CoNIC Challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting

Supplementary Material

S1. Detailed summary of the challenge algorithms

We received 26 and 24 submissions to the final segmentation and classification and cellular composition leaderboards, respectively. At the time of submission, we required all participants to submit a short paper outlining their approach. These can be viewed by visiting the final test leaderboards at the following web page: https://warwick.ac.uk/conic-challenge. While each technique is described in detail in the provided manuscripts, we also outline a summary of the submitted methods below. Here, we give an overview of the model architecture, the instance segmentation target, the loss function and whether a strategy was used to overcome the class imbalance present in the dataset. In the descriptions below, **SC** denotes segmentation and classification, while **CC** denotes cellular composition.

S1.1 EPFL | StarDist: **SC = 1st, CC = 3rd**. StarDist [1] is based on U-Net [2] and it predicts an object probability map and 64 radial distance maps. The conventional StarDist model does not perform nuclear classification – therefore for this challenge, a second upsampling branch was added to perform semantic segmentation. To deal with the class imbalance, patches that contained minority classes were oversampled during training. Geometric and H&E-based augmentations were used and multiple models were ensembled with test-time augmentation to obtain the final prediction.

S1.2 MDC Berlin | IFP Bern: **SC = 2nd, CC = 9th**. A U-Net style architecture was used with an EfficientNet [3] backbone and two upsampling branches. The first branch performed instance segmentation and the second branch performed semantic segmentation. The instance segmentation branch predicted each pixel to be either: the interior of the nucleus; the nuclear boundary; or the background. In addition, regression of the nuclear centroids was performed as an auxiliary task. For class imbalance, oversampling of patches containing underrepresented nuclear classes was performed in combination with utilisation of a weighted focal loss. Geometric, blur, noise and H&E-based augmentations were used and multiple models were ensembled.

S1.3 Pathology AI: **SC = 3rd, CC = 1st**. A HoVer-Net [4] was used with a SE-ResNeXt101 [5] backbone and heavy dropout layers [6] in the upsampling branches. Despite describing the concept of diagonal distance maps in their method description paper, standard horizontal and vertical distance maps [4] were used in the final submission. A combination of dice and weighted cross entropy loss was used to help overcome the class imbalance. Geometric, blur and colour augmentations were used during training. The final model was trained on several splits of the data and the results ensembled for submission.

S1.4 LSLL000UD: **SC = 4th, CC = 6th**. A HoVer-Net [4] with a DenseNet-121 [7] backbone was used for instance segmentation, but without the classification branch for simultaneous prediction. Diagonal distance maps were used to improve the performance. A second lightweight U-Net [2] was used to perform boundary refinement, which takes probability maps cropped at each nucleus as input. Then, a devoted network for pixel-wise nuclear classification was used with the same base architecture as the HoVer-Net for instance segmentation. A combination of standard cross entropy and dice loss were used to deal with the class imbalance. Geometric, blur, noise and colour augmentations were performed during training and test-time augmentations were used to obtain the final result.

S1.5 AI_medical: **SC = 5th, CC = 2nd**. A HoVer-Net [4] was utilised with an SE-ResNeXt50 [5] backbone and a Coordinate Attention module [8] in the decoder. Conventional horizontal and vertical distance maps were used to perform instance segmentation. To counter the class imbalance, the submission utilised a both dice and weighted cross entropy loss in the classification branch. Geometric and colour augmentations were used during optimisation. For the final submission several models were ensembled and test-time augmentation were used.

S1.6 Arontier: **SC = 6th, CC = 7th**. A HoVer-Net [4] was used, with skip connections inspired by U-Net++ [9] and an EfficientNet [3] backbone. An interesting approach was used for dealing with the class imbalance, consisting of copy and paste augmentation [10]. Geometric, blur, noise, Cutout [11], Cutmix [12] and colour augmentations were used and an ensemble of 5 models trained on different data splits was considered for the final submission.

S1.7 CIA Group: **SC = 7th, CC = 4th**. An ensemble of a conventional HoVer-Net [4] and a Cascaded Mask-RCNN [13], both with ResNeXt-152 [14] backbones, was used for the challenge. However, despite this strategy being used during the preliminary submission phase, it exceeded the maximum 60-minute allotted time during the final submission phase. Therefore, the final submission comprised of the Cascaded Mask-RCNN by itself. No method was documented for dealing with the class imbalance in the dataset. Geometric, blur and noise augmentations were used during training and model ensembling was performed when making the submission.

S1.8 MAIIA: **SC = 8th, CC = 17th**. A StarDist [1] model with a U-Net [2] architecture was implemented but using more convolutional filters than the standard approach. Here, the off-the-shelf StarDist repository was used to see how it performed with minimal modification. The model predicted the star convex polygons for each nucleus by outputting an object probability map, along with 32 radial distance maps.

Fairly conventional augmentations, consisting of geometric transformations and additive noise were used. No specific strategy was utilised for dealing with the class imbalance and no form of ensembling was performed.

S1.9 ciscNet: **SC = 9th, CC = 11th**. A regular U-Net [2] was used, but with group normalisation [15] and mish activation [16]. The normalised Euclidean distance maps of nuclear pixels to their nearest boundary was predicted to enable instance segmentation. For this, a separate distance map was considered per nuclear type to enable simultaneous classification. To counter the class imbalance, a weighted summation of the per-class regression losses was utilised, where more weight was given to minority classes. Geometric, blur, noise and colour augmentations were used during training and test-time augmentation was performed to yield the final result.

S1.10 MBZUAI_CoNIC: **SC = 10th, CC = 8th**. A HoVer-Net [4] with a ConvNeXt-Small [17] backbone was used with standard horizontal and vertical distance maps as the instance segmentation target. With an aim of learning more discriminative features, each image was converted to various colour spaces and concatenated with the original RGB image before input to the network. A unified focal loss [18] was used during training, which aimed to counter the class imbalance. Geometric, noise and blur augmentations were used during training.

S1.11 Denominator: **SC = 11th, CC = 10th**. Similar to above, HoVer-Net [4] with a ConvNeXt-Tiny [17] backbone was used. Separation of the Haematoxylin and eosin stains was performed before input to the network. A combination of focal [19] and dice loss was used to help combat the imbalance of classes in the data. No model ensembling was used during submission of the algorithm.

S1.12 Softsensor_Group: **SC = 12th, CC = 5th**. A fusion of HoVer-Net [4] and Triple U-Net [20] was used that considered both the original RGB image and the Haematoxylin stain channel as input. Each input was processed by a separate encoder, which were then fused using a progressive dense feature aggregation block. Following HoVer-Net, the model predicted the horizontal and vertical maps, binary nuclear segmentation map, and the multi-class semantic segmentation map. All RGB input patches used Reinhard normalisation [21] to combat differences in the stain appearance and geometric augmentations were introduced during training.

S1.13 BMS_LAB: **SC = 13th, CC = 12th**. A Swin-Transformer [22] with a Hybrid Task Cascade model [23] was used. The model did not use a strategy to deal with the class imbalance. Geometric augmentations were performed and Macenko stain normalisation [24] was used to help reduce the variability of the image appearance across the dataset. For submission, input images were resized to five different scales before processing and the results were then merged together.

S1.14 GDPH_HC: **SC = 14th, CC = 13th**. HoVer-Net [4] was used without any modification to the original architecture. To enhance the available data for training, a generative adversarial network [25] was used to create synthetic images as an augmentation strategy. In addition, a self-supervised technique called RestainNet [26] was used to perform stain normalisation and geometric transformations were applied to all input images. To help train with the presence imbalanced data, a class-weighted loss function was incorporated at the output of the classification branch.

S1.15 conic-challenge-inescteam: **SC = 15th**. CenterNet [27] was used, which is a probabilistic two-stage object detection model. This model allows the reduction of proposals from the Region Proposal Network (RPN), which could be important in this application where each image has many objects. Like Mask-RCNN [28], the original CenterNet approach was extended so that is also produced a segmentation mask for each nucleus. No specific method was used to deal with class imbalance and geometric augmentation was used during model training.

S1.16 Aman: **SC = 16th**. A subtly modified HoVer-Net [4] model was used for the challenge, where major focus given to the data preprocessing step. Copy and paste augmentation [10] of neutrophil and eosinophil nuclei was utilised in addition to performing geometric augmentation of the images. Also, a transformation of the colour space of images was applied to increase the variability of the stain appearance in the training set. Following this, weighted cross entropy and weighted Dice loss functions were used to help counter the class imbalance in the dataset.

S1.17 Bin: **SC = 17th, CC = 19th**. A HoVer-Net [4] approach was used, but each convolution was swapped with a multiple filter block. Here, multiple filter sizes were utilised in parallel during each operation and the results were merged. This was repeated throughout the network. A combination of cross entropy and Dice loss were used, like in the original HoVer-Net implementation and no augmentation was performed.

S1.18 DH-Goods: **SC = 18th, CC = 16th**. Two separate HoVer-Nets [4] with the same architecture were used that aimed to tackle the class imbalance present in the dataset – one that considered epithelial, lymphocyte and connective tissue cell classes, and the other that considered plasma cell, neutrophil and eosinophil classes. The intuition was that separating out the minority classes may lead to better performance. Each HoVer-Net used a HR-Net backbone [29] with an Atrous Spatial Pyramid Pooling (ASPP) unit [30] after the encoder. In addition, a YOLOv5 [31] was trained for nuclear detection and classification, where a U-Net model was used to generate the segmentation masks within the bounding boxes. For tackling the class imbalance, equalised focal loss was used during optimisation of YOLOv5 and mosaic augmentation was used as a way of introducing underrepresented classes into input images. HoVer-Net and YOLOv5 results were then merged using a custom strategy. Geometric transformations of input images were performed and test-time augmentation used to generate the final submission.

S1.19 VNIT: **SC = 19th, CC = 22nd**. A hybrid approach was implemented incorporating handcrafted features, such as local binary patterns and histogram of oriented gradients, into a HoVer-Net [4] model. Specifically, handcrafted features were combined with the deep features after passing input images through the encoder and are then upsampled via three separate upsampling branches, in the same way as the original

HoVer-Net approach. The same loss strategy as the original implementation was used and so no proposed technique was used to deal with the class imbalance. Blur augmentation and colour jitter was used during training.

S1.20 Sk: **SC = 20th, CC = 8th**. An Eff-UNet [32] was used, which combines the effectiveness of EfficientNet [3] as the encoder with U-Net [2] as the decoder. The approach outputs two prediction maps: 1) direction map for instance segmentation and 2) semantic segmentation map for classification. The direction map divides each nucleus into N segments around the centroid. For this submission, N was set to be 4 and therefore divided each nucleus into quadrants. Each quadrant was then treated as a separate class to predict and instance segmentation was performed using a purpose-built post-processing pipeline. Colour, geometric and blur augmentation was used during training. A combination of cross entropy and dice loss was used for the semantic segmentation map, which may partly help alleviate the difficulty in dealing with the class imbalance. Rather than using the segmentation and classification output to predict cellular composition, a separate branch was added to the encoder that directly regressed the nuclear counts.

S1.21 Jiffy Labs and CET CV Lab: **SC = 22nd, CC = 15th**. A HoVer-Net [4] with a ConvNeXt [17] backbone was used for the challenge submission. To help deal with the class imbalance, a combination of Dice and focal loss [19] was used.

S1.22 TIA Warwick: **SC = 23rd, CC = 14th**. ALBRT [33] with an Xception [34] backbone was used for directly predicting the cellular composition from the input image, without performing nuclear segmentation. As opposed to the original approach that used a ranking loss, a Huber loss was used, that aimed to directly maximise the R^2 score. Due to the difficulty in predicting underrepresented classes, a separate network was trained for predicting the counts of eosinophils. Then, each network was trained multiple times and the per-class nuclear counts averaged for the final submission. A standard HoVer-Net [4] model was trained for the segmentation and classification task on a single split of the data.

S1.23 QuIIL: **CC = 23rd**. A YOLOv5 [31] with a Cross Stage Partial Network [35] backbone was used to perform the task of nuclear detection and classification, where the results were then utilised to perform cellular composition. Geometric, colour and mosaic augmentations were used during training.

S2. Hyperparameter search for downstream tasks

We performed a random search over the XGBoost hyperparameters for each downstream clinical task to select the best model. This search space is defined as in Table S1.

S3. Downstream results

The details are provided in Tables S2 to S8.

S4. Complete feature description for downstream tasks

The details are provided in Table S9.

