

Supplementary Material

CytoDiff: AI-Driven Cytomorphology Image Synthesis for Medical Diagnostics

Jan Carreras Boada^{1,4,*}, Rao Muhammad Umer^{1,*}, Carsten
Marr^{1,2,3}

¹ Institute of AI for Health, Helmholtz Zentrum München -
German Research Center
for Environmental Health, Neuherberg 85764, Germany

² Department of Medicine III, Ludwig-Maximilian-University
Hospital, Munich, Germany

³ DKTK, German Cancer Consortium, Germany

⁴ Escola Superior de Comerç Internacional, Universitat Pompeu
Fabra (ESCI-UPF),
Barcelona 08003, Spain

Supplementary Figures

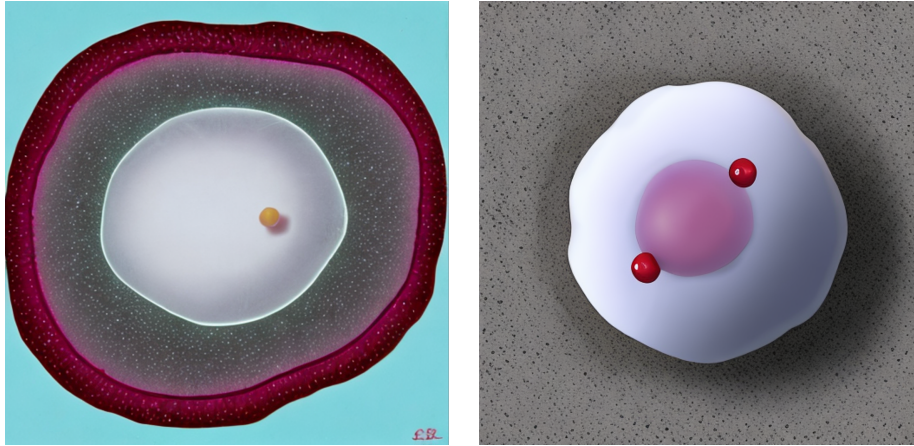


Figure S1: Synthetic eosinophil images generated with stable diffusion 2.1

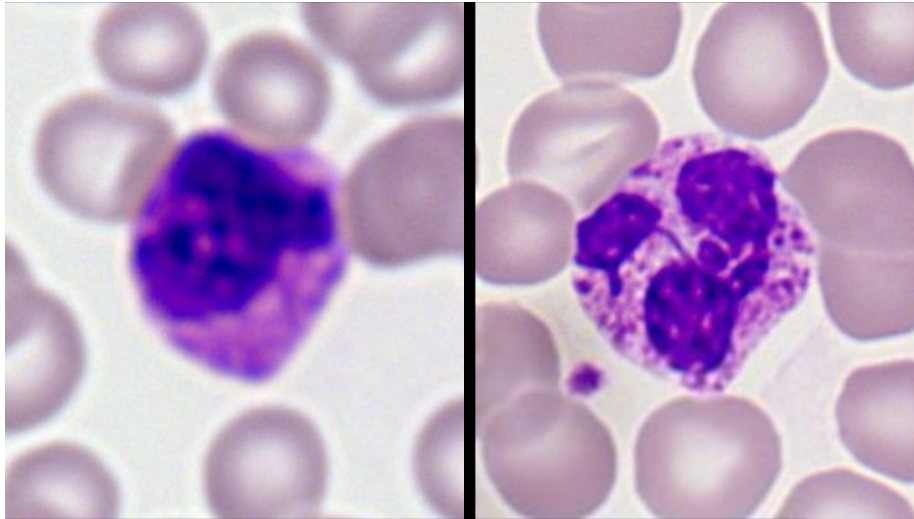


Figure S2: Quality comparison of real images from the Munich AML Morphology Dataset. This figure shows the first two images from the Basophils folder. The images demonstrate a significant quality difference, which reinforces the idea of doing a manual selection.

