Explain Any Pathological Concept: Discovering Hierarchical Explanations for Pathology Foundation Models

Shuting Xu[1†], Junlin Hou[2†], and Hao Chen[2,3,4,5,6✉]

[1] School of Medicine, Nankai University, Tianjin, China.
[2] Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.
[3] Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.
[4] Division of Life Science, The Hong Kong University of Science and Technology, Hong Kong, China.
[5] HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China.
[6] State Key Laboratory of Nervous System Disorders, The Hong Kong University of Science and Technology, Hong Kong, China.

Abstract. Foundation models have demonstrated significant promise in medical image analysis, particularly in pathology. However, their black-box nature makes it challenging for clinicians to understand their decision-making processes. In this paper, we evaluate the explainability of existing pathology foundation models based on visual concepts. Considering the hierarchical structure of pathological anatomy, comprising of regions, units, and cells, we introduce a novel Hierarchical Concept-based Explanation (HCE) method to illuminate how concepts at different levels influence the model's predictions. Specifically, our approach begins with the utilization of a specialist-generalist collaborative segmentation model to perform instance segmentation across various levels. We then employ a surrogate model to approximate the target foundation model and compute the Shapley values for each concept. Finally, we visualize these contributions through a comprehensive global ShapMap. We evaluate several state-of-the-art pathology foundation models, including CONCH, UNI, and Virchow, on an adenoma classification task. The findings reveal that the explanations provided by CONCH and UNI show better composability, suggesting they draw from a wider contextual understand demonstrate great separability, reflecting a reliance on specific regions. Additionally, we explore the consistency of concept explanations across different foundation models.

Keywords: Explainable AI · Concept-based Explanation · Foundation Model · Pathology Image.

† These authors contributed equally to this work.

1 Introduction

Pathological examinations play a crucial role in cancer diagnosis [1], subtyping [2], and survival prediction [3]. Recently, pathology foundation models [4–9] have emerged as promising tools in image analysis, achieving impressive performance on various downstream tasks through pre-training on large-scale pathology images or image-report pairs. However, their highly complex architectures and vast number of parameters result in limited interpretability, undermining doctors' trust in clinical use. Therefore, investigating the interpretability of pathology foundation models is crucial.

To tackle this challenge, explainable artificial intelligence (XAI) has emerged to ensure that a model's decision is both transparent and comprehensible to humans [10, 11]. Furthermore, post-hoc XAI methods [12] analyze and interpret the decision process after a trained model makes predictions, thereby providing insights into how outputs are predicted. Attribution-based approaches [13–16], a classical XAI method, aim to quantify the contribution of each input to the network's output. However, they only offer explanations at the pixel or superpixel level, failing to reveal the specific visual features underpinning the model's decisions and thus lacking a deeper level of understanding [17]. Recently, concept-based XAI has attracted considerable attention for its ability to provide more understandable explanations [18, 19]. However, it often rely on human annotations or specific concept discovery methods. Recent studies such as Explain Any Concept (EAC) [17] and Lesion Concept Explainer (LCE) [20] have attempted to use the strong zero-shot segmentation capabilities of SAM [21] to automatically discover all concepts in an image, computing the Shapley value [22] for each concept region to indicate its contribution to the final prediction. While these models perform well for images with a single dominant target, they generate suboptimal explanations for pathology images containing complex or overlapping structures.

In pathology images, complex semantic information is typically organized into three hierarchical structures, ranging from broad regions (e.g., tissue) to specific functional units (e.g., glands) and individual cells [23]. These hierarchical layers integrate macro-level tissue organization and micro-level cellular details, offering a more holistic view of pathological changes. Clinicians integrate concepts across these different levels to make diagnoses and provide treatment recommendations. However, existing models like EAC [17] are usually limited to interpreting a single tissue type. To address this issue, we propose a Hierarchical Concept-based Explanation (HCE) method, which enhances the interpretability of foundation models, while also aligning with the clinical significance of pathology images. Specifically, we first adopt a hierarchical segmentation approach for pathology images by integrating the specialist models' robust segmentation abilities with the strong generalization performance of a segmentation foundation model. This method automatically produces a highly accurate and clinically meaningful set of concepts from pathology images. Next, we compute the Shapley value for each concept at every hierarchical level, quantifying its contribution to the model's predictions. However, computing the Shapley value under this approach can

be computationally expensive. We employ a lightweight surrogate model [17] by employing binary encoding instead of complex image inputs and replacing large models with shallow neural networks. In addition, we present ShapMap, a Shapley-based importance visualization method that illustrates the model's attention across different feature levels. In the experiments, we perform interpretability analyses of pathology foundation models (UNI [4], CONCH [5], Virchow [6]) on the pathology image classification task. The results reveal that these foundation models effectively leverage local features but lack comprehensive feature integration. We further examine the similarity in focal areas across different large models, offering new insights for optimizing pathology image analysis.

