Interpretable point cloud classification using multiple instance learning

Supplementary Material

9. Transformer block feature extractor

9.1. Group features through k-nearest neighbours:

Formally, we constructed a k-NN graph on $\mathbf{P}$ with the graph including a self-loop to point-level features:

$$\begin{aligned} \mathcal{N}(\mathbf{p}_i) &= \mathrm{KNN}(\mathbf{P}, ||\mathbf{p}_i - \mathbf{p}_j||_2^2), \mathbf{p}_i, \mathbf{p}_j \in \mathbf{P}, \\ \mathbf{f}_i' &= [(\mathbf{f}_j - \mathbf{f}_i), \mathbf{f}_i]_{j \in \mathcal{N}(\mathbf{p}_i)} \in \mathbb{R}^{k \times 2d_{in}}, \end{aligned} \tag{7}$$

where KNN$(\cdot)$ is the k-NN function, $[\cdot, \cdot]$ is concatenation, k is the hyperparameter of the k-NN graph, $\mathcal{N}(\mathbf{p}_i)$ is the set of neighbours of $\mathbf{p}_i$, and $\mathbf{f}_i'$ is the point feature augmented with local contextual information.

9.2. Learned relative positional encoding:

To encode spatial configurations per point cloud neighbourhood, we incorporated positional embeddings, $\mathbf{h}_i$ such that:

$$\mathbf{h}_i \in \mathbb{R}^{k \times d_h} = \phi_{pos}([\mathbf{p}_i - \mathbf{p}_j]_{j \in \mathcal{N}(\mathbf{p}_i)}), \tag{8}$$

where ϕ_{pos} is an MLP and d_h is the output channel dimension of ϕ_{pos}. The features were then further augmented with this positional encoding to give:

$$\mathbf{f}_i'' = [\mathbf{f}_i', \mathbf{h}_i]. \tag{9}$$

Thus, we obtained a new feature set $\mathbf{F}'' \in \mathbb{R}^{N \times k \times (2d_{in} + d_h)} = \{\mathbf{f}_i''\}_{i=1}^{N}$.

9.3. Attention on the augmented features:

The resulting features, $\mathbf{F}''$, were then fed into a transformer with EdgeConv as the query operation. Recall that EdgeConv [55] computes graph features for each point using the following equation:

$$\mathbf{e}_i \in \mathbb{R}^{d_e} = \max_{j \in \mathcal{N}(\mathbf{p}_i)}(\phi_{edge}(\mathbf{p}_i, \mathbf{p_j} - \mathbf{p}_i)), \tag{10}$$

where ϕ_{edge} is an MLP with output dimension d_e. The $\mathbf{F}''$ were then transformed using attention [51]:

$$\begin{aligned} \mathbf{Q} \in \mathbb{R}^{N \times d_k} &= \mathrm{EdgeConv}\left(\mathbf{F}''\right) W_q \\ \mathbf{K} \in \mathbb{R}^{(N \times k) \times d_k} &= \mathrm{Flatten}\left(\mathbf{F}''\right) W_k \\ \mathbf{V} \in \mathbb{R}^{(N \times k) \times d_v} &= \mathrm{Flatten}\left(\mathbf{F}''\right) W_v, \end{aligned} \tag{11}$$

where $\mathbf{W}_q \in \mathbb{R}^{d_e \times d_k}$, $\mathbf{W}_k \in \mathbb{R}^{(2d_{in} + d_h) \times d_k}$ and $\mathbf{W}_v \in \mathbf{R}^{(2d_{in} + d_h) \times d_v}$ are learnable weight matrices. The final point-level output features from the transformer block were then given by:

$$\mathbf{z}_i \in \mathbb{R}^{N \times d_v} = \mathbf{q}_i(\mathrm{softmax}(\mathbf{k}_i)^\mathsf{T} \mathbf{v}_i). \tag{12}$$

For all experiments, we used two transformer layers such that the final feature vector for each point was of size 256.

10. CurveNet adaptation

CurveNet [59] uses sampling and grouping. Our only adaptation to CurveNet was to use the same number of sampled points as input into the farthest point sampling algorithm. We kept everything else the same as the original paper. We replaced the original adaptive max, adaptive mean pooling, and the classification head with MIL pooling. The final feature vector for each point was 1024 in size.

11. PointMLP adaptation

Similar to CurveNet, PointMLP [35] employs sampling and grouping techniques. We adapted PointMLPElite by simply changing the reducers from 2 to 1 to avoid point sampling. We kept everything else the same as the original paper. We replaced the original adaptive max, adaptive mean pooling, and the classification head with MIL pooling. The final feature vector for each point was 256 in size.

