AIDA: AI-Driven Intelligent Diagnostics and Analytics

Zanxi Ruan[1], Stefano Gobbo[2], Luca Cima[3], Mattia Mondo[1], Shakiba Sharifi[1], Enrico Munari[3], Aldo Scarpa[2,4], Omar Shahbaz Khan[5], Stefán Ólafsson[5], Andrea Giachetti[1], Francesco Setti[1], Yiming Wang[6], Björn Þór Jónsson[5], and Marco Cristani[1,5*]

[1] Department of Engineering for Innovation Medicine, University of Verona, Italy
{zanxi.ruan, marco.cristani}@univr.it
[2] Department of Diagnostic and Public Health, University of Verona, Italy
[3] Department of Pathology and Diagnostics, University and Hospital Trust of Verona, Italy
[4] ARC-Net Research Centre, Verona, Italy
[5] Department of Computer Science, Reykjavik University, Iceland
[6] Fondazione Bruno Kessler (FBK), Trento, Italy

Abstract. We present AIDA (AI-Driven Intelligent Diagnostics and Analytics), an ongoing initiative aimed at advancing the field of digital pathology through the integration of artificial intelligence. The aim of AIDA is twofold: 1) to create a collaborative repository of annotated data on rare pathologies, where specialists from around the world contribute expert insights through both text and graphical explanations of complex cases; and 2) to leverage this annotated data to train machine learning tools, with an emphasis on Visual Language Models (VLMs), that can enhance and accelerate the diagnosis of such conditions. The project adopts a human-in-the-loop paradigm supported by retrieval and exploration of large-scale, high-dimensional medical data. This position paper outlines the current status of the AIDA project, highlights key challenges, and explores possible directions for its ongoing and future evolution.

Keywords: Digital Pathology · Human in the Loop · VLMs

1 Introduction

Digital pathology (DP) has rapidly advanced with the progress in Artificial Intelligence (AI), thanks to the long-lasting efforts in automated analysis of histopathological images [35] and the recent rise of large multimodal models [3,36]. Despite significant progress, this field still faces substantial challenges related to the inherent complexity and ambiguity of tissue image interpretation, inter-annotator variability, and inconsistent terminologies [2, 8, 23, 40]. Consequentially, the diagnostic pipeline (from image acquisition to a final medical assessment)

*Corresponding author: marco.cristani@univr.it

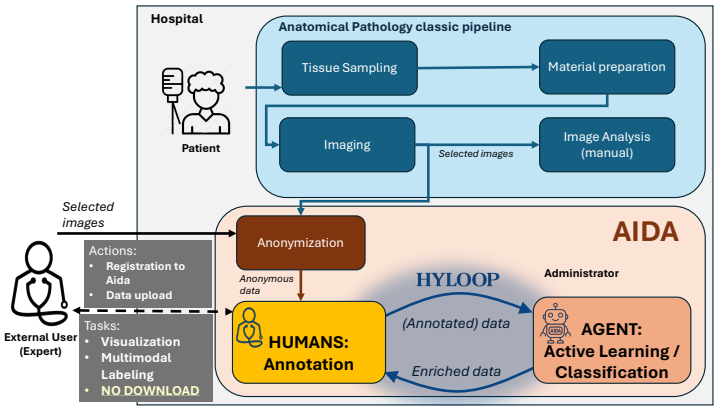

Fig. 1: Overview of the AIDA platform. The upper part shows the classic anatomical pathology workflow, while the lower part illustrates the AIDA pipeline with the HYLOOP reasoning engine.

can take hours, that is not only physically and cognitively demanding for the medics, but also limiting timely responses to patients at scale.

