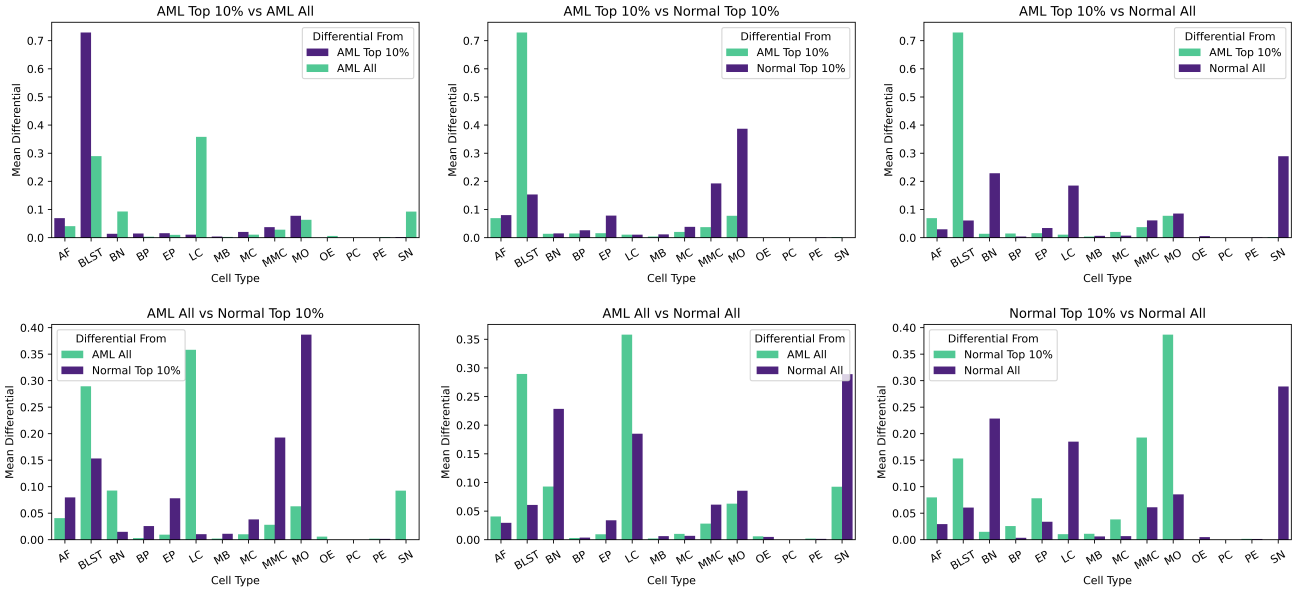

Figure Supplementary 1. Comparative attention barplots In each bar plot we are comparing the mean differentials of the patients taken just from the top 10% attention cells vs all the cells.