12. Interpretability metrics

AOPCR does not require instance labels, whereas NDCG@n does. AOPCR works by removing the most important instances in sequence and observing the impact on prediction accuracy. The faster the prediction declines, the better the ordering, as the most influential instances are removed earlier. When point clouds are annotated, NDCG@n evaluates how closely the model's interpretability ranking matches the true order. It rewards rankings that prioritise relevant instances, with higher scores indicating better alignment and interpretability.

13. Datasets

13.1. IntrA

IntrA is an open source dataset of 3D intracranial aneurysm [63]. The task is to classify blood vessels as healthy or aneurysmal. There are a total of 1909 blood vessel segments, including 1694 healthy vessel segments and 215 aneurysm segments for diagnostic purposes. 116 of the aneurysm segments are expertly annotated. We use IntrA to evaluate interpretability, classification, and segmentation.

13.2. Red Blood Cell

We used another dataset of 3D red blood cells (RBC; [45]) for classification. This dataset includes 825 3D red blood cells imaged using confocal microscopy, grouped into nine expertly annotated shape classes. Blood samples were collected from healthy donors and patients using finger-prick

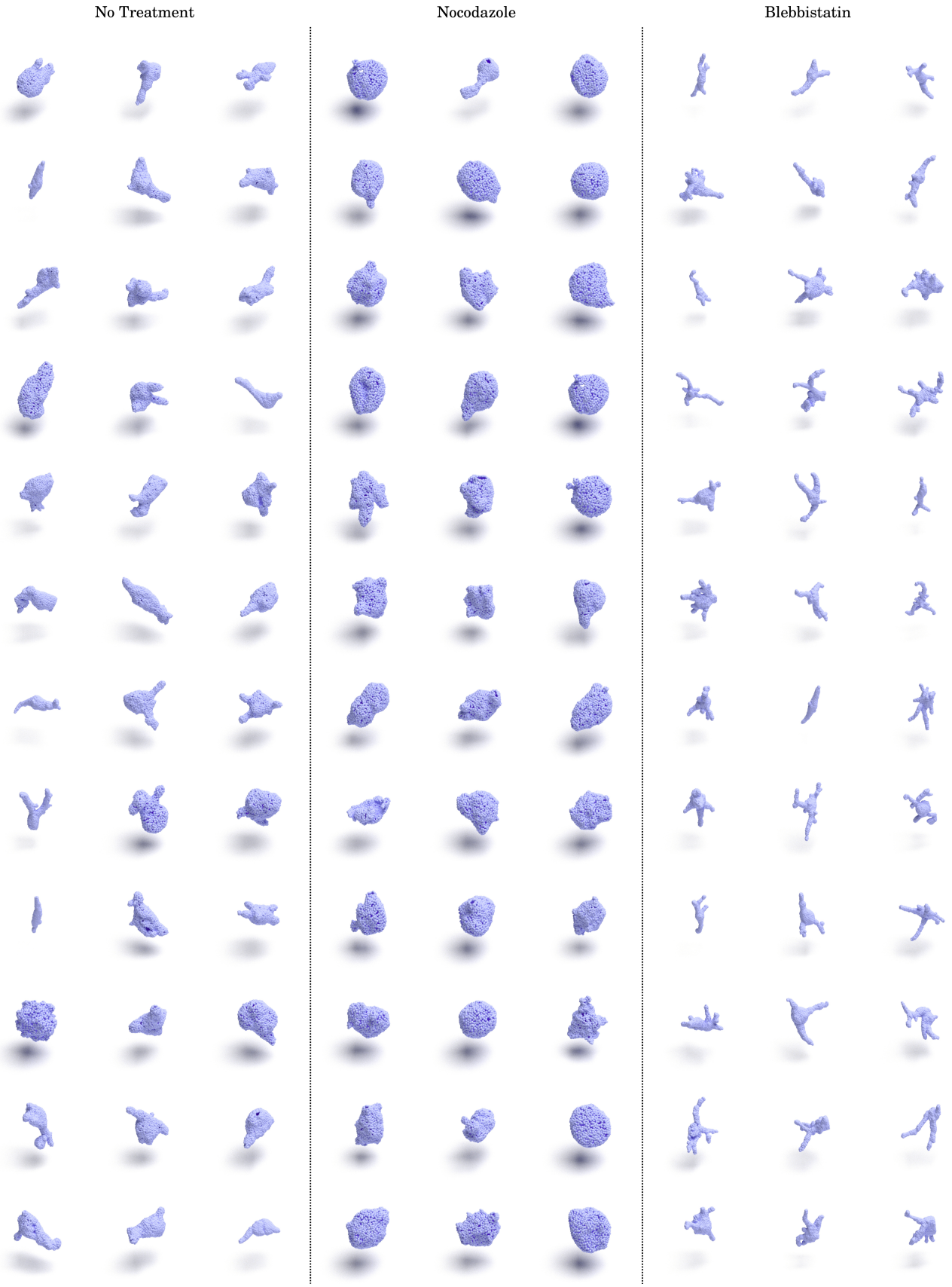

Figure 8. Rendered point clouds from the ATLAS-1 dataset.

