Establishing research priorities for AI implementations in Upper Aerodigestive Tract Endoscopy: a modified Delphi study

Yeo Eun Ashley Kim, BA[1], Christopher Holsinger, MD[2], Alberto Paderno, MD PhD[3], Anita Rau, PhD[2], Michael Chang, MD[2], Matthew Crowson, MD[4], James Curtis, PhD[1], Omer Ahmad, MBBS[5], Dong Chul Cha, MD[6], Daniel Donoho, MD[7], Michael Dunham, MD[8], Necati Enver, MD[9], Al-Rahim Habib, MD[10], Aaron Johnson, PhD[11], Eva Kiaer, MD[12], Jinrang Li,MD[13], Andrés Bur, MD[14], Matthew Naunheim, MD[4], Xiao-Guang Ni,MD[15], Rita Patel, PhD[16], Mette Pedersen, MD PhD[17], Rakesh Srivastava, MS[18], Andrew Thamboo, MD[19], Anaïs Rameau, MD[1]

[1]Sean Parker Institute for the Voice, Weill Cornell Medical College/New York Presbyterian Hospital, [2]Stanford University, [3]Humanitas University, [4]Massachusetts Eye and Ear, 5 University College London, [6]NAVER Corp., Seoul, South Korea,[7]Children's National Hospital, [8]Louisiana State University Health Sciences Center, [9]Marmara University Pendik Research and Training Hospital, Istanbul, Turkey, [10]Gold Coast University Hospital, Westmead, Australia, [11]New York University Grossman School of Medicine, [12]Copenhagen University Hospital, Copenhagen, Denmark, [13]PLA General Hospital, Beijing, China, [14]University of Kansas, [15]Cancer Hospital of Chinese Academy of Medical Sciences, Beijing, China, [16]Indiana University, [17]The Medical Center, Copenhagen, Denmark, [18]SIPS Super-specialty Hospital, Lucknow, India, [19]University of British Columbia

Introduction

- AI applications in videoendoscopy for the upper aerodigestive tract (UAT)are emerging in research but have yet to become part of routine clinical practice in otolaryngology.
- While AI-based systems in gastrointestinal endoscopy, such as colonoscopy, have received regulatory approval, the integration of AI in UAT endoscopy remains in its early stages[1].
- This study introduces an international expert consensus statement (ECS) that outlines research priorities focused on addressing key challenges as AI moves closer to clinical application.

Methods

Study Design
A modified Delphi process was utilized to establish research priorities for AI implementations in upper aerodigestive tract endoscopy.[2]

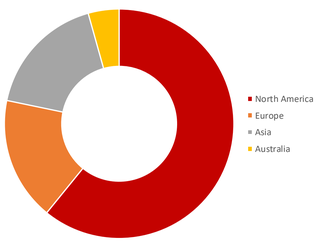
Steering Committee and Expert Participants
A steering committee of otolaryngologists and computer scientists with expertise in AI and endoscopy guided the study. The committee invited 19 international experts, selected based on their involvement in AI research and their publication history. Geographic diversity was ensured, with participants from North America, Europe, Australia, and Asia (Figure 1).

Stage 1: Literature Search and Question Generation
A preliminary list of 58 topic questions was generated, addressing critical research areas in AI integration for upper aerodigestive tract endoscopy.

Stage 2: Delphi Survey 1 Development and Rating
Delphi Survey 1 was developed using Qualtrics XM. Each statement was rated on a 9-point Likert scale, with ratings ranging from "strongly disagree" (1) to "strongly agree" (9). Categorization of consensus statements are as described in Table 1.

Stage 3: Final Consensus and Prioritization
In Delphi Survey 2, near-consensus statements were revised and distributed to the expert group for re-evaluation. The final top 11 research priorities were identified based on mean scores and expert discussion.

A



B

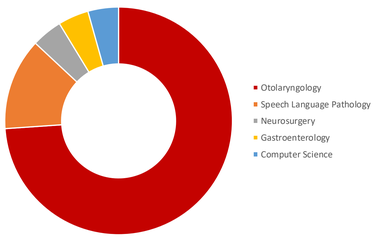


Figure 1A. Geographical distribution of experts; **B.** Specialty distribution of experts

Table 1. Categorization of consensus statements.