2 Methodology

As illustrated in Fig. 1, our proposed HCE consists of two main stages: a pathology image classification stage and a hierarchical concept-based explanation stage. In the classification stage, we use target foundation models Φ_t as the feature extractor to produce an embedding of the input image, which is then passed through a fully connected (FC) layer Ψ to obtain the classification prediction. In the explanation stage, we first perform hierarchical instance segmentation. Next, we train a surrogate model Φ_s to approximate the behavior of the target foundation model Φ_t. We simplify the segmentation into binary features as input to Φ_s, which outputs feature vector, then passed through the same FC layer Ψ to obtain classification probabilities. Finally, we apply Monte Carlo [24] simulations to compute the Shapley value for each concept at every hierarchical level, representing its contribution to the classification. Through this process, we construct a hierarchical ShapMap to visualize and explain the model's predictions. The following sections provide a detailed description of our model.

Hierarchical Segmentation. In clinical practice, pathologists rely on concepts at multiple hierarchical levels to diagnose pathology images. Inspired by this process, we perform hierarchical segmentation to isolate semantic targets at different scales, including cells (epithelial cells, fibroblasts, and inflammatory cells, and other cells), units (glands), and regions (tissue regions). Traditional segmentation models, trained via supervised learning for specific pathology segmentation tasks, often achieve superior segmentation performance. Meanwhile, the latest segmentation foundation model SAM2 [25], trained on large-scale datasets, offers stronger generalization capabilities. We therefore combine the advantages of both approaches, proposing a specialist-generalist collaborative segmentation model. Specifically, we first use specialized models for each scale to generate a series of masks. For the foreground tissue region, we apply a segmentation method based on grayscale values [26]. For the gland region, we employ a U-Net model [27] trained on the GlaS dataset [28]. For the cell region, we utilize HoverNet [29], a model known for its strong performance on cell segmentation datasets. Based on the generated masks from these specialized models, we extract the bounding boxes of each instance as spatial prompts to SAM2. Finally, SAM2 generates a set of concepts for each hierarchical level. Here, we represent each

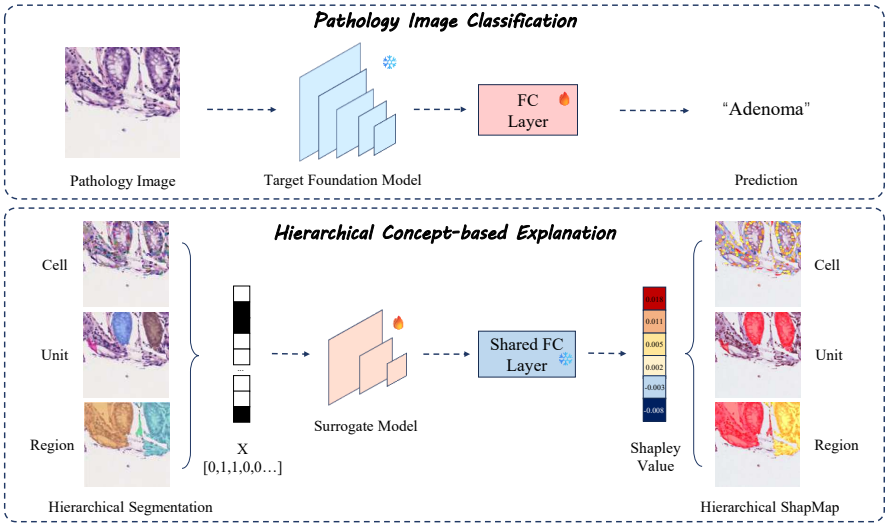

Fig. 1. The overall framework of our proposed HCE, including a pathology image classification stage and a hierachical concept-based explanation stage.

set of concepts as $C_l = \{c_l^1, c_l^2, \ldots, c_l^n\}$, where n denotes the number of concepts at a given hierarchical level l.

