FViM: Frequency Vision Mamba for Label-Free Cell Death Pathway Prediction in Lung Cancer Chemotherapy

Zhaoyi Ye[1], Shubin Wei[1], Liye Mei[1,2,*], Yueyun Weng[1], Qing Geng[3], Du Wang[1], and Cheng Lei[1,4,5,*]

[1] The Institute of Technological Sciences, Wuhan University, Wuhan, China
liyemei@whu.edu.cn, leicheng@whu.edu.cn
[2] School of Computer Science, Hubei University of Technology, Wuhan, China
[3] Department of Thoracic Surgery, Renmin Hospital of Wuhan University, Wuhan, China
[4] Suzhou Institute of Wuhan University, Suzhou, China
[5] Shenzhen Institute of Wuhan University, Shenzhen, China

Abstract. Chemotherapy is the standard first-line treatment for lung cancer, and cellular death is an inevitable consequence of the process. However, current methods lack high-throughput, label-free approaches for accurately assessing cell death, and existing techniques struggle to capture cellular heterogeneity, complicating the prediction of lung cancer prognosis. Therefore, we propose frequency vision Mamba (FViM) for label-free cell death pathway prediction in lung cancer chemotherapy. Specifically, we introduce multi-dimensional optical time-stretch imaging flow cytometry (OTS-IFC) to capture high-throughput, multi-dimensional cell images under various cell death states. To effectively extract key features that are highly indicative of cellular heterogeneity, we propose FViM that integrates modeling remote dependencies of Mamba alongside frequency domain analysis of Fourier Transform. FViM first employs the frequency guided enhancement (FGE) module to enhance cellular detail features in the high-frequency domain, while reinforcing global contextual features in the low-frequency domain. The enhanced features are then processed through the Mamba-based visual state space block, which models the intricate relationships between different visual states, achieving a holistic prediction of cell death states. Experimental results demonstrate that FViM outperforms existing state-of-the-art (SOTA) methods. Notably, FViM successfully predicts cell death pathways in response to increasing cisplatin concentrations, demonstrating its potential for effective and promising applications in lung cancer chemotherapy. Our code is available at https://github.com/yzygit1230/FViM.

Keywords: Lung Cancer · Cell Death · Frequency Vision Mamba · Multi-Dimensional Optical Time-Stretch Imaging

1 Introduction

Lung cancer remains one of the most serious global health challenges, causing an estimated 1.8 million annual deaths globally, which is 18% of all deaths attributed to cancer [10]. Among them, approximately 85% of lung cancer cases are non-small cell lung cancer (NSCLC), which is the most prevalent subtype [8]. Chemotherapy remains the cornerstone of lung cancer treatment, utilizing a variety of chemotherapeutic agents specifically formulated to induce targeted cell death through mechanisms such as autophagy and apoptosis, both of which are immunologically "silent" forms of cell death [4]. Cell death is a double-edged sword: it eliminates damaged organelles and promotes cell survival, but dysregulated autophagy can facilitate cell death by triggering apoptosis [26]. The interplay between autophagy and apoptosis is intrinsic, and any disruption in this dynamic balance can lead to uncontrolled proliferation or tumorigenesis [2]. Therefore, it is critical to develop a reliable method for predicting cell death pathways in lung cancer chemotherapy.

Nowadays, scientists have developed various methods to investigate cell death pathways. Flow cytometry [9] is a widely utilized technique for cell populations analysis, but its reliance on staining agents may alter cellular states, and it lacks spatial resolution. Transmission electron microscopy [15] offers exceptional resolution for observing ultrastructural changes, but its complex sample preparation limits its suitability for high-throughput applications. Fluorescence microscopy [1] enables real-time monitoring of cell death events, yet its throughput is limited, and it often depends on fluorescent markers that may affect cellular behavior. Western blotting [16] is a standard technique for analyzing proteins involved in cell death pathways, but it provides only static, end-point data and cannot track real-time progression. Optical time-stretch imaging flow cytometry (OTS-IFC) [20] is an emerging, label-free technique with high-throughput capabilities. However, current implementations primarily focus on intensity-based cell image analysis, limiting their ability to reveal detailed internal cellular structures [21,25]. Despite these advances, extracting and characterizing cell death features remains challenging due to the inherent complexity and heterogeneity of the cellular environment. State space models (SSMs) [7], exemplified by Mamba [6], have garnered considerable attention for their superior ability to capture long-range dependencies through state transitions. Furthermore, Vision Mamba [27] demonstrates the potential of Mamba in image analysis by its application in natural images. However, Mamba often ignores fine-grained features, which are essential for accurate cell feature extraction, particularly in the context of cellular heterogeneity and dynamic processes such as autophagy and apoptosis.