We devise the AIDA (Artificial Intelligence for Diagnostic Assistance) project within this context, aiming to develop an assistive platform with tools that enable faster and more accurate diagnoses. As shown in Fig. 1, AIDA represents a spin-off from the classic anatomical pathology pipeline (top). Selected images are anonymized and transferred to the AIDA platform (bottom), where the *HYLOOP* (Human-in-the-Hyperloop) is the main reasoning engine. HYLOOP is based on two main parts: HUMANS, and AGENT. HUMANS is a collaborative platform for storing, visualizing, annotating, and sharing digital pathology data. Designed for expert users such as pathologists and medical researchers, it features role-based access control and supports high-resolution image uploads, manual annotations, and structured labeling using controlled vocabularies such as SNOMED CT[1]. The platform also fosters interaction through social-feedback tools like comments, likes/dislikes, and discussion threads. It is built to:

- Handle massive image datasets with high redundancy and access performance;
- Support secure, credentialed access and strict anonymization policies;
- Foster community engagement through gamified mechanisms (user roles, activity levels, and thematic challenges);
- Provide innovative annotation tools which integrate text and graphics;
- Ensure annotation consistency by using AI tools to perform uni-modal and cross-modal checks between pathological images and expert tags/comments;
- Provide case recommendation, as the system suggests relevant cases based on user profiles and diagnostic interest; and

The AIDA AGENT is the AI-driven engine of AIDA. It leverages machine learning techniques to provide assistive diagnostic services and advanced data curation tools. Among its core functionalities:

[1] `https://www.snomed.org/`

- Image classification and visual explainability: pathology images are automatically interpreted, with saliency maps highlighting key diagnostic features;
- Image classification and textual explainability: classification outcomes are augmented by textual explanation. The text is labeled so that the confidence in the textual explanations could be evaluated quantitatively;
- Semantic segmentation: produces segmentation maps where each segment is associated to a specific medical meaning.

The HYLOOP mechanism transforms AIDA to a hybrid human-AI platform, enabling faster, more consistent, expert-supervised diagnostics, augmenting clinical judgment. The two modules communicate with each other: the AGENT classification module identifies low-confidence predictions and sends this information to the HUMANS annotation tool, which then requests additional labeled data or more detailed annotations. This enriched supervision is expected to improve training and, in turn, lead to better classification performance.

The AIDA project is led by the University of Verona (UNIVR), in collaboration with Reykjavik University (RU), bringing together complementary expertise in multimodal learning and clinical pathology. At UNIVR, computer engineers from the Department of Engineering for Innovation Medicine [1] work closely with clinical pathologists from the Department of Diagnostics and Public Health [2], ensuring tight integration between system development and clinical validation [5, 29, 30]. UNIVR also leads research in Vision-Language Models and confident learning [14, 17, 37–39], which are essential for building trustworthy AI in medical imaging. RU contributes complementary expertise in large-scale interactive retrieval and active learning through the Exquisitor platform [16, 34], along with experience in processing unstructured clinical text via NLP pipelines [12, 13], and in designing human-agent interaction frameworks for healthcare [27]. The collaboration also involves peers from other research centers, such as Fondazione Bruno Kessler (FBK), contributing to robust AI methods and domain-specific applications.

The paper is organized as follows: Sec. 2 analyzes the related literature; Sec. 3 details the current state of AIDA; Sec. 4 show what are the open challenge, and what is our position for addressing them. Finally, Sec. 5 concludes the paper.

2 Related Work

We cover main related works on DP platforms and relevant techniques below.

AI in Digital Pathology. The integration of AI in DP has been an area of intensive research and innovation. Significant efforts have been dedicated to addressing the challenges of interobserver variability and inconsistent diagnoses among pathologists [2, 8, 23, 40]. Numerous platforms such as PathPresenter, KiKo, ExpertPath, and PathElective have provided essential tools for sharing, annotating, and discussing histopathological images, facilitating collaborative diagnostics and educational efforts [9, 20, 28, 32].

[1] `https://www.dimi.univr.it`

[2] `https://www.ddsp.univr.it`

Vision-Language Models. Advances in AI have increasingly leveraged multimodal data, combining visual and linguistic information to enhance the interpretability and accuracy of diagnostic models. Techniques such as VisualBERT, ViLBERT, MDETR, and UniTAB have demonstrated promising capabilities in vision-language integration, providing robust foundations for detailed image annotation and interpretation [15, 19, 21].
Confident Learning. Additionally, methodologies like Confident Learning have emerged to effectively estimate uncertainties and reduce label noise in training data, crucial for building reliable AI-driven diagnostic systems [26, 43].
Active Learning. Active learning methodologies have also become increasingly important in DP, optimizing annotation efforts by intelligently selecting the most informative cases for expert review [11, 18, 33]. This approach significantly improves annotation efficiency, particularly in specialized medical fields where expert input is scarce and costly.

