.tiff
6.tiff

2.tiff

Label	Key	Count	Color	Example
Mitotic Figure, recognizable only in PHH3	1	1	🟩 ☑	
Mitotic Figure, recognizable in H&E and PHH3	2	4	🟥 ☑	
Mitotic Figure, recognizable only in H&E	3	0	☑	

Total: 0

Mitotic Figure, recognizable only in PHH3 (1)

Fig. S4. a The experts were able to view the slides at different magnification levels. In the second phase of the study, they could also display the PHH3-stained slide as a transparent overlay over the H&E-stained slide. In the second phase of the study (P2), different annotation categories were available. Cells could be annotated by clicking on them. The bounding boxes had a default size of 50 by 50 pixels, but the size could be adjusted manually. b Since the registration was not always perfect, there was a slight displacement between the H&E- and the PHH3-stained slides. In this case, the experts were instructed to annotate the position of the respective cell in the H&E-stained slide.