Surrogate Model. Our aim is to assess the contribution of each segmented concept to the classification outcome using Shapley value [22]. However, the structural complexity and large number of parameters in pathology foundation models pose significant computational challenges for Shapley value analysis. To address this issue, we adopt a lightweight surrogate model Φ_s to approximate the target foundation model Φ_t. Inspired by the EAC model [17], we employ the per-input equivalence method to construct the surrogate model Φ_s. Specifically, Φ_s employs a two-layer linear network for feature extraction that achieves competing prediction accuracy while maintaining relatively low computational complexity. Given a segmented pathology image, Φ_s processes a binary concept vector, each element corresponding to a concept, i.e., a value of 0 indicates that the concept is masked, while 1 indicates retained. The extracted features are then projected through a shared fully-connected layer Ψ to produce classification probabilities Y_s. During optimization, we freeze the FC layer Ψ and minimize the cross-entropy loss between the surrogate predictions Y_s and target probabilities Y_t from the classification stage:

$$\mathcal{L} = -(Y_t \log Y_s + (1 - Y_t) \log(1 - Y_s)). \tag{1}$$

Hierarchical ShapMap. Shapley value has been widely adopted in machine learning to quantify the contribution of individual features or concepts. We compute Shapley value based on the trained surrogate model to obtain the contribution of each concept, which is formulated as a enumerated weighted sum of

all possible marginal contributions from each concept. We define the marginal contribution of the i-th concept c_l^i at level l as the difference between the model prediction on $S \cup \{c_l^i\}$ and S, where $S \subseteq C_l \setminus \{c_l^i\}$:

$$\Delta_{c_l^i} = v(S \cup \{c_l^i\}) - v(S). \tag{2}$$

Here, $v(S)$ represents the prediction of the surrogate model Φ_s on the image x, using only the concepts in S with the remaining concepts masked. Then, the Shapley value of concept c_l^i is defined as:

$$\varphi_{c_l^i}(x) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\binom{n-1}{k-1}} \sum_{S \in S_k(i)} \Delta_{c_l^i}(S), \tag{3}$$

where n represents the number of concepts at each level l, $S_k(i)$ represents the set of all coalitions of size k that do not include c_l^i. Given the large number of possible coalitions, we approximate the Shapley value using Monte Carlo sampling [24]. Specifically, for each concept, we sample K coalitions and approximate the Shapley value as:

$$\varphi'_{c_l^i}(x) = \frac{1}{K} \sum_{k=1}^{K} \Delta_{c_l^i(S_k)}. \tag{4}$$

To further visualize how different concepts contribute to classification, we propose ShapMap at each hierarchical level based on Shapley values. For each scale, we define the mask corresponding to the positive Shapley value as red to yellow, indicating its positive contribution to the classification, while the mask corresponding to the negative Shapley value is defined as deep blue to white.

3 Experiment

Implementation Details and Dataset. For validation and analysis, we select three state-of-the-art pathology models: (1) UNI [4] is a vision foundation model, trained on over 200 million pathology H&E and IHC images; (2) CONCH [5] is a vision-language foundation model, trained on over 1.17 million image-caption pairs; (3) Virchow [6] is a vision foundation model, trained on 1.5 million whole slide images. For the binary classification task, we employ the image encoder from three models as a frozen feature extractor, training only the classifier. The FC layer is trained for 100 epochs using the AdamW optimizer with weight decay of 0.001. The surrogate model takes a (1000, X) binary matrix as input, where X denotes concept count per level. All images are resized to 512×512 pixels, and the batch size is set to 32. All experiments are performed using Pytorch on an NVIDIA GeForce RTX 4090 GPU.

We use the publicly available Chaoyang dataset [30] for colonoscopy classification tasks. Images of normal and adenoma classes are selected, resulting in the training set of 1,111 normal images and 1,404 adenoma images, and the test set of 705 normal images and 840 adenoma images. Each image contains 286

cells (e.g., epithelial cell), 4 units (e.g., gland), and 2 regions (e.g., foreground) on average. The model's interpretability is evaluated on the test set.

Evaluation Metric. For the classification task, we use accuracy as the metric to evaluate classification performance. For model interpretation, we use the deletion AUC [31] as the metric to evaluate the model's separability and composability. The evaluation starts with an unmasked image and gradually removes concepts in descending order of Shapley values. Composability reflects the model's ability to integrate multiple concepts from different local regions, and a higher deletion AUC within the same scale suggests the model maintains good prediction accuracy despite some concept removal, indicating stronger composability and reliance on multiple local features. While separability indicates reliance on a few local regions, which a lower deletion AUC value indicates.