3 Current Status of the AIDA Project

The AIDA system has reached a stable and operational stage for several of its core modules, which are now ready for use and integration. The development of the full platform is progressing in phases, with the completion of the first full release currently planned for the end of June 2026. In the following, we report the completed modules.
Data Preparation. The AIDA project, in collaboration with the University Hospital of Verona and the ARC-Net [1] cancer research center, has established a compliant and stable data access pipeline to obtain anonymized DP images derived from real clinical practice. All patient data used in this study are anonymized in compliance with the EU GDPR and relevant national regulations. This real-world data forms a solid foundation for model development, validation, and research on clinical translatability. In addition, the project team has systematically integrated several internationally recognized public datasets (MHIST [41], NuInsSeg [22], HISTAI [25]) to supplement the development phase with diverse samples and standardized benchmarks.
Platform Development. At the current stage, AIDA is equipped with interactive annotation tools that support whole-slide image (WSI) visualization and basic labeling, and these tools are already in active use by the team. In parallel, a mature image retrieval platform, Exquisitor [16], developed by Reykjavik University, provides fast, content-based search over large-scale pathology archives, enabling integrated decision support within the system. The integration of Exquisitor with AIDA is the ongoing work. AIDA is deployed on multiple *Samsung Galaxy Book4 Pro 360* laptops for annotation, while model training and inference run on a cluster of servers equipped with NVIDIA RTX 5090 GPUs.
Annotation. The AIDA team has conducted exploratory research on a representative cross-modal task: how to align and generate multimodal information by integrating coarse sketch-like inputs with fine-grained, structured visual–linguistic signals. Based on a large-scale fashion image dataset, we proposed a novel mul-

timodal fusion framework [10] that effectively integrates heterogeneous inputs, including global and local visual features, sketch structures, free-form text, and attribute descriptions.

Although the initial validation was conducted in the domain of general visual data, the core modeling principles and algorithmic strategies are directly transferable to the pathology setting in AIDA. In clinical practice, pathologists also use sketches to mark regions of interest and describe findings using structured terminology. Our proposed approach provides both the theoretical foundation and technical feasibility for integrating sketch + text + local/global multimodal annotation in the AIDA framework.

Medical Retrieval and Exploration. Exquisitor has competed successfully in the Video Browser Showdown [34], which includes a collection of Gynecologic Laparoscopy surgery videos [24] among its datasets. While textually describing a query for surgical videos is difficult for non-experts, the use of a visual relevance feedback approach showed more success.

4 Challenges

Several challenges remain critical in the development of AIDA:

- **Interobserver Variability**: Persistent differences among pathologists' interpretations lead to inconsistent diagnostic outcomes, complicating the creation of standardized AI models [2, 40].
- **Data Quality and Annotation Noise**: Ensuring high-quality annotations and managing label noise are vital for training reliable AI systems, yet difficult due to variability and complexity in histopathological data [26, 42].
- **Scalability of Expert Annotations**: Limited availability of expert pathologists poses significant constraints on the scalability and effectiveness of active learning and annotation validation processes [18, 33].
- **Model Interpretability and Trust**: Developing AI models that are transparent, interpretable, and trustworthy in their decision making, to facilitate trust and acceptance within clinical practice remains a significant challenge [11, 43].
- **Integration into Clinical Workflow**: Seamlessly integrating advanced AI tools into existing clinical workflows without disruption is essential but poses substantial logistical and practical challenges [9, 28].

Collectively, these advancements and challenges underscore the critical role of sophisticated AI methodologies and interactive learning frameworks in overcoming current limitations and shaping the future of digital anatomical pathology.

To systematically address the five core challenges previously outlined in the field of DP, the AIDA framework proposes a deeply integrated and synergistic suite of solutions. In the following subsections, we detail how AIDA's methodology is specifically designed to tackle each of these core challenges.