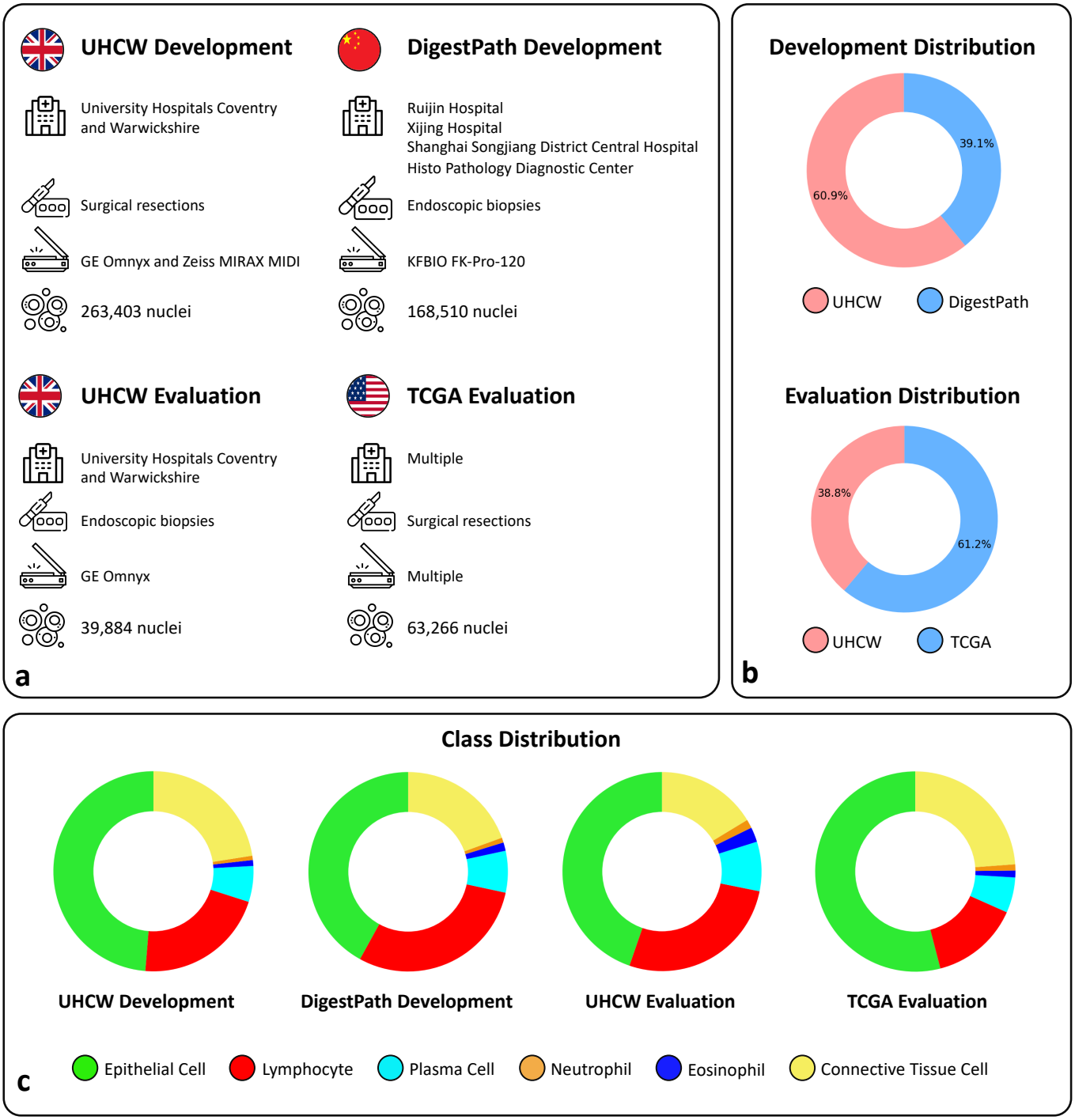


Fig. S1. Summary of the datasets used in the challenge. a, Information regarding the data source, specimen type, scanner manufacturer and number of labelled nuclei. **b**, Distribution of data in the development and evaluation sets. **c**, Distribution of nuclear classes across the different data subsets.

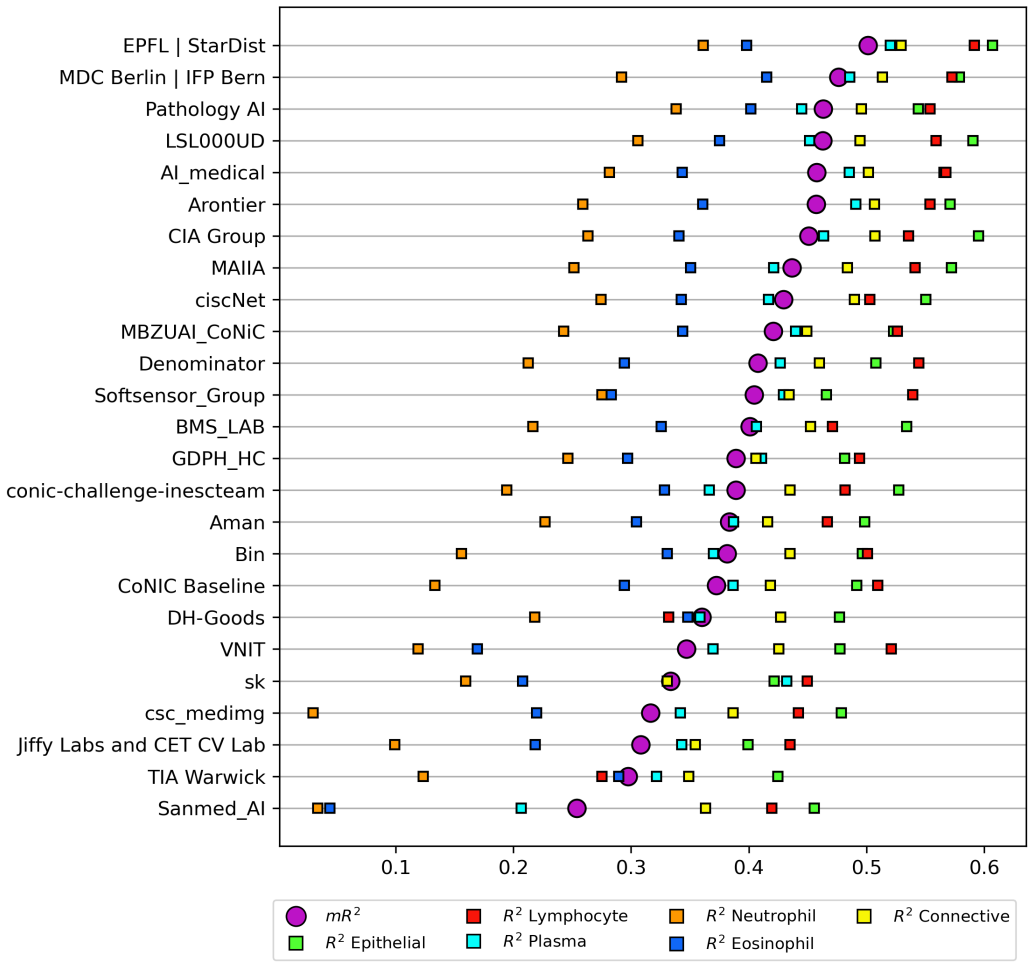


Fig. S2. Segmentation and classification results on the final test set. These results are the same as provided in Fig. 3 of the main manuscript, but are shown as a point plot as an alternative form of visualisation.

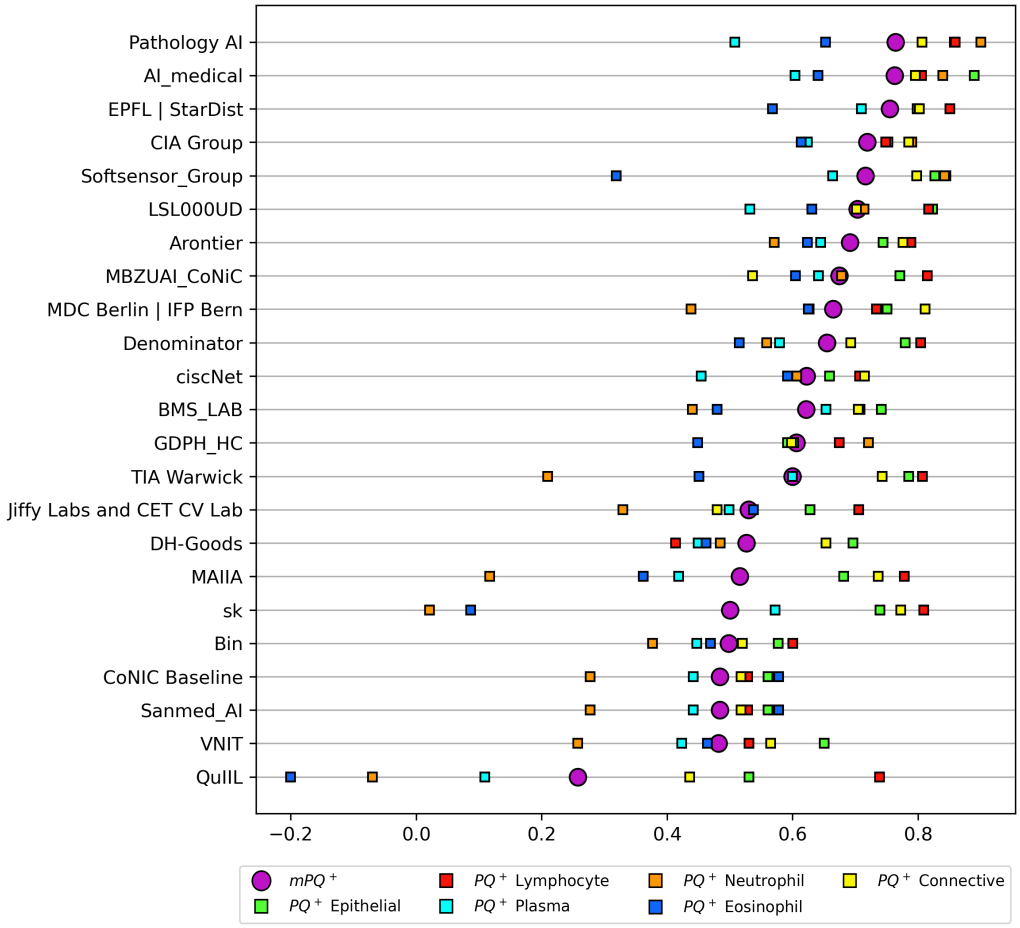


Fig. S3. Cellular composition results on the final test set. These results are the same as provided in Fig. 4 of the main manuscript, but are shown as a point plot as an alternative form of visualisation

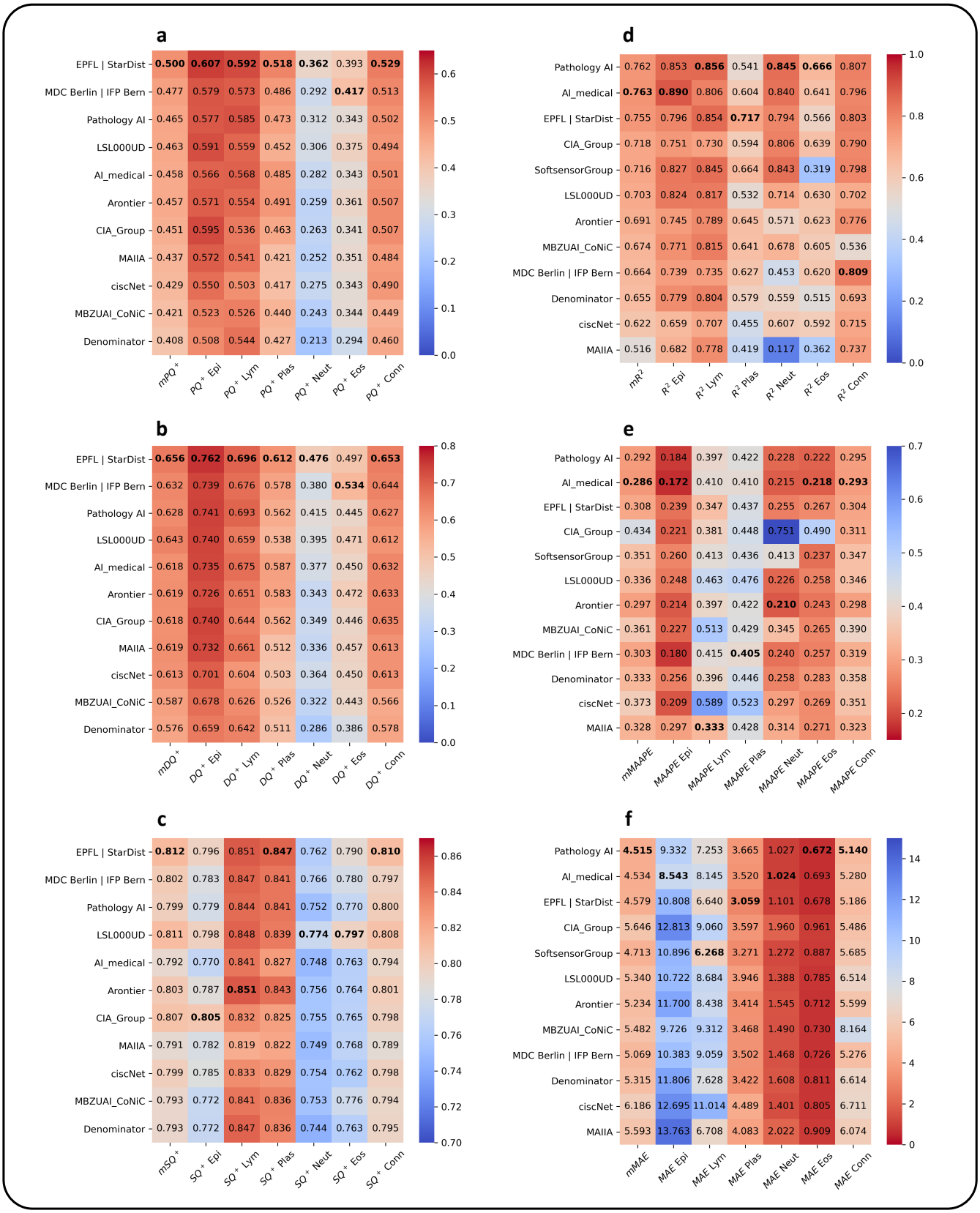


Fig. S4. Additional results for each task using alternative metrics. We computed these results using the algorithms that were sent by the participants, which explains slight differences in the results compared to the original standings. The left side (**a**, **b**, **c**) show segmentation and classification results. **a**, mPQ^{+}, **b**, mDQ^{+} and **c**, mSQ^{+}. The right side (**d**, **e**, **f**) show cellular composition results. **d**, mR^2, **e**, $mMAE$ and **f**, $mMAAPE$.