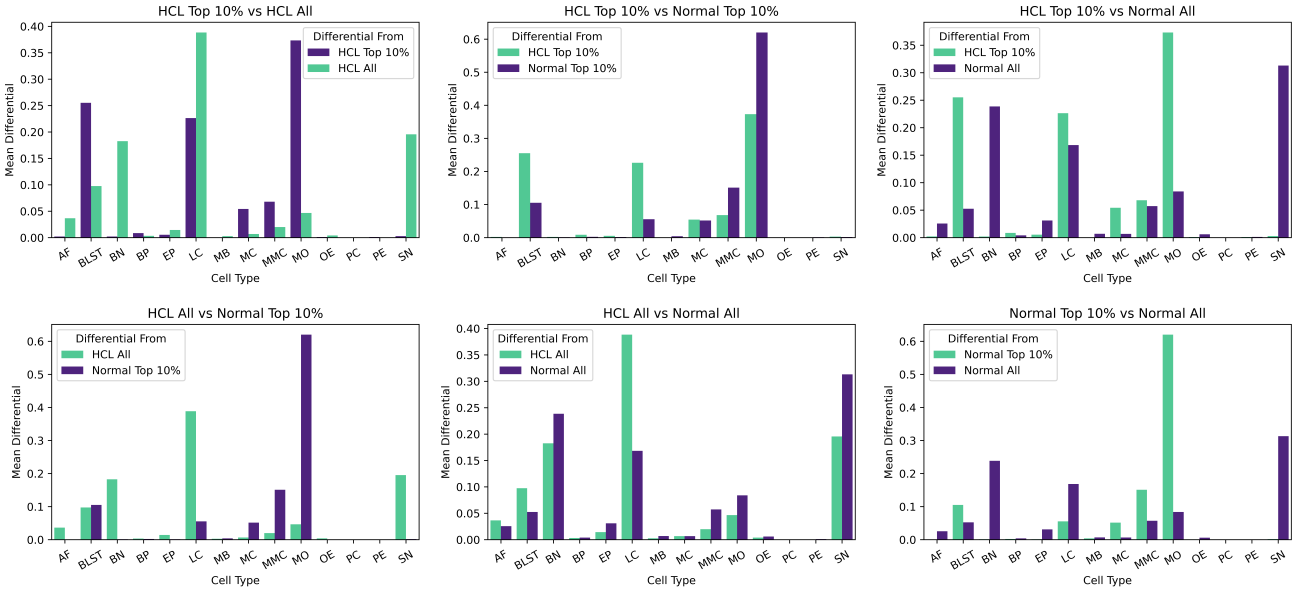

Figure Supplementary 2. Comparative attention barplots In each bar plot we are comparing the mean differentials of the patients taken just from the top 10% attention cells vs all the cells.

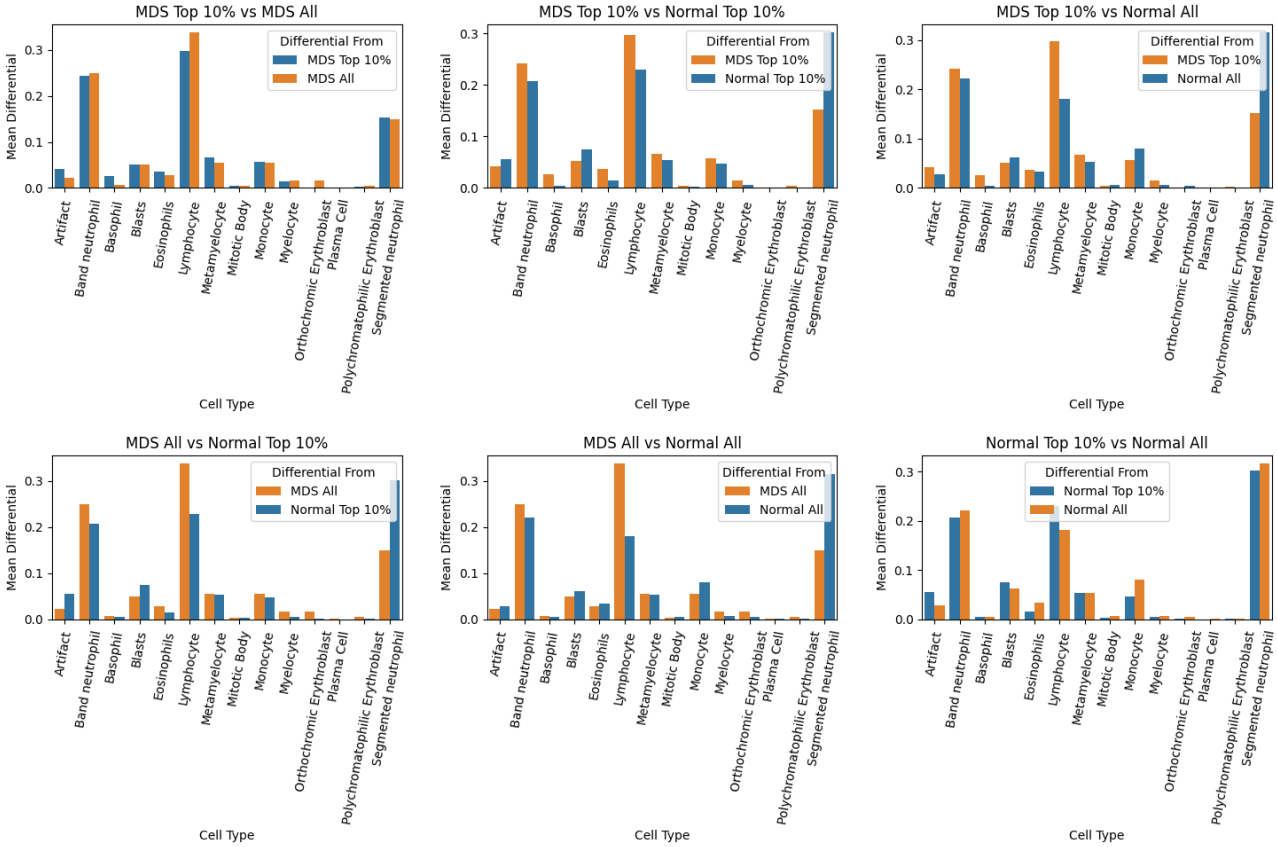

Figure Supplementary 3. Comparative attention barplots In each bar plot we are comparing the mean differentials of the patients taken just from the top 10% attention cells vs all the cells.