4.1 Addressing Interobserver Variability

Interobserver variability is a critical issue in anatomical pathology, characterized by persistent differences among pathologists' interpretations of histopathological findings on whole slide images. These differences reflect the inherent limits of human understanding and can lead to inconsistent diagnostic outcomes, affecting patient management, prognostic stratification, and eligibility for targeted therapies. Even among experts, variability arises from subjective assessment, heterogeneous training, and difficulties in applying standardized criteria, especially in borderline cases such as prostate cancer grading or PD-L1 evaluation [6, 7, 31]. This variability poses significant challenges in developing and validating AI models in DP. For AI to reach clinically useful performance, it requires high-quality ground truth labels, but when expert annotations diverge, it becomes difficult to establish a reference standard for training and benchmarking.

Furthermore, AIDA incorporates advanced generative AI techniques that integrate visual and conceptual reasoning, enabling the production of highly interpretable and transparent diagnostic reports. These reports are accompanied by explicit confidence scores that indicate the reliability of annotations, thereby reducing uncertainty and facilitating consensus among pathologists.

AIDA's HYLOOP engine, an iterative learning mechanism utilizing sophisticated active learning strategies, promotes continuous interaction and validation with human experts. This ensures that any discrepancies in annotations are systematically identified and resolved, leading to refined and harmonized datasets. By actively engaging pathologists through this iterative feedback loop, AIDA progressively minimizes interobserver variability, thereby enhancing the reliability and consistency of AI-driven diagnostic outcomes and supporting the development of standardized and robust AI models for clinical practice.

4.2 Addressing Data Quality and Annotation Noise

High-quality annotations are essential for training reliable AI models. However, in histopathology, obtaining accurate ground truth labels is extremely challenging, and label noise is almost inevitable. This difficulty stems from several factors: observer variability causes inconsistent labels; tumor heterogeneity and subtle features create ambiguity; and free-text pathology reports introduce semantic noise through inconsistent terminology or contradictions (e.g., a conclusion stating "no cancer" while describing "atypical cells"). If such noisy data are used without careful handling, models may learn incorrect associations, reducing generalization and clinical reliability.

To systematically address label inconsistency and noise, AIDA employs an active human-in-the-loop (HYLOOP) framework to identify, refine, and iteratively purify structured labels. The goal is to build a high-quality, low-bias training set that strengthens model robustness. AIDA consists of three core components: (1) cross-modal disagreement detection, (2) confidence-aware expert-machine collaboration, and (3) an iterative purification and expansion loop.

In disagreement detection, AIDA compares expert annotations with AI-generated suggestions on two levels: *visual evidence* and *conceptual inference*. The AI model analyzes each image independently and generates region proposals with semantic descriptions. The system then matches AI and expert annotations bi-directionally using spatial overlap and semantic similarity. Three key disagreement types are identified: (1) mismatched annotations with spatial alignment but conflicting labels, (2) system-discovered entities absent in expert labels, and (3) human-only annotations the AI fails to detect.

To assess disagreement severity, AIDA trains a noise-aware classifier that estimates prediction correctness. A sample is marked as high-risk when the model shows high confidence while the expert reports low confidence, signalling annotation issues that merit review. High-risk conflicts are routed to an interactive refinement interface, where a prioritized queue guides experts to address the most critical cases first. Experts operate across two semantic layers: *visual spans*, referring to text segments linked to observable structures (e.g., "increased mitotic figures"); and *conceptual spans*, capturing higher-level reasoning without direct localization. Within this interface, experts can revise masks, adjust diagnostic logic, and update confidence scores.

These refined annotations are incorporated into an iterative loop combining label purification and active sample selection. In each cycle, the model identifies new conflicts in labeled data and selects high-value unlabeled samples based on uncertainty or representativeness. Updated annotations are merged to retrain the system, improving visual feature extractors, vision-language alignment, and diagnostic reasoning components. Through this dual-path strategy, AIDA promotes the co-evolution of data quality and model capability, establishing a trustworthy learning loop for clinical deployment.