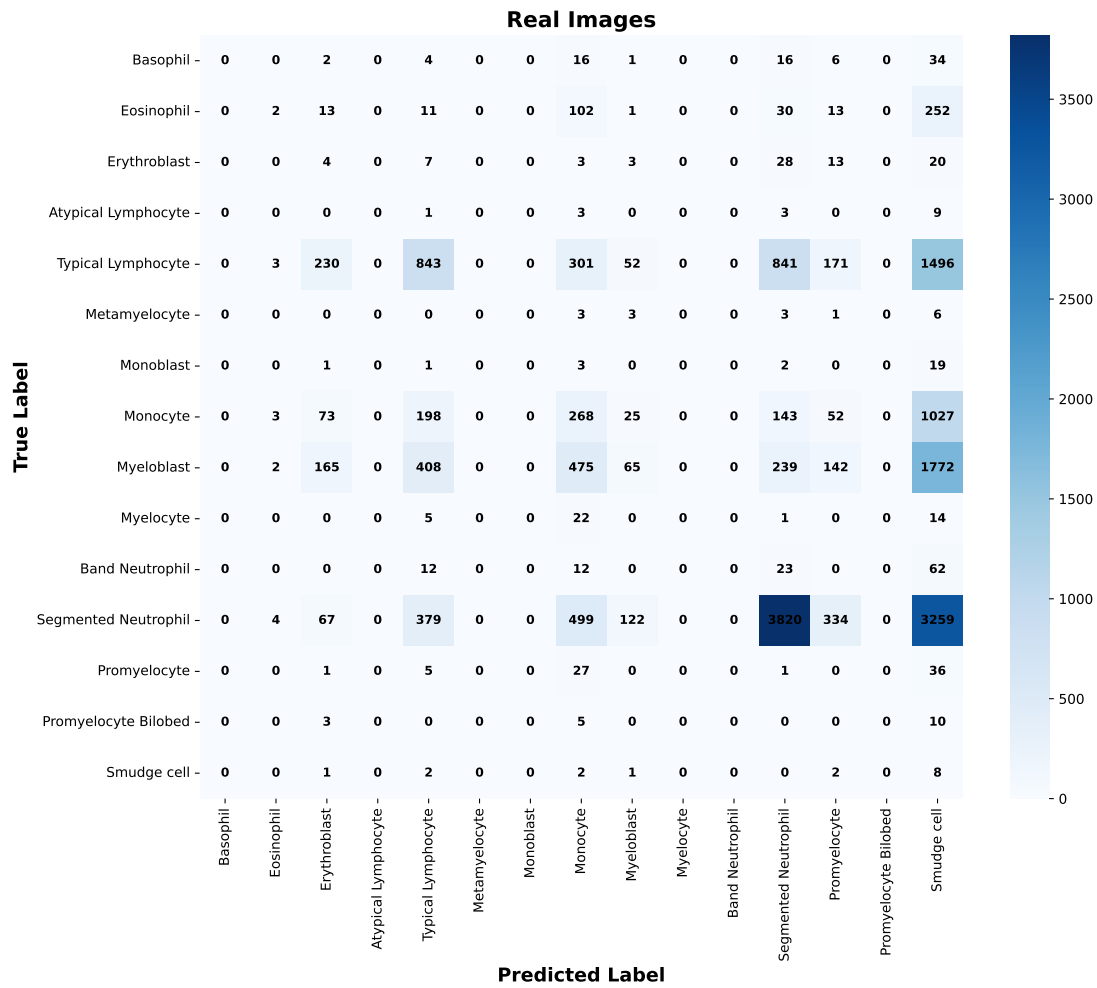


Figure S3: Confusion matrix of ResNet-50 trained on all real images. Results are aggregated over the test sets of all 5 cross-validation folds.

