SF-DINO: Spatial-Frequency Adapted Foundation Model for Multiple Myeloma Diagnosis

Zhaoyi Ye[1,†], Hui Shen[2,†], Fuling Zhou[2,†], Xinglong Hu[1], Du Wang[1], Liye Mei[3,4,5,*], and Cheng Lei[1,2,6,*]

[1] School of Integrated Circuits, Wuhan University, Wuhan, China
liyemei@whu.edu.cn, leicheng@whu.edu.cn
[2] Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan, China
[3] The Institute of Technological Sciences, Wuhan University, Wuhan, China
[4] School of Computer Science and Artificial Intelligence, Hubei University of Technology, Wuhan, China
[5] Xinguang Zhiying (Wuhan) Technology Co., Ltd., Wuhan, China
[6] Shenzhen Institute of Wuhan University, Shenzhen, China

Abstract. Microscopic analysis of bone marrow aspirate smears is a critical routine procedure for the initial screening and diagnosis of multiple myeloma (MM). However, manual interpretation is subjective and labor-intensive, further complicated by the fine-grained morphological heterogeneity and inter-class similarities of bone marrow cells. While vision foundation models like DINOv3 have revolutionized natural image understanding, they lack the domain-specific inductive bias required to capture subtle pathological textures in medical cytology. To address these challenges, we first construct a comprehensive, expert-annotated MM diagnosis dataset encompassing six fine-grained cell categories to reflect real-world clinical scenarios. We then propose SF-DINO, a spatial-frequency adapted foundation model that efficiently adapts DINOv3 for MM diagnosis. Specifically, we introduce a parallel spatial-morphology (SM) adapter to inject high-frequency local morphological cues into the backbone without disrupting pre-trained semantics. Additionally, we devise a frequency-selective hashing attention (FSHA) module to model global long-range dependencies efficiently via spectral clustering. Extensive experiments demonstrate that SF-DINO significantly outperforms six state-of-the-art methods, showing impressive generalization and establishing a new benchmark for automated, fine-grained MM diagnosis. Our code is available at https://github.com/yzygit1230/SF-DINO.

Keywords: Multiple Myeloma · Spatial-Morphology Adapter · Frequency-Selective Hashing Attention

1 Introduction

Multiple myeloma (MM) is the second most frequent hematologic malignancy globally, accounting for approximately 10% of all blood cancers [15]. Despite

[†] These authors contributed equally to this work.

advances in therapeutic strategies, MM remains largely incurable with high relapse rates, making early and accurate diagnosis pivotal for patient survival [14]. While cytomorphological examination of bone marrow aspirate smears remains the gold standard for initial screening and staging, manual interpretation is notoriously labor-intensive, time-consuming, and prone to inter-observer variability [21]. This diagnostic process is further complicated by the fine-grained nature of bone marrow cytology: cells within the same lineage exhibit significant morphological heterogeneity due to diverse maturation stages, while different lineages often share subtle visual similarities [2,6]. Consequently, there is an urgent imperative to develop an intelligent, automated method capable of parsing these complex morphological details to assist clinicians in achieving objective and efficient MM diagnosis.

Deep learning has revolutionized automated hematological analysis, showing remarkable potential for reducing manual workload [11,12,23]. While object detection-based frameworks have been employed to localize and count plasma cells directly from whole-slide images, these methods often prioritize localization speed over the high-resolution feature extraction required for fine-grained morphological differentiation [13,8]. Furthermore, most existing studies simplify the diagnostic task into binary classification (normal cells and myeloma cells) or focus on a limited set of cell lineages, failing to account for the complex confounding factors introduced by erythroid and granulocytic series in the real-world bone marrow microenvironment [16,5]. Recently, vision foundation models, particularly the DINOv3 [18] utilizing self-supervised learning, have demonstrated dominance in natural image understanding. DinoBloom [7] has emerged as a pioneering leukocyte classification foundation model, However, its scope leaves the specialized, fine-grained differentiation of monoclonal plasma cells required for MM diagnosis largely unexplored. Furthermore, while clinical MM diagnosis depends on plasma cell proportions, accurate fine-grained cell classification remains a prerequisite for reliable quantification. However, the direct deployment of foundation models for this task is challenging: their pre-training on natural scenes lacks the domain-specific inductive bias necessary to capture high-frequency nuclear chromatin textures, often leading to suboptimal generalization when confronting the subtle morphological variances of pathological cells.