4.3 Addressing Scalability of Expert Annotations

In DP, the scarcity of expert annotators remains a central bottleneck to acquiring large-scale, high-quality training data. To address this, the AIDA framework introduces an Optimized Expert-in-the-Loop Protocol, powered by its HYLOOP engine, designed to improve the efficiency and impact of expert annotations. Rather than merely determining which samples to label, AIDA jointly optimizes who should annotate each sample, combining task selection with annotator routing to enhance data quality and resource allocation.

For sample selection, AIDA employs a multi-dimensional active learning mechanism that prioritizes the most informative unlabeled instances by integrating model uncertainty, diversity relative to the labeled set, and expected influence on model performance. The weighting of these factors can be adjusted to reflect specific task priorities, ensuring that expert effort focuses on representative and challenging examples most beneficial for model improvement. Once high-value samples are identified, AIDA activates an annotator routing mechanism. Each expert is associated with a dynamic capability profile reflecting their competence across tasks, and sample-expert assignments are optimized so that valuable tasks

are routed to the most qualified experts, reducing resource consumption and maximizing learning.

Experts interact with AIDA through an annotation interface that supports two complementary workflows: required supplementation and proactive supplementation. Required supplementation involves system-initiated tasks triggered by uncertainty or semantic conflicts and routed to the appropriate expert. Proactive supplementation enables experts to contribute critical annotations voluntarily. In both workflows, media retrieval and exploration tools help experts compare images and annotations and detect potential issues. All annotations are incorporated into the same refinement process and used to update the model, creating a continuous and evolving data-model feedback loop.

4.4 Addressing Model Interpretability and Trust

The application of AI, particularly deep learning, in pathology is rapidly expanding, but its black-box nature poses critical challenges in clinical practice. Even with high predictive performance, models lacking interpretability are difficult for clinicians to verify or correct, limiting their adoption and conflicting with medical principles of responsibility and transparency. Therefore, building interpretable and auditable AI models is essential for widespread clinical deployment.

The AIDA framework addresses this by fundamentally restructuring model outputs. Rather than producing an opaque classification label, AIDA leverages large multimodal models [4] to generate structured diagnostic reports that decompose decisions into a chain of multimodal reasoning supported by visual evidence and conceptual inferences. By explicitly separating "what is seen," "what is known," and "what is inferred," the system makes its decision process transparent and reviewable.

First, AIDA outputs a structured two-layer textual report, dividing content into Visual Spans (VSs) and Conceptual Spans (CSs). VSs describe observable morphological features grounded in specific image regions, while CSs capture higher-level diagnostic reasoning. This layered explanation makes the decision process traceable for medical experts. Second, the system generates reports at both global and local levels. Each includes a global summary and detailed analyses of sub-regions. The model first segments the region of interest (ROI), produces fine-grained VSs and CSs for each part, and then synthesizes them into a global diagnosis. This bottom-up structure helps clinicians examine both overall conclusions and supporting evidence. Finally, to further enhance trust, AIDA attaches calibrated confidence scores to each report element. Every observation, inference, and label is associated with a numerical certainty measure, enabling clinicians to identify ambiguous predictions and prioritize human review during collaborative diagnosis.

4.5 Addressing Integration into Clinical Workflow

Seamlessly integrating AI tools into existing, highly optimized clinical workflows is a critical step for translating technical capabilities into real clinical value. Even

a powerful system may be rejected if it disrupts established routines or adds operational burden. Any AI system designed for clinical use must thus prioritize workflow integration from the outset. To address this challenge, our AIDA framework adopts a workflow-centric system architecture. Instead of introducing a new standalone platform that clinicians must adapt to, AIDA embeds AI capabilities as integrated services within familiar clinical environments through the collaboration of two specialized platforms:

AIDA Infrastructure. This is the interactive front-end for clinical users, providing unified access to annotation, review, and report visualization tools. It supports structured label refinement and the generation of explainable reports, forming the core workspace for expert interaction.

Exquisitor Platform. Serving as a high-performance backend engine, this platform enables scalable image retrieval and interactive exploration, supporting real-time access to large-scale historical image databases.