To address the above dilemma, we propose a multi-dimensional OTS-IFC that enables the acquisition of multimodal cell death images. We further design a frequency guided enhancement (FGE) module, integrated with a visual state space (VSS) block, to propose a frequency vision Mamba (FViM), which effectively extracts key features indicative of cellular heterogeneity. Comparative experimental results demonstrate that FViM achieves superior feature extraction for distinguishing cell death states. Our main contributions are as follows: **1)**

We devise a multi-dimensional OTS-IFC system for label-free, high-throughput acquisition of intensity and phase-contrast cell images, enabling detailed morphological analysis of diverse cell death populations. **2)** We propose FViM, which integrates the remote modeling strengths of Vision Mamba with the powerful capabilities of the fast Fourier transform (FFT) to capture both local textures and global semantics of cells. **3)** FViM effectively performs both qualitative and quantitative cell death pathway prediction as cisplatin concentration increases. These promising results offer significant potential for clinical applications.

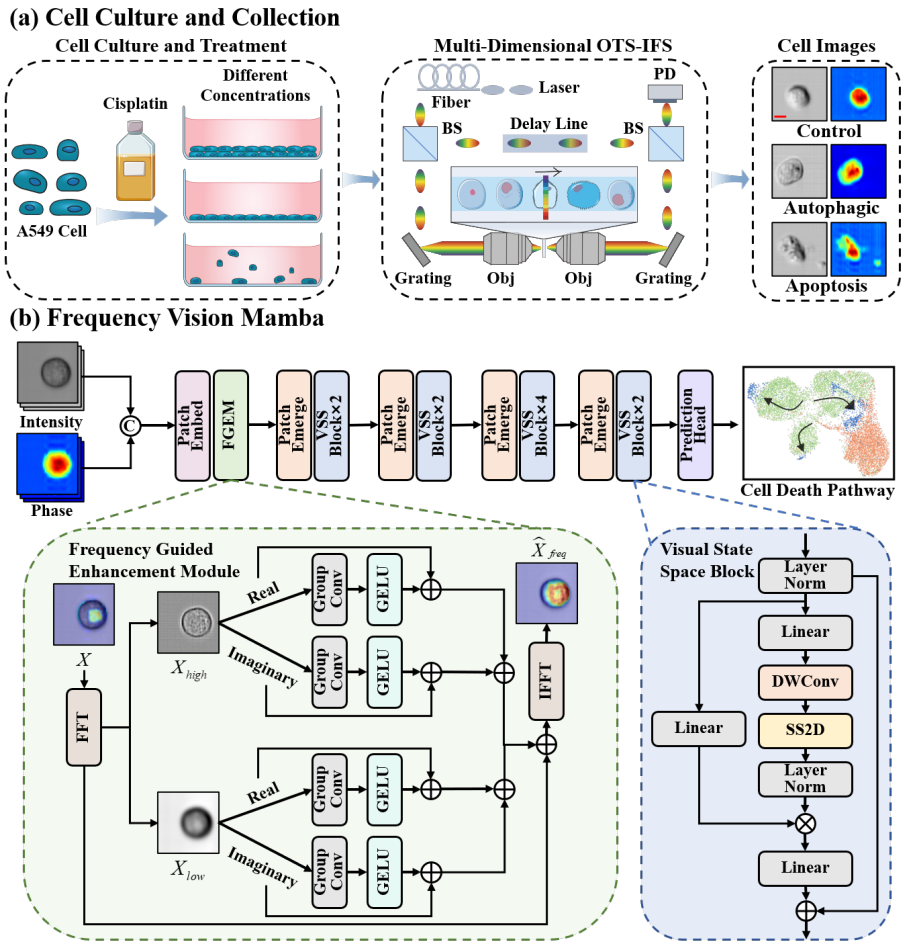

Fig. 1. Overview of the proposed pipeline. (a) Flow diagram of cell sample preparation and analysis. Scale bar: 10 µm. (b) Flow diagram of the frequency vision Mamba.