blood sampling. To induce transitions in RBC shapes, the blood cells of five healthy donors were treated with NaCl solutions of varying concentrations to create different RBC shapes. Specific shape classes were expertly annotated according to particular motifs. Thus, similar to IntrA, RBC was suitable for evaluating interpretability by the ability to identify these motifs. Segmentation masks are publicly available. We converted the segmentation to mesh objects using marching cubes with Laplacian smoothing, and then sampled points from the vertices of these mesh objects. We have also made these point clouds available.

13.3. ModelNet40

ModelNet40 [58] is the *de-facto* benchmark for point cloud classification containing 9,843 training and 2,468 testing meshed CAD models belonging to 40 different object classes.

13.4. ScanObjectNN

ScanObjectNN [50] is a point cloud benchmark comprising 15,000 objects categorised into 15 classes, with 2,902 unique real-world object instances. The presence of background clutter, noise, and occlusions makes this benchmark particularly challenging for existing point cloud analysis methods. For this paper, we used the hardest perturbed variant (PB_T50_RS) of ScanObjectNN.

14. ATLAS-1

ATLAS-1 is a new dataset of 3D drug-treated cancer cells in physiologically relevant environments. The dataset comprises 1500 single-cell point clouds, derived from WM266-4 human melanoma cells embedded in tissue-like collagen matrices and imaged in 3D using oblique-plane light-sheet microscopy. This is a subset of the data used in De Vries et al. [9]. Cells were treated with one of the following conditions:

- **Nocodazole (500 cells)**: Microtubule inhibitor, causing cells to adopt a round morphology.
- **Blebbistatin (500 cells)**: Non-Muscle-Myosin II inhibitor, leading to elongated, spindle-like cell shapes.
- **Control (500 cells)**: Untreated cells exhibiting a range of normal human melanoma morphological variability.

The 3D point clouds were generated via the following pipeline: (1) automatic segmentation of individual cells, (2) surface reconstruction using the marching cubes algorithm [33] with Laplacian smoothing, and (3) point cloud sampling of 1024 points per cell from the reconstructed surface mesh. Figure 8 shows the point cloud renderings of the ATLAS-1 dataset.

14.1. Dataset availability and usage

The dataset is available under the Creative Commons Attribution 4.0 (CC-BY 4.0) license. This dataset can be used for:

- Benchmarking point cloud classification models on real-world biological data.
- Studying the 3D morphological effects of small-molecule inhibitors on cancer cells.

A metadata file ('metadata.csv') accompanies the dataset, providing labels and experimental details. Additionally, we provide Python scripts for loading and visualising the dataset.

The dataset, along with download instructions and documentation, is available through the project page: `https://Sentinal4D.github.io/PointMIL`

15. Trainig splits

For IntrA, RBC, and ATLAS-1, we used a five-fold cross-validation and reported the average test metrics across folds. For ModelNet40 and ScanObjectNN, we utilised the provided train and test splits and reported the test results.

16. Visual interpretation examples

Figure 9 shows additional interpretability visualisations on ModelNet40.

17. Comparison between backbones

Figure 10 presents the interpretability results for different backbones when classifying a *Bed* with `Additive` pooling (top row) and a *Plant* with `Conjunctive` pooling (middle row) from the ModelNet40 dataset. The perturbation curves are shown in the bottom row. Interestingly, DGCNN, CurveNet, PointMLP, and Transformer backbones consistently highlight similar regions of importance on the *Bed*, particularly focusing on the frame and headboard of the bed, which are key features distinguishing it from other objects. All backbones focussed on the leaves in the *Plant* as opposed to the pot. This consistency across the backbones demonstrates the robustness of PointMIL in identifying informative regions. Additionally, the agreement among backbones suggests that PointMIL effectively leverages the feature representations generated by each model, ensuring the interpretability results are meaningful and aligned with the task.

18. Additional results

This section contains additional results of individual pooling methods.

18.1. Interpretability

Tables 3 show the IntrA interpretability results for each of the pooling methods using the Transformer backbone, showing the effect of contextual attention. The mean and standard deviations on the test sets across the five folds are shown.