To address the aforementioned dilemmas, we first construct a comprehensive MM diagnostic dataset reflecting the real-world heterogeneity of the bone marrow microenvironment, encompassing six fine-grained categories: monoclonal plasma, erythroid, granulocytic, lymphoid, monocytic, and other cells. Building upon the robust feature representation capabilities of the DINOv3 model, we propose a spatial-frequency adapted foundation model (SF-DINO) for fine-grained MM diagnosis. Specifically, to bridge the domain gap between natural and cell morphology images, we devise a lightweight spatial-morphology (SM) adapter to capture local spatial textures critical for cell differentiation without disrupting pre-trained semantic knowledge. Furthermore, we introduce a frequency-selective hashing attention (FSHA) module to efficiently model global semantic dependencies and long-range interactions via hash-based clustering. Extensive experiments

demonstrate that our proposed SF-DINO significantly outperforms six state-of-the-art competitive methods, establishing a new benchmark for automated MM diagnosis. The main contributions are summarized as follows:

1) We formally model the task of fine-grained multiple myeloma diagnosis, a clinically realistic yet challenging setting that accounts for complex confounding factors in the bone marrow, which has been largely unexplored. **2)** We construct a high-quality, expert-annotated dataset comprising 6,961 single-cell images, providing a solid foundation for evaluating algorithmic robustness in real-world scenarios. **3)** We propose SF-DINO, a parameter-efficient framework that synergizes local morphological adaptation and global semantic reasoning. By integrating the SM adapter and FSHA module, our method effectively unleashes the power of vision foundation models for cytopathology.

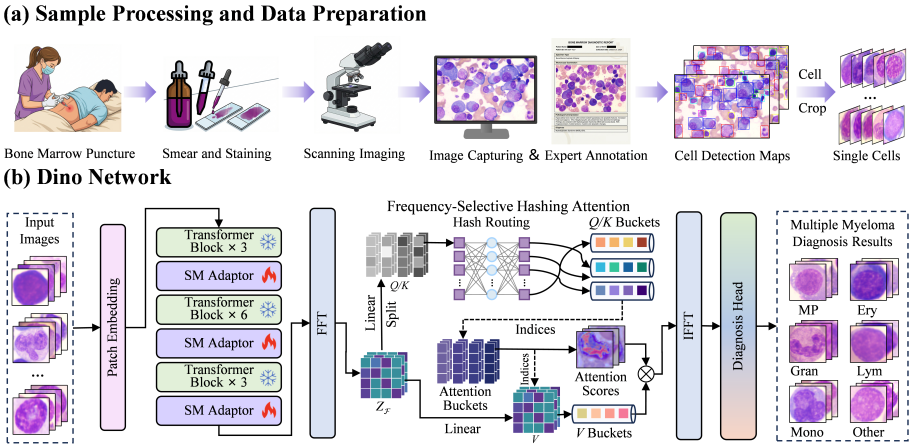

Fig. 1. Overview of the proposed pipeline. (a) Sample processing and data preparation of multiple myeloma cell images. (b) Flow diagram of the SF-DINO model.

2 Materials and Methods

2.1 Sample Processing and Data Preparation

In this study, we retrospectively collected bone marrow aspirate samples from clinical patients diagnosed with MM at Zhongnan Hospital of Wuhan University, under the approval of the institutional Medical Ethics Committee (No. 2019065). The inclusion criteria require a confirmed diagnosis of MM, active bone marrow hyperplasia, and a bone marrow plasma cell proportion $\geq 10\%$. We adhere strictly to the principles of the Declaration of Helsinki and obtain written informed consent from all participants for the clinical use of their microscopy images.

As illustrated in Fig. 1 (a), the data preparation workflow consists of three main stages: sample preparation, scanning imaging, and dataset construction. First, we perform bone marrow aspiration to extract less than 0.2 ml of bone marrow fluid and prepare approximately 7 smears per patient. After the smears air-dry, we add 2 ml of Wright-Giemsa staining solution followed by 2-3 ml of phosphate buffer saline (pH 6.4–6.8). We mix the fluids thoroughly and allow the slides to stain for 15–20 minutes at room temperature before rinsing and drying them. Next, we insert the dried, stained smears into an Olympus BX53 upright biological microscope. We scan the samples using a 100× oil immersion objective (NA: 1.25, WD: 0.15). Through this microscopic imaging process, we acquire 400 bone marrow images with a resolution of 2376×1824 pixels.