2 Materials and Methods

2.1 Cell Culture and Treatment

We cultured A549 cells, a well-established *in vitro* model for NSCLC, under standard conditions (37°C, 5% CO_2, and 95% humidity) in growth medium supplemented with 90% Ham's F12, 10% fetal bovine serum, and 1% penicillin-streptomycin. To induce autophagy, we cultured the cells with 10 µM rapamycin, a well-known inhibitor commonly used to induce autophagy, for 12 hours (Solarbio, Beijing, China). To induce apoptosis, we cultured the cells with 50 µM Carbonylcyanide-3-chlorophenylhydrazone (GLPBIO, CA, USA) for 12 hours.

As depicted in Fig. 1 (a), we cultured A549 cells (ATCC, Manassas, VA, USA) with cisplatin (GLPBIO, CA, USA), a platinum-based chemotherapeutic agent widely employed in clinical oncology, at concentrations of 7.5, 15, 30, and 60 mM for 24 hours. We used Dimethyl sulfoxide as the control group. Post-treatment, we collected cells for subsequent cell death analyses.

2.2 Multi-Dimensional Optical Time-Stretch Imaging Flow Cytometry

To analyze post-culture cells, we employed multi-dimensional OTS-IFC, which enables high-throughput acquisition of intensity and phase cell images in a label-free manner. As shown in Fig. 1 (a), multi-dimensional OTS-IFC consists of time-domain stretching and phase interference imaging. Specifically, the light source is a Titanium-sapphire laser (Coherent Inc., Santa Clara, CA, USA, model Vitara-T-HP), generating femtosecond pulses centered at 800 nm with a spectral bandwidth of 40 nm and a repetition frequency of 80 MHz. These pulses are stretched to 2.8 µs by a single-mode fiber spanning 1.55 kilometers, with a group velocity dispersion of -186 ps/nm. Subsequently, the pulses are split into two independent stretched laser beams by a beam splitter (Thorlabs Inc., Newton, NJ, USA, model BS026) to create object and reference lights. The reference light achieves signal interference by introducing a time delay, while the object light passes through a diffraction grating (Thorlabs, GR25-1210, New Jersey, USA). The diffraction grating generates a beam of one-dimensional (1D) rainbow light, which is focused onto a cell within the microchannel of the microfluidic device using an objective lens (Olympus, LCPlanN, 50×, NA=0.65, Japan). Then, the 1D rainbow light carrying cellular information converges through a symmetrical structure, including an identical objective lens and diffraction grating, and combines into object light. The reference and object lights interfere and are captured by a photodetector (Newport, 1544-B, California, USA). The interference signal is processed by an oscilloscope at a sampling rate of 40 GSa/s into a digital signal, further transformed into intensity and phase cell images via MATLAB.

2.3 Frequency Vision Mamba

Fig. 1 (b) provides an overview of the proposed FViM. FViM first concatenates the input intensity and phase cell images at the channel level. Then, the FGE

module enhances both local and global characteristics in the frequency domain. Following this, FViM employs consecutive stacked VSS blocks to further extract and refine feature representations. Finally, FViM achieves cell state recognition and cell death pathway prediction.