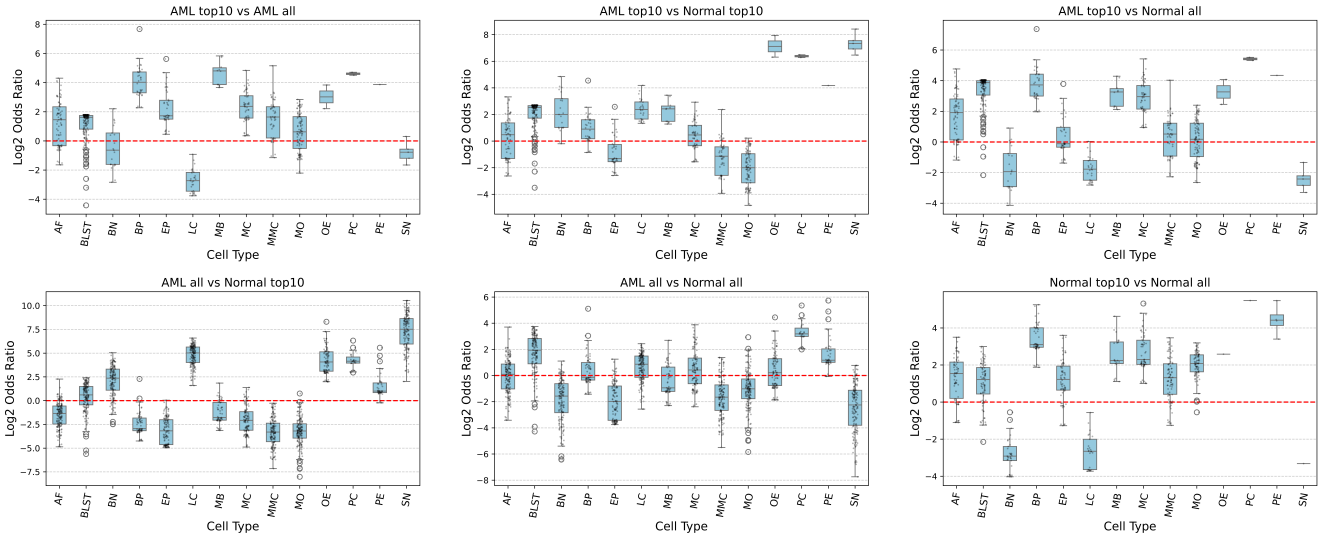

Figure Supplementary 4. Comparative Log odds ratios In each plot we are comparing the mean differentials of the patients taken just from the the two differential types and plotting the log odds ratios of the two.