To construct the fine-grained MM diagnosis dataset, we recruit clinical experts from Zhongnan Hospital of Wuhan University to annotate bone marrow images, specifying the location and category of each cell. Note that single cell images do not allow patient-specific tracking. Subsequently, we utilize the YOLO algorithm [19] to automatically identify and crop individual cell images based on expert annotations. This process yields 6,961 single-cell images, which hematopathology experts review and verify to ensure reliability. The final dataset consists of six categories: monoclonal plasma (2,489), erythroid (1,081), granulocytic (1,673), lymphoid (528), monocytic (159), and other cells (1,031).

2.2 Spatial-Frequency Adapted Foundation Model

Due to the subtle intra-class variances and high inter-class similarities inherent in morphological cell images, standard vision foundation models often struggle to capture the fine-grained details for accurate diagnosis. As illustrated in Fig. 1(b), our proposed SF-DINO framework synergizes the generalized semantic ability of a frozen DINOv3 backbone with domain-specific adaptability for fine-grained MM diagnosis. Specifically, we first transform input cell image into a sequence of tokens via patch embedding. Subsequently, we introduce a parallel SM adapter to each transformer block, utilizing depth-wise convolutions to inject local morphological cues into the semantic stream without disrupting pretrained knowledge. To further capture global context efficiently, the aggregated features are processed by a FSHA module, which models long-range dependencies via spectral clustering in the frequency domain. Finally, we feed refined representations into a diagnosis head that dynamically fuses learnable prototype similarities with linear classification scores to ensure robust decision boundaries for the six fine-grained cell categories.

Spatial-Morphology Adapter As shown in Fig. 1(b), we employ the frozen DINOv3 ViT-B backbone to leverage its robust semantic priors. However, its global self-attention mechanism inherently lacks the inductive bias to capture high-frequency morphological details essential for fine-grained cell differentiation. To bridge this gap efficiently, we introduce a lightweight SM adapter inserted in parallel with each transformer block. The SM-Adapter utilizes a bottleneck structure with depth-wise convolution (DWConv) to explicitly inject local

spatial cues into the semantic stream. Formally, let Z_{l-1} be the input to the l-th layer. The output Z_l fuses the frozen transformer branch $\mathcal{T}_l$ and the learnable SM adapter branch as follows:

$$Z_l = \mathcal{T}_l(Z_{l-1}) + \lambda_l \cdot \mathcal{W}_{up}\left(\text{GELU}\left(\text{DWConv}\left(\mathcal{W}_{down}(Z_{l-1})\right)\right)\right) \tag{1}$$

where $\mathcal{W}_{down}$ and $\mathcal{W}_{up}$ denote linear projections for dimension reduction and restoration, GELU$(\cdot)$ is the GELU activation, and λ_l is a learnable scaling factor initialized to 10^{-4} to ensure training stability.

Frequency-Selective Hashing Attention Module Complementing the local SM-Adapter, we design a FSHA module to capture global structural dependencies and periodic textural patterns intrinsic to cell pathology. As illustrated in Fig. 1(b), FSHA operates in the spectral domain to disentangle complex chromatin textures. Firstly, the input feature Z is transformed into the frequency domain via Fast Fourier Transform (FFT). The spectral representation $Z_{\mathcal{F}}$ is then processed through a dual-path mechanism to perform content-aware filtering:

Hash Routing Branch: A linear split projection maps $Z_{\mathcal{F}}$ into query (Q) and key (K) embeddings. To identify spectrally correlated components, these embeddings undergo hash routing by projecting vectors onto a random rotation matrix R. The process of hash routing $h(x)$ for a spectral token x is generated by determining its hyper-plane region, which is calculated as follows:

$$h(x) = \arg\max([xR; -xR]) \tag{2}$$

where $[;]$ denotes concatenation covering the full hypersphere. Tokens sharing identical hash codes are clustered into Q/K buckets. Crucially, this sorting process generates a set of permutation indices, which record the spatial mapping of spectral components to their respective buckets. By applying these indices to reorganize the spectral embeddings, we formally obtain attention buckets, which serve as localized receptive fields for the subsequent sparse attention scores.

