Table A 1. Per-class validation performance of region-attention embedding in combination with Xception network, in terms of precision, recall and f1-score. Additionally, the support of the corresponding classification values is presented in the right most column, to provide a better contextualization of the obtained values and their dependency to the number of images in each class

Cell subtype	Mean Precision	Mean recall	Mean f1-score	Support
ABE	0,44	0,28	0,33	2
ART	0,8	0,82	0,81	3448
BAS	0,71	0,56	0,61	90
BLA	0,76	0,77	0,76	2171
EBO	0,89	0,89	0,89	3596
EOS	0,84	0,76	0,80	1129
FGC	0,59	0,43	0,47	10
HAC	0,74	0,63	0,66	94
KSC	0,57	0,39	0,42	11
LYI	0,78	0,38	0,46	10
LYT	0,88	0,91	0,89	5669
MMZ	0,73	0,69	0,71	610
MON	0,71	0,73	0,72	783
MYB	0,73	0,71	0,72	1301
NGB	0,78	0,80	0,79	1979
NGS	0,89	0,90	0,89	3897
NIF	0,75	0,64	0,68	691
OTH	0,66	0,49	0,53	51
PEB	0,78	0,64	0,69	487
PLM	0,80	0,75	0,77	1402
PMO	0,76	0,78	0,77	2399

Table A 2. Complementary feature organization experiment results. In the second row, all averaged results are presented given the combination of a randomized region-attention embedding arrangement and Xception network, while in the third row averaged results are shown when combining region-attention embedding and a linear neural network with no convlutional layers

Complementary Experiment	Average score type	Accuracy	Precision	Recall	F1-score	Specificity
Randomizing feature organization within region-attention embedding	Weighted	0,78	0,77	0,78	0,77	0,98
	Micro	0,78	0,78	0,78	0,78	0,99
	Macro	0,59	0,67	0,59	0,61	0,99
Linear neural network (no convolution layers)	Weighted	0,54	0,54	0,55	0,54	0,95
	Micro	0,54	0,54	0,54	0,54	0,98
	Macro	0,30	0,34	0,30	0,31	0,98

Table A 3. Data ablation experiment results. Comparison between the classification performance of Xception network trained by using consecutively 2000 **and** 1000 **images represented as: region-attention embeddings (second row), and original RGB images (third row), together with imageNet weights , and** 2000 **and** 1000 **RGB images, but trained from scratch. The best classification results are presented in bold, for both cases, using** 2000 **and** 1000 **images**

Classification strategy	**Number of images**	**Averaging type**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Region-attention embedding + Xception	2000 images	Micro	0,73	0,73	0,73	0,73
		Macro	0,64	0,67	0,64	0,65
		Weighted	**0,73**	**0,73**	**0,73**	**0,73**
	1000 images	Micro	0,72	0,72	0,72	0,72
		Macro	0,66	0,68	0,66	0,67
		Weighted	**0,72**	**0,72**	**0,72**	**0,72**
Xception (imageNet weights) + RGB image input	2000 images	Micro	0,3	0,94	0,36	0,28
		Macro	0,3	0,92	0,33	0,28
		Weighted	0,3	0,93	0,36	0,28
	1000 images	Micro	0,37	0,84	0,04	0,37
		Macro	0,35	0,83	0,036	0,36
		Weighted	0,35	0,83	0,036	0,36
Xception (trained from scratch) + RGB image input	2000 images	Micro	0,63	0,67	0,59	0,62
		Macro	0,60	0,68	0,53	0,59
		Weighted	0,63	0,67	0,59	0,62
	1000 images	Micro	0,46	0,86	0,18	0,42
		Macro	0,45	0,84	0,18	0,41
		Weighted	0,46	0,86	0,18	0,42