Meanwhile, different foundation models may focus on the similar regions across the same hierarchical level of pathology images. To quantify the similarity of positively contributing concepts across models, we compute the similarity between two models as follows:

$$\text{Similarity} = \frac{num(C_l^A \cap C_l^B)}{num(C_l^B)}, \tag{5}$$

where C_l^A and C_l^B represent the sets of positively contributing concepts for models A and B in each level l, respectively. num represents the number of concepts in the set.

Classification Performance of Foundation Models. Considering the multi-level semantic information in pathology images, we investigate the classification performance of different foundation models based on hierarchical segmentation results. Fig. 2 presents the visualization of segmentation results across different hierarchical levels of the pathology images.

As shown in Table 1, the classification accuracy of all models increases across three levels from cell, unit to region in pathology images. This result suggests that as the scale expands from individual cells to larger regions, models gain richer contextual information, leading to more accurate predictions. The region level integrates not only local details from cells and units but also broader spatial features, providing the most substantial contribution to classification decisions.

When comparing the three pathology foundation models, CONCH consistently achieves the highest classification accuracy across all levels. This indicates its superior ability to capture multi-scale information. It effectively extracts pathological features from larger structures of unit and region, while also preserving fine-grained details at the cell level. These results highlight CONCH's stronger representational capacity and robustness.

Separability and Composability of Foundation Models. Table 2 presents the deletion AUC of different foundation models in adenoma images. Across three hierarchical levels, Virchow exhibits the lowest deletion AUC, indicating higher separability, meaning it relies more on single concepts for decision-making. In

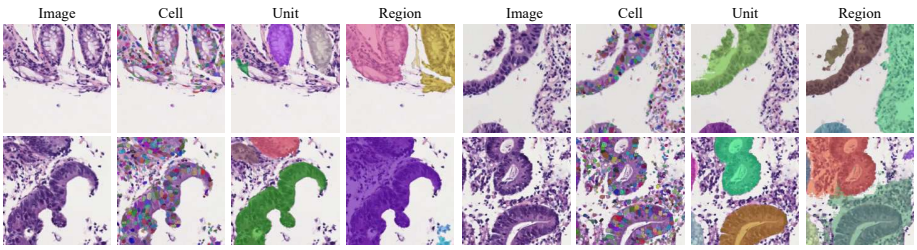

Fig. 2. Segmentation results of hierarchical levels in pathology images.

Table 1. Accuracy (%) of UNI, CONCH, and Virchow across different levels for adenoma classification.

Method	Cell	Unit	Region
UNI [4]	2.02	86.78	97.14
CONCH [5]	8.33	98.21	99.04
Virchow [6]	3.09	87.14	94.16

Table 2. Composability (%) of UNI, CONCH, and Virchow across different levels for adenoma classification.

Method	Cell	Unit	Region
UNI [4]	88.04	47.74	47.76
CONCH [5]	80.54	74.31	61.80
Virchow [6]	74.51	43.93	30.56

contrast, UNI and CONCH show higher AUC values, suggesting better composability, as they integrate multiple concepts to form predictions.

At different levels, UNI achieves the highest deletion AUC at the cell level, indicating strong composability by integrating multiple cellular features into its decisions. Meanwhile, CONCH demonstrates the best composability at the unit and region levels, effectively capturing structural features by integrating spatial concepts at larger scales. Clinically, high separability implies that a model behaves like a "key abnormality detector", aiding doctors in quickly identifying the precise location of abnormalities, which is an essential factor for accurate diagnosis. In contrast, high composability suggests a more holistic "pattern reasoning" approach by integrating information across different levels or regions. This capability helps clinicians gain a comprehensive understanding of the lesion within its larger pathological landscape, facilitating a more holistic diagnosis.

ShapMap. To further illustrate what features foundation models focus on across hierarchical levels in pathology images, we present several case studies, as shown in Fig. 3. In the visualization, HCE reveals the key concepts that model decisions primarily rely on, such as regions, units, and cells, which are highlighted in red. It can be observed that at different hierarchical levels, CONCH mainly focuses on true pathological regions, which aligns with its higher classification accuracy. At the cell level, all models exhibit omissions in identifying abnormal cells, further validating their lower accuracy at this scale. Furthermore, across all three levels, a coarse-to-fine relationship is observed. This method allows pathologists to first identify suspicious lesion areas at the region level, then analyze abnormal glands at the unit level, and finally examine abnormal cells.