Frequency Guided Enhancement Module Since the Fourier Transform [19] reveals that high-frequency components capture fine-grained details, while low-frequency components provide global semantic information, we design the FGE module to enhance feature representation in the frequency domain. As shown in Fig. 1 (b), the FGE module first applies the FFT to the input feature map X, obtaining the frequency domain features X_{freq}. By multiplying with high- and low-frequency masks, we separate the high-frequency features X_{high} and low-frequency features X_{low}. Next, we modulate the real and imaginary parts of both the high- and low-frequency components using two group convolution weight matrices W_1, W_2 and biases b_1, b_2, resulting in enhanced high-frequency features $\widehat{X}_{high}$ and low-frequency features $\widehat{X}_{low}$. Finally, after applying residual connections and performing inverse FFT, the enhanced feature map $\widehat{X}_{freq}$ is obtained. The above processes can be calculated as follows:

$$
\begin{aligned}
&X_{freq} = \mathrm{FFT}\left(X\right), \\
&\widehat{X}_{high}^{real} = \mathrm{GELU}\left(X_{high}^{real}W_1 - X_{high}^{imag}W_2 + b_1\right) + X_{high}^{real}, \\
&\widehat{X}_{high}^{imag} = \mathrm{GELU}\left(X_{high}^{imag}W_2 - X_{high}^{real}W_1 + b_2\right) + X_{high}^{imag}, \\
&\widehat{X}_{low}^{real} = \mathrm{GELU}\left(X_{low}^{real}W_1 - X_{low}^{imag}W_2 + b_1\right) + X_{low}^{real}, \\
&\widehat{X}_{low}^{imag} = \mathrm{GELU}\left(X_{low}^{imag}W_2 - X_{low}^{real}W_1 + b_2\right) + X_{low}^{imag}, \\
&\widehat{X}_{high} = \widehat{X}_{high}^{real} + \widehat{X}_{high}^{imag}, \widehat{X}_{low} = \widehat{X}_{low}^{real} + \widehat{X}_{low}^{imag}, \\
&\widehat{X}_{freq} = \mathrm{FFT}^{-1}\left(X_{freq} + \widehat{X}_{high} + \widehat{X}_{low}\right)
\end{aligned}
\tag{1}
$$

Vision State Space Block Building upon the insights from VMamba [12], we leverage the VSS block to model the intricate relationships between remote dependencies. As shown in Fig. 1 (b), the VSS block consists of two parallel branches. The first branch applies a linear layer to expand the channels, followed by DWConv, SSM, and LayerNorm. The second branch also uses a linear layer to expand the channels. Finally, the features from both branches are aggregated to generate the output feature map.

3 Experiments and Results

3.1 Datasets and Evaluation Metrics

We obtained three types of high-throughput cells through cell culture and multi-dimensional OTS-IFC: control, autophagy, and apoptosis, with cell counts of 2636, 3983, and 1381, respectively. To train the network to recognize different

cell death states, we divided the data into training, validation, and testing sets in a 3:1:1 ratio. We utilize Accuracy (Acc), Precision, Recall, F1, and Kappa coefficient to evaluate the performance of cell death state recognition [3,22].

Additionally, for cell death pathway prediction, we cultured cells treated with four different concentrations of cisplatin: 7.5 mM, 15 mM, 30 mM, and 60 mM, with cell counts of 2730, 2505, 3109, and 2596, respectively. Note that the specific cell death states at each concentration are not pre-labeled in the experiment.

3.2 Implementation Details

We conduct all experiments on the PyTorch framework with an NVIDIA A6000 GPU. We maintain consistent experimental settings across all methods, including a batch size of 64. Training is conducted for 100 epochs, using the Adam optimizer with an initial learning rate of 0.00008.

3.3 Comparative Analysis

To validate the effectiveness of FViM in distinguishing the cell death states, we conduct a comparative analysis with seven SOTA methods, including: InceptionNeXt [23], MedViT [13], TransNeXt [18], RMT [5], VMamba [12], MedMamba [24], and Swin-UMamba [11]. As shown in Table 1, Swin-UMamba achieves the highest Precision (98.21%), demonstrating its effectiveness in reducing false detection. Inspiringly, FViM outperforms the comparative methods across the other five metrics, highlighting its robust ability to capture key features indicative of cellular heterogeneity among control, autophagy, and apoptosis cells.

Table 1. Comparative results from the cell death state recognition dataset. The best and second best results are shown in **bold** and underline, respectively.