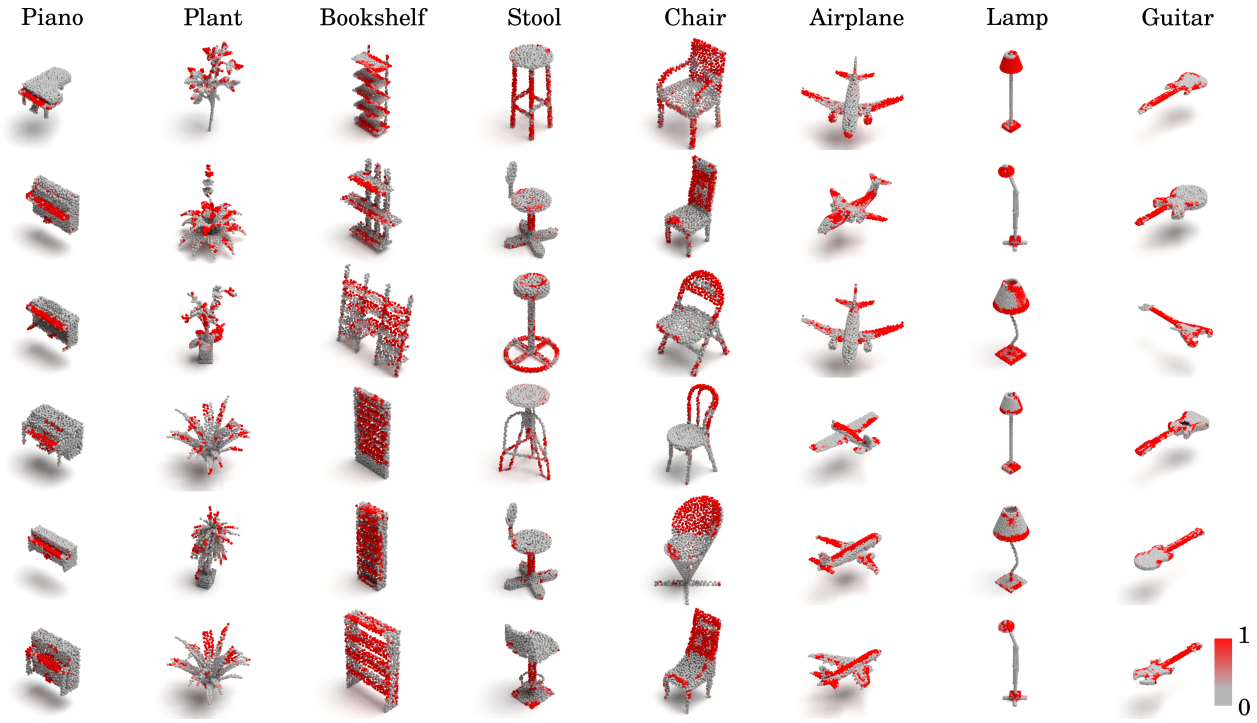

Figure 9. Examples of POINTMIL interpretations for correctly classified shapes from ModelNet40.

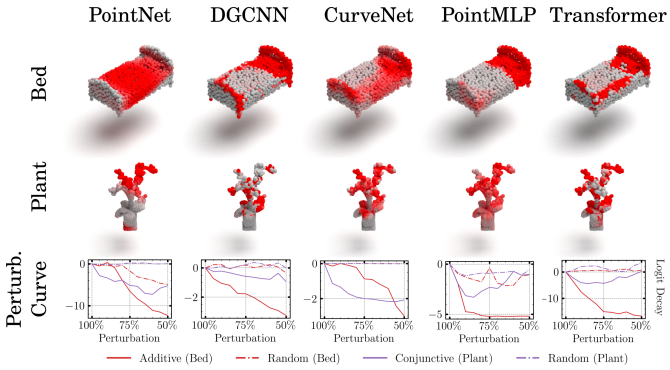

Figure 10. Interpretability of POINTMIL with different backbones on an example *Bed* (top row) and *Plant* (middle row) from ModelNet40. Perturbation curves are shown in the bottom row.

18.2. Assessing misclassifications

Finally, we demonstrated how POINTMIL could be used to assess where the model went wrong. For example, Figure 11 shows example confusion plots in which the attention of the predicted class is shown in red. Interestingly, for classifying plants, the model only focused on the plant, although when classifying flower pots, the model focused on both the flower and the pot.

Table 3. Additional POINTMIL interpretability results on IntrA using the transformer backbone. We also show the effect of the best contextual attention for each attention-based method.