AIDA provides embedded, real-time decision support directly within the clinical workflow. When reviewing challenging or rare cases, a pathologist can highlight any ROI within the AIDA interface. This action immediately triggers a content-based image retrieval process via the Exquisitor platform, searching across institutional archives for visually similar cases with consensus diagnoses, and subsequently presenting the retrieved cases to the pathologist for comparison and review. Through modular design, platform-level coordination, and embedded diagnostic assistance, AIDA enables deep integration of AI capabilities into real-world clinical practice, enhancing both efficiency and trust.

5 Conclusions

We presented the AIDA platform for digital pathology, whose core idea is to setup its innovative HYLOOP human-AI collaboration engine to deeply integrate expert knowledge with advanced artificial intelligence, strengthening rather than replacing clinicians' expertise. AIDA directly addresses key challenges in digital pathology by establishing a continuous feedback loop of AI analysis and expert validation. This loop systematically reduces inter-observer variability and annotation noise, improving data quality and model performance. To build clinical trust, AIDA generates transparent diagnostic reports that link each conclusion to its underlying visual evidence, making the model's reasoning directly reviewable by clinicians. AIDA's core tools have already been validated at top-tier conferences such as CVPR, ICCV, and MMM. In summary, AIDA pioneers a new *multimodal* paradigm of efficient human–AI collaboration in medical diagnostics, with the potential to deliver trustworthy decision support, and ultimately to improve overall patient care with earlier interventions and more effective treatments.

References

1. Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.M., Wu, J., et al.: Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature **491**(7424), 399–405 (2012)
2. Corona, R., Mele, A., Amini, M., De Rosa, G., Coppola, G., Piccardi, P., Fucci, M., Pasquini, P., Faraggiana, T.: Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. Journal of clinical Oncology **14**(4), 1218–1223 (1996)
3. Dai, D., Zhang, Y., Yang, Q., Xu, L., Shen, X., Xia, S., Wang, G.: Pathologyvlm: A large vision–language model for pathology image understanding. Artificial Intelligence Review **58**, 186 (2025)
4. Deitke, M., et al.: Molmo and pixmo: Open weights and open data for state-of-the-art multimodal models. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (2025)
5. Eccher, A., L'Imperio, V., Pantanowitz, L., Cazzaniga, G., Del Carro, F., Marletta, S., Gambaro, G., Barreca, A., Becker, J.U., Gobbo, S., et al.: Galileo—an artificial intelligence tool for evaluating pre-implantation kidney biopsies. Journal of Nephrology pp. 1–7 (2024)
6. van Eekelen, L., Spronck, J., Looijen-Salamon, M., Vos, S., Munari, E., Girolami, I., Eccher, A., Acs, B., Boyaci, C., de Souza, G.S., et al.: Comparing deep learning and pathologist quantification of cell-level pd-l1 expression in non-small cell lung cancer whole-slide images. Scientific Reports **14**(1), 7136 (2024)
7. Egevad, L., Ahmad, A.S., Algaba, F., Berney, D.M., Boccon-Gibod, L., Compérat, E., Evans, A.J., Griffiths, D., Grobholz, R., Kristiansen, G., et al.: Standardization of gleason grading among 337 european pathologists. Histopathology **62**(2), 247–256 (2013)
8. Elmore, J.G., Longton, G.M., Carney, P.A., Geller, B.M., Onega, T., Tosteson, A.N., Nelson, H.D., Pepe, M.S., Allison, K.H., Schnitt, S.J., et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. Jama **313**(11), 1122–1132 (2015)
9. ExpertPath: Comprehensive anatomic and clinical pathology decision support. Elsevier (2025), https://www.elsevier.com/products/expertpath
10. Federico, G., Davide, T., Ziyue, L., Zanxi, R., Yiming, W., Cristani, M.: LOTS of Fashion! multi-conditioning for image generation via sketch-text pairing. In: Proc. Int'l Conf. on Computer Vision (2025)
11. Goh, H.W., Mueller, J.: Activelab: Active learning with re-labeling by multiple annotators. arXiv preprint arXiv:2301.11856 (2023)
12. Harðarson, Þ.H., Loftsson, H., Ólafsson, S.: Aligning language models for Icelandic legal text summarization. In: Johansson, R., Stymne, S. (eds.) Proc. Joint Nordic Conf. on Computational Linguistics and Baltic Conf. on Human Language Technologies (Mar 2025)
13. Jóhannsson, Ó.A., Arndal, B.H., Jónsson, E.Ö., Ólafsson, S., Loftsson, H.: Evaluating icelandic sentiment analysis models trained on translated data. In: Proc. Annual Meeting of the Special Interest Group on Under-resourced Languages@LREC-COLING (2024)
14. Joppi, C., Skenderi, G., Cristani, M.: Pop: Mining potential performance of new fashion products via webly cross-modal query expansion. In: Proc. European Conf. on Computer Vision (2022)