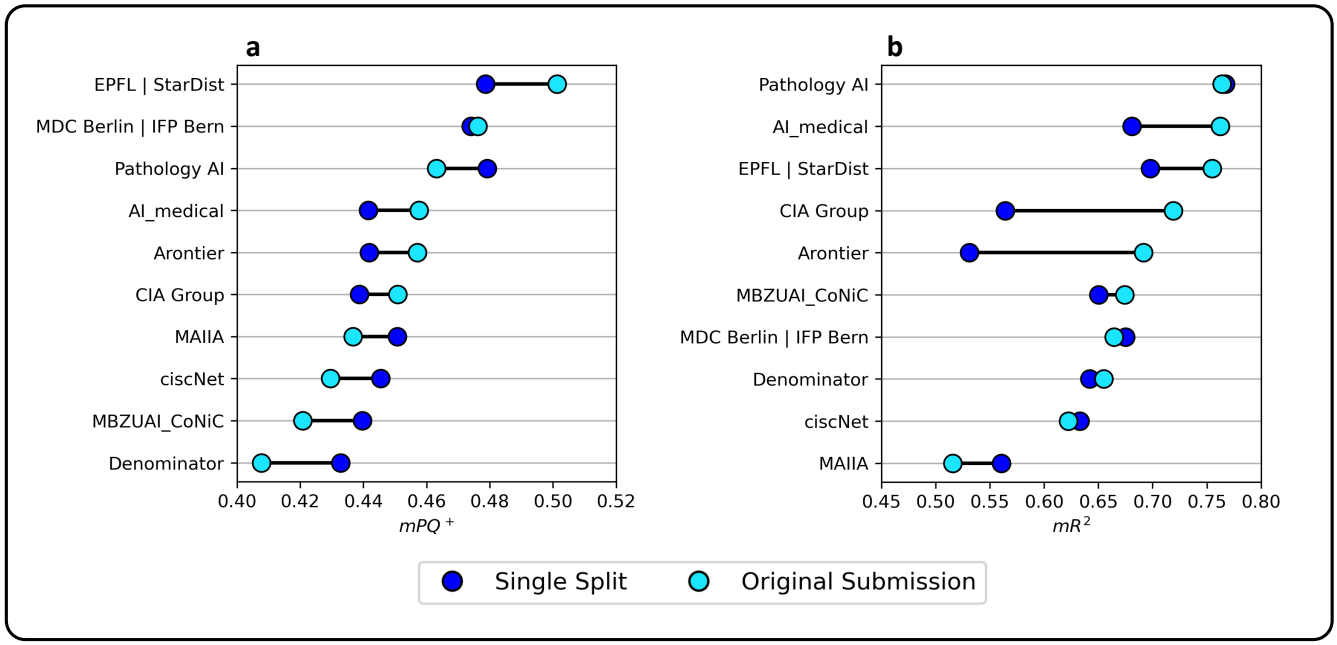


Fig. S5. Difference in results of original submission compared to those obtained using the models trained on a single split of the data and without ensembling. a, Segmenation and classification results, in terms of mPQ^{+} and **b**, cellular composition results, in terms of mR^2.

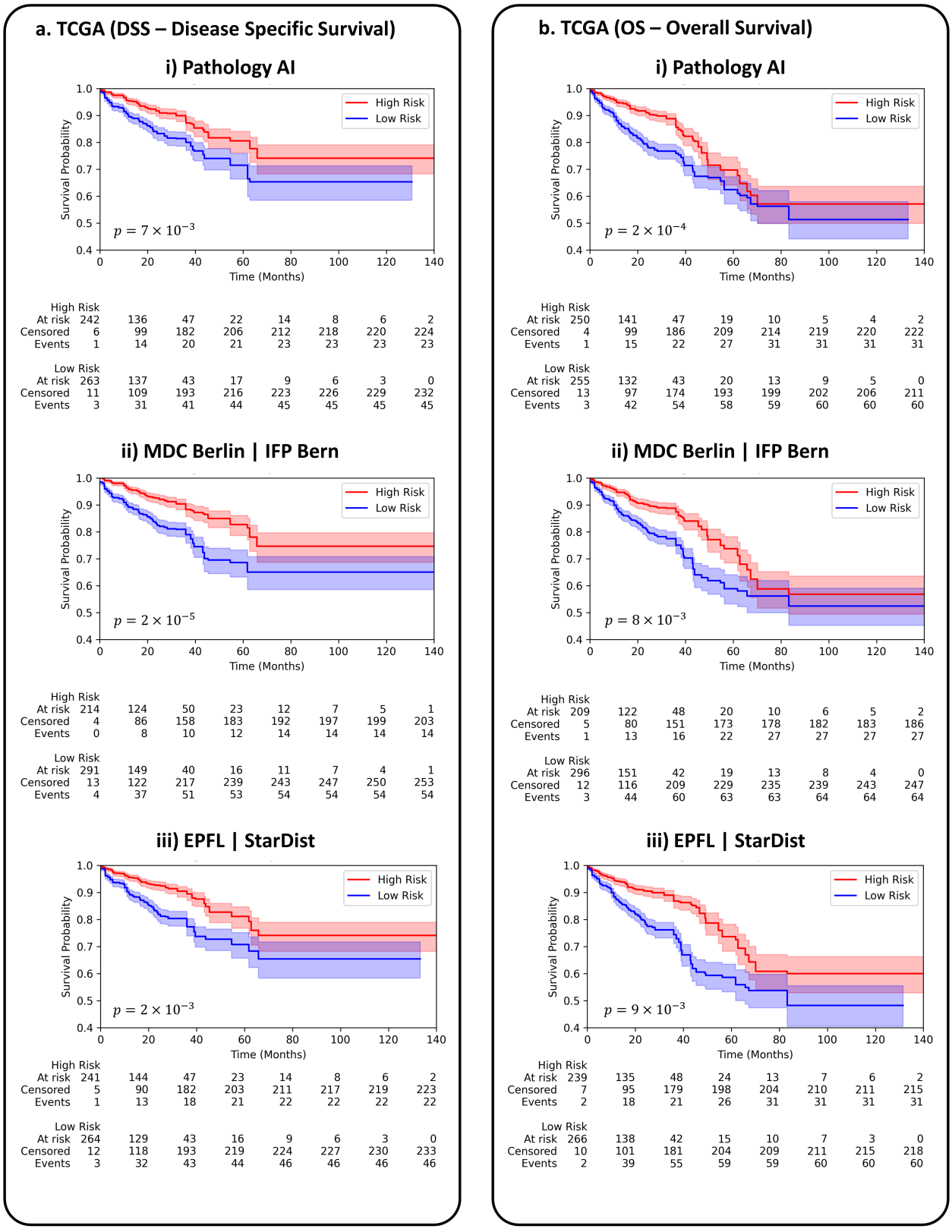


Fig. S6. Kaplan–Meier curves and statistical tests for survival analysis on TCGA. Risk scores on the testing portions from each model were aggregated and utilised to stratify patients into high-risk group and low-risk groups. The threshold criteria is the median value of risk scores obtained from the validation portion. Log-rank tests were conducted and reported to evaluate the degree of separation between two populations. **a**, Disease Specific Survival and **b** Overall Survival.

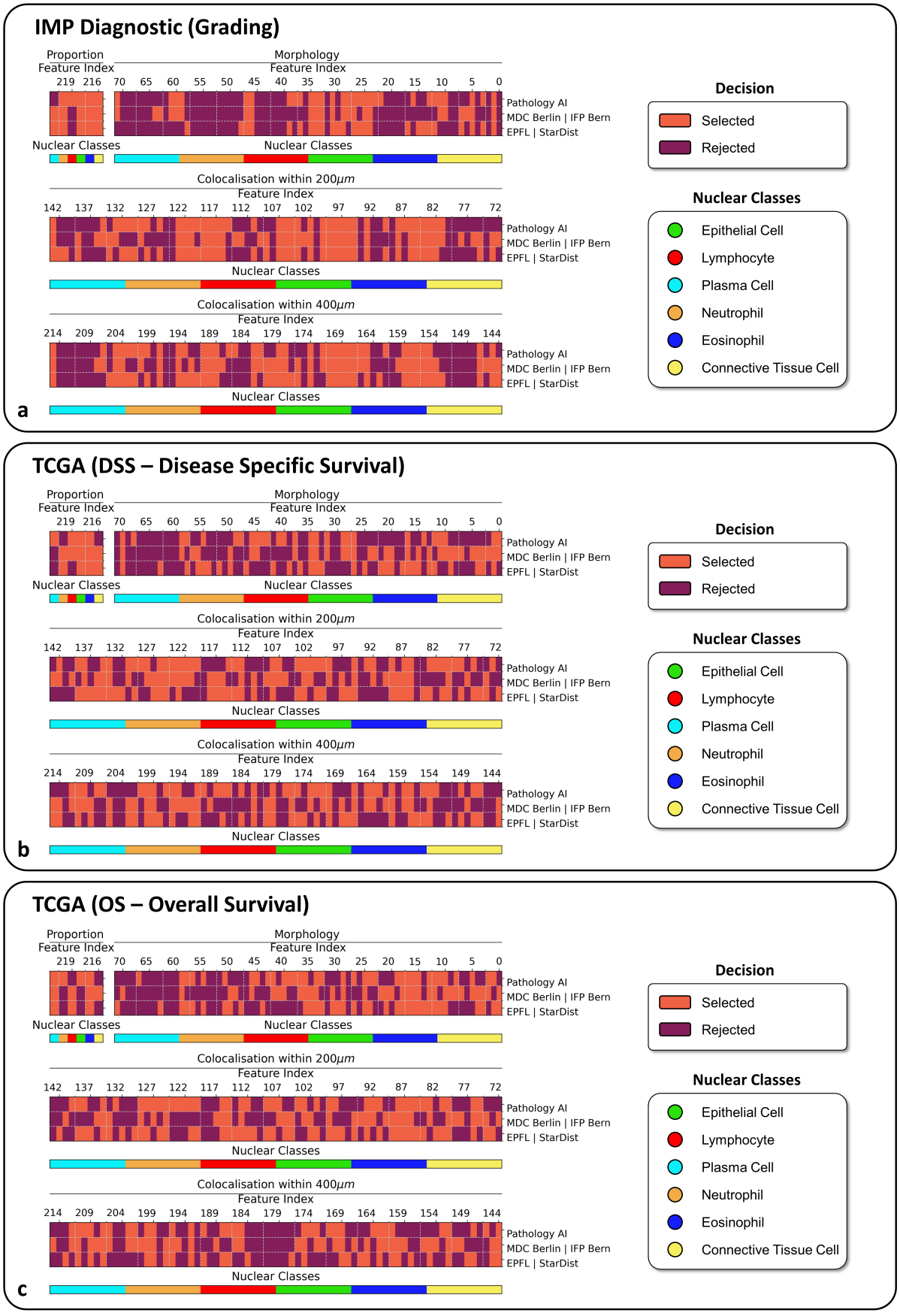


Fig. S7. Summary of which features (out of a possible 222) were utilised for subsequent analyses on grading, disease specific survival or overall survival tasks. **a**, **b** and **c** show the selected features for dysplasia grading, disease specific survival analysis and overall survival analysis, respectively.

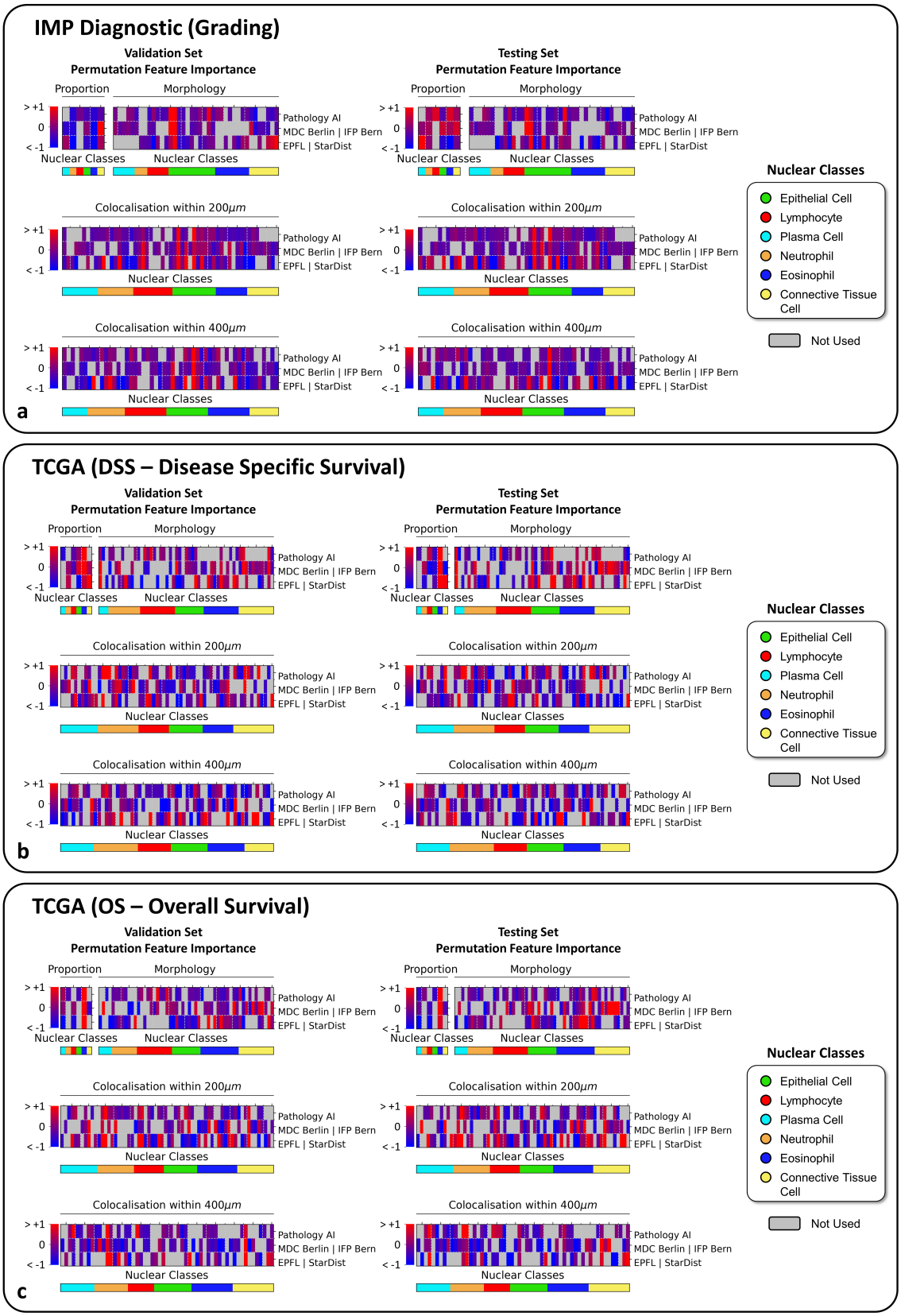
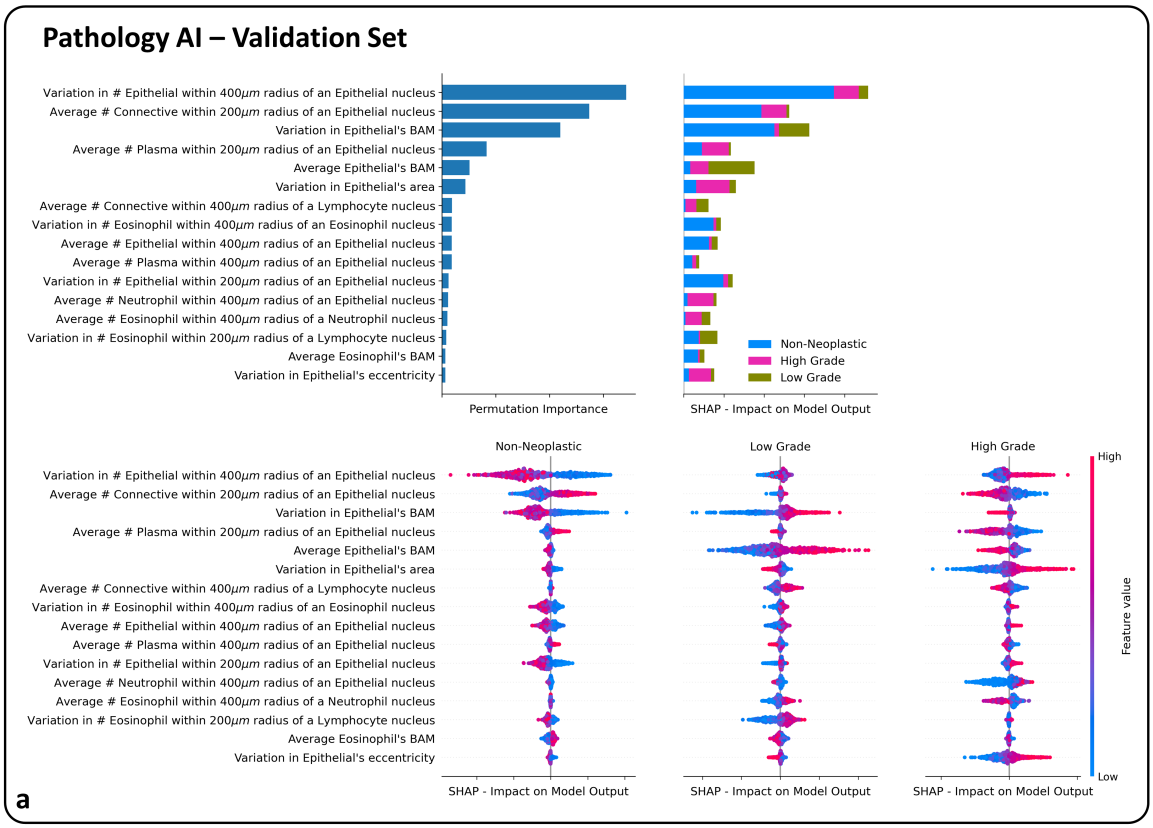