Value Content Branch: Simultaneously, a parallel linear layer projects $Z_{\mathcal{F}}$ into value V embeddings. To align with routing logic, these values are re-ordered using the indices derived from the hash routing branch, resulting in aligned V buckets.

Within the attention buckets, we compute sparse attention scores S to quantify the spectral relevance between tokens. These scores are then utilized to adaptively aggregate the aligned V buckets, effectively filtering out high-frequency noise while focusing on dominant global patterns. To synthesize robust features and mitigate boundary effects from hashing, we fuse the outputs from n_{hash} rounds using a confidence-weighted sum. Finally, the refined spectral features are transformed back to the spatial domain via inverse FFT (IFFT). The above processes can be calculated as follows::

$$Z_{out} = \text{IFFT}\left(\sum_{r=1}^{n_{hash}} \text{Unsort}_{\mathcal{I}_r}(S_r \cdot V_r)\right) \tag{3}$$

where S_r and V_r denote the attention scores and value buckets in the r-th round, respectively, and $\text{Unsort}_{\mathcal{I}_r}(\cdot)$ restores the original spatial order based on the permutation indices.

3 Experiments and Results

3.1 Implementation Details

We conduct experiments on the PyTorch framework with an NVIDIA A6000 GPU. To ensure robust and statistically reliable results, we adopt a 5-fold cross-validation strategy for our constructed MM diagnosis dataset. To evaluate cross-domain transferability, we retrain all models on the external Acevedo dataset [1], partitioning it into training, validation, and testing sets with a ratio of 3:1:1. We resize the input images to a resolution of 448×448 pixels. we use data augmentations including random horizontal and vertical flips, as well as random rotations. We train the model for 300 epochs with a batch size of 24. Furthermore, we utilize the Adam optimizer to update the network parameters, with an initial learning rate of 5×10^{-5}.

We utilize accuracy, precision, recall, and F1 to evaluate the performance of MM diagnosis [3,22].

3.2 Performance of Multiple Myeloma Diagnosis

Fig. 2 presents the multi-dimensional effectiveness of SF-DINO in fine-grained MM diagnosis. The t-SNE visualization [10] in Fig. 2(a) shows a clear separation of the six different clusters from the others, confirming that SF-DINO learns highly discriminative representations to maximize inter-class variance. Fig. 2(b) details the class-wise performance. SF-DINO achieves a remarkable accuracy of 95.4% for monoclonal plasma (MP) cells, significantly minimizing false negative risks. While slight confusion exists between the morphologically similar lymphoid (Lym) and monocytic (Mono) cells, the model maintains competitive performance (over 71%) for these challenging classes. Besides, the precision-recall (PR) and receiver operating characteristic (ROC) curves in Fig. 2(c-d) demonstrate robust MM diagnostic capability, yielding an impressive macro-average area under curve (AUC) of 95.90% and a dominant 95.97% AUC for granulocytic (Gran) cells. To validate the model's interpretability, we visualize the class activation maps using Grad-CAM [17] in Fig. 2(e). Obviously, the high-attention regions precisely align with biologically salient structures, such as the eccentric nuclei of MP cells and the cytoplasmic granules of granulocytic cells.

3.3 Comparative Analysis

As shown in Table 1, we benchmark SF-DINO against six SOTA methods, including InceptionNeXt [24], RMT [4], LSNet [20], MedMamba [25], Swin-UMamba [9], and DINOv3 [18], training each comparative method according to the original