15. Kamath, A., Singh, M., LeCun, Y., Synnaeve, G., Misra, I., Carion, N.: Mdetr-modulated detection for end-to-end multi-modal understanding. In: Proc. IEEE/CVF Int'l Conf. on Computer Vision (2021)
16. Khan, O.S., Jónsson, B.Þ., Rudinac, S., Zahálka, J., Ragnarsdóttir, H., Þorleiksdóttir, Þ., Guðmundsson, G.Þ., Amsaleg, L., Worring, M.: Interactive learning for multimedia at large. In: Proc. European Conf. on IR Research (2020)
17. Kong, Q., Chen, J., Jiang, J., Ruan, Z., Kang, L.: Dual-branch fusion with style modulation for cross-domain few-shot semantic segmentation. In: Proc. ACM Int'l Conf. on Multimedia (2024)
18. Li, D., Wang, Z., Chen, Y., Jiang, R., Ding, W., Okumura, M.: A survey on deep active learning: Recent advances and new frontiers. IEEE Transactions on Neural Networks and Learning Systems **36**(4), 5879–5899 (2025)
19. Li, L.H., Yatskar, M., Yin, D., Hsieh, C.J., Chang, K.W.: Visualbert: A simple and performant baseline for vision and language. arXiv preprint arXiv:1908.03557 (2019)
20. Lilley, C.M., Arnold, C.A., Arnold, M.A., Booth, A.L., Gardner, J.M., Jiang, X., Loghavi, S., Mirza, K.M.: Virtual pathology elective, real education: the pathelective.com experience as a model for novel pathology pedagogy and a primer for curricular evolution. Archives of Pathology & Laboratory Medicine **148**(5), 595–602 (2024)
21. Lu, J., Batra, D., Parikh, D., Lee, S.: Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. Advances in neural information processing systems **32** (2019)
22. Mahbod, A., Polak, C., Feldmann, K., et al.: Nuinsseg: A fully annotated dataset for nuclei instance segmentation in h&e-stained histological images. Scientific Data **11**(1), 295 (2024)
23. McCoy, C.A., Coleman, H.G., McShane, C.M., McCluggage, W.G., Wylie, J., Quinn, D., McMenamin, Ú.C.: Factors associated with interobserver variation amongst pathologists in the diagnosis of endometrial hyperplasia: A systematic review. Plos one **19**(4), e0302252 (2024)
24. Nasirihaghighi, S., Ghamsarian, N., Husslein, H., Schoeffmann, K.: Event recognition in laparoscopic gynecology videos with hybrid transformers. In: Proc. MultiMedia Modeling (MMM) (2024)
25. Nechaev, D., Pchelnikov, A., Ivanova, E.: Histai: An open-source, large-scale whole slide image dataset for computational pathology. arXiv preprint arXiv:2505.12120 (2025)
26. Northcutt, C., Jiang, L., Chuang, I.: Confident learning: Estimating uncertainty in dataset labels. Journal of Artificial Intelligence Research **70**, 1373–1411 (2021)
27. Olafsson, S., Wallace, B.C., Bickmore, T.W.: Towards a computational framework for automating substance use counseling with virtual agents. In: Proc. Int'l Conf. on Autonomous Agents and Multiagent Systems (2020)
28. PathPresenter: Pathpresenter: Digital pathology - image management simplified (2025), `https://www.pathpresenter.com/`
29. Pigaiani, N., Oliva, A., Cirielli, V., Grassi, S., Arena, V., Solari, L.M., Tatriele, N., Raniero, D., Brunelli, M., Gobbo, S., et al.: iForensic, multicentric validation of digital whole slide images (wsi) in forensic histopathology setting according to the college of american pathologists guidelines. International Journal of Legal Medicine pp. 1–8 (2025)
30. Rizzo, P.C., Caputo, A., Maddalena, E., Caldonazzi, N., Girolami, I., Dei Tos, A.P., Scarpa, A., Sbaraglia, M., Brunelli, M., Gobbo, S., et al.: Digital pathology world tour. Digital Health **9**, 20552076231194551 (2023)