Fig. S8. Summary of the overall importance of each selected feature in Fig. S7 from each team for each task. The importance is measured by a permutation test and combined across all data splits. A feature is important if the changes in its value heavily impact the *evaluation results* (C-Index for survival tasks and QWK for grading). Features that are important to the final XGBoost model performance are coloured in red. On the other hand, a feature being blue indicates a gain in performance when its value becomes more noisy (undesirable or significantly less important). Grey cells show features not selected by this team, but are selected by others.



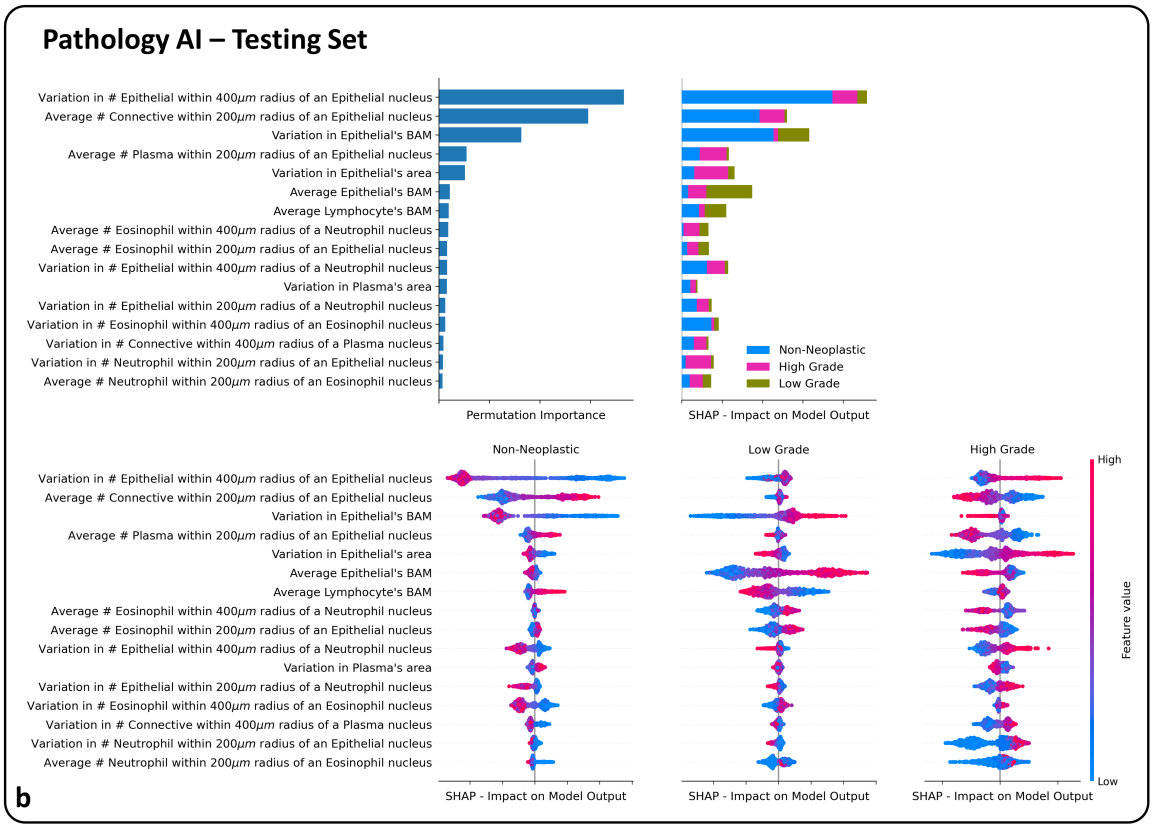


Fig. S9. Contribution of the top 16 features from Pathology AI (taken from Fig. S8) for the grading task on IMP Diagnostics dataset. The first column reports the permutation importance of the feature on the evaluation results (QWK). On the other hand, other columns (SHAP) reflect how the changes in the feature value affect the model predictions (the predicted probabilities of each class by XGBoost). The reported importances are combined across all data splits.

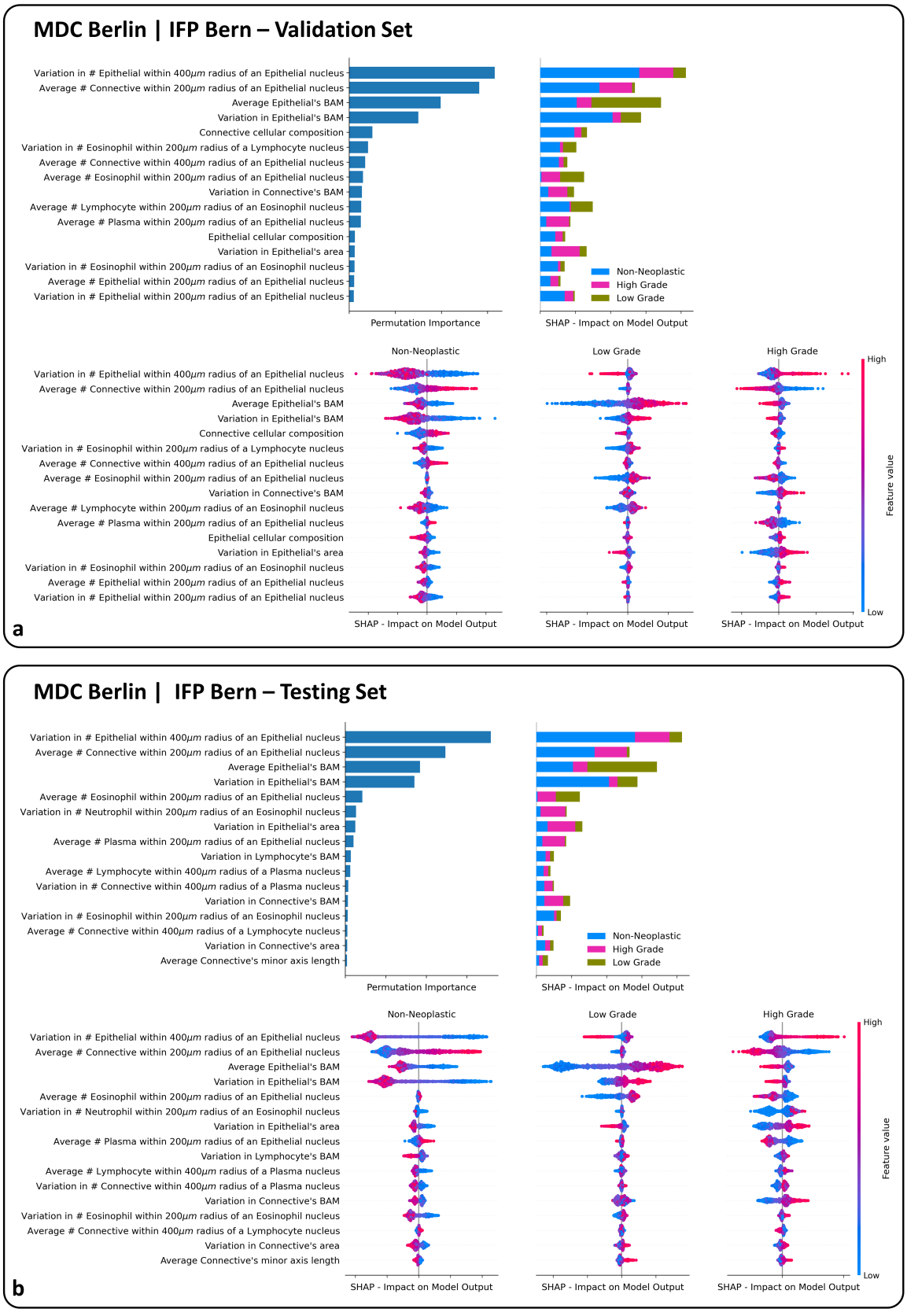
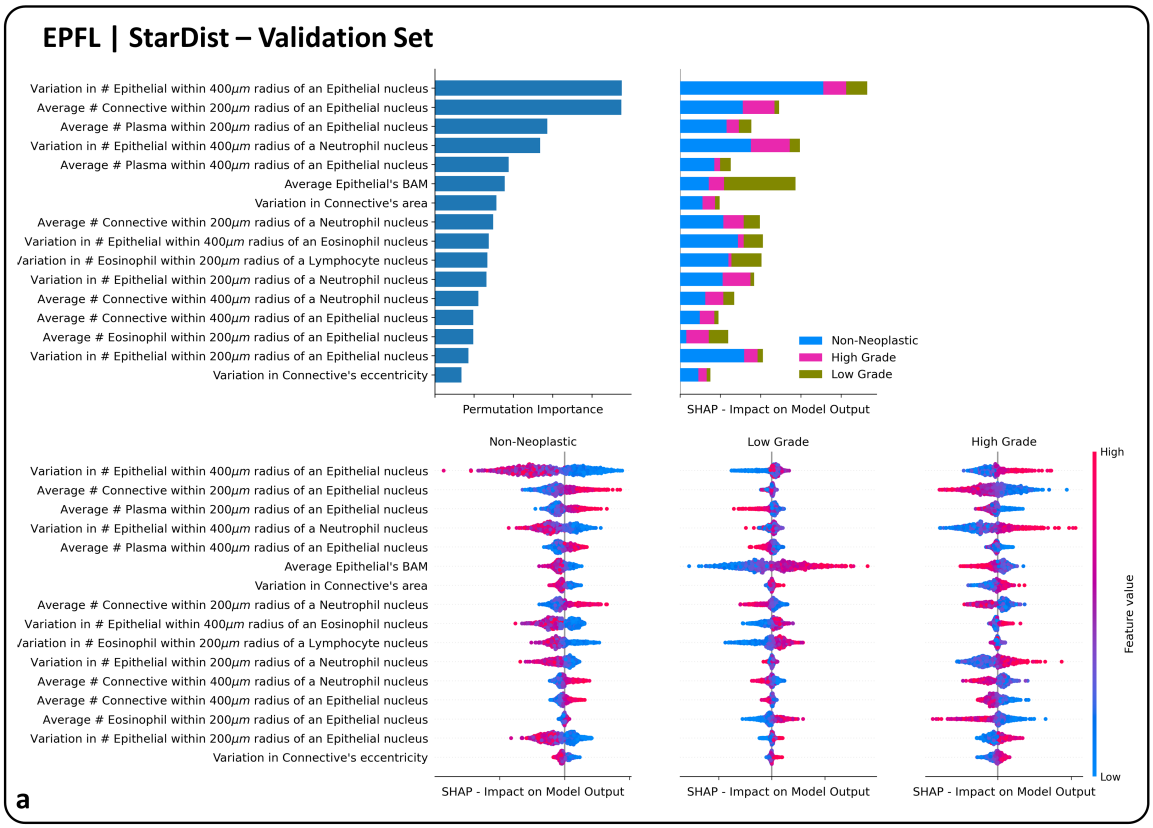


Fig. S10. Contribution of the top 16 features from MDC Berlin | IFP Bern (taken from Fig. S8) for the grading task on IMP Diagnostics dataset. The first column reports the permutation importance of the feature on the evaluation results (QWK). On the other hand, other columns (SHAP) reflect how the changes in the feature value affect the model predictions (the predicted probabilities of each class by XGBoost). The reported importances are combined across all data splits.