Method	Acc (%)	Precision (%)	Recall (%)	F1 (%)	Kappa (%)
InceptionNeXt [23]	97.88	97.30	97.20	97.25	96.54
MedViT [13]	97.69	<u>98.14</u>	95.93	97.01	96.20
TransNeXt [18]	96.50	96.42	95.28	95.85	94.30
RMT [5]	96.44	96.70	94.35	95.51	94.15
VMamba [12]	97.82	97.60	96.41	97.00	96.43
MedMamba [24]	<u>98.31</u>	97.78	<u>97.92</u>	<u>97.85</u>	<u>97.26</u>
Swin-UMamba [11]	98.19	**98.21**	96.72	97.46	97.04
FViM	**98.56**	97.92	**98.25**	**98.09**	**97.66**

3.4 Abaltion Results

As shown in Table 2 and Table 3, we conduct ablation study to validate the effectiveness of the FViM from the cell death state recognition dataset. Table 2 reveals that combining intensity and phase cell images allows for the extraction

Table 2. Effectiveness of multi-dimensional cell images.

Image Dimension		Acc (%)	F1 (%)
Intensity	Phase		
✓		96.63	95.73
	✓	<u>97.69</u>	<u>97.14</u>
✓	✓	**98.56**	**98.09**

Table 3. Effectiveness of FGE module and VSS block in FViM.

Components		Acc (%)	F1 (%)
FGE	VSS		
		95.32	94.02
✓		<u>97.82</u>	<u>97.12</u>
✓	✓	**98.56**	**98.09**

of crucial morphological and thickness information, which enables the effective discrimination of cell death states. In Table 3, we observe that removing the FGE module and replacing the VSS block with a simple 3×3 convolution layer results in suboptimal performance. The integration of these two components is crucial, as it allows FViM to effectively capture key distinguishing features that are highly indicative of cellular heterogeneity.

3.5 Performance of Cell Death Pathway Prediction

Accurate prediction of cell death pathways is crucial for understanding cellular responses to chemotherapy and aiding personalized treatment strategies. Fig. 2 provides a comprehensive analysis of FViM in predicting cell death pathways across different cisplatin concentrations. By projecting the deep feature maps of FViM into a 2-D space using UMAP [14], we visualize the clustering trends and dynamic evolution of control, autophagy, and apoptosis cells under varying cisplatin concentrations, as illustrated in Fig. 2 (a). Notably, as cisplatin concentration increases, cell clusters shift toward the apoptosis cluster, with autophagic cells predominating at lower concentrations, reflecting a protective mechanism. However, as cisplatin concentration rises, autophagy transitions to apoptosis, marking a shift from survival to cell death. In Fig. 2 (b), the quantitative results further corroborate the findings in the UMAP diagram, showing a gradual increase in apoptosis with higher cisplatin concentrations. As the concentration increases, a marked increase in apoptotic cells is observed, with apoptosis cells reaching as high as 54% at 60 mM cisplatin. These results highlight the critical role of autophagy in inducing apoptosis under chemotherapy stress. Through Grad-CAM analysis [17], Fig. 2 (c) demonstrates FViM's ability to identify biologically significant morphological patterns: normal cells maintain regular contours, autophagic cells exhibit characteristic size expansion and cytoplasmic accumulation, while apoptotic cells display typical structural disintegration. As a post-hoc interpretability tool, it is evident from these attention patterns that FViM's predictions are grounded in pathologically relevant features, validating the model's capability to predict cell death pathways in response to increasing cisplatin concentrations.