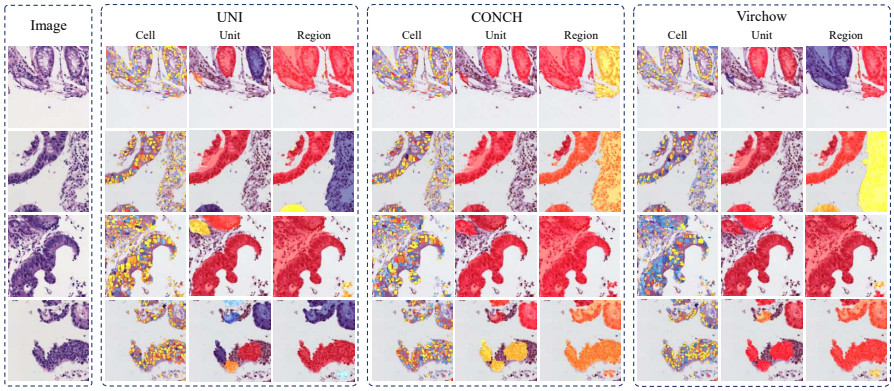

Fig. 3. Case Studies illustrating the hierarchical focus of foundation models in pathology image classification.

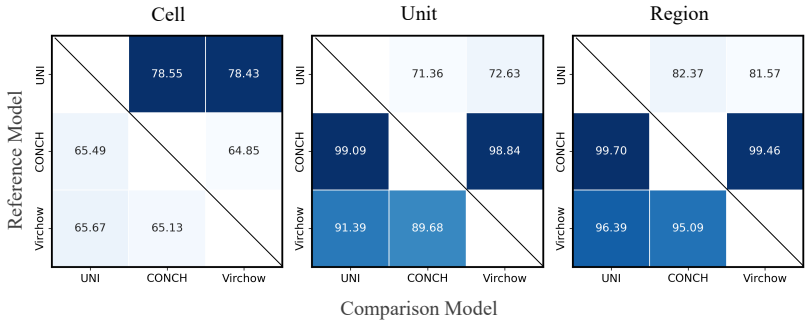

Fig. 4. Similarity of features of UNI, CONCH, and Virchow across different levels.

Similarity of Features Focused by Foundation Models. Fig. 4 presents the similarity scores among three models across each level. At broader levels (unit and region), most features attended by UNI are also captured by CONCH, but CONCH identifies additional regions, contributing to its superior performance over UNI. This insight highlights a key direction for model improvement, identifying relevant regions attended by high-performing models but overlooked by others. At finer hierarchical levels, the similarity among the three models is lower, possibly due to the high diversity of cell-level concepts. None of the models appear to have learned a consistent set of meaningful features, resulting in less overlap in their areas of attention. Furthermore,the common features that the model focuses on are also shown in ShapMap obviously, which may require special attention from doctors.

4 Conclusion

In this study, we introduced a Hierarchical Concept-based Explanation (HCE) method for pathology image analysis. By using the specialist-generalist collabo-