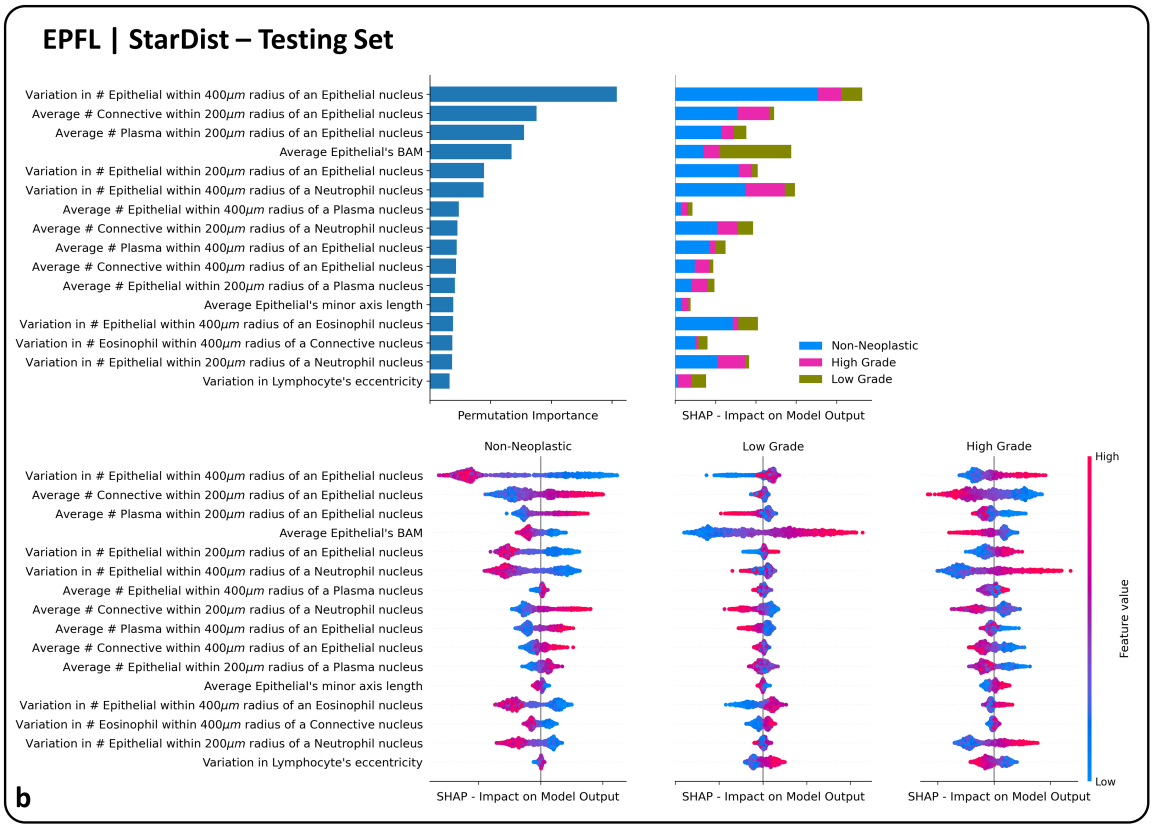
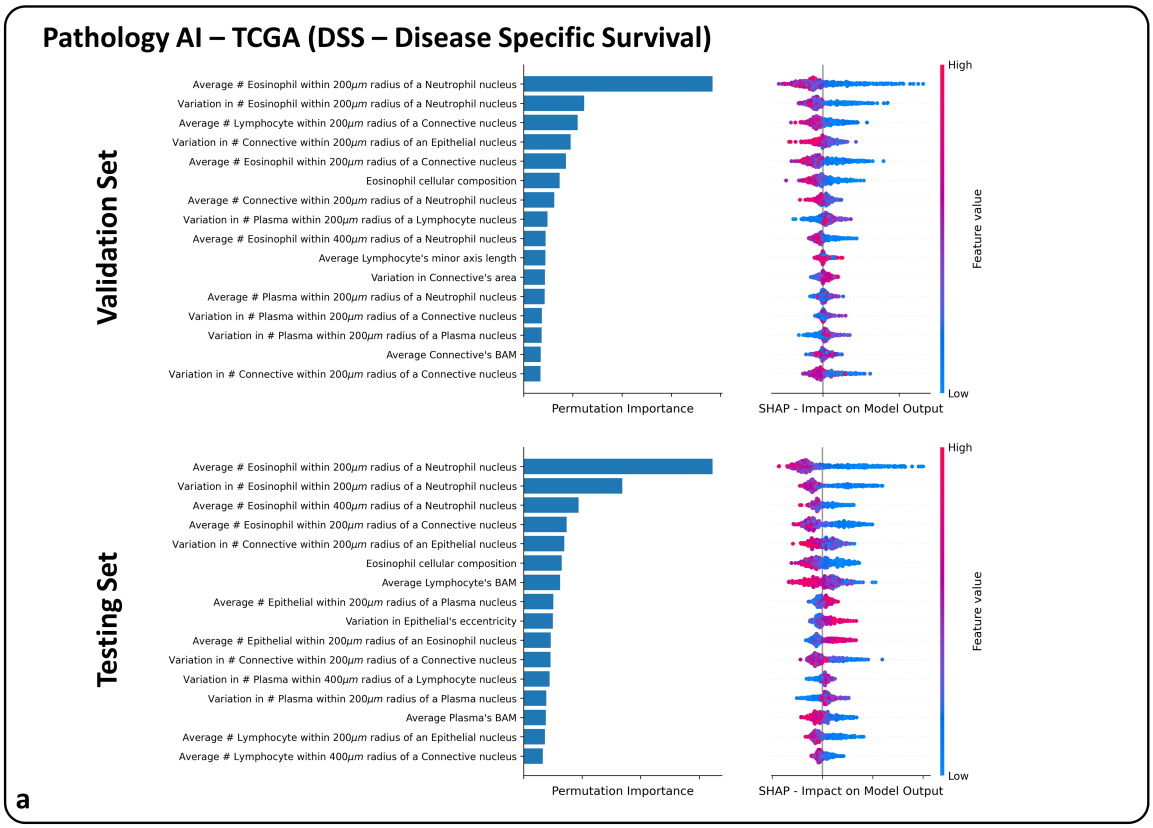


Fig. S11. Contribution of the top 16 features from EPFL | StarDist (taken from Fig. S8) for the grading task on IMP Diagnostics dataset. The first column reports the permutation importance of the feature on the evaluation results (QWK). On the other hand, other columns (SHAP) reflect how the changes in the feature value affect the model predictions (the predicted probabilities of each class by XGBoost). The reported importances are combined across all data splits.



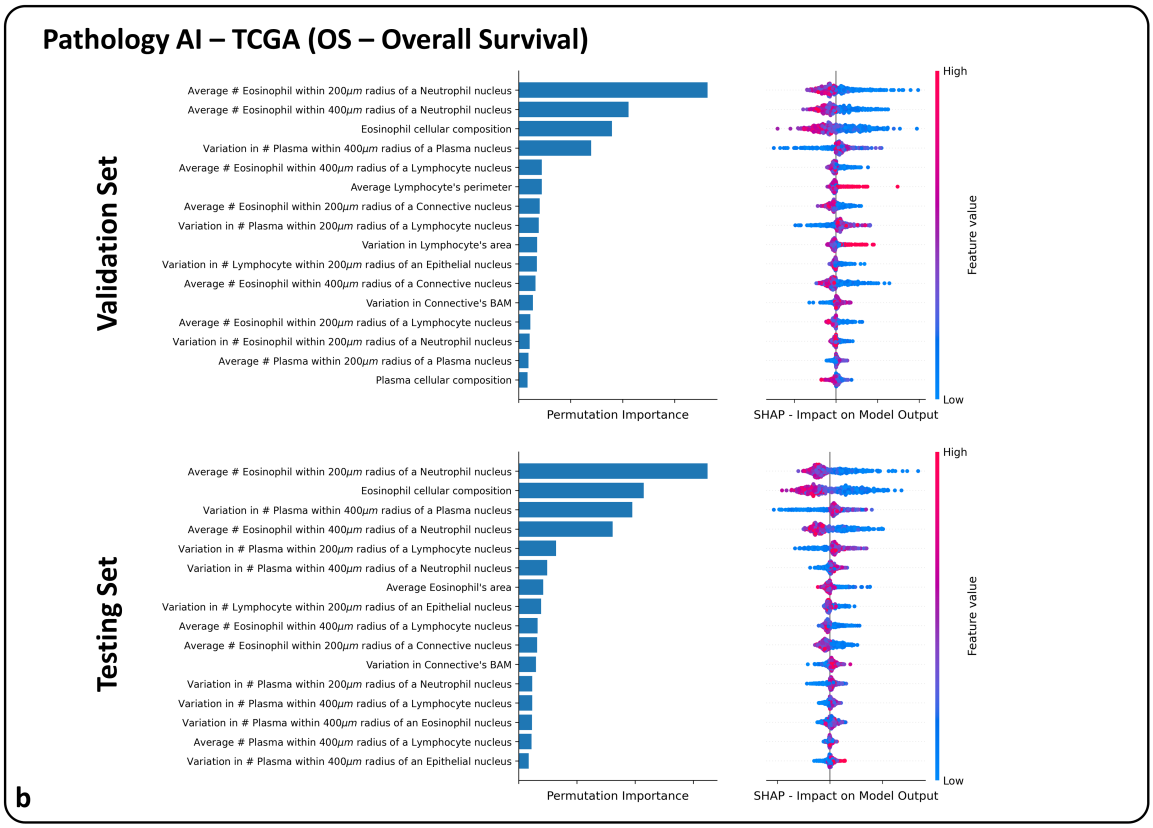
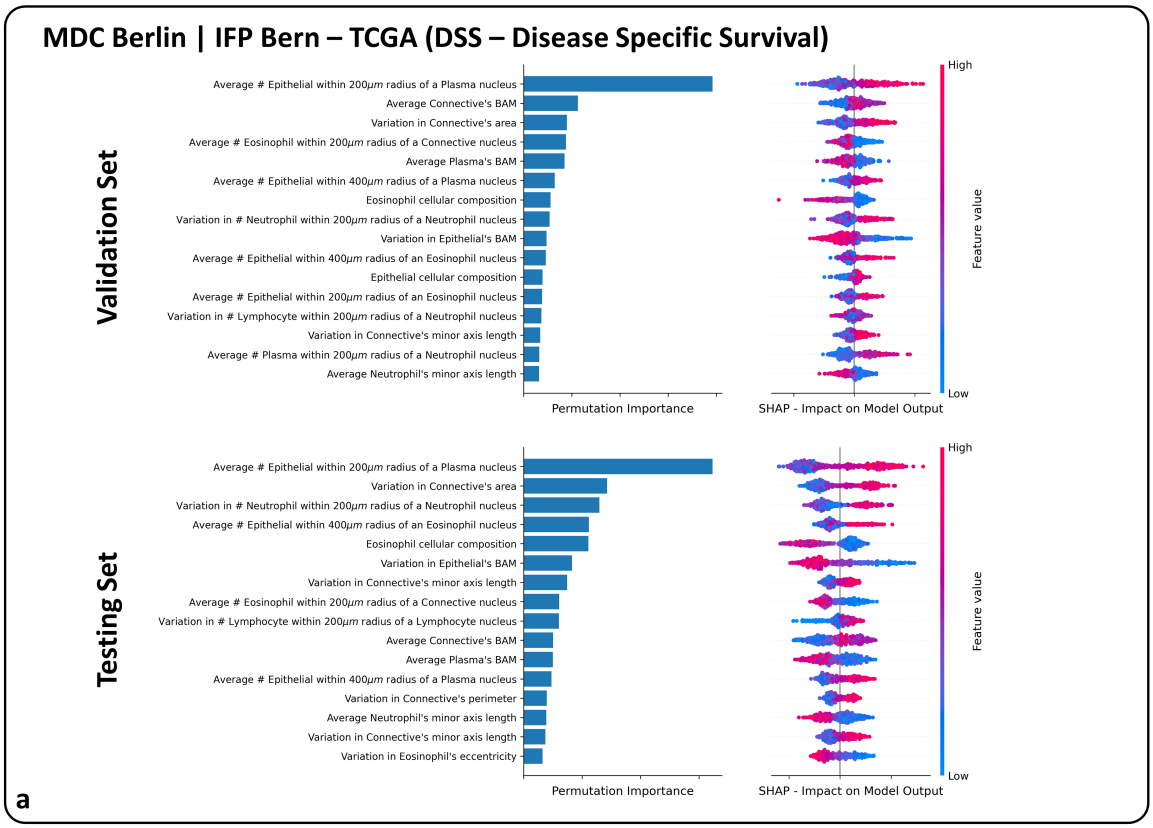


Fig. S12. Contribution of the top 16 features from Pathology AI (taken from Fig. S8) for the survival analyses on TCGA dataset. Within each Validation and Testing subset, the first column reports the permutation importance of the feature on the evaluation results (C-Index). On the other hand, other columns (SHAP) reflect how the changes in the feature value affect the model predictions (the predicted risk scores by XGBoost). The reported importances are combined across all data splits.



