Vision-Language Models for Biomedical Applications

Surendrabikram Thapa
surendrabikram@vt.edu
Virginia Tech
Blacksburg, Virginia, USA

Usman Naseem
usman.naseem@mq.edu.au
Macquarie University
Sydney, New South Wales, Australia

Luping Zhou
luping.zhou@sydney.edu.au
The University of Sydney
Sydney, New South Wales, Australia

Jinman Kim
jinman.kim@sydney.edu.au
The University of Sydney
Sydney, New South Wales, Australia

Abstract

Vision-language models (VLMs) are transforming the landscape of biomedical research and healthcare by enabling the seamless integration and interpretation of complex multimodal data, including medical images and clinical texts. Recognizing the growing impact of these models, the first international workshop on Vision-Language Models for Biomedicine (VLM4Bio) was held in conjunction with ACM Multimedia 2024. The workshop aimed to address the critical need for advanced techniques that can leverage VLMs in applications such as medical imaging, diagnostics, and personalized treatment. As healthcare data increasingly involves both visual and textual information, VLM4Bio provided a platform for interdisciplinary collaboration between experts in natural language processing, computer vision, biomedical engineering, and AI ethics. This paper provides an overview of the inaugural edition of the VLM4Bio workshop, summarizing the key discussions, contributions, and future directions for expanding the workshop's scope and influence in subsequent editions.

CCS Concepts

• **Information systems** → **Multimedia information systems**; *Decision support systems*; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Applied computing** → **Bioinformatics**; **Health informatics**.

Keywords

Vision-Language Models (VLMs), Multimodal Biomedical AI, Visual Question Answering (VQA), Clinical Decision Support Systems, Healthcare Applications

ACM Reference Format:
Surendrabikram Thapa, Usman Naseem, Luping Zhou, and Jinman Kim. 2024. Vision-Language Models for Biomedical Applications. In *Proceedings of the First International Workshop on Vision-Language Models for Biomedical Applications (VLM4Bio '24), October 28-November 1, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3689096.3690770

VLM4Bio '24, October 28-November 1, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-1207-4/24/10
https://doi.org/10.1145/3689096.3690770

1 Introduction

Vision-Language Models (VLMs) are becoming increasingly important in biomedical research and healthcare [4]. These models can process and understand both visual data, like medical images, and also textual data, such as clinical reports depicting the impressions from interpreting the images . In biomedical applications, where medical images and clinical texts often need to be interpreted together, these models offer new opportunities for enhancing diagnostic accuracy, treatment planning, and personalized medicine [1, 7]. VLMs has been demonstrated to address challenges in tasks like medical image captioning, visual question answering (VQA), and integrating multimodal biomedical data.

As healthcare data is inherently multimodal, the role of VLMs in extracting meaningful insights from diverse sources of information is growing rapidly [5, 6]. From analyzing radiology reports paired with imaging scans to supporting automated diagnostics, VLMs bridge the gap between different data types, leading to more comprehensive decision support systems, improved patient outcomes, and new applications in drug discovery, pathology, and beyond [2, 3].

Recognizing this emerging trend, the first Vision-Language Models for Biomedical Applications (VLM4Bio) workshop was held at ACM Multimedia 2024. The workshop brought together researchers from natural language processing, computer vision, biomedical engineering, and healthcare to discuss the latest developments, challenges, and future directions for VLMs in biomedical research and practice. This paper provides an overview of the workshop's goals, contributions, and key discussions while outlining potential areas for growth in future editions.

2 Call for Papers

We invited original contributions on a wide range of topics to explore the latest developments, challenges, and opportunities in vision-language models in biomedical applications. The workshop invited submissions on topics including, but not limited to:

- VLM4Bio: Biomedical image understanding and captioning using VLMs.
- VLM4Bio: Visual Question Answering (VQA) in biomedical applications.
- VLM4Bio: Integration of multimodal biomedical data for enhanced decision support systems.
- VLM4Bio: Applications of VLMs in medical imaging, pathology, radiology, and histology.

- VLM4Bio: Approaches based on VLM for drug discovery, pharmacogenomics, and personalized medicine.
- VLM4Bio: VLM-based approaches for biological imaging including cells and transcriptomics.
- VLM4Bio: Clinical applications of VLMs in disease diagnosis, prognosis, and treatment planning.
- VLM4Bio: Benchmark datasets, evaluation metrics, and reproducibility challenges in VLM research for biomedicine.
- VLM4Bio: Development, scalibility, and optimization of VLM architectures for biomedical data analysis.
- VLM4Bio: Case studies, real-world applications, and considerations for the deployment of VLMs in healthcare settings.
- VLM4Bio: Ethical considerations, bias mitigation, and interpretability in VLMs for biomedical applications.