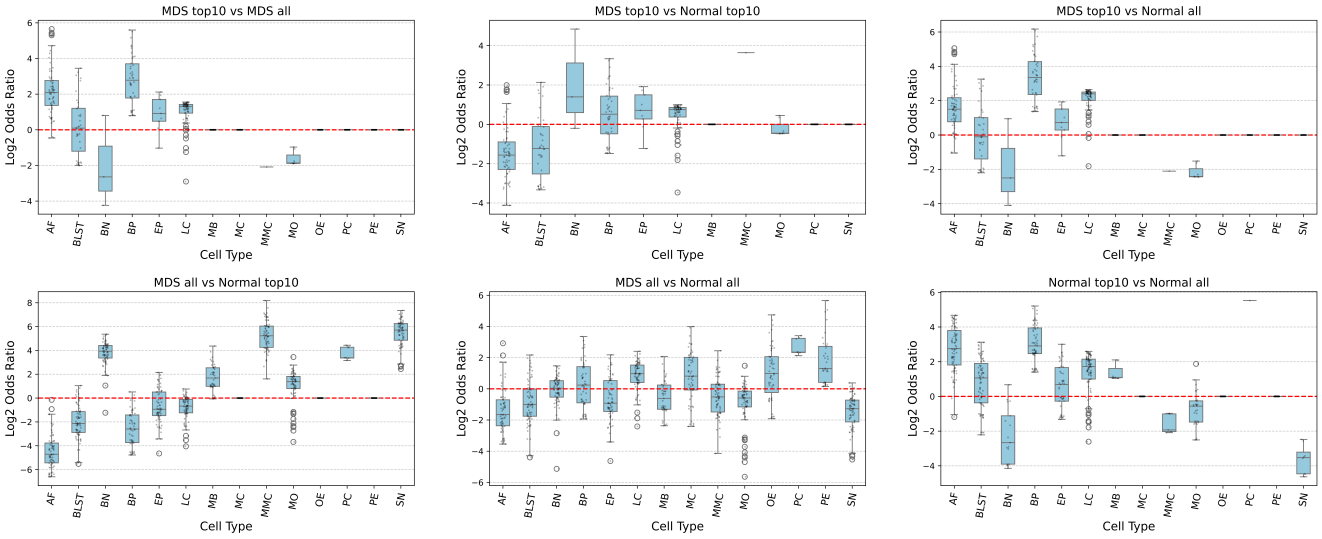

Figure Supplementary 5. Comparative Log odds ratios In each plot we are comparing the mean differentials of the patients taken just from the the two differential types and plotting the log odds ratios of the two.

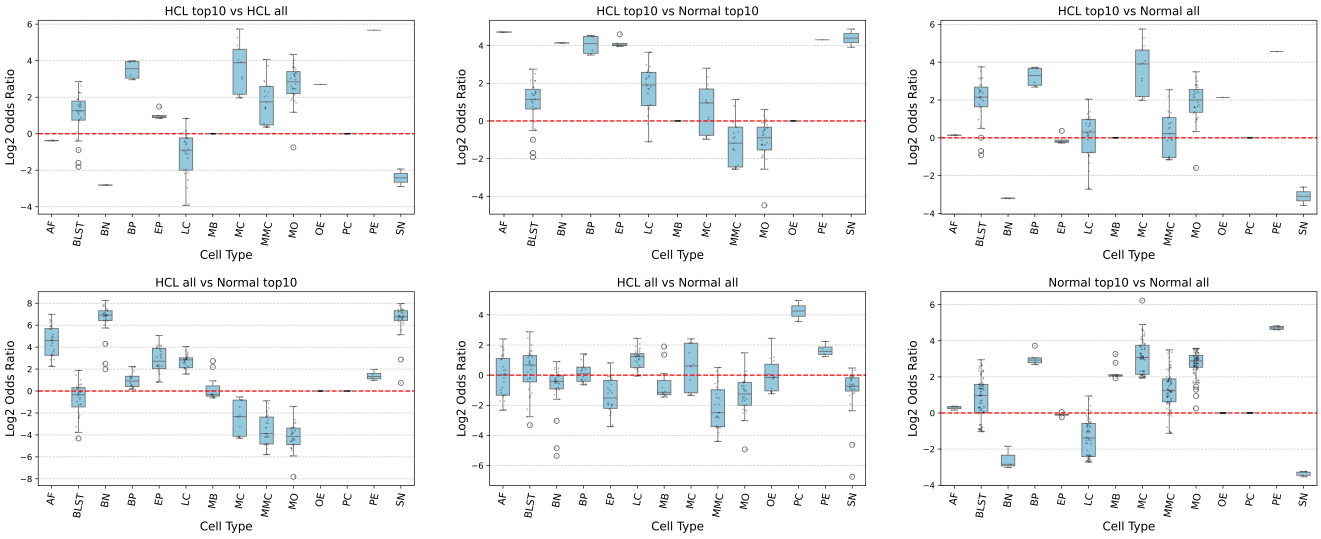

Figure Supplementary 6. Comparative Log odds ratios In each plot we are comparing the mean differentials of the patients taken just from the the two differential types and plotting the log odds ratios of the two.