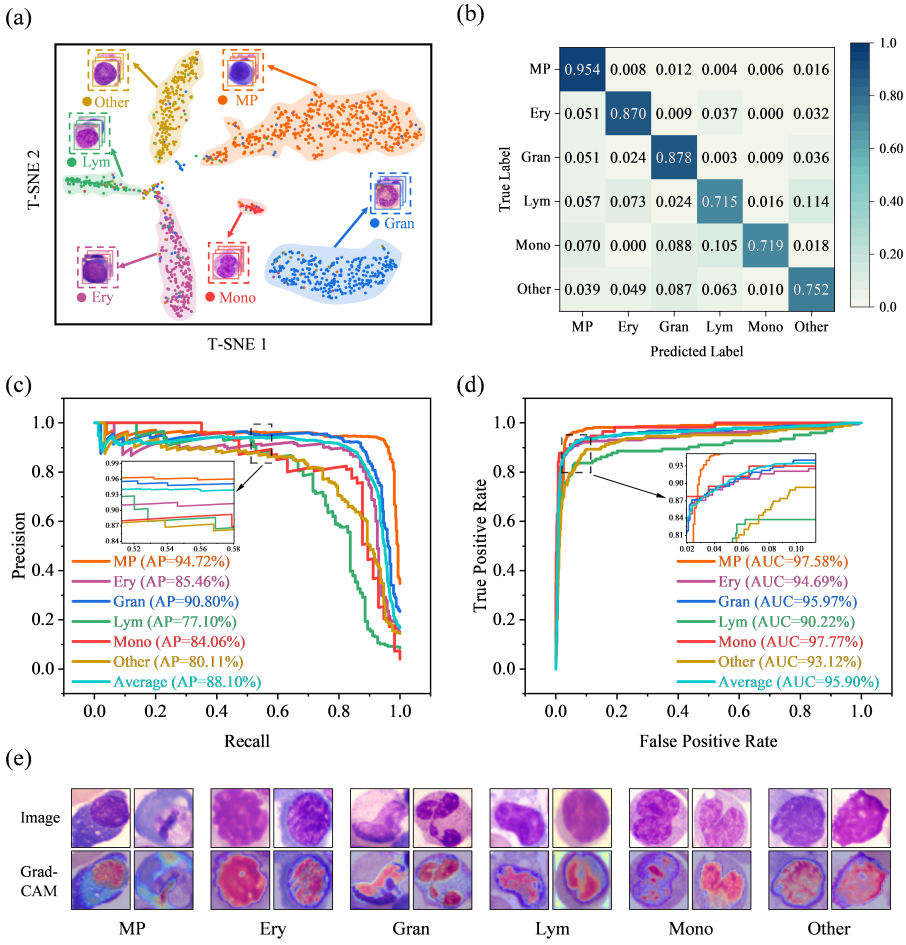

Fig. 2. MM diagnosis performance of SF-DINO. (a) t-SNE plot. (b) Confusion matrix. (c) PR curves. (d) ROC curves. (e) Grad-CAM visualization. Best viewed in color.

Table 1. Comparative results from the multiple myeloma diagnosis dataset using 5-fold cross-validation. The best and second-best results are shown in **bold** and underline, respectively (mean ± standard deviation).

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Speed (img/s)
InceptionNeXt	87.50 ± 2.32	86.09 ± 6.54	79.28 ± 14.78	82.54 ± 9.33	67.94
RMT	88.79 ± 1.32	87.19 ± 3.53	81.16 ± 13.01	84.07 ± 8.19	13.25
LSNet	85.06 ± 1.59	84.00 ± 3.58	80.18 ± 9.03	82.05 ± 5.41	**77.30**
MedMamba	89.37 ± 2.17	**87.71** ± 7.26	84.34 ± 9.43	85.99 ± 6.26	40.62
Swin-UMamba	89.08 ± 2.09	86.94 ± 5.46	85.28 ± 7.76	86.10 ± 4.91	24.41
DINOv3	87.28 ± 1.85	82.34 ± 9.15	81.92 ± 8.68	82.13 ± 8.73	26.60
SF-DINO	**89.95** ± 1.92	87.02 ± 5.54	**87.24** ± 6.81	**87.13** ± 4.93	22.69

parameter configurations. SF-DINO achieves the best performance in Accuracy, Recall, and F1-score. Crucially for MM diagnosis where sensitivity is paramount, our method attains an impressive recall value of 87.24% ± 6.81, outperforming the runner-up Swin-UMamba by 1.96%. These results validate the efficacy of our SM-Adapter and FSHA module in capturing fine-grained pathological details. Although LSNet offers higher speeds, SF-DINO maintains a competitive inference rate of 22.69 image/s, ensuring an optimal trade-off between diagnostic precision and computational efficiency.

3.4 Ablation Analysis

As shown in Table 2, we validate the contribution of each component. The SM-adapter improves baseline (DINOv3) F1 by 4.18% with negligible trainable parameters (+1.16M), confirming its efficiency in capturing local morphology. Meanwhile, the FSHA module boosts accuracy to 89.08%. Finally, the SF-DINO achieves the best performance, demonstrating that local morphological cues and global spectral features are mutually complementary for robust MM diagnosis.

