

Nutzung von künstlicher Intelligenz (KI) in der endoskopischen Schluckdiagnostik. Erste Ergebnisse zur Genauigkeit der KI-basierten Aspirations-Detektionsleistung

Jürgen Konradi^{1a}, Milla Zajber^{2a}, Svenja Stegner¹, Carina Czysch¹, Sabine Corsten², Ulrich Betz¹, Constantin Disch³, Annika Hänsch³, Hans Meine³

¹Institut für Physikalische Therapie, Prävention und Rehabilitation, Universitätsmedizin der Johannes Gutenberg-Universität Mainz. juergen.konradi@unimedizin-mainz.de

²Katholische Hochschule Mainz, Fachbereich Gesundheit & Pflege

³Fraunhofer-Institut für Digitale Medizin MEVIS, Bremen

^a gleichwertige Beteiligung, geteilte Erstautorenschaft

Zitation:

Konradi, J., Zajber, M., Stegner, S., Czysch, C., Corsten, S., Betz, U., Disch, C., Hänsch, A. & Meine, H. (2022). Nutzung von künstlicher Intelligenz (KI) in der endoskopischen Schluckdiagnostik. Erste Ergebnisse zur Genauigkeit der KI-basierten Aspirations-Detektionsleistung. Sprachtherapie aktuell: Forschung - Wissen - Transfer 9(1): XXXIV. Workshop Klinische Linguistik. e2022-02

Thema und Ziel. Schluckstörungen (Dysphagien) führen aufgrund von Aspirationen zu einer erhöhten Pneumoniegefahr und gesteigerten Mortalität, wenn keine Dysphagiediagnostik und -therapie erfolgt (Steele et al, 2019). Die Flexible Endoskopische Evaluation des Schluckens (FEES) gilt dabei als Goldstandard. Die Bildaufnahmen müssen gespeichert und für den Diagnosebericht erneut geprüft werden, um eine höhere Zuverlässigkeit bei der Erkennung von Aspiration zu erreichen (Echtzeit-Krippendorffs Alpha $\sim .775$ vs. zweite Videoinspektion Bild für Bild $\sim .865$) (Hey et al, 2015). Neben binären Kontrasten (Aspiration / keine Aspiration) können auch detailliertere Skalen wie die Penetrations-Aspirations-Skala (PAS) zur Beschreibung der Ergebnisse verwendet werden. In diesen Fällen wird die allgemeine Inter-Rater-Reliabilität (IRR) zwischen Kliniker:innen unabhängig von der spezifischen Erfahrung mit $.35$ (PAS Score 5; Colodny, 2002) und $.85$ (Butler et al, 2015) angegeben. Je differenzierter die Diagnostik sein soll, desto ungenauer werden die menschlichen Entscheidungen. Der Einsatz einer künstlichen Intelligenz (KI) könnte hier hilfreich sein. Die wichtigste Voraussetzung wäre, dass eine KI in der Lage ist, Aspiration zu erkennen. Während die Erkennung von Aspiration (Steele et al, 2019) auf der Grundlage von Signalen bereits mit Hilfe künstlicher neuronaler Netze untersucht wurde, ist dies bei der Erkennung von Aspiration anhand von Videos/Bildern nicht der Fall. Ersteres führte im besten Fall zu einer noch unbefriedigenden Genauigkeit von $81,5\%$ (Steele et al, 2019). Der KI-Einsatz bei Bildmaterial verspricht weitaus mehr, was sich in benachbarten Anwendungen (gastrointestinalen Endoskopie) bereits eine hohe Sensitivität und Genauigkeit gezeigt hat, selbst im Vergleich zu erfahrenen menschlichen Endoskopiker:innen (Alagappan, 