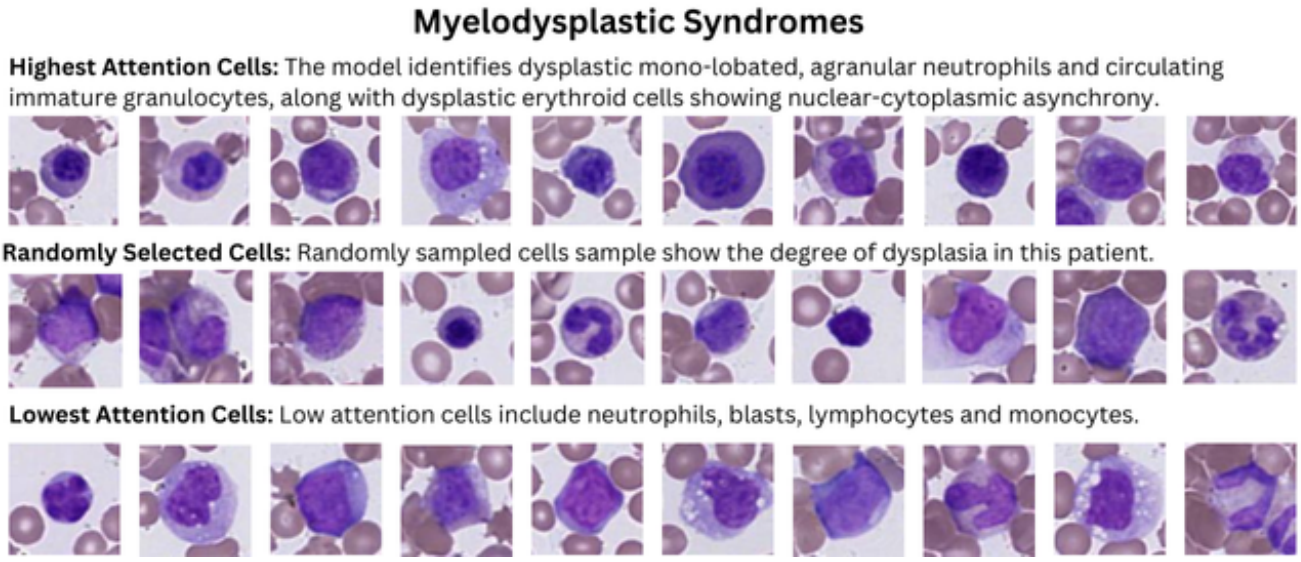

Figure Supplementary 7. MDS Attention Panel High-attention cells include dysplastic hypogranular neutrophils, monolobated neutrophils, dysplastic erythroblasts showing nuclear-cytoplasmic dyssynchrony, and abnormal immature granulocytes circulating in peripheral blood.

Table Supplementary 1. Different cytometry-based ML classifiers and their hyperparameters

Classifiers	Hyperparameters	Values
SVC	kernel	rbf (radial basis function)
	error (c)	[5,10, 20, 50, 100, 200]
	Decision Functions	OvO, OvR
Random Forest	n_estimators	range(100,300,30)
	max_depth	[6,8,10,12]
	min_sample_split	[4,6,8,10]
Logistic Regression	Objective Function	[newton_cg, liblinear]
	C	[1,10,20,30]
Multinomial Naive Bayes	α	[0.1,1,10]
	Prior Probabilities	[True, False]
XG Boost	n_estimators	range(90,160,20)
	max_depth	range(3,8,1)
	learning rate	[0.02,0.05]
	γ	[0.05,0.1,0.2]
	Objective Function	softmax
Multi-Layer Perceptron	Activation Function	Relu
	Optimizer	Adam
	Hidden Layers	(32,64,32)

Table Supplementary 2. Hyperparameters for MIL models training

Hyperparameters	**Values**
Learning Rate	1e-4
Epochs	300
Learning Rate Scheduler	CyclicLR, mode=triangular2
Loss Function	Cross Entropy Loss
Batch Size	16
Optimizer	Adagrad
Train, Validation, Test Split	0.55, 0.2, 0.25
GPUs	3 (RTX 3090 each)

[b]

(A)

Diagnosis	Number of Patients
Acute Leukemia	307
Normal	130
Myelodysplastic syndromes	95
Hairy Cell Leukemia	41

[b]

(B)

Standard Deviation

$$g\left(\{p_j\}\right) = \sqrt{\frac{1}{|j|}\sum_j \left(p_j - \bar{p}\right)^2}$$

Mean

$$g\left(\{p_j\}\right) = \frac{1}{|j|}\sum_j p_j$$

Generalized Mean

$$g\left(\{p_j\}\right) = \left(\frac{1}{|j|}\sum_j p_j^r\right)^{\frac{1}{r}}$$

Log Sum Exponentiation

$$g(\{p_j\}) = \frac{1}{r}\log\left(\frac{1}{|j|}\sum_j e^{r\cdot p_j}\right)$$

[b]

(C)

Aggregation group	Aggregation function
Normal	Mean, Standard Deviation, Max, Min
GM	Generalized mean with $r = 1.0, 2.5$ and 5.0
LSE	Log Sum Exponentiation with $r = 2.5$ and 5.0

Figure Supplementary 8. Combined tables of analysis. (A) Patient cohort by diagnosis. (B) Aggregation functions used in the analysis. (C) Summary of aggregation function experiments.Three groups of functions were experimented with for aggregation of cell embeddings. The performance difference across the different groups was negligible and attributed to stochasticity.