rative segmentation model, we achieve automatic concept discovery. To reduce computational overhead, we employ a lightweight surrogate model with binary encoding. We use Monte Carlo simulations to compute the Shapley value for each concept and introduce a visualization method, ShapMap. We evaluate the classification features of different pathology foundation models. Experimental results indicate that these foundation models effectively leverage local features but lack comprehensive feature integration.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 62202403), Innovation and Technology Commission (Project No. MHP/002/22 and ITCPD/17-9), Research Grants Council of the Hong Kong Special Administrative Region, China (Project No: T45-401/22-N) and National Key R&D Program of China (Project No. 2023YFE0204000).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Claire Mounier-Vehier, Christophe Lions, Patrick Devos, Olivier Jaboureck, Serge Willoteaux, Alain Carre, and Jean-Paul Beregi. Cortical thickness: an early morphological marker of atherosclerotic renal disease. *Kidney international*, 61(2):591–598, 2002.
2. B Weigelt, HM Horlings, B Kreike, MM Hayes, M Hauptmann, LFA Wessels, D De Jong, MJ Van de Vijver, LJ Van't Veer, and JL Peterse. Refinement of breast cancer classification by molecular characterization of histological special types. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 216(2):141–150, 2008.
3. John A Kellum. Acute kidney injury. *Critical care medicine*, 36(4):S141–S145, 2008.
4. Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.
5. Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024.
6. Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Siqi Liu, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, et al. Virchow: A million-slide digital pathology foundation model. *arXiv preprint arXiv:2309.07778*, 2023.
7. Jinxi Xiang, Xiyue Wang, Xiaoming Zhang, Yinghua Xi, Feyisope Eweje, Yijiang Chen, Yuchen Li, Colin Bergstrom, Matthew Gopaulchan, Ted Kim, et al. A vision–language foundation model for precision oncology. *Nature*, pages 1–10, 2025.
8. Jiabo Ma, Zhengrui Guo, Fengtao Zhou, Yihui Wang, Yingxue Xu, Jinbang Li, Fang Yan, Yu Cai, Zhengjie Zhu, Cheng Jin, et al. Towards a generalizable pathology foundation model via unified knowledge distillation. *arXiv preprint arXiv:2407.18449*, 2024.

9. Yingxue Xu, Yihui Wang, Fengtao Zhou, Jiabo Ma, Cheng Jin, Shu Yang, Jinbang Li, Zhengyu Zhang, Chenglong Zhao, Huajun Zhou, et al. A multimodal knowledge-enhanced whole-slide pathology foundation model. *arXiv preprint arXiv:2407.15362*, 2024.
10. Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.
11. Junlin Hou, Sicen Liu, Yequan Bie, Hongmei Wang, Andong Tan, Luyang Luo, and Hao Chen. Self-explainable ai for medical image analysis: A survey and new outlooks. *arXiv preprint arXiv:2410.02331*, 2024.
12. Carl O Retzlaff, Alessa Angerschmid, Anna Saranti, David Schneeberger, Richard Roettger, Heimo Mueller, and Andreas Holzinger. Post-hoc vs ante-hoc explanations: xai design guidelines for data scientists. *Cognitive Systems Research*, 86:101243, 2024.
13. 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*, pages 1135–1144, 2016.
14. Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
15. Amirata Ghorbani, James Wexler, James Y Zou, and Been Kim. Towards automatic concept-based explanations. *Advances in neural information processing systems*, 32, 2019.
16. Ruth C Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of the IEEE international conference on computer vision*, pages 3429–3437, 2017.
17. Ao Sun, Pingchuan Ma, Yuanyuan Yuan, and Shuai Wang. Explain any concept: Segment anything meets concept-based explanation. *Advances in Neural Information Processing Systems*, 36:21826–21840, 2023.
18. Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, et al. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In *International conference on machine learning*, pages 2668–2677. PMLR, 2018.
19. Junlin Hou, Jilan Xu, and Hao Chen. Concept-attention whitening for interpretable skin lesion diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 113–123. Springer, 2024.
20. Weiji Kong, Xun Gong, and Juan Wang. Lce: A framework for explainability of dnns for ultrasound image based on concept discovery. *arXiv preprint arXiv:2408.09899*, 2024.
21. Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.
22. Lloyd S Shapley et al. A value for n-person games. 1953.
23. Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Juming Xiong, Shunxing Bao, Hao Li, Mengmeng Yin, Yu Wang, Shilin Zhao, et al. Hats: Hierarchical adap-

tive taxonomy segmentation for panoramic pathology image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 155–166. Springer, 2024.
24. Eunhye Song, Barry L Nelson, and Jeremy Staum. Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083, 2016.
25. Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.
26. Daniel Bug, Friedrich Feuerhake, and Dorit Merhof. Foreground extraction for histopathological whole slide imaging. In *Bildverarbeitung für die Medizin 2015: Algorithmen-Systeme-Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2015 in Lübeck*, pages 419–424. Springer, 2015.
27. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
28. Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017.
29. Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019.
30. Chuang Zhu, Wenkai Chen, Ting Peng, Ying Wang, and Mulan Jin. Hard sample aware noise robust learning for histopathology image classification. *IEEE transactions on medical imaging*, 41(4):881–894, 2021.
31. Vitali Petsiuk, Abir Das, and Kate Saenko. Rise: Randomized input sampling for explanation of black-box models. *arXiv preprint arXiv:1806.07421*, 2018.