Model	NDCG@n	AOPCR
`Additive`	$0.613_{0.033}$	$18.108_{4.374}$
`Additive` + context 12	$0.608_{0.035}$	$18.162_{3.013}$
`Attention`	$0.426_{0.030}$	$10.336_{1.065}$
`Attention` + context 12	$0.539_{0.019}$	$14.541_{1.821}$
`Conjunctive`	$0.592_{0.018}$	$12.526_{2.960}$
`Conjunctive` + context 12	$0.610_{0.024}$	$16.305_{5.859}$
`Instance`	$0.587_{0.022}$	$16.166_{3.794}$

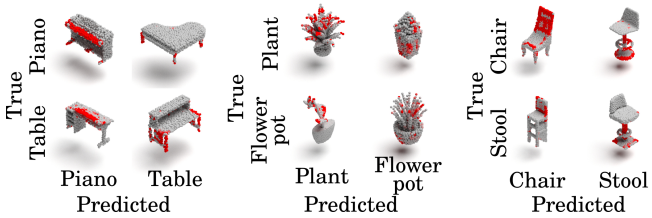

Figure 11. Interpretability visualisations of incorrect classifications from POINTMIL with Transformer backbone on ModelNet40.

19. Robustness to noise

Similar to the methods described by Xiang et al. [59] and Yan et al. [61], we assessed the robustness of POINTMIL to noisy inputs. Specifically, we measured the F1 score of models trained on clean (raw) inputs when subjected to noisy inputs during inference. This approach allowed us to evaluate the model's ability to maintain performance in the presence of input perturbations. The F1 score (left) and the mACC (right) is plotted against the number of noisy points introduced during inference for different POINTMIL methods with the DGCNN backbone and the original DGCNN model in Figure 12. POINTMIL methods demonstrate higher robustness to noise compared to baseline models, with `Additive` and `Conjunctive` maintaining consistently higher F1 and mACC scores than the original DGCNN without MIL.

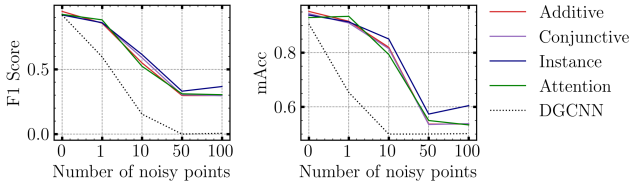

Figure 12. Robustness evaluation of models to noisy inputs.

References

[1] Perpetual Hope Akwensi, Ruisheng Wang, and Bo Guo. Preformer: A memory-efficient transformer for point cloud semantic segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 128:103730, 2024. 2

[2] Saad Ali and Mubarak Shah. Human action recognition in videos using kinematic features and multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(2):288–303, 2010. 3

[3] Nicholas I. Arnold, Plamen Angelov, and Peter M. Atkinson. An improved explainable point cloud classifier (xpcc). *IEEE Transactions on Artificial Intelligence*, 4(1):71–80, 2023. 2

[4] Chris Bakal, John Aach, George Church, and Norbert Perrimon. Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science*, 316 (5832):1753–1756, 2007. 4

[5] Le Cheng, Cuijuan An, Yu Gao, Yinfeng Gao, and Dawei Ding. Point mlp-former: Combining local and global receptive fields in point cloud classification. In *2022 China Automation Congress (CAC)*, pages 4895–4900, 2022. 2

[6] H.T. Cheung and D.S. Terry. Effects of nocodazole, a new synthetic microtubule inhibitor, on movement and spreading of mouse peritoneal macrophages. *Cell Biology International Reports*, 4(12):1125–1129, 1980. 6

[7] Jaesung Choe, Chunghyun Park, Francois Rameau, Jaesik Park, and In So Kweon. Pointmixer: Mlp-mixer for point cloud understanding. In *Computer Vision – ECCV 2022*, pages 620–640, Cham, 2022. Springer Nature Switzerland. 2

[8] Matt De Vries, Reed Naidoo, Olga Fourkioti, Lucas G. Dent, Nathan Curry, Christopher Dunsby, and Chris Bakal. Interpretable phenotypic profiling of 3D cellular morphodynamics . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. Springer Nature Switzerland, 2024. 4

[9] Matt De Vries, Lucas G. Dent, Nathan Curry, Leo Rowe-Brown, Vicky Bousgouni, Olga Fourkioti, Reed Naidoo, Hugh Sparks, Adam Tyson, Chris Dunsby, and Chris Bakal. Geometric deep learning and multiple-instance learning for 3d cell-shape profiling. *Cell Systems*, 16(3):101229, 2025. 3, 5

[10] Lucas G. Dent, Nathan Curry, Hugh Sparks, Vicky Bousgouni, Vincent Maioli, Sunil Kumar, Ian Munro, Francesca Butera, Ian Jones, Mar Arias-Garcia, Leo Rowe-Brown, Chris Dunsby, and Chris Bakal. Environmentally dependent and independent control of 3d cell shape. *Cell Reports*, page 114016, 2024. 1, 4

[11] Thomas G. Dietterich, Richard H. Lathrop, and Tomás Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1):31–71, 1997. 2, 3