Table Supplementary 3. Table of experiments: Summary of diagnostic experiments

Experiments	Description
AL vs NL	Binary classifier of acute leukemia vs normal
MDS vs NL	Binary classifier of myelodysplastic syndromes vs normal
HCL vs NL	Binary classifier of hairy cell leukemia
Multiclass	4-way classification between acute leukemia, myelodysplastic syndrome, hairy cell leukemia and normal

(A)

Model Name	Description
Cytometry-Based ML Classifiers	The DeepHeme classifier sorts the cells into respective classes to create a cytometric vector of the relative proportion of each cell class. This vector is used to train ML classifiers to predict the diagnosis.
Gated MIL - ImageNet	The cell embeddings are created from an ImageNet-trained model and passed to Gated MIL for predictions.
CAREMIL - ImageNet	The cell embeddings are created from an ImageNet-trained model and passed to CAREMIL for predictions.
Gated MIL - Deep Heme	The cell embeddings from the DeepHeme model are passed to the Gated MIL model for predictions.
CAREMIL - Deep Heme	The cell embeddings from the DeepHeme model are passed to the CAREMIL model for predictions.
Gated MIL - UNI2-h	The cell embeddings from the UNI2-h model and passed to Gated MIL for predictions.
CAREMIL - UNI2-h	The cell embeddings from the UNI2-h model and passed to CAREMIL for predictions.
Gated MIL - Virchow2	The cell embeddings from the Virvchow2 model are passed to the Gated MIL model for predictions.
CAREMIL - Virchow2	The cell embeddings from the Virchow2 model are passed to the CAREMIL model for predictions.

Table Supplementary 4. Overview of Model Architectures and Performances. Model Architectures: A summary description of each model architecture used in the study.

Figure Supplementary 9. Cell types seen in the blood. Five prototypical cell types seen in peripheral blood smears are shown. Blasts are stem-cell-like cells that are increased in number in acute myeloid and lymphoid leukemias. The purple nucleus makes up nearly the entire cell (high nuclear:cytoplasmic ratio). The nuclear chromatin is fine, and small circles (nucleoli) can be seen within it. The cell is relatively large with almost no visible blue cytoplasm. Lymphocytes have darker, more mature nuclear chromatin. Monocytes have visible folds in the nucleus and spongy chromatin, blue-gray cytoplasm, and sometimes visible white vacuoles in the cytoplasm. Neutrophils and eosinophils have segmented nuclei. The eosinophil is notable for its bright pink cytoplasm.