Category	Mean Score[a]		Outliers[b]
Consensus	≥7.00	and	≤1
Near consensus	≥6.50	and	≤2
No consensus	<6.50	or	≥3

[a] Nine-point Likert scale from *disagree strongly* (1) to *agree strongly* (9).
[b] Outliers are defined as any rating ≥2 Likert points from the mean (in either direction)

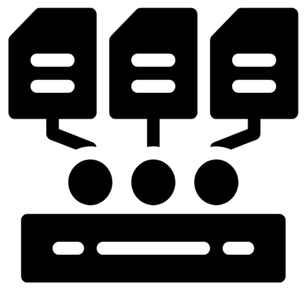
Contact

Ashley Yeo Eun Kim
yek4001@med.cornell.edu

Top Research Priorities

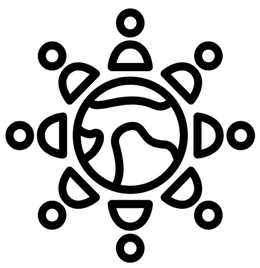
All participants completed each of the three rounds of the study, achieving a 100% response rate. The final top 11 consensus statements fell under six themes, shown below. Consensus statements and their mean scores are shown in Table 2.

Variability Reduction & Verification Accuracy

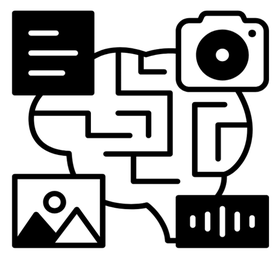


Data Collection & Annotation

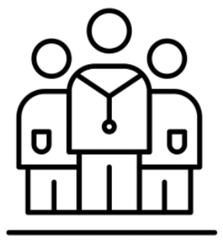
Clinical Use Cases & Applications



Equity & Accessibility



Multimodal Data Integration



Training & Education

Consensus Statements

Reduction in Variability and Accuracy in Verification
Videomics may reduce intra- and inter-rater variability in the assessment of UAT endoscopy.
Videomics in UAT endoscopy may fail in cases of expert disagreement regarding the ground truth annotation.
Algorithm verification and transparency can be achieved by the high quality, quantity, and diversity of training data. It involves rigorous testing, validation, and ongoing monitoring. Data and code sharing are crucial for this purpose.
Data Collection and Annotation
To develop high-quality, standardized, real-world clinical datasets for computer vision AI models, datasets for training and validation would ideally be sourced from multiple clinicians and multiple institutions, and be representative of a diverse patient population.
The challenge of limited data availability can be addressed by promoting collaborative sharing of data and AI models among healthcare entities, leveraging synthetic data generation, and employing federated learning to train models without compromising data privacy.
In applications of videomics, the level of certainty around ground truth and details about the annotation process are important to specify.
Clinical Use Cases and Applications
Videomics in UAT endoscopy may be applied in clinical decision support of diagnosis, objective measurements, and co-piloting during procedures.
One of the aims of videomics is to help clinicians improve patient safety and augment the value of their interventions.
Equity and Accessibility
Geographic and sociodemographic diversity in data collection and model evaluation and deployment are important to minimize disparities in videomics benefits.
Multimodal Data Integration
Incorporating multimodal data may enhance videomics in UAT endoscopy. However, further investigation is needed to evaluate if such integration will improve model performance.
Training and Education
One of the goals of videomics in UAT endoscopy is to aid training clinicians in diagnostic and procedural interventions.

Conclusion

This is the first international expert consensus on implementation priority setting in AI research on UAT endoscopy. These statements can be used to guide future research and catalyze the implementation of AI in UAT endoscopy.

References

1. Kudo S, Mori Y, Misawa M et al. Artificial intelligence and colonoscopy: Current status and future perspectives. Dig Endosc. 2019;31:363–371
2. Rosenfeld RM, Nnacheta LC, Corrigan MD. Clinical Consensus Statement Development Manual. Otolaryngol Head Neck Surg. 2015 Nov;153(2 Suppl):S1-S14.