[12] Meghan K. Driscoll, Erik S. Welf, Andrew R. Jamieson, Kevin M. Dean, Tadamoto Isogai, Reto Fiolka, and Gaudenz Danuser. Robust and automated detection of subcellular morphological motifs in 3D microscopy images. *Nature Methods*, 16(10):1037–1044, 2019. 1

[13] C. Dunsby. Optically sectioned imaging by oblique plane microscopy. *Optics Express*, 16(25):20306–20316, 2008. 4, 5

[14] Joseph Early, Christine Evers, and SArvapali Ramchurn. Model agnostic interpretability for multiple instance learning. In *International Conference on Learning Representations*, 2022. 4

[15] Joseph Early, Gavin Cheung, Kurt Cutajar, Hanting Xie, Jas Kandola, and Niall Twomey. Inherently interpretable time series classification via multiple instance learning. In *The Twelfth International Conference on Learning Representations*, 2024. 3, 4

[16] Feng-Lei Fan, Jinjun Xiong, Mengzhou Li, and Ge Wang. On interpretability of artificial neural networks: A survey. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(6):741–760, 2021. 2

[17] Zhiwen Fan, Yi Yang, and Mohan S. Kankanhalli. Point 4d transformer networks for spatio-temporal modeling in point cloud videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 852–861, 2021. 2

[18] Tuo Feng, Ruijie Quan, Xiaohan Wang, Wenguan Wang, and Yi Yang. Interpretable3d: An ad-hoc interpretable classifier for 3d point clouds. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(2):1761–1769, 2024. 2

[19] Olga Fourkioti, Matt De Vries, and Chris Bakal. CAMIL: Context-aware multiple instance learning for cancer detection and subtyping in whole slide images. In *The Twelfth In-*

ternational Conference on Learning Representations, 2024. 3, 4

[20] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R. Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7(2):187–199, 2021. 2, 6

[21] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 1

[22] Pedro Gómez-Gálvez, Pablo Vicente-Munuera, Antonio Tagua, Cristina Forja, Ana M. Castro, Marta Letrán, Andrea Valencia-Expósito, Clara Grima, Marina Bermúdez-Gallardo, Óscar Serrano-Pérez-Higueras, Florencia Cavodeassi, Sol Sotillos, María D. Martín-Bermudo, Alberto Márquez, Javier Buceta, and Luis M. Escudero. Scutoids are a geometrical solution to three-dimensional packing of epithelia. *Nature Communications*, 9(1):1–14, 2018. Number: 1 Publisher: Nature Publishing Group. 1

[23] Shikun Huang, Binbin Zhang, Wen Shen, and Zhihua Wei. A claim approach to understanding the pointnet. In *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, page 97–103, New York, NY, USA, 2020. Association for Computing Machinery. 2, 4

[24] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2127–2136. PMLR, 2018. 3

[25] Syed Ashar Javed, Dinkar Juyal, Harshith Padigela, Amaro Taylor-Weiner, Limin Yu, and aaditya prakash. Additive MIL: Intrinsically interpretable multiple instance learning for pathology. In *Advances in Neural Information Processing Systems*, 2022. 3

[26] Alexandr A. Kalinin, Xinhai Hou, Alex S. Ade, Gordon-Victor Fon, Walter Meixner, Gerald A. Higgins, Jonathan Z. Sexton, Xiang Wan, Ivo D. Dinov, Matthew J. O'Meara, and Brian D. Athey. Valproic acid-induced changes of 4D nuclear morphology in astrocyte cells. *Molecular Biology of the Cell*, 32(18):1624–1633, 2021. Publisher: American Society for Cell Biology (mboc). 1

[27] Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean, Jess Leung, Vasudev Sharma, Ayla Khan, Jia Balakrishnan, Safiye Celik, Dominique Beaini, Maciej Sypetkowski, Chi Vicky Cheng, Kristen Morse, Maureen Makes, Ben Mabey, and Berton Earnshaw. Masked autoencoders for microscopy are scalable learners of cellular biology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11757–11768, 2024. 4

[28] Thibault Laugel, Marie-Jeanne Lesot, Christophe Marsala, Xavier Renard, and Marcin Detyniecki. The dangers of post-hoc interpretability: unjustified counterfactual explanations. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, page 2801–2807. AAAI Press, 2019. 2

[29] Oscar Li, Hao Liu, Chaofan Chen, and Cynthia Rudin. Deep learning for case-based reasoning through prototypes: a neural network that explains its predictions. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, 2018. 2

[30] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2018. 2

[31] Zhenan Liu, Leo A. van Grunsven, Elke Van Rossen, Ben Schroyen, Jean-Pierre Timmermans, Albert Geerts, and Hendrik Reynaert. Blebbistatin inhibits contraction and accelerates migration in mouse hepatic stellate cells. *British Journal of Pharmacology*, 159(2):304–315, 2010. 6

