

ACS 2024 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

O-01

Research Abstracts

Recent Advances of Transformers in Endoscopic Video Analysis: A Review

Abhisri Ramesh; Abdulla Ahmed; Myra Zaheer; Nishanth Gowda; Arhum Naeem; and Daniel Donoho.

George Washington University School of Medicine and Health Sciences, Washington DC, DC; Children's National Hospital, Washington DC, DC

Methods: We indexed articles from the literature of the US National Library of Medicine (PubMed/MEDLINE), Taylor & Francis Online, SpringerLink, and Institute of Electrical and Electronics Engineers from January 2000 to May 2023. We selected studies that developed a novel Transformer architecture for endoscopic video analysis and excluded studies that utilized Transformer architecture for medical image analysis. The risk of bias was evaluated with the Newcastle Ottawa Quality assessment tool. Data from the studies were visually presented in table using the SPIDER tool.

Results: Of the 426 articles identified, 12 were selected for review. The data was extracted by two reviewers based on the surgical analysis category. The published works were sub-categorized into four key areas: (1) surgical workflow recognition (50.0%), (2) surgical gesture recognition (16.7%), (3) surgical instrument segmentation (16.7%), and (4) surgical skill assessment (8.3%). Transformer architecture consistently outperformed state-of-the-art neural networks in endoscopic workflow recognition, gesture recognition, instrument segmentation and skill assessment.

Conclusions: This research serves as a novel roadmap that showcases the unique strengths and promising potential of Transformers in the field of endoscopic video analysis. Key contributions include that Transformers capture long-range dependencies, overcome the limitations of sequential processing, and excel in information fusion and feature extraction. The inherent sequence modeling, attention mechanism, multimodal integration, and transfer learning capabilities of Transformer architecture make it well-suited for patient outcomes prediction, surgical training, and clinical decision support systems in the near future.