3.5 Cross-Domain Transferability Analysis

To evaluate the transferability of our method across different domains, we conduct experiments on the external Acevedo dataset [1] for peripheral white blood cell classification. As presented in Table 3, SF-DINO achieves the best accuracy of 95.35% and F1 of 94.91% in identifying eight different peripheral blood cells. Notably, compared to the DINOv3 baseline, SF-DINO yields a significant improvement of 1.96% in F1 metric. This demonstrates that the SM adapter effectively captures universal morphological cues, while the FSHA module extracts domain-invariant frequency features, enabling the model to adapt robustly to distinct staining and acquisition conditions of peripheral blood samples.

Table 2. Effectiveness of SM adapter and FSHA module in SF-DINO.

Components		Acc (%)	F1 (%)	Trainable
Adapter	FSHA			Params (M)
		87.28	82.13	**0.65**
✓		88.65	<u>86.31</u>	<u>1.81</u>
	✓	<u>89.08</u>	85.87	10.04
✓	✓	**89.95**	**87.13**	11.23

Table 3. White blood cell classification results from the Acevedo dataset.

Method	Accuracy (%)	F1 (%)
InceptionNeXt	92.16	92.07
RMT	94.56	<u>94.71</u>
MedMamba	<u>95.03</u>	94.68
DINOv3	93.62	93.51
SF-DINO	**95.35**	**94.91**

4 Conclusion

Automated MM diagnosis is complicated by fine-grained morphological heterogeneity and high inter-class similarities of bone marrow cells, which challenges

standard Large Vision Models. To address this, we propose SF-DINO, a spatial-frequency synergistic framework. It integrates an SM-Adapter to capture local morphological cues and a FSHA module to model global spectral dependencies. Experiments on our constructed fine-grained MM diagnosis dataset demonstrate impressive performance, particularly for identifying monoclonal plasma cells. Crucially, while definitive MM diagnosis remains a patient-level clinical decision, SF-DINO automates the critical first step of morphology-based diagnostic assessment. By providing highly accurate single-cell classifications, it delivers the essential morphological evidence required for reliable plasma-cell quantification. Furthermore, cross-domain transferability analysis on the external Acevedo dataset confirms that SF-DINO achieves the best white blood cell identification results across different staining and acquisition conditions. Ultimately, by adapting vision foundation models to highly specialized medical domains in a parameter-efficient manner, SF-DINO provides a robust and interpretable solution for automated cytopathology. Future work will aim at validating SF-DINO across multi-center cohorts and exploring patient-level evidence aggregation to further align with comprehensive clinical workflows.

Acknowledgments

The research work of this paper was supported by National Key R&D Program of China (2023YFF0723300); National Natural Science Foundation of China (62475198); Hubei Provincial Natural Science Foundation of China (2025AFB055); Shenzhen Science and Technology Program (JCYJ20230807090207014); Science and Technology Research Project of Education Department of Hubei Province (Q20241403); Graduate Education Reform Research Project of Hubei University of Technology (2025025).

Disclosure of Interests

The authors declare that they have no conflicts of interest.

References

1. Acevedo, A., Merino, A., Alférez, S., Molina, Á., Boldú, L., Rodellar, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in Brief **30**, 105474 (2020)
2. Bishop, R.T., Miller, A.K., Froid, M., Nerlakanti, N., Li, T., Frieling, J.S., Nasr, M.M., Nyman, K.J., Sudalagunta, P.R., Canevarolo, R.R., et al.: The bone ecosystem facilitates multiple myeloma relapse and the evolution of heterogeneous drug resistant disease. Nature Communications **15**(1), 2458 (2024)
3. Chilukoti, S.V., Shan, L., Tida, V.S., Maida, A.S., Hei, X.: A reliable diabetic retinopathy grading via transfer learning and ensemble learning with quadratic weighted kappa metric. BMC Medical Informatics and Decision Making **24**(1), 37 (2024)