[32] A. J. Lomakin, C. J. Cattin, D. Cuvelier, Z. Alraies, M. Molina, G. P. F. Nader, N. Srivastava, P. J. Sáez, J. M. Garcia-Arcos, I. Y. Zhitnyak, A. Bhargava, M. K. Driscoll, E. S. Welf, R. Fiolka, R. J. Petrie, N. S. De Silva, J. M. González-Granado, N. Manel, A. M. Lennon-Duménil, D. J. Müller, and M. Piel. The nucleus acts as a ruler tailoring cell responses to spatial constraints. *Science*, 370(6514):eaba2894, 2020. 1

[33] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, page 163–169, New York, NY, USA, 1987. Association for Computing Machinery. 3

[34] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5(6):555–570, 2021. 3

[35] Xu Ma, Can Qin, Haoxuan You, Haoxi Ran, and Yun Fu. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. In *International Conference on Learning Representations*, 2022. 2, 3, 6, 1

[36] Vincent Maioli, George Chennell, Hugh Sparks, Tobia Lana, Sunil Kumar, David Carling, Alessandro Sardini, and Chris Dunsby. Time-lapse 3-d measurements of a glucose biosensor in multicellular spheroids by light sheet fluorescence microscopy in commercial 96-well plates. *Scientific Reports*, 6 (1):37777, 2016. 4, 5

[37] Reed Naidoo, Olga Fourkioti, Matt De Vries, and Chris Bakal. Survivmil: A multimodal, multiple instance learning pipeline for survival outcome of neuroblastoma patients. In *Proceedings of the MICCAI Workshop on Computational Pathology*, pages 131–141. PMLR, 2024. 3

[38] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 3, 6

[39] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Informa-*

tion Processing Systems. Curran Associates, Inc., 2017. 2, 6

[40] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. In *Advances in Neural Information Processing Systems*, pages 23192–23204. Curran Associates, Inc., 2022. 6, 8

[41] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. 2

[42] Cynthia Rudin, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, and Chudi Zhong. Interpretable machine learning: Fundamental principles and 10 grand challenges. *CoRR*, abs/2103.11251, 2021. 2

[43] Wojciech Samek, Alexander Binder, Grégoire Montavon, Sebastian Lapuschkin, and Klaus-Robert Müller. Evaluating the visualization of what a deep neural network has learned. *IEEE Transactions on Neural Networks and Learning Systems*, 2017. 4

[44] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing Zhang. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. In *Advances in Neural Information Processing Systems*, 2021. 3

[45] Greta Simionato, Konrad Hinkelmann, Revaz Chachanidze, Paola Bianchi, Elisa Fermo, Richard van Wijk, Marc Leonetti, Christian Wagner, Lars Kaestner, and Stephan Quint. Red blood cell phenotyping from 3d confocal images using artificial neural networks. *PLOS Computational Biology*, 17(5):1–17, 2021. 4, 1

[46] Andrew H. Song, Mane Williams, Drew F.K. Williamson, Sarah S.L. Chow, Guillaume Jaume, Gan Gao, Andrew Zhang, Bowen Chen, Alexander S. Baras, Robert Serafin, Richard Colling, Michelle R. Downes, Xavier Farré, Peter Humphrey, Clare Verrill, Lawrence D. True, Anil V. Parwani, Jonathan T.C. Liu, and Faisal Mahmood. Analysis of 3d pathology samples using weakly supervised ai. *Cell*, 187 (10):2502–2520.e17, 2024. 1

[47] Saeid Asgari Taghanaki, Kaveh Hassani, Pradeep Kumar Jayaraman, Amir Hosein Khasahmadi, and Tonya Custis. Pointmask: Towards interpretable and bias-resilient point cloud processing. *arXiv preprint arXiv:2007.04525*, 2020. 2

[48] Hanxiao Tan and Helena Kotthaus. Surrogate model-based explainability methods for point cloud nns. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2927–2936, 2022. 2

[49] Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6410–6419, 2019. 2

[50] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1588–1597, 2019. 2, 4, 7, 3