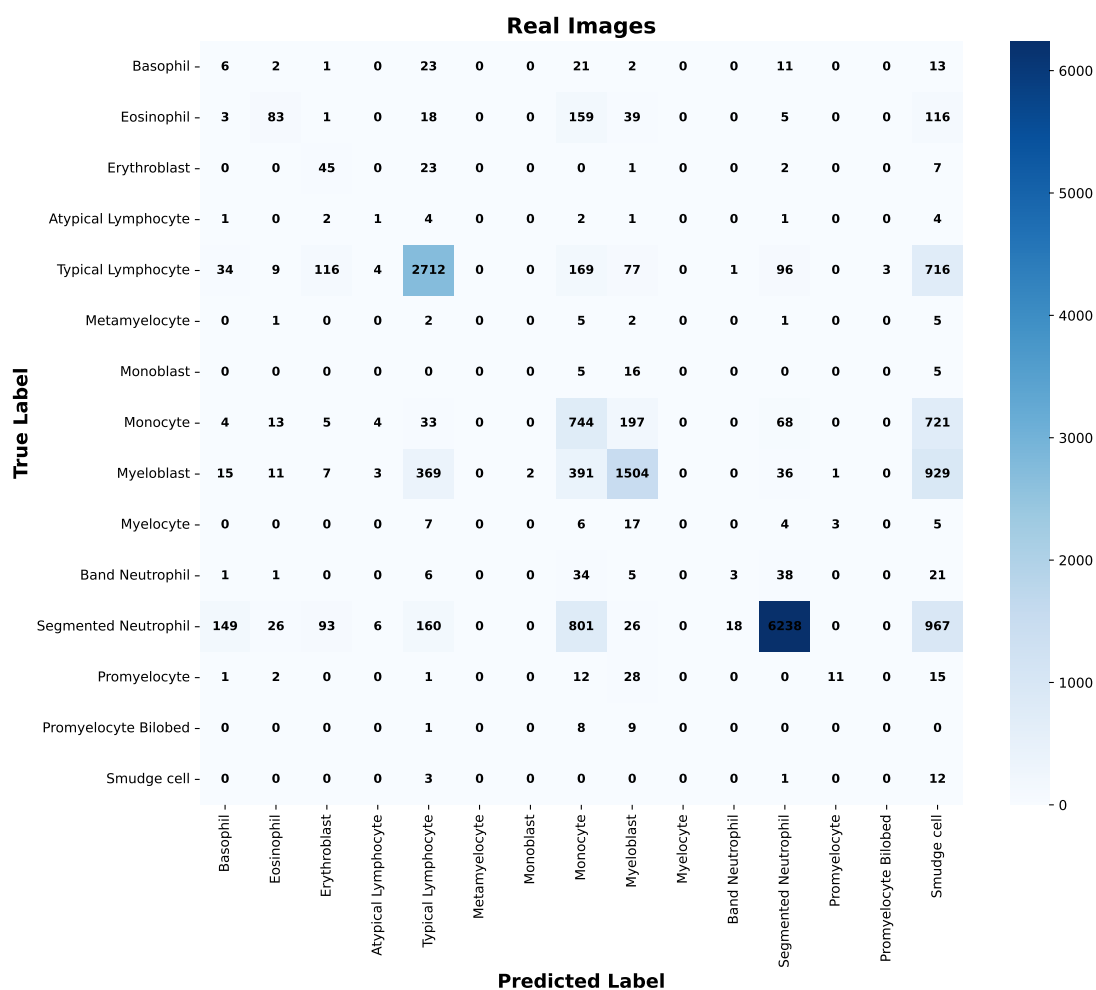


Figure S4: Confusion matrix of the fine-tuned CLIP model trained on all real images. Results are aggregated over the test sets of all 5 cross-validation folds.

Promyelocyte

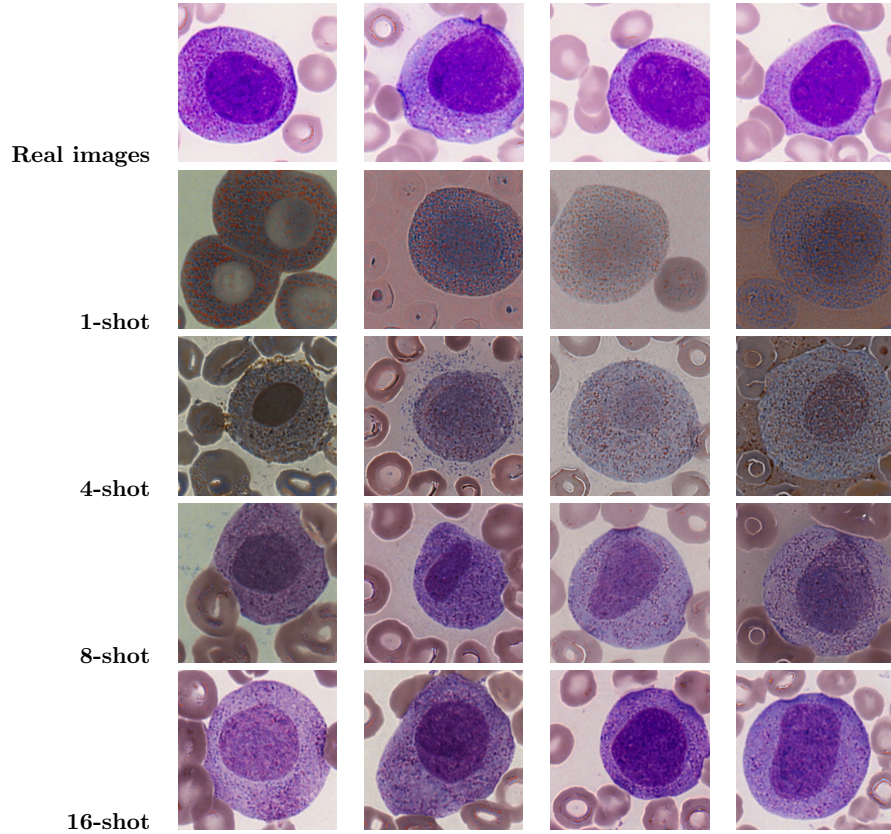


Figure S5: Real and synthetic Promyelocyte images generated using different shot conditions. Prompt used for generation: *"Photorealistic promyelocyte in peripheral blood smear, large round nucleus with visible nucleoli, deep blue cytoplasm with prominent azurophilic granules, Wright-Giemsa stain, ultra-detailed, medical cytology, 40x magnification, sharp focus, high-resolution."*