4. Fan, Q., Huang, H., Chen, M., Liu, H., He, R.: Rmt: Retentive networks meet vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5641–5651 (2024)
5. Hagos, Y.B., Lecat, C.S., Patel, D., Mikolajczak, A., Castillo, S.P., Lyon, E.J., Foster, K., Tran, T.A., Lee, L.S., Rodriguez-Justo, M., et al.: Deep learning enables spatial mapping of the mosaic microenvironment of myeloma bone marrow trephine biopsies. Cancer Research **84**(3), 493–508 (2024)
6. Kamal, M., Shishido, S.N., Mason, J., Patel, K., Manasanch, E.E., Orlowski, R.Z., Kuhn, P.: Single-cell proteomic analysis reveals multiple myeloma heterogeneity and the dynamics of the tumor immune microenvironment in precursor and advanced states. Neoplasia **66**, 101189 (2025)
7. Koch, V., Wagner, S.J., Kazeminia, S., Sancar, E., Hehr, M., Schnabel, J.A., Peng, T., Marr, C.: Dinobloom: a foundation model for generalizable cell embeddings in hematology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 520–530 (2024)
8. LB Andrade, C., Ferreira, M.V., M. Alencar, B., SB Filho, J.L., A. Guimaraes, M., Porto Cruz Moraes, I., S. Lopes, T.J., S. dos Santos, A., M. dos Santos, M., CS e Silva, M.I., et al.: Pcmmd: A novel dataset of plasma cells to support the diagnosis of multiple myeloma. Scientific Data **12**(1), 161 (2025)
9. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., et al.: Swin-umamba: Mamba-based unet with imagenet-based pretraining. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 615–625 (2024)
10. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**, 2579–2605 (2008)
11. Matek, C., Krappe, S., Münzenmayer, C., Haferlach, T., Marr, C.: Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. Blood **138**(20), 1917–1927 (2021)
12. Mei, L., Lian, C., Han, S., Ye, Z., Hua, Y., Sun, M., He, J., Ye, Z., Mei, M., Yalikun, Y., et al.: High-efficiency spatially guided learning network for lymphoblastic leukemia detection in bone marrow microscopy images. Computers in Biology and Medicine **196**, 110860 (2025)
13. Mei, L., Shen, H., Yu, Y., Weng, Y., Li, X., Zahid, K.R., Huang, J., Wang, D., Liu, S., Zhou, F., et al.: High-throughput and high-accuracy diagnosis of multiple myeloma with multi-object detection. Biomedical Optics Express **13**(12), 6631–6644 (2022)
14. Mohty, M., Avet-Loiseau, H., Malard, F., Harousseau, J.L.: Potential future direction of measurable residual disease evaluation in multiple myeloma. Blood **142**(18), 1509–1517 (2023)
15. Rajkumar, S.V.: Multiple myeloma: 2020 update on diagnosis, risk-stratification and management. American Journal of Hematology **95**(5), 548–567 (2020)
16. Salvi, M., Michielli, N., Meiburger, K.M., Cattelino, C., Cotrufo, B., Giacosa, M., Giovanzana, C., Molinari, F.: cyto-knet: An instance segmentation approach for multiple myeloma plasma cells using conditional kernels. International Journal of Imaging Systems and Technology **34**(1), e22984 (2024)
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626 (2017)

18. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
19. Tian, Y., Ye, Q., Doermann, D.: Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524 (2025)
20. Wang, A., Chen, H., Lin, Z., Han, J., Ding, G.: Lsnet: See large, focus small. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 9718–9729 (2025)
21. Wobst, J., Wossidlo, N., Truger, M., Meggendorfer, M., Kern, W., Haferlach, T., Haferlach, C.: Clonal independence of chip-associated mutations in multiple myeloma. Leukemia & Lymphoma pp. 1–4 (2026)
22. Ye, Z., Wei, S., Mei, L., Weng, Y., Geng, Q., Wang, D., Lei, C.: Fvim: Frequency vision mamba for label-free cell death pathway prediction in lung cancer chemotherapy. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 230–239 (2025)
23. Ye, Z., Zhang, Y., Huang, J., Wang, D., Liu, S., Mei, L., Lei, C.: Rethinking multi-center semi-supervised breast cancer ultrasound image segmentation: An intermediate-domain perspective. IEEE Journal of Biomedical and Health Informatics **30**(6), 5156–5166 (2026)
24. Yu, W., Zhou, P., Yan, S., Wang, X.: Inceptionnext: When inception meets convnext. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5672–5683 (2024)
25. Yue, Y., Li, Z.: Medmamba: Vision mamba for medical image classification. arXiv preprint arXiv:2403.03849 (2024)