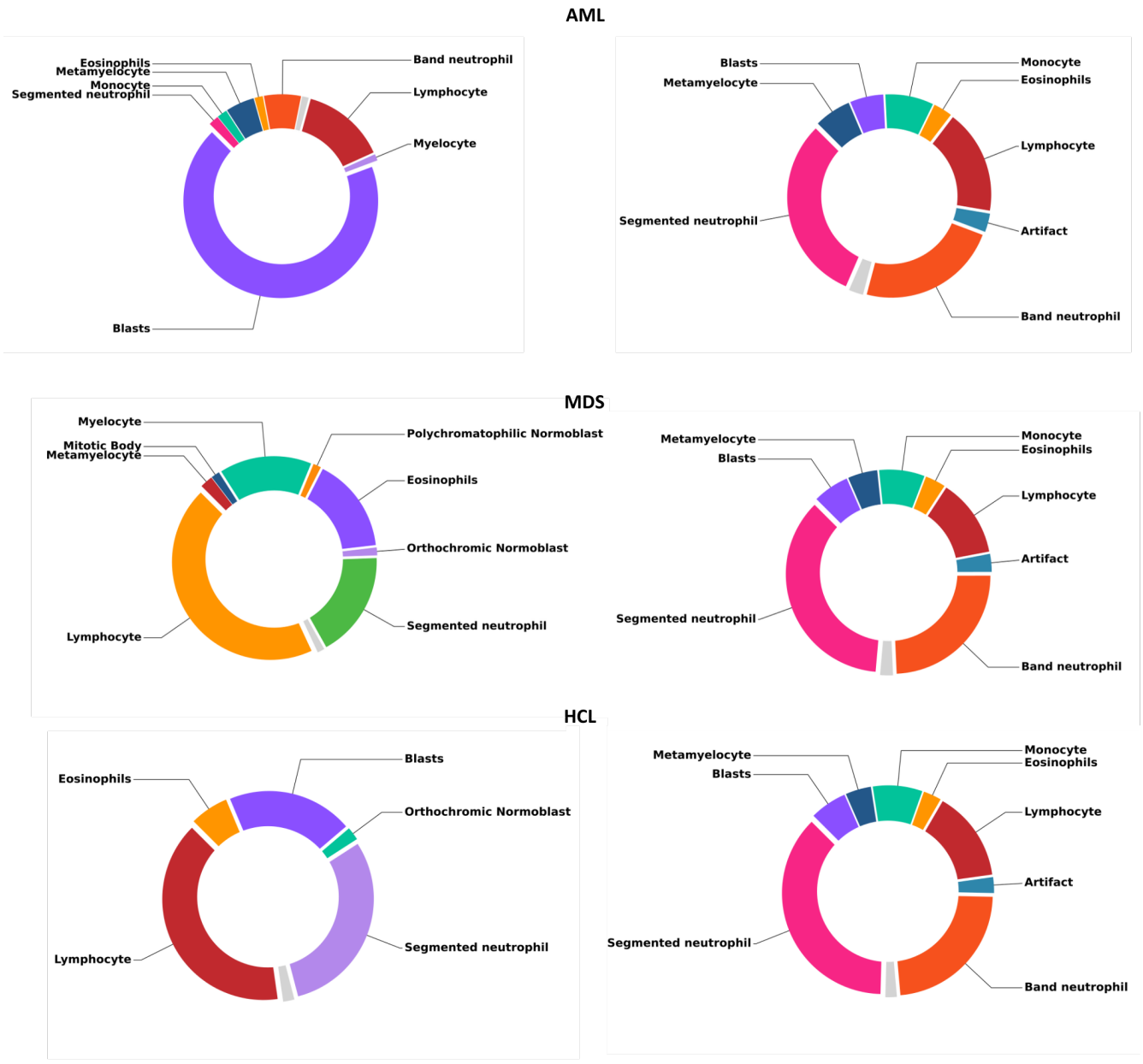

Figure Supplementary 10. <Greg>

(B)

Model Type - Encoder	AML vs NL	MDS vs NL	HCL vs NL	Multiclass
Cytometry-Based ML Classifiers	0.991 ± 0.00	0.877 ± 0.00	0.714 ± 0.00	0.892 ± 0.000
CAREMIL - Deep Heme	$\mathbf{0.999 \pm 0.002}$	0.891 ± 0.024	$\mathbf{0.945 \pm 0.029}$	0.916 ± 0.027
CAREMIL - Imagenet	0.750 ± 0.040	0.863 ± 0.066	0.627 ± 0.089	0.802 ± 0.046
CAREMIL - UNI2-h	0.884 ± 0.017	$\mathbf{0.896 \pm 0.009}$	0.917 ± 0.004	$\mathbf{0.923 \pm 0.023}$
CAREMIL - Virchow2	0.923 ± 0.016	0.758 ± 0.026	0.859 ± 0.014	0.873 ± 0.019
Gated MIL - Deep Heme	0.957 ± 0.010	0.829 ± 0.064	0.878 ± 0.019	0.919 ± 0.016
Gated MIL - Imagenet	0.770 ± 0.073	0.787 ± 0.066	0.616 ± 0.097	0.599 ± 0.062
Gated MIL - UNI2-h	0.773 ± 0.062	0.816 ± 0.088	0.715 ± 0.034	0.819 ± 0.033
Gated MIL - Virchow2	0.866 ± 0.005	0.823 ± 0.036	0.672 ± 0.038	0.806 ± 0.012

Table Supplementary 5. Model Performance: AUROC scores with standard deviations for each model architecture across diagnostic experiments. The highest-performing models for each experiment are highlighted in bold.