Segmented Neutrophil

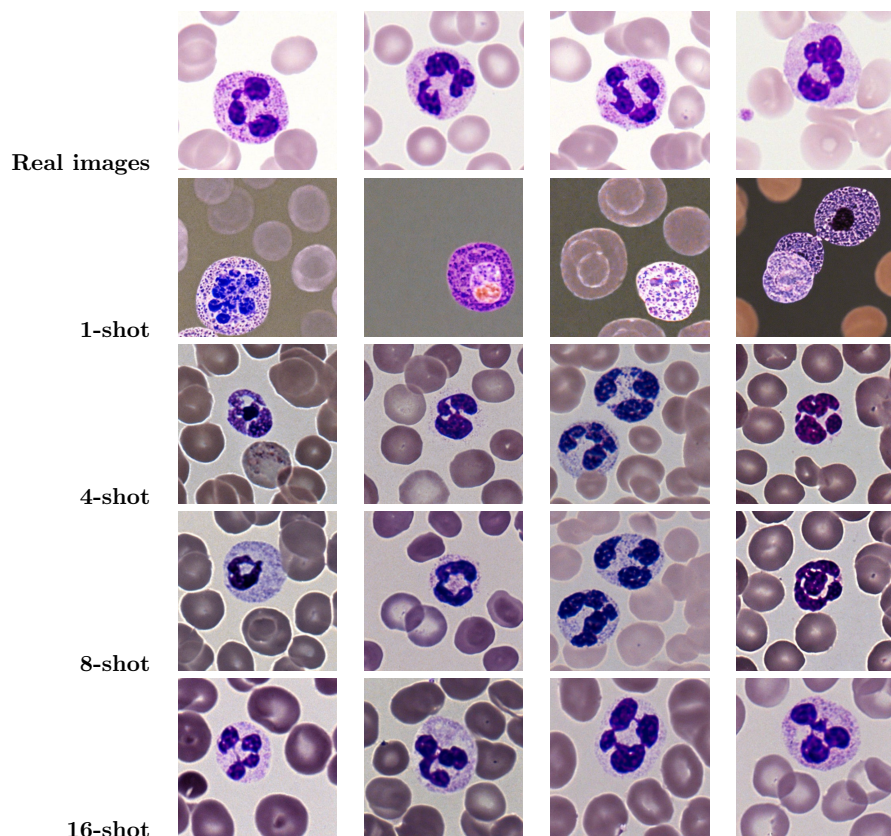


Figure S6: Real and synthetic Segmented Neutrophil images generated using different shot conditions. Prompt used for generation: *"Photorealistic segmented neutrophil cell under microscope, peripheral blood smear, multi-lobed nucleus (3/5 lobes) with thin chromatin strands, pale pink cytoplasm with fine granules, surrounded by red blood cells, medical cytology, high detail, clinical pathology, Wright-Giemsa stain, white background, soft lighting, 40x magnification, scientific illustration style, sharp focus, ultra-detailed, macro lens."*