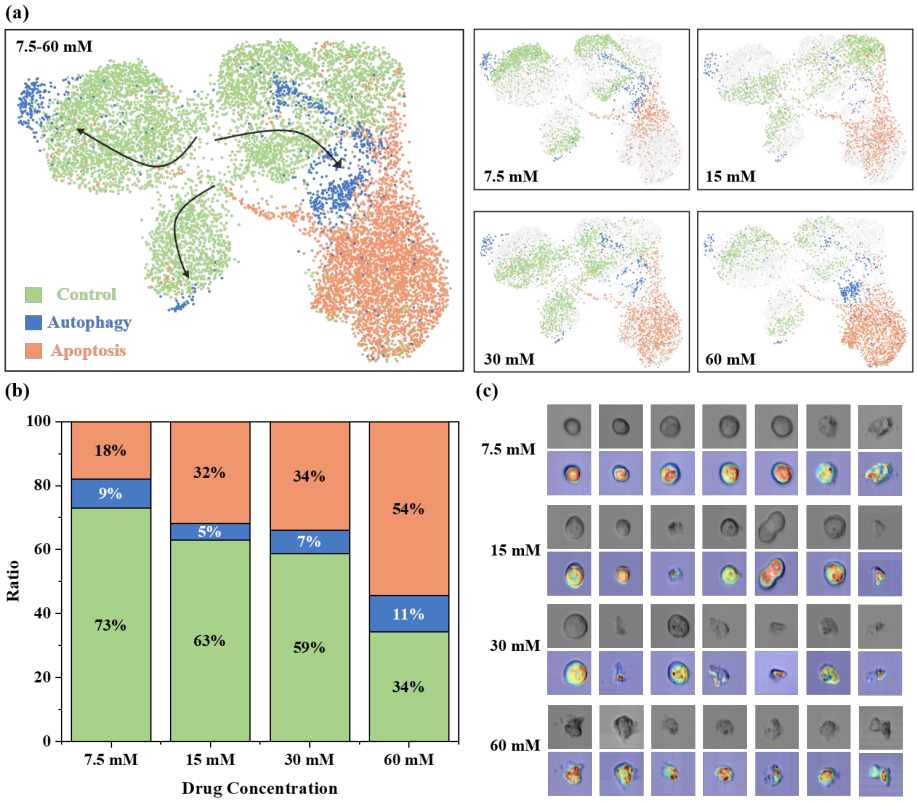

Fig. 2. Qualitative and quantitative performance of FViM in cell death pathway prediction at different drug concentrations. (a) UMAP diagram of control, autophagy, and apoptosis cells clustering trend. (b) Quantitative detection results of control, autophagy, and apoptosis cells. (c) Grad-CAM results.

4 Discussion and Conclusion

Lung cancer remains one of the most prevalent and lethal malignancies worldwide, with chemotherapy being the cornerstone of its treatment. However, the efficacy of chemotherapeutic agents, such as cisplatin, is often limited by the complex and heterogeneous responses of cancer cells, including variations in cell death pathways. Understanding these pathways—autophagy and apoptosis—is critical for optimizing treatment strategies and improving patient outcomes.

In this study, we propose FViM, a novel method combining multi-dimensional OTS-IFS for label-free cell death pathway prediction in lung cancer chemotherapy. The multi-dimensional OTS-IFS provides a robust platform for acquiring and analyzing multimodal cell images, while FViM leverages the long-range modeling capabilities of Mamba and frequency domain analysis to effectively extract key heterogeneity features. Experimental results demonstrate that FViM surpasses seven SOTA methods in cell death state recognition. Furthermore, both quantitative and qualitative analyses reveal its capability to predict cell death pathways at increasing cisplatin concentrations, offering promising clinical implications.

Acknowledgments

The research work of this paper was supported by National Natural Science Foundation of China (62475198); Hubei Provincial Natural Science Foundation of China (2025AFB055); Hubei Province Young Science and Technology Talent Morning Hight Lift Project (202319); Jiangsu Science and Technology Program (BK20221257); Shenzhen Science and Technology Program (JCYJ20220530140601003, JCYJ20230807090207014); Science and Technology Research Project of Education Department of Hubei Province; Doctoral Starting Up Foundation of Hubei University of Technology (XJ2023007301).

Disclosure of Interests

The authors declare that they have no conflicts of interest.