[51] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 1

[52] Matheus P. Viana, Jianxu Chen, Theo A. Knijnenburg, Ritvik Vasan, Calysta Yan, Joy E. Arakaki, Matte Bailey, Ben Berry, Antoine Borensztejn, Eva M. Brown, Sara Carlson, Julie A. Cass, Basudev Chaudhuri, Kimberly R. Cordes Metzler, Mackenzie E. Coston, Zach J. Crabtree, Steve Davidson, Colette M. DeLizo, Shailja Dhaka, Stephanie Q. Dinh, Thao P. Do, Justin Domingus, Rory M. Donovan-Maiye, Alexandra J. Ferrante, Tyler J. Foster, Christopher L. Frick, Griffin Fujioka, Margaret A. Fuqua, Jamie L. Gehring, Kaytlyn A. Gerbin, Tanya Grancharova, Benjamin W. Gregor, Lisa J. Harrylock, Amanda Haupt, Melissa C. Hendershott, Caroline Hookway, Alan R. Horwitz, H. Christopher Hughes, Eric J. Isaac, Gregory R. Johnson, Brian Kim, Andrew N. Leonard, Winnie W. Leung, Jordan J. Lucas, Susan A. Ludmann, Blair M. Lyons, Haseeb Malik, Ryan McGregor, Gabe E. Medrash, Sean L. Meharry, Kevin Mitcham, Irina A. Mueller, Timothy L. Murphy-Stevens, Aditya Nath, Angelique M. Nelson, Sandra A. Oluoch, Luana Paleologu, T. Alexander Popiel, Megan M. Riel-Mehan, Brock Roberts, Lisa M. Schaefbauer, Magdalena Schwarzl, Jamie Sherman, Sylvain Slaton, M. Filip Sluzewski, Jacqueline E. Smith, Youngmee Sul, Madison J. Swain-Bowden, W. Joyce Tang, Derek J. Thirstrup, Daniel M. Toloudis, Andrew P. Tucker, Veronica Valencia, Winfried Wiegraebe, Thushara Wijeratna, Ruian Yang, Rebecca J. Zaunbrecher, Ramon Lorenzo D. Labitigan, Adrian L. Sanborn, Graham T. Johnson, Ruwanthi N. Gunawardane, Nathalie Gaudreault, Julie A. Theriot, and Susanne M. Rafelski. Integrated intracellular organization and its variations in human iPS cells. *Nature*, 613(7943):345–354, 2023. 1, 4

[53] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10288–10297, 2019. 2

[54] Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, and Wenyu Liu. Revisiting multiple instance neural networks. *Pattern Recognition*, 74:15–24, 2018. 3

[55] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019. 2, 3, 6, 1

[56] Gregory P. Way, Ted Natoli, Adeniyi Adeboye, Lev Litichevskiy, Andrew Yang, Xiaodong Lu, Juan C. Caicedo, Beth A. Cimini, Kyle Karhohs, David J. Logan, Mohammad H. Rohban, Maria Kost-Alimova, Kate Hartland, Michael Bornholdt, Srinivas Niranj Chandrasekaran, Marzieh Haghighi, Erin Weisbart, Shantanu Singh, Aravind Subramanian, and Anne E. Carpenter. Morphology and gene

expression profiling provide complementary information for mapping cell state. *Cell Systems*, 13(11):911–923.e9, 2022. 1

[57] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9613–9622, 2019. 2, 6

[58] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 4, 3

[59] Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 915–924, 2021. 2, 3, 6, 1, 5

[60] Qiangeng Xu, Xudong Sun, Cho-Ying Wu, Panqu Wang, and Ulrich Neumann. Grid-gcn for fast and scalable point cloud learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5660–5669, 2020. 2

[61] Xu Yan, Chaoda Zheng, Zhen Li, Sheng Wang, and Shuguang Cui. Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 5

[62] Cheng-Kun Yang, Ji-Jia Wu, Kai-Syun Chen, Yung-Yu Chuang, and Yen-Yu Lin. An mil-derived transformer for weakly supervised point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11830–11839, 2022. 3

[63] Xi Yang, Ding Xia, Taichi Kin, and Takeo Igarashi. Intra: 3d intracranial aneurysm dataset for deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 4, 1

[64] Jianhui Yu, Chaoyi Zhang, Heng Wang, Dingxin Zhang, Yang Song, Tiange Xiang, Dongnan Liu, and Weidong Cai. 3d medical point transformer: Introducing convolution to attention networks for medical point cloud analysis, 2021. 2, 3, 6

[65] Binbin Zhang, Shikun Huang, Wen Shen, and Zhihua Wei. Explaining the pointnet: What has been learned inside the pointnet? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 2

[66] Gege Zhang, Qinghua Ma, Licheng Jiao, Fang Liu, and Qigong Sun. Attan: Attention adversarial networks for 3d point cloud semantic segmentation. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 789–796. International Joint Conferences on Artificial Intelligence Organization, 2020. Main track. 2

[67] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16239–16248, 2021. 2

[68] T. Zheng, C. Chen, J. Yuan, B. Li, and K. Ren. Pointcloud saliency maps. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1598–1606, Los Alamitos, CA, USA, 2019. IEEE Computer Society. 2, 4

[69] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning Deep Features for Discriminative Localization . In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2921–2929, Los Alamitos, CA, USA, 2016. IEEE Computer Society. 2, 4