31. Robert, M.E., Rüschoff, J., Jasani, B., Graham, R.P., Badve, S.S., Rodriguez-Justo, M., Kodach, L.L., Srivastava, A., Wang, H.L., Tang, L.H., et al.: High interobserver variability among pathologists using combined positive score to evaluate pd-l1 expression in gastric, gastroesophageal junction, and esophageal adenocarcinoma. Modern Pathology **36**(5) (2023)
32. Schukow, C.P., McKee, P.H.: Knowledge in knowledge out: A next-generation platform intersecting social media with digital pathology. Archives of Pathology & Laboratory Medicine **147**(4) (2023)
33. Settles, B.: Active learning literature survey. Tech. rep. (2009), https://burrsettles.com/pub/settles.activelearning.pdf
34. Sharma, U., Khan, O.S., Rudinac, S., Jónsson, B.Þ.: Exquisitor at the Video Browser Showdown 2025: Unifying conversational search and user relevance feedback. In: Proc. Multimedia Modeling (2025)
35. Song, A.H., Jaume, G., Williamson, D.F., Lu, M.Y., Vaidya, A., Miller, T.R., Mahmood, F.: Artificial intelligence for digital and computational pathology. Nature Reviews Bioengineering **1**(12), 930–949 (2023)
36. Sun, Y., Zhang, Y., Si, Y., Zhu, C., Shui, Z., Zhang, K., Li, J., Lyu, X., Lin, T., Yang, L.: Pathgen-1.6 m: 1.6 million pathology image-text pairs generation through multi-agent collaboration. arXiv preprint arXiv:2407.00203 (2024)
37. Taioli, F., Giuliari, F., Wang, Y., Berra, R., Castellini, A., Del Bue, A., Farinelli, A., Cristani, M., Setti, F.: Unsupervised active visual search with monte carlo planning under uncertain detections. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
38. Taioli, F., Rosa, S., Castellini, A., Natale, L., Del Bue, A., Farinelli, A., Cristani, M., Wang, Y.: Mind the error! detection and localization of instruction errors in vision-and-language navigation. In: Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (2024)
39. Talon, D., Girella, F., Liu, Z., Cristani, M., Wang, Y.: Seeing the abstract: Translating the abstract language for vision language models. In: Proc. Computer Vision and Pattern Recognition Conf. (2025)
40. Van Den Einden, L.C., De Hullu, J.A., Massuger, L.F., Grefte, J.M., Bult, P., Wiersma, A., Van Engen-van Grunsven, A.C., Sturm, B., Bosch, S.L., Hollema, H., et al.: Interobserver variability and the effect of education in the histopathological diagnosis of differentiated vulvar intraepithelial neoplasia. Modern pathology **26**(6), 874–880 (2013)
41. Wei, J., Suriawinata, A., Ren, B., Liu, X., Lisovsky, M., Vaickus, L., Brown, C., Baker, M., Tomita, N., Torresani, L., et al.: A petri dish for histopathology image analysis. In: Proc. Int'l Conf. on Artificial Intelligence in Medicine (2021)
42. Xu, Z., Lu, D., Luo, J., Wang, Y., Yan, J., Ma, K., Zheng, Y., Tong, R.K.Y.: Anti-interference from noisy labels: Mean-teacher-assisted confident learning for medical image segmentation. IEEE Transactions on Medical Imaging **41**(11), 3062–3073 (2022)
43. Zhang, Y., Li, B., Ling, Z., Zhou, F.: Mitigating label bias in machine learning: Fairness through confident learning. In: Proc. AAAI Conf. on Artificial Intelligence (2024)