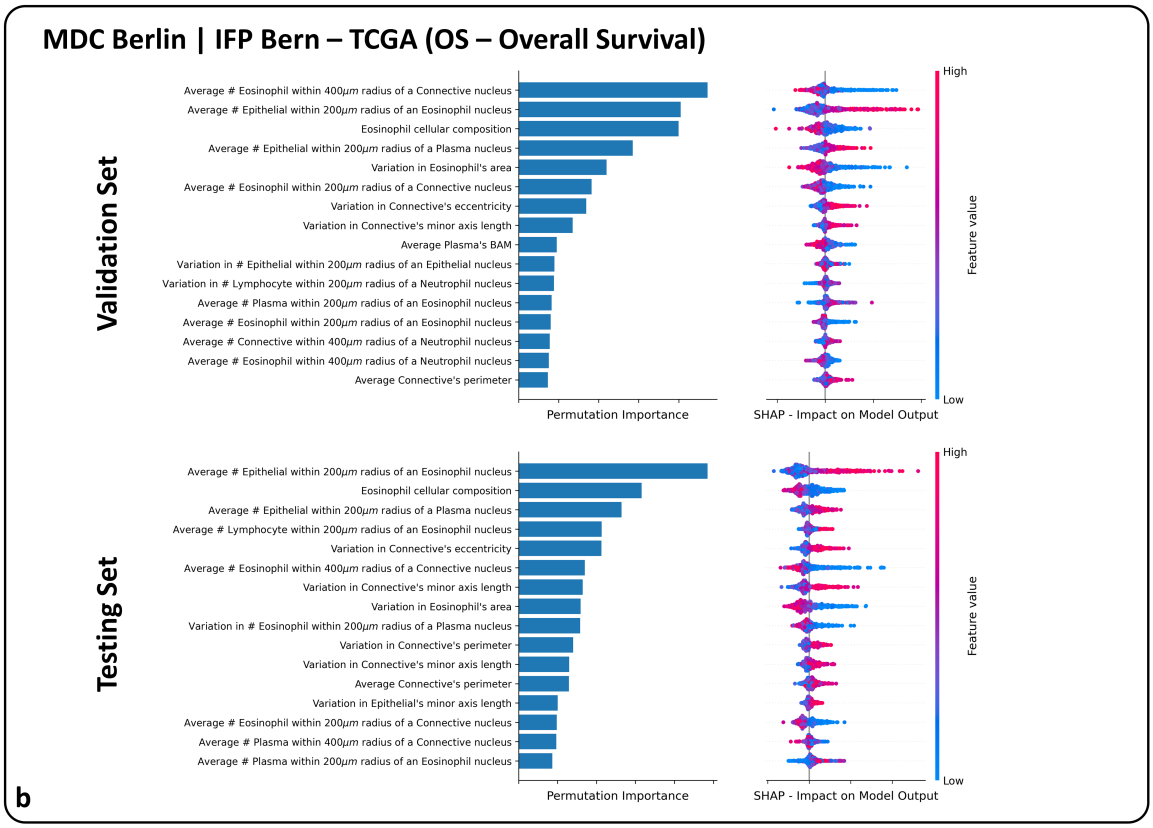
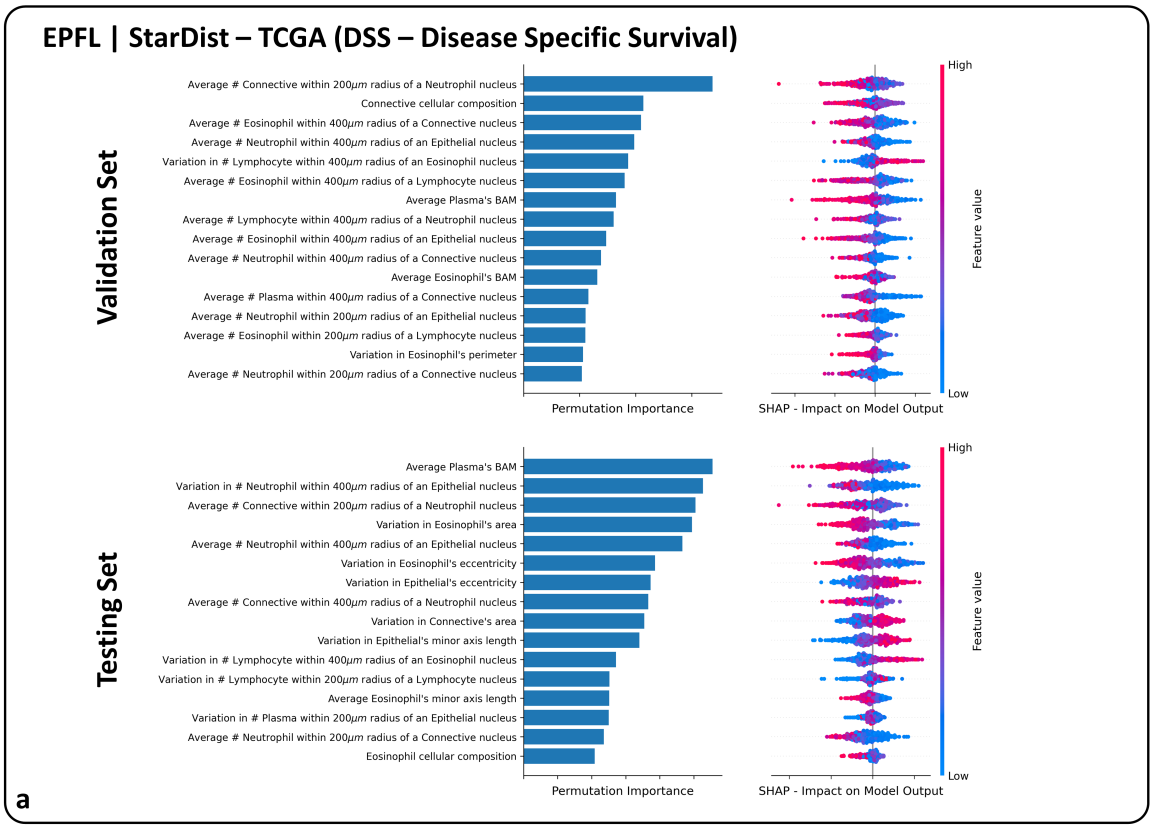


Fig. S13. Contribution of the top 16 features from MDC Berlin | IFP Bern (taken from Fig. S8) for the survival analyses on TCGA dataset. Within each Validation and Testing subset, the first column reports the permutation importance of the feature on the evaluation results (C-Index). On the other hand, other columns (SHAP) reflect how the changes in the feature value affect the model predictions (the predicted risk scores by XGBoost). The reported importances are combined across all data splits.



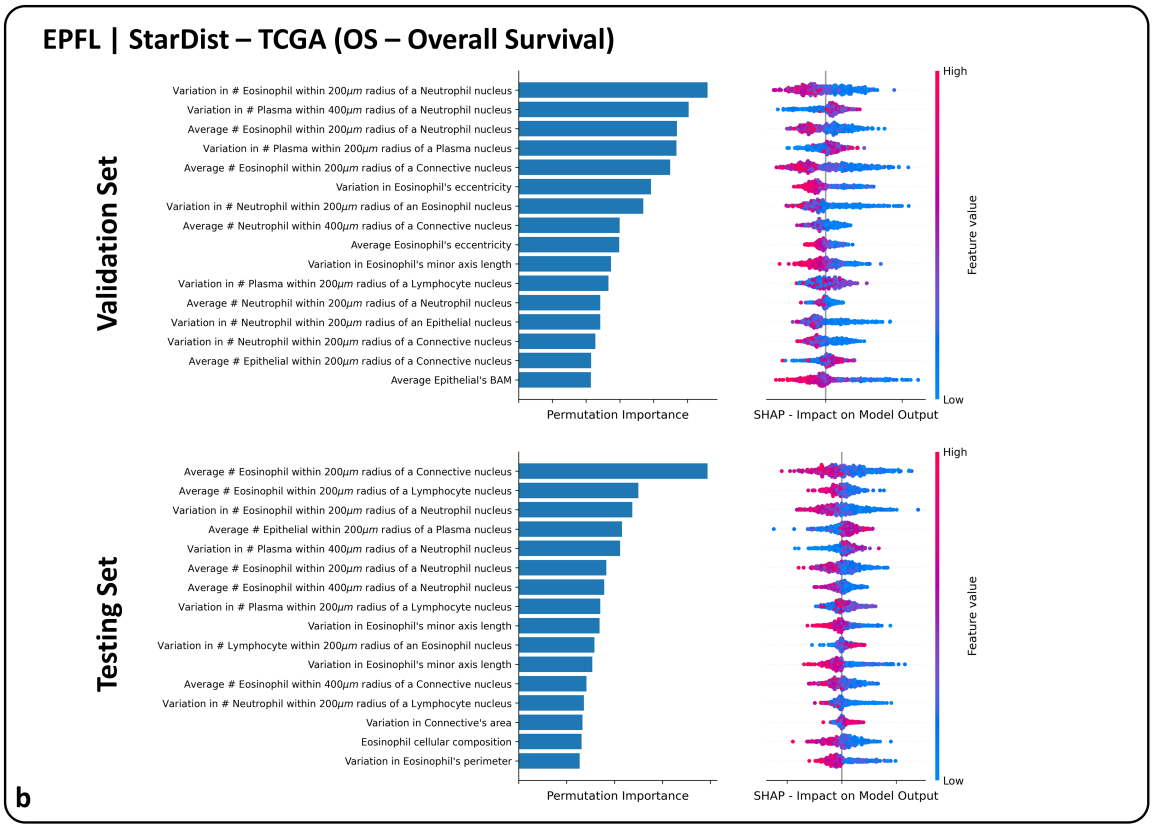


Fig. S14. Contribution of the top 16 features from EPFL | StarDist (taken from Fig. S8) for the survival analyses on TCGA dataset. Within each Validation and Testing subset, the first column reports the permutation importance of the feature on the evaluation results (C-Index). On the other hand, other columns (SHAP) reflect how the changes in the feature value affect the model predictions (the predicted risk scores by XGBoost). The reported importances are combined across all data splits.

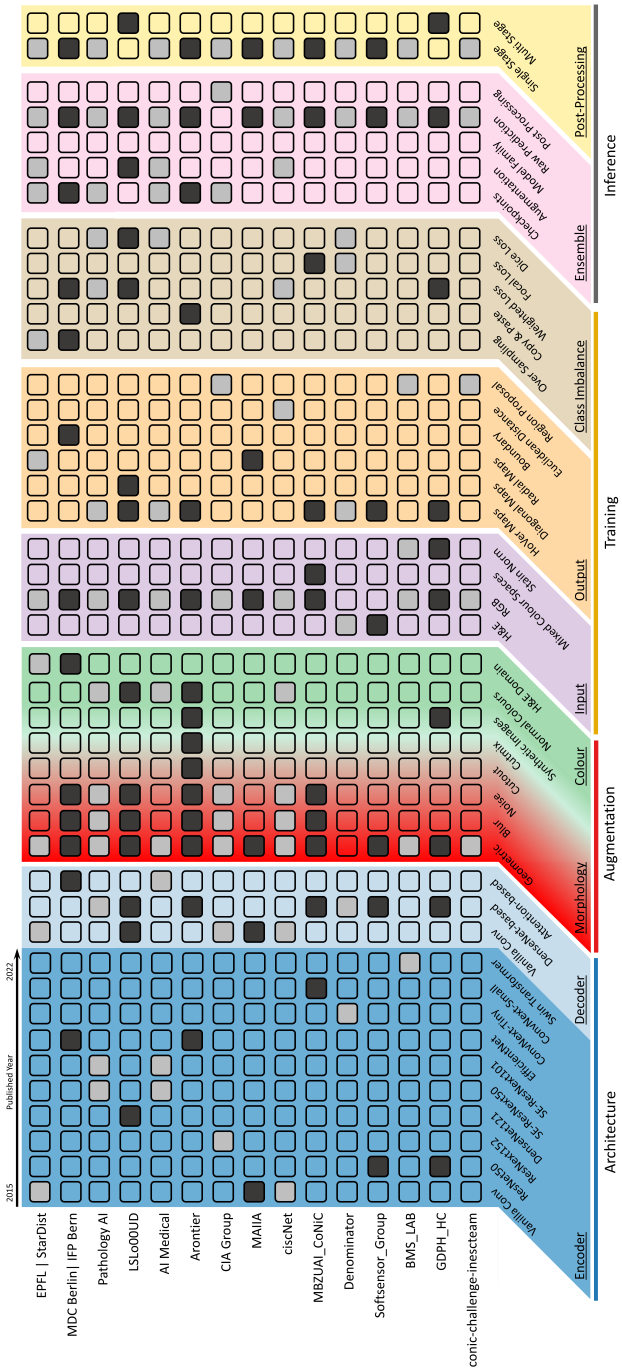


Fig. S15. Summary of the top 15 participant algorithms. The figure is split into various segments to better understand the differences between each team. We identify the network architecture of each submission, including the encoder and decoder design. We determine the augmentation strategy, distinguishing between morphology-based and colour-based augmentation. We indicate the training strategy, consisting of the input type, the output type and whether a technique was used to deal with the class imbalance. We also identified the inference strategy, denoting whether ensembling was used and the post-processing technique. The colour within each box (grey or black) is insignificant – it is used to help distinguish between teams on each row.

TABLE S1. Hyperparameter space when performing Random Search. We provide the name of the parameter, as used in the Python implementation (https://xgboost.readthedocs.io/en/stable/parameter.html), along with the range of values that we randomly sample from.