3 Workshop Overview

The workshop received 10 submissions, of which 6 were accepted, leading to an acceptance rate of 60%. Each paper was evaluated by at least three members of the program committee. The inaugural edition of the workshop attracted global attention, with submissions coming from a diverse set of countries, including Australia, China, India, Mexico, Nepal, Norway, Saudi Arabia, the United Kingdom, and the United States. The workshop program featured poster sessions, oral presentations, and keynote speeches.

4 Program Committee

Our workshop was supported by various program committee members across the globe. The program committee is listed below (in no specific order):

- Surabhi Adhikari (Columbia University, USA)
- Lakshmi Nair (Lightmater Inc., USA)
- Sandesh Jain (Virginia Tech, USA)
- Mira Moukheiber (Massachusetts Institute of Technology, USA)
- Farhan Ahmad Jafri (Samsung Research, India)
- Shuvam Shiwakoti (Delhi Technological University, India)
- Kritesh Rauniyar (Delhi Technological University, India)
- Siddhant Bikram Shah (Northeastern University, USA)
- Sushant Gautam (Simula Metropolitan Center for Digital Engineering, Oslo)
- Shah Nawaz (Johannes Kepler Universität Linz, Austria)
- Zehua Cheng (University of Oxford, UK)
- Wei Dai (Chinese Academy of Sciences, China)
- Yongpei Ma (University of Sydney, Australia)
- MSVPJ Sathvik (Raickers AI, India)
- H M Dipu Kabir (Charles Sturt University, Australia)
- Shuchang Ye (University of Sydney, Australia)
- Mingyuan Meng (University of Sydney, Australia)
- Mohammad Ali Moni (Charles Sturt University, Australia)

5 Future Workshops

In future iterations, the VLM4Bio workshop aims to broaden its influence in the field of understanding the intersection of vision-language models and biomedical applications. Building on the success of the inaugural edition, the organizers plan to introduce more interactive elements, such as expert panels, hands-on tutorials, and focused roundtable discussions, to foster deeper engagement and knowledge sharing among participants. These sessions will emphasize emerging challenges and opportunities in biomedical AI, including advancing the development of scalable and interpretable vision-language models, addressing ethical concerns and bias in biomedical AI, and exploring novel applications of multimodal data integration in healthcare and life sciences. Additionally, as large vision-language models (LVLMs) continue to show significant promise in enhancing medical imaging, diagnostics, and personalized treatment, future workshops will focus on their evolving roles and applications in biomedical research and clinical practice.

Acknowledgments

We would like to express our sincere gratitude to all those who contributed to the successful organization of the first Vision-Language Models for Biomedical Applications (VLM4Bio) workshop. We are especially grateful to the program committee members for their valuable time and effort in reviewing the submissions and ensuring a high-quality program. We extend our thanks to the keynote speakers for sharing their expertise and insights, which enriched the discussions at the workshop.

We also wish to acknowledge the support provided by ACM Multimedia 2024 for hosting the workshop, and the organizers who worked tirelessly to bring this event to fruition. Lastly, we appreciate the contributions of the participants and presenters who shared their research and ideas, making this workshop a fruitful exchange of knowledge.

References

[1] Yakoub Bazi, Mohamad Mahmoud Al Rahhal, Laila Bashmal, and Mansour Zuair. 2023. Vision–language model for visual question answering in medical imagery. *Bioengineering* 10, 3 (2023), 380.

[2] Jean-benoit Delbrouck, Khaled Saab, Maya Varma, Sabri Eyuboglu, Pierre Chambon, Jared Dunnmon, Juan Zambrano, Akshay Chaudhari, and Curtis Langlotz. 2022. ViLMedic: a framework for research at the intersection of vision and language in medical AI. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 23–34.

[3] Qi Li. 2023. Harnessing the power of pre-trained vision-language models for efficient medical report generation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 1308–1317.

[4] Usman Naseem, Surendrabikram Thapa, Anum Masood, et al. 2024. Advancing Accuracy in Multimodal Medical Tasks Through Bootstrapped Language-Image Pretraining (BioMedBLIP): Performance Evaluation Study. *JMIR Medical Informatics* 12, 1 (2024), e56627.

[5] Ziyuan Qin, Hua Hui Yi, Qicheng Lao, and Kang Li. 2023. Medical image understanding with pretrained vision language models: A comprehensive study. In *The Eleventh International Conference on Learning Representations*.

[6] Corentin Royer, Anjany Sekuboyina, et al. [n. d.]. MultiMedEval: A Benchmark and a Toolkit for Evaluating Medical Vision-Language Models. In *Medical Imaging with Deep Learning*.

[7] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, et al. 2024. Multimodal healthcare AI: identifying and designing clinically relevant vision-language applications for radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.