Promyelocyte Bilobed

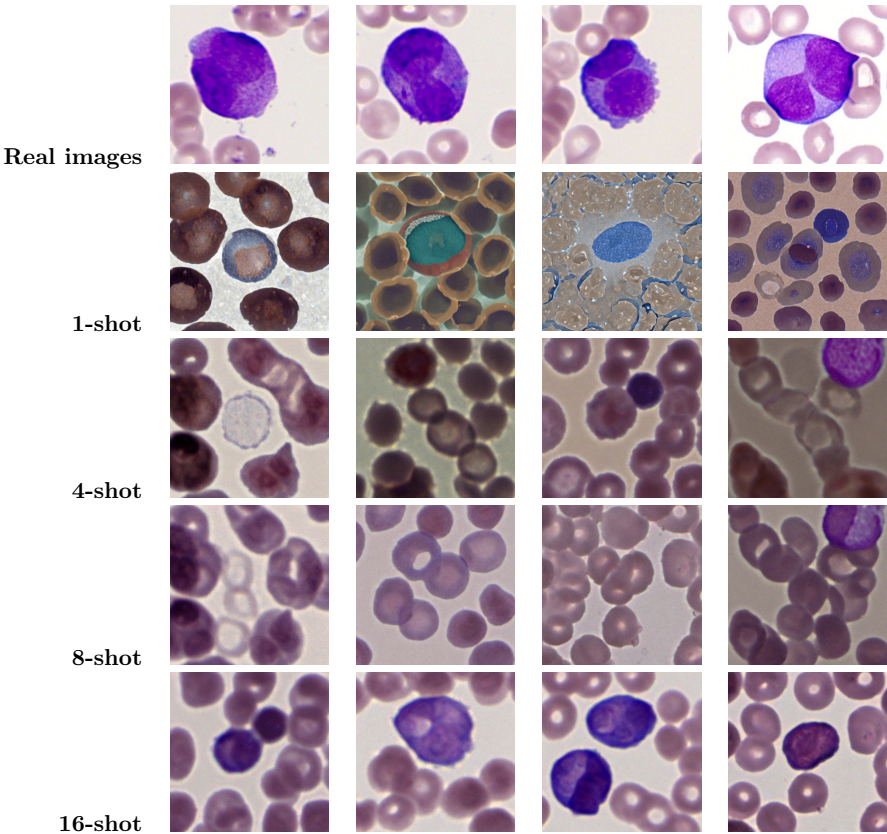


Figure S7: Real and synthetic Promyelocyte Bilobed images generated using different shot conditions. Prompt used for generation: *"Detailed microscopic image of a bilobed promyelocyte, peripheral blood smear, two distinct nuclear lobes, cytoplasm filled with azurophilic granules, Wright-Giemsa stain, high-resolution, medical cytology, 40x magnification, sharp focus, realistic texture."*

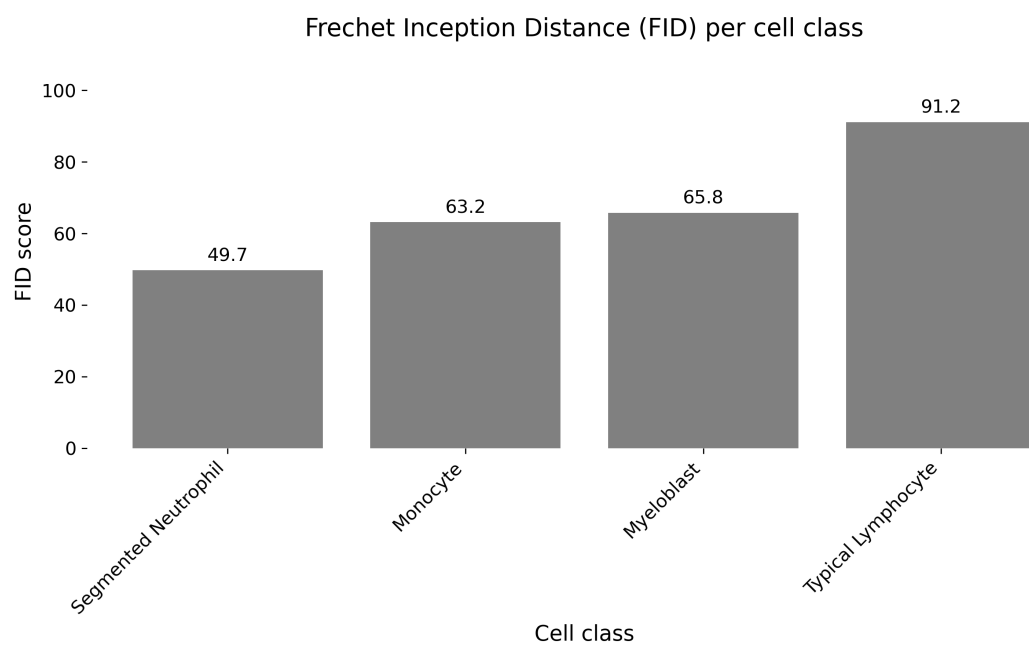


Figure S8: Bar plot showing the Fréchet inception distance (FID) of classes with more than 1,000 samples.