Parameter Name	Value Ranges
num_boost_round	8 to 256
learning_rate	0.001 to 0.1
max_depth	1 to 16
subsample	One of [0.3, 0.4, 0.5, 0.6, 0.7, 0.8]
colsample_bytree	One of [0.3, 0.4, 0.5, 0.6, 0.7, 0.8]
colsample_bylevel	One of [0.3, 0.4, 0.5, 0.6, 0.7, 0.8]
colsample_bynode	One of [0.3, 0.4, 0.5, 0.6, 0.7, 0.8]
min_child_weight	0.01 to 3.0
reg_lambda	0.1 to 2.0
reg_alpha	0.1 to 2.0
booster	"booster" or "dart"
rate_drop	0.1 to 0.7

TABLE S2. Performance of the XGBoost on TCGA dataset for the DSS (disease specific survival) task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of C-index. C denotes, clinical features, D_d denotes density-based features, D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	C	D_d	D_m	D_c	D	$\overline{D}$
N/A	0.7665±0.0600	-	-	-	-	-
Baseline	-	0.6270±0.0631	0.6018±0.0511	0.5922±0.0690	0.5981±0.0464	0.6228±0.0601
Pathology AI	-	0.6437±0.0648	0.5879±0.0683	0.6443±0.0492	0.6366±0.0650	0.6672±0.0583
MDC Berlin | IFP Bern	-	0.6333±0.0482	0.6242±0.0732	0.6346±0.0591	0.6413±0.0604	0.6686±0.0443
EPFL | StarDist	-	0.6418±0.0543	0.6087±0.0676	0.6334±0.0526	0.6354±0.0561	0.6685±0.0628

Testing Set

	C	D_d	D_m	D_c	D	$\overline{D}$
N/A	0.7662±0.0527	-	-	-	-	-
Baseline	-	0.6320±0.0772	0.5785±0.0747	0.5537±0.0782	0.5781±0.0634	0.5744±0.0738
Pathology AI	-	0.6284±0.0734	0.5742±0.0697	0.6263±0.0781	0.6106±0.0671	0.6450±0.0703
MDC Berlin | IFP Bern	-	0.6081±0.0792	0.6358±0.0591	0.6088±0.0566	0.6160±0.0571	0.6518±0.0582
EPFL | StarDist	-	0.6068±0.0826	0.6144±0.0722	0.5976±0.0978	0.6114±0.0827	0.6554±0.0631

TABLE S3. Performance of the XGBoost on TCGA dataset for the OS (overall survival) task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of C-index. C denotes, clinical features, D_d denotes density-based features, D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	C	D_d	D_m	D_c	D	$\overline{D}$
N/A	0.7354±0.0481	-	-	-	-	-
Baseline	-	0.5888±0.0631	0.5478±0.0674	0.5769±0.0768	0.5748±0.0348	0.6015±0.0580
Pathology AI	-	0.6291±0.0543	0.5855±0.0623	0.6340±0.0558	0.6399±0.0426	0.6716±0.0526
MDC Berlin | IFP Bern	-	0.6160±0.0476	0.6049±0.0727	0.6014±0.0613	0.6156±0.0644	0.6453±0.0533
EPFL | StarDist	-	0.6140±0.0685	0.6051±0.0778	0.6275±0.0514	0.6349±0.0617	0.6589±0.0563

Testing Set

	C	D_d	D_m	D_c	D	$\overline{D}$
N/A	0.7251±0.0411	-	-	-	-	-
Baseline	-	0.5641±0.0731	0.5401±0.0757	0.5550±0.0649	0.5586±0.0647	0.5729±0.0650
Pathology AI	-	0.6185±0.0685	0.5838±0.0653	0.6218±0.0620	0.6187±0.0639	0.6418±0.0565
MDC Berlin | IFP Bern	-	0.6019±0.0640	0.5808±0.0458	0.5876±0.0587	0.6115±0.0797	0.6176±0.0616
EPFL | StarDist	-	0.5811±0.0719	0.5930±0.0688	0.5846±0.0755	0.6105±0.0653	0.6456±0.0614

TABLE S4. Performance of the XGBoost on IMP Diagnostic for the grading task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of mF_1. D_d denotes density-based features, D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.6958±0.0212	0.7564±0.0265	0.7781±0.0230	0.8228±0.0185	0.8307±0.0208
Pathology AI	0.7565±0.0246	0.7702±0.0318	0.8520±0.0204	0.8669±0.0168	0.8720±0.0130
MDC Berlin | IFP Bern	0.7110±0.0265	0.7442±0.0344	0.8545±0.0242	0.8705±0.0220	0.8765±0.0214
EPFL | StarDist	0.7716±0.0234	0.7390±0.0265	0.8461±0.0169	0.8533±0.0186	0.8573±0.0184

Testing Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.6862±0.0379	0.7402±0.0287	0.7729±0.0255	0.8203±0.0316	0.8227±0.0273
Pathology AI	0.7492±0.0349	0.7618±0.0317	0.8451±0.0296	0.8649±0.0282	0.8664±0.0280
MDC Berlin | IFP Bern	0.6885±0.0297	0.7472±0.0304	0.8520±0.0264	0.8698±0.0256	0.8739±0.0218
EPFL | StarDist	0.7529±0.0299	0.7375±0.0340	0.8367±0.0282	0.8439±0.0340	0.8463±0.0341

TABLE S5. Performance of the XGBoost on IMP Diagnostic for the grading task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of mAP. D_d denotes density-based features, D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.7502±0.0263	0.8271±0.0261	0.8574±0.0164	0.8981±0.0183	0.8997±0.0175
Pathology AI	0.8234±0.0310	0.8472±0.0260	0.9209±0.0151	0.9302±0.0149	0.9349±0.0120
MDC Berlin | IFP Bern	0.7770±0.0277	0.8114±0.0330	0.9188±0.0170	0.9357±0.0165	0.9385±0.0146
EPFL | StarDist	0.8261±0.0278	0.8167±0.0253	0.9131±0.0146	0.9186±0.0137	0.9232±0.0134

Testing Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.7456±0.0314	0.8189±0.0290	0.8547±0.0243	0.8921±0.0237	0.8956±0.0242
Pathology AI	0.8171±0.0328	0.8380±0.0302	0.9141±0.0235	0.9262±0.0227	0.9268±0.0219
MDC Berlin | IFP Bern	0.7624±0.0295	0.8166±0.0322	0.9154±0.0189	0.9324±0.0191	0.9373±0.0179
EPFL | StarDist	0.8175±0.0307	0.8136±0.0355	0.9058±0.0251	0.9142±0.0245	0.9185±0.0243

TABLE S6. Performance of the XGBoost on IMP Diagnostic for the grading task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of QWK (Quadratic Weighted Kappa). D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.5892±0.0427	0.6539±0.0496	0.7074±0.0413	0.7600±0.0258	0.7696±0.0300
Pathology AI	0.6829±0.0308	0.7052±0.0445	0.8066±0.0349	0.8333±0.0299	0.8392±0.0243
MDC Berlin | IFP Bern	0.6178±0.0447	0.6689±0.0538	0.8259±0.0361	0.8413±0.0335	0.8501±0.0330
EPFL | StarDist	0.6974±0.0385	0.6786±0.0439	0.8082±0.0287	0.8193±0.0312	0.8248±0.0309

Testing Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.5720±0.0580	0.6328±0.0551	0.6990±0.0406	0.7506±0.0551	0.7574±0.0492
Pathology AI	0.6696±0.0562	0.6904±0.0468	0.8051±0.0423	0.8302±0.0404	0.8354±0.0367
MDC Berlin | IFP Bern	0.5842±0.0400	0.6685±0.0506	0.8212±0.0361	0.8436±0.0382	0.8463±0.0319
EPFL | StarDist	0.6732±0.0467	0.6744±0.0480	0.7943±0.0383	0.8038±0.0477	0.8051±0.0477

TABLE S7. Performance of the XGBoost on IMP Diagnostic for the grading task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of *Sensitivity* averaged across 3 classes. D_d denotes density-based features, D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.6882±0.0217	0.7541±0.0276	0.7730±0.0235	0.8189±0.0202	0.8275±0.0213
Pathology AI	0.6829±0.0308	0.7052±0.0445	0.8066±0.0349	0.8333±0.0299	0.8392±0.0243
MDC Berlin | IFP Bern	0.7057±0.0251	0.7357±0.0342	0.8510±0.0248	0.8674±0.0227	0.8732±0.0217
EPFL | StarDist	0.7660±0.0251	0.7290±0.0268	0.8421±0.0165	0.8491±0.0203	0.8535±0.0182

Testing Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.6789±0.0393	0.7379±0.0287	0.7691±0.0282	0.8174±0.0317	0.8206±0.0280
Pathology AI	0.7416±0.0357	0.7544±0.0316	0.8427±0.0309	0.8608±0.0297	0.8622±0.0296
MDC Berlin | IFP Bern	0.6847±0.0309	0.7402±0.0317	0.8490±0.0282	0.8669±0.0285	0.8711±0.0243
EPFL | StarDist	0.7475±0.0317	0.7272±0.0357	0.8334±0.0305	0.8413±0.0366	0.8432±0.0357

TABLE S8. Performance of the XGBoost on IMP Diagnostic for the grading task when using a feature set obtained from different nuclear recognition methods. The reported values are mean ± std of *Specificity* averaged across 3 classes. D_d denotes density-based features, D_m denotes morphology-based features and D_c denotes colocalisation features. D refers to the combination of all types of features (excluding clinical) and $\overline{D}$ is a subset of D after feature selection.

Validation Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.8423±0.0106	0.8778±0.0130	0.8845±0.0116	0.9091±0.0101	0.9133±0.0109
Pathology AI	0.8734±0.0135	0.8788±0.0166	0.9234±0.0107	0.9302±0.0086	0.9329±0.0071
MDC Berlin | IFP Bern	0.8529±0.0126	0.8658±0.0172	0.9226±0.0127	0.9316±0.0114	0.9346±0.0113
EPFL | StarDist	0.8815±0.0126	0.8600±0.0136	0.9193±0.0087	0.9225±0.0098	0.9248±0.0094

Testing Set

	D_d	D_m	D_c	D	$\overline{D}$
Baseline	0.8377±0.0196	0.8693±0.0137	0.8820±0.0143	0.9084±0.0151	0.9095±0.0134
Pathology AI	0.8697±0.0182	0.8754±0.0156	0.9195±0.0155	0.9291±0.0146	0.9295±0.0149
MDC Berlin | IFP Bern	0.8419±0.0163	0.8680±0.0155	0.9216±0.0141	0.9310±0.0137	0.9334±0.0118
EPFL | StarDist	0.8715±0.0159	0.8592±0.0181	0.9145±0.0153	0.9183±0.0182	0.9195±0.0180

TABLE S9. Complete list of features considered in the downstream pipelines. Here, we give a description of the feature, along with the category in which it belongs.

ID	Feature Names	Category
0	Average Connective's area	Morphology
1	Variation in Connective's area	Morphology
2	Average Connective's eccentricity	Morphology
3	Variation in Connective's eccentricity	Morphology
4	Average Connective's perimeter	Morphology
5	Variation in Connective's perimeter	Morphology
6	Average Connective's minor axis length	Morphology
7	Variation in Connective's minor axis length	Morphology
8	Average Connective's minor axis length	Morphology
9	Variation in Connective's minor axis length	Morphology
10	Average Connective's BAM	Morphology
11	Variation in Connective's BAM	Morphology
12	Average Eosinophil's area	Morphology
13	Variation in Eosinophil's area	Morphology
14	Average Eosinophil's eccentricity	Morphology
15	Variation in Eosinophil's eccentricity	Morphology
16	Average Eosinophil's perimeter	Morphology
17	Variation in Eosinophil's perimeter	Morphology
18	Average Eosinophil's minor axis length	Morphology
19	Variation in Eosinophil's minor axis length	Morphology
20	Average Eosinophil's minor axis length	Morphology
21	Variation in Eosinophil's minor axis length	Morphology
22	Average Eosinophil's BAM	Morphology
23	Variation in Eosinophil's BAM	Morphology
24	Average Epithelial's area	Morphology
25	Variation in Epithelial's area	Morphology
26	Average Epithelial's eccentricity	Morphology
27	Variation in Epithelial's eccentricity	Morphology
28	Average Epithelial's perimeter	Morphology
29	Variation in Epithelial's perimeter	Morphology
30	Average Epithelial's minor axis length	Morphology
31	Variation in Epithelial's minor axis length	Morphology
32	Average Epithelial's minor axis length	Morphology
33	Variation in Epithelial's minor axis length	Morphology
34	Average Epithelial's BAM	Morphology
35	Variation in Epithelial's BAM	Morphology
36	Average Lymphocyte's area	Morphology
37	Variation in Lymphocyte's area	Morphology
38	Average Lymphocyte's eccentricity	Morphology
39	Variation in Lymphocyte's eccentricity	Morphology
40	Average Lymphocyte's perimeter	Morphology
41	Variation in Lymphocyte's perimeter	Morphology
42	Average Lymphocyte's minor axis length	Morphology
43	Variation in Lymphocyte's minor axis length	Morphology
44	Average Lymphocyte's minor axis length	Morphology
45	Variation in Lymphocyte's minor axis length	Morphology
46	Average Lymphocyte's BAM	Morphology
47	Variation in Lymphocyte's BAM	Morphology
48	Average Neutrophil's area	Morphology
49	Variation in Neutrophil's area	Morphology
50	Average Neutrophil's eccentricity	Morphology
51	Variation in Neutrophil's eccentricity	Morphology
52	Average Neutrophil's perimeter	Morphology