References

1. Awasthi, K., Li, S.P., Zhu, C.Y., Hsu, H.Y., Ohta, N.: Fluorescence microscopic approach for detection of two different modes of breast cancer cell death induced by nanosecond pulsed electric field. Sensors and Actuators B: Chemical **378**, 133199 (2023)
2. Biswas, U., Roy, R., Ghosh, S., Chakrabarti, G.: The interplay between autophagy and apoptosis: its implication in lung cancer and therapeutics. Cancer Letters p. 216662 (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. Dongsar, T.S., Dongsar, T.T., Nasir, N., Wahab, S., Sahebkar, A., Gupta, G., Kesharwani, P.: Cuproptosis engineered nanomaterials: The silent warrior against cancer. Applied Materials Today **39**, 102305 (2024)
5. 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)
6. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
7. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)
8. Hendriks, L.E., Remon, J., Faivre-Finn, C., Garassino, M.C., Heymach, J.V., Kerr, K.M., Tan, D.S., Veronesi, G., Reck, M.: Non-small-cell lung cancer. Nature Reviews Disease Primers **10**(1), 71 (2024)
9. Jiang, L., Tixeira, R., Caruso, S., Atkin-Smith, G.K., Baxter, A.A., Paone, S., Hulett, M.D., Poon, I.K.: Monitoring the progression of cell death and the disassembly of dying cells by flow cytometry. Nature Protocols **11**(4), 655–663 (2016)
10. Leiter, A., Veluswamy, R.R., Wisnivesky, J.P.: The global burden of lung cancer: current status and future trends. Nature Reviews Clinical Oncology **20**(9), 624–639 (2023)
11. 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. Springer (2024)
12. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. Advances in Neural Information Processing Systems **37**, 103031–103063 (2025)
13. Manzari, O.N., Ahmadabadi, H., Kashiani, H., Shokouhi, S.B., Ayatollahi, A.: Medvit: a robust vision transformer for generalized medical image classification. Computers in Biology and Medicine **157**, 106791 (2023)
14. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
15. Neikirk, K., Vue, Z., Katti, P., Rodriguez, B.I., Omer, S., Shao, J., Christensen, T., Garza Lopez, E., Marshall, A., Palavicino-Maggio, C.B., et al.: Systematic transmission electron microscopy-based identification and 3d reconstruction of cellular degradation machinery. Advanced Biology **7**(6), 2200221 (2023)
16. Qi, Z., Huang, X., Jing, J., Feng, W., Xu, M., Yan, L., Gao, M., Liu, S., Yu, X.F.: Amino-modified ionps potentiates ferroptotic cell death due to the release of fe ion in the lysosome. Journal of Environmental Sciences **150**, 1–13 (2025)
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. Shi, D.: Transnext: Robust foveal visual perception for vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17773–17783 (2024)
19. Vaish, P., Wang, S., Strisciuglio, N.: Fourier-basis functions to bridge augmentation gap: Rethinking frequency augmentation in image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17763–17772 (2024)

20. Wei, S., Song, C., Ye, Z., Weng, Y., Mei, L., Li, R., Yan, R., Deng, Y., Liu, X., Xu, X., et al.: Label-free prediction of immunotherapy response in lung cancer. ACS Photonics **11**(11), 5000–5011 (2024)
21. Weng, Y., Shen, H., Mei, L., Liu, L., Yao, Y., Li, R., Wei, S., Yan, R., Ruan, X., Wang, D., et al.: Typing of acute leukemia by intelligent optical time-stretch imaging flow cytometry on a chip. Lab on a Chip **23**(6), 1703–1712 (2023)
22. Yilmaz, A.E., Demirhan, H.: Weighted kappa measures for ordinal multi-class classification performance. Applied Soft Computing **134**, 110020 (2023)
23. 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)
24. Yue, Y., Li, Z.: Medmamba: Vision mamba for medical image classification. arXiv preprint arXiv:2403.03849 (2024)
25. Zhou, J., Mei, L., Yu, M., Ma, X., Hou, D., Yin, Z., Liu, X., Ding, Y., Yang, K., Xiao, R., et al.: Imaging flow cytometry with a real-time throughput beyond 1,000,000 events per second. Light: Science & Applications **14**(1), 76 (2025)
26. Zhou, J.c., Wang, J.l., Ren, H.z., Shi, X.l.: Autophagy plays a double-edged sword role in liver diseases. Journal of Physiology and Biochemistry pp. 1–9 (2022)
27. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)