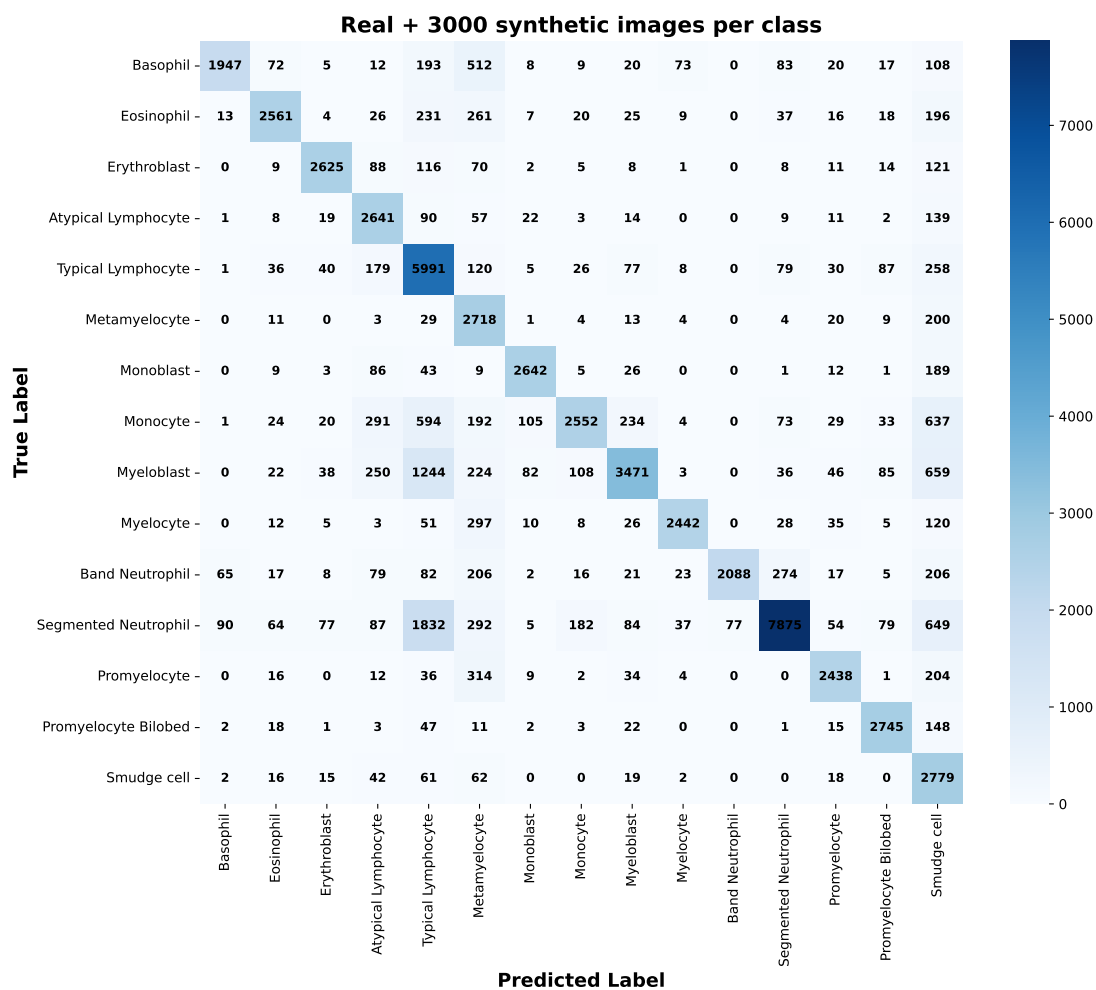


Figure S9: Confusion matrix of ResNet-50 trained on all real images and 3000 additional synthetic images per class. Results are aggregated over the test sets of all 5 cross-validation folds.