53	Variation in Neutrophil's perimeter	Morphology
54	Average Neutrophil's minor axis length	Morphology
55	Variation in Neutrophil's minor axis length	Morphology
56	Average Neutrophil's minor axis length	Morphology
57	Variation in Neutrophil's minor axis length	Morphology
58	Average Neutrophil's BAM	Morphology
59	Variation in Neutrophil's BAM	Morphology
60	Average Plasma's area	Morphology
61	Variation in Plasma's area	Morphology
62	Average Plasma's eccentricity	Morphology
63	Variation in Plasma's eccentricity	Morphology
64	Average Plasma's perimeter	Morphology
65	Variation in Plasma's perimeter	Morphology
66	Average Plasma's minor axis length	Morphology
67	Variation in Plasma's minor axis length	Morphology
68	Average Plasma's minor axis length	Morphology
69	Variation in Plasma's minor axis length	Morphology
70	Average Plasma's BAM	Morphology
71	Variation in Plasma's BAM	Morphology
72	Average # Neutrophil within 200um radius of a Connective nucleus	Colocalisation
73	Variation in # Neutrophil within 200um radius of a Connective nucleus	Colocalisation
74	Average # Epithelial within 200um radius of a Connective nucleus	Colocalisation
75	Variation in # Epithelial within 200um radius of a Connective nucleus	Colocalisation
76	Average # Lymphocyte within 200um radius of a Connective nucleus	Colocalisation
77	Variation in # Lymphocyte within 200um radius of a Connective nucleus	Colocalisation
78	Average # Plasma within 200um radius of a Connective nucleus	Colocalisation
79	Variation in # Plasma within 200um radius of a Connective nucleus	Colocalisation
80	Average # Eosinophil within 200um radius of a Connective nucleus	Colocalisation
81	Variation in # Eosinophil within 200um radius of a Connective nucleus	Colocalisation
82	Average # Connective within 200um radius of a Connective nucleus	Colocalisation
83	Variation in # Connective within 200um radius of a Connective nucleus	Colocalisation
84	Average # Neutrophil within 200um radius of an Eosinophil nucleus	Colocalisation
85	Variation in # Neutrophil within 200um radius of an Eosinophil nucleus	Colocalisation
86	Average # Epithelial within 200um radius of an Eosinophil nucleus	Colocalisation
87	Variation in # Epithelial within 200um radius of an Eosinophil nucleus	Colocalisation
88	Average # Lymphocyte within 200um radius of an Eosinophil nucleus	Colocalisation
89	Variation in # Lymphocyte within 200um radius of an Eosinophil nucleus	Colocalisation
90	Average # Plasma within 200um radius of an Eosinophil nucleus	Colocalisation
91	Variation in # Plasma within 200um radius of an Eosinophil nucleus	Colocalisation
92	Average # Eosinophil within 200um radius of an Eosinophil nucleus	Colocalisation
93	Variation in # Eosinophil within 200um radius of an Eosinophil nucleus	Colocalisation
94	Average # Connective within 200um radius of an Eosinophil nucleus	Colocalisation
95	Variation in # Connective within 200um radius of an Eosinophil nucleus	Colocalisation
96	Average # Neutrophil within 200um radius of an Epithelial nucleus	Colocalisation
97	Variation in # Neutrophil within 200um radius of an Epithelial nucleus	Colocalisation
98	Average # Epithelial within 200um radius of an Epithelial nucleus	Colocalisation
99	Variation in # Epithelial within 200um radius of an Epithelial nucleus	Colocalisation
100	Average # Lymphocyte within 200um radius of an Epithelial nucleus	Colocalisation
101	Variation in # Lymphocyte within 200um radius of an Epithelial nucleus	Colocalisation
102	Average # Plasma within 200um radius of an Epithelial nucleus	Colocalisation
103	Variation in # Plasma within 200um radius of an Epithelial nucleus	Colocalisation
104	Average # Eosinophil within 200um radius of an Epithelial nucleus	Colocalisation
105	Variation in # Eosinophil within 200um radius of an Epithelial nucleus	Colocalisation
106	Average # Connective within 200um radius of an Epithelial nucleus	Colocalisation
107	Variation in # Connective within 200um radius of an Epithelial nucleus	Colocalisation
108	Average # Neutrophil within 200um radius of a Lymphocyte nucleus	Colocalisation
109	Variation in # Neutrophil within 200um radius of a Lymphocyte nucleus	Colocalisation
110	Average # Epithelial within 200um radius of a Lymphocyte nucleus	Colocalisation

111	Variation in # Epithelial within 200um radius of a Lymphocyte nucleus	Colocalisation
112	Average # Lymphocyte within 200um radius of a Lymphocyte nucleus	Colocalisation
113	Variation in # Lymphocyte within 200um radius of a Lymphocyte nucleus	Colocalisation
114	Average # Plasma within 200um radius of a Lymphocyte nucleus	Colocalisation
115	Variation in # Plasma within 200um radius of a Lymphocyte nucleus	Colocalisation
116	Average # Eosinophil within 200um radius of a Lymphocyte nucleus	Colocalisation
117	Variation in # Eosinophil within 200um radius of a Lymphocyte nucleus	Colocalisation
118	Average # Connective within 200um radius of a Lymphocyte nucleus	Colocalisation
119	Variation in # Connective within 200um radius of a Lymphocyte nucleus	Colocalisation
120	Average # Neutrophil within 200um radius of a Neutrophil nucleus	Colocalisation
121	Variation in # Neutrophil within 200um radius of a Neutrophil nucleus	Colocalisation
122	Average # Epithelial within 200um radius of a Neutrophil nucleus	Colocalisation
123	Variation in # Epithelial within 200um radius of a Neutrophil nucleus	Colocalisation
124	Average # Lymphocyte within 200um radius of a Neutrophil nucleus	Colocalisation
125	Variation in # Lymphocyte within 200um radius of a Neutrophil nucleus	Colocalisation
126	Average # Plasma within 200um radius of a Neutrophil nucleus	Colocalisation
127	Variation in # Plasma within 200um radius of a Neutrophil nucleus	Colocalisation
128	Average # Eosinophil within 200um radius of a Neutrophil nucleus	Colocalisation
129	Variation in # Eosinophil within 200um radius of a Neutrophil nucleus	Colocalisation
130	Average # Connective within 200um radius of a Neutrophil nucleus	Colocalisation
131	Variation in # Connective within 200um radius of a Neutrophil nucleus	Colocalisation
132	Average # Neutrophil within 200um radius of a Plasma nucleus	Colocalisation
133	Variation in # Neutrophil within 200um radius of a Plasma nucleus	Colocalisation
134	Average # Epithelial within 200um radius of a Plasma nucleus	Colocalisation
135	Variation in # Epithelial within 200um radius of a Plasma nucleus	Colocalisation
136	Average # Lymphocyte within 200um radius of a Plasma nucleus	Colocalisation
137	Variation in # Lymphocyte within 200um radius of a Plasma nucleus	Colocalisation
138	Average # Plasma within 200um radius of a Plasma nucleus	Colocalisation
139	Variation in # Plasma within 200um radius of a Plasma nucleus	Colocalisation
140	Average # Eosinophil within 200um radius of a Plasma nucleus	Colocalisation
141	Variation in # Eosinophil within 200um radius of a Plasma nucleus	Colocalisation
142	Average # Connective within 200um radius of a Plasma nucleus	Colocalisation
143	Variation in # Connective within 200um radius of a Plasma nucleus	Colocalisation
144	Average # Neutrophil within 400um radius of a Connective nucleus	Colocalisation
145	Variation in # Neutrophil within 400um radius of a Connective nucleus	Colocalisation
146	Average # Epithelial within 400um radius of a Connective nucleus	Colocalisation
147	Variation in # Epithelial within 400um radius of a Connective nucleus	Colocalisation
148	Average # Lymphocyte within 400um radius of a Connective nucleus	Colocalisation
149	Variation in # Lymphocyte within 400um radius of a Connective nucleus	Colocalisation
150	Average # Plasma within 400um radius of a Connective nucleus	Colocalisation
151	Variation in # Plasma within 400um radius of a Connective nucleus	Colocalisation
152	Average # Eosinophil within 400um radius of a Connective nucleus	Colocalisation
153	Variation in # Eosinophil within 400um radius of a Connective nucleus	Colocalisation
154	Average # Connective within 400um radius of a Connective nucleus	Colocalisation
155	Variation in # Connective within 400um radius of a Connective nucleus	Colocalisation
156	Average # Neutrophil within 400um radius of an Eosinophil nucleus	Colocalisation
157	Variation in # Neutrophil within 400um radius of an Eosinophil nucleus	Colocalisation
158	Average # Epithelial within 400um radius of an Eosinophil nucleus	Colocalisation
159	Variation in # Epithelial within 400um radius of an Eosinophil nucleus	Colocalisation
160	Average # Lymphocyte within 400um radius of an Eosinophil nucleus	Colocalisation
161	Variation in # Lymphocyte within 400um radius of an Eosinophil nucleus	Colocalisation
162	Average # Plasma within 400um radius of an Eosinophil nucleus	Colocalisation
163	Variation in # Plasma within 400um radius of an Eosinophil nucleus	Colocalisation
164	Average # Eosinophil within 400um radius of an Eosinophil nucleus	Colocalisation
165	Variation in # Eosinophil within 400um radius of an Eosinophil nucleus	Colocalisation
166	Average # Connective within 400um radius of an Eosinophil nucleus	Colocalisation
167	Variation in # Connective within 400um radius of an Eosinophil nucleus	Colocalisation
168	Average # Neutrophil within 400um radius of an Epithelial nucleus	Colocalisation

169	Variation in # Neutrophil within 400um radius of an Epithelial nucleus	Colocalisation
170	Average # Epithelial within 400um radius of an Epithelial nucleus	Colocalisation
171	Variation in # Epithelial within 400um radius of an Epithelial nucleus	Colocalisation
172	Average # Lymphocyte within 400um radius of an Epithelial nucleus	Colocalisation
173	Variation in # Lymphocyte within 400um radius of an Epithelial nucleus	Colocalisation
174	Average # Plasma within 400um radius of an Epithelial nucleus	Colocalisation
175	Variation in # Plasma within 400um radius of an Epithelial nucleus	Colocalisation
176	Average # Eosinophil within 400um radius of an Epithelial nucleus	Colocalisation
177	Variation in # Eosinophil within 400um radius of an Epithelial nucleus	Colocalisation
178	Average # Connective within 400um radius of an Epithelial nucleus	Colocalisation
179	Variation in # Connective within 400um radius of an Epithelial nucleus	Colocalisation
180	Average # Neutrophil within 400um radius of a Lymphocyte nucleus	Colocalisation
181	Variation in # Neutrophil within 400um radius of a Lymphocyte nucleus	Colocalisation
182	Average # Epithelial within 400um radius of a Lymphocyte nucleus	Colocalisation
183	Variation in # Epithelial within 400um radius of a Lymphocyte nucleus	Colocalisation
184	Average # Lymphocyte within 400um radius of a Lymphocyte nucleus	Colocalisation
185	Variation in # Lymphocyte within 400um radius of a Lymphocyte nucleus	Colocalisation
186	Average # Plasma within 400um radius of a Lymphocyte nucleus	Colocalisation
187	Variation in # Plasma within 400um radius of a Lymphocyte nucleus	Colocalisation
188	Average # Eosinophil within 400um radius of a Lymphocyte nucleus	Colocalisation
189	Variation in # Eosinophil within 400um radius of a Lymphocyte nucleus	Colocalisation
190	Average # Connective within 400um radius of a Lymphocyte nucleus	Colocalisation
191	Variation in # Connective within 400um radius of a Lymphocyte nucleus	Colocalisation
192	Average # Neutrophil within 400um radius of a Neutrophil nucleus	Colocalisation
193	Variation in # Neutrophil within 400um radius of a Neutrophil nucleus	Colocalisation
194	Average # Epithelial within 400um radius of a Neutrophil nucleus	Colocalisation
195	Variation in # Epithelial within 400um radius of a Neutrophil nucleus	Colocalisation
196	Average # Lymphocyte within 400um radius of a Neutrophil nucleus	Colocalisation
197	Variation in # Lymphocyte within 400um radius of a Neutrophil nucleus	Colocalisation
198	Average # Plasma within 400um radius of a Neutrophil nucleus	Colocalisation
199	Variation in # Plasma within 400um radius of a Neutrophil nucleus	Colocalisation
200	Average # Eosinophil within 400um radius of a Neutrophil nucleus	Colocalisation
201	Variation in # Eosinophil within 400um radius of a Neutrophil nucleus	Colocalisation
202	Average # Connective within 400um radius of a Neutrophil nucleus	Colocalisation
203	Variation in # Connective within 400um radius of a Neutrophil nucleus	Colocalisation
204	Average # Neutrophil within 400um radius of a Plasma nucleus	Colocalisation
205	Variation in # Neutrophil within 400um radius of a Plasma nucleus	Colocalisation
206	Average # Epithelial within 400um radius of a Plasma nucleus	Colocalisation
207	Variation in # Epithelial within 400um radius of a Plasma nucleus	Colocalisation
208	Average # Lymphocyte within 400um radius of a Plasma nucleus	Colocalisation
209	Variation in # Lymphocyte within 400um radius of a Plasma nucleus	Colocalisation
210	Average # Plasma within 400um radius of a Plasma nucleus	Colocalisation
211	Variation in # Plasma within 400um radius of a Plasma nucleus	Colocalisation
212	Average # Eosinophil within 400um radius of a Plasma nucleus	Colocalisation
213	Variation in # Eosinophil within 400um radius of a Plasma nucleus	Colocalisation
214	Average # Connective within 400um radius of a Plasma nucleus	Colocalisation
215	Variation in # Connective within 400um radius of a Plasma nucleus	Colocalisation
216	Connective cellular composition	Density
217	Eosinophil cellular composition	Density
218	Epithelial cellular composition	Density
219	Lymphocyte cellular composition	Density
220	Neutrophil cellular composition	Density
221	Plasma cellular composition	Density

REFERENCES

[1] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*. Springer, 2018, pp. 265–273.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[3] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[4] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, vol. 58, p. 101563, 2019.

[5] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[8] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 713–13 722.

[9] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

[10] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2918–2928.

[11] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[12] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[13] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6154–6162.

[14] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[15] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[16] D. Misra, "Mish: A self regularized non-monotonic activation function," *arXiv preprint arXiv:1908.08681*, 2019.

[17] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 976–11 986.

[18] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[20] B. Zhao, X. Chen, Z. Li, Z. Yu, S. Yao, L. Yan, Y. Wang, Z. Liu, C. Liang, and C. Han, "Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Medical Image Analysis*, vol. 65, p. 101786, 2020.

[21] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.

[22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[23] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang *et al.*, "Hybrid task cascade for instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4974–4983.

[24] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2009, pp. 1107–1110.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[26] B. Zhao, C. Han, X. Pan, J. Lin, Z. Yi, C. Liang, X. Chen, B. Li, W. Qiu, D. Li *et al.*, "Restainnet: a self-supervised digital re-stainer for stain normalization," *Computers and Electrical Engineering*, vol. 103, p. 108304, 2022.

[27] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "Centernet: Keypoint triplets for object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6569–6578.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[29] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703.

[30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[32] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Eff-unet: A novel architecture for semantic segmentation in unstructured environment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 358–359.

[33] M. Dawood, K. Branson, N. M. Rajpoot, and F. Minhas, "Albrt: Cellular composition prediction in routine histology images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 664–673.

[34] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[35] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.