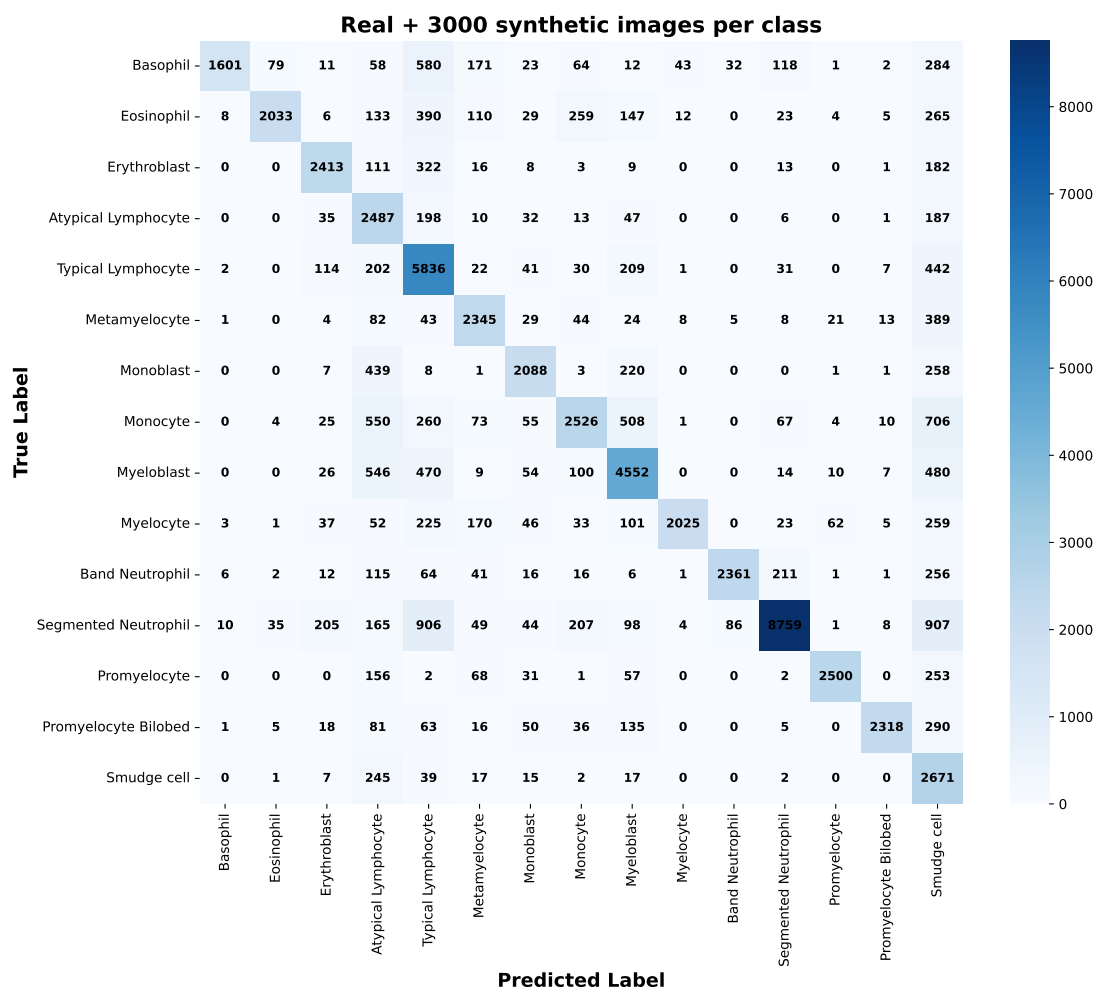


Figure S10: Confusion matrix of the fine-tuned CLIP model trained on all real images and 3000 additional synthetic images per class. Results are aggregated over the test sets of all 5 cross-validation folds.

Table S1: Classification performance of ResNet-50 and CLIP models trained on 45,000 synthetic images only (3,000 per class generated by CytoDiff) of 15 blood cell classes. The dataset is split into 60% training, 20% validation, and 20% test sets. Results are averaged over 5 cross-validation folds.

Model (Split)	Accuracy	F1 macro	AUC
ResNet-50 (Validation)	0.93 $\pm$ 0.01	0.93 $\pm$ 0.01	0.99 $\pm$ 0.00
ResNet-50 (Test)	0.81 $\pm$ 0.01	0.82 $\pm$ 0.00	0.99 $\pm$ 0.00
CLIP (Validation)	0.87 $\pm$ 0.01	0.87 $\pm$ 0.01	0.99 $\pm$ 0.00
CLIP (Test)	0.77 $\pm$ 0.01	0.78 $\pm$ 0.01	0.98 $\pm$ 0.00

Table S2: Classification performance of ResNet-50 trained with different amounts of synthetic images per class to the real dataset. A random classifier baseline is included for comparison. Results are averaged over 5 cross-validation folds.

# Images	Accuracy		F1 Macro		AUC	
	Validation	Test	Validation	Test	Validation	Test
Random	0.07 $\pm$ 0.00	0.07 $\pm$ 0.00	0.07 $\pm$ 0.00	0.07 $\pm$ 0.00	0.50 $\pm$ 0.00	0.50 $\pm$ 0.00
0	0.40 $\pm$ 0.07	0.27 $\pm$ 0.09	0.08 $\pm$ 0.03	0.07 $\pm$ 0.01	0.57 $\pm$ 0.07	0.56 $\pm$ 0.03
100	0.62 $\pm$ 0.06	0.49 $\pm$ 0.05	0.41 $\pm$ 0.02	0.35 $\pm$ 0.05	0.90 $\pm$ 0.01	0.82 $\pm$ 0.02
200	0.65 $\pm$ 0.02	0.52 $\pm$ 0.03	0.53 $\pm$ 0.02	0.46 $\pm$ 0.02	0.93 $\pm$ 0.01	0.87 $\pm$ 0.01
300	0.65 $\pm$ 0.02	0.53 $\pm$ 0.03	0.59 $\pm$ 0.02	0.50 $\pm$ 0.02	0.94 $\pm$ 0.01	0.89 $\pm$ 0.01
400	0.67 $\pm$ 0.02	0.52 $\pm$ 0.01	0.61 $\pm$ 0.04	0.51 $\pm$ 0.02	0.95 $\pm$ 0.01	0.89 $\pm$ 0.01
500	0.69 $\pm$ 0.02	0.55 $\pm$ 0.04	0.66 $\pm$ 0.03	0.56 $\pm$ 0.02	0.96 $\pm$ 0.01	0.91 $\pm$ 0.01
600	0.71 $\pm$ 0.02	0.58 $\pm$ 0.02	0.69 $\pm$ 0.02	0.58 $\pm$ 0.01	0.97 $\pm$ 0.01	0.92 $\pm$ 0.01
700	0.71 $\pm$ 0.01	0.60 $\pm$ 0.03	0.70 $\pm$ 0.01	0.61 $\pm$ 0.01	0.97 $\pm$ 0.01	0.93 $\pm$ 0.01
800	0.73 $\pm$ 0.01	0.62 $\pm$ 0.01	0.72 $\pm$ 0.02	0.61 $\pm$ 0.02	0.97 $\pm$ 0.01	0.94 $\pm$ 0.01
900	0.75 $\pm$ 0.01	0.66 $\pm$ 0.03	0.76 $\pm$ 0.02	0.66 $\pm$ 0.01	0.98 $\pm$ 0.01	0.94 $\pm$ 0.01
1000	0.77 $\pm$ 0.02	0.64 $\pm$ 0.02	0.77 $\pm$ 0.02	0.65 $\pm$ 0.02	0.98 $\pm$ 0.01	0.94 $\pm$ 0.01
2000	0.82 $\pm$ 0.02	0.72 $\pm$ 0.01	0.84 $\pm$ 0.01	0.74 $\pm$ 0.01	0.99 $\pm$ 0.01	0.96 $\pm$ 0.01
3000	0.86 $\pm$ 0.01	0.75 $\pm$ 0.01	0.87 $\pm$ 0.01	0.77 $\pm$ 0.01	0.99 $\pm$ 0.01	0.98 $\pm$ 0.01
4000	0.87 $\pm$ 0.01	0.76 $\pm$ 0.01	0.88 $\pm$ 0.01	0.77 $\pm$ 0.01	0.99 $\pm$ 0.01	0.98 $\pm$ 0.01
5000	0.87 $\pm$ 0.01	0.77 $\pm$ 0.01	0.89 $\pm$ 0.01	0.78 $\pm$ 0.01	0.99 $\pm$ 0.01	0.98 $\pm$ 0.01

Table S3: Classification performance of the CLIP model trained with different amounts of synthetic images per class to the real dataset. Results are averaged over 5 cross-validation folds.

# Images	Accuracy		F1 Macro		AUC	
	Validation	Test	Validation	Test	Validation	Test
0	0.78 $\pm$ 0.01	0.62 $\pm$ 0.03	0.27 $\pm$ 0.05	0.23 $\pm$ 0.02	0.88 $\pm$ 0.02	0.83 $\pm$ 0.02
100	0.71 $\pm$ 0.03	0.62 $\pm$ 0.02	0.44 $\pm$ 0.03	0.43 $\pm$ 0.02	0.92 $\pm$ 0.01	0.88 $\pm$ 0.01
200	0.73 $\pm$ 0.02	0.65 $\pm$ 0.02	0.59 $\pm$ 0.02	0.54 $\pm$ 0.02	0.95 $\pm$ 0.01	0.91 $\pm$ 0.01
300	0.73 $\pm$ 0.01	0.65 $\pm$ 0.01	0.62 $\pm$ 0.04	0.58 $\pm$ 0.01	0.95 $\pm$ 0.01	0.92 $\pm$ 0.01
400	0.73 $\pm$ 0.02	0.67 $\pm$ 0.02	0.65 $\pm$ 0.03	0.60 $\pm$ 0.02	0.96 $\pm$ 0.01	0.93 $\pm$ 0.01
500	0.74 $\pm$ 0.02	0.68 $\pm$ 0.01	0.69 $\pm$ 0.03	0.64 $\pm$ 0.02	0.96 $\pm$ 0.01	0.94 $\pm$ 0.01
600	0.75 $\pm$ 0.02	0.69 $\pm$ 0.02	0.71 $\pm$ 0.02	0.65 $\pm$ 0.01	0.97 $\pm$ 0.01	0.94 $\pm$ 0.01
700	0.77 $\pm$ 0.01	0.70 $\pm$ 0.01	0.73 $\pm$ 0.02	0.66 $\pm$ 0.02	0.97 $\pm$ 0.01	0.95 $\pm$ 0.01
800	0.76 $\pm$ 0.02	0.69 $\pm$ 0.01	0.73 $\pm$ 0.02	0.67 $\pm$ 0.01	0.97 $\pm$ 0.01	0.95 $\pm$ 0.01
900	0.78 $\pm$ 0.01	0.71 $\pm$ 0.01	0.75 $\pm$ 0.02	0.69 $\pm$ 0.01	0.97 $\pm$ 0.01	0.95 $\pm$ 0.01
1000	0.77 $\pm$ 0.01	0.71 $\pm$ 0.01	0.75 $\pm$ 0.01	0.69 $\pm$ 0.01	0.97 $\pm$ 0.01	0.95 $\pm$ 0.01
2000	0.82 $\pm$ 0.01	0.72 $\pm$ 0.01	0.81 $\pm$ 0.01	0.72 $\pm$ 0.01	0.98 $\pm$ 0.01	0.96 $\pm$ 0.01
3000	0.83 $\pm$ 0.01	0.73 $\pm$ 0.01	0.83 $\pm$ 0.01	0.74 $\pm$ 0.01	0.99 $\pm$ 0.01	0.97 $\pm$ 0.01
4000	0.85 $\pm$ 0.01	0.75 $\pm$ 0.01	0.86 $\pm$ 0.01	0.76 $\pm$ 0.01	0.99 $\pm$ 0.01	0.97 $\pm$ 0.01
5000	0.85 $\pm$ 0.02	0.77 $\pm$ 0.01	0.85 $\pm$ 0.02	0.77 $\pm$ 0.01	0.99 $\pm$ 0.01	0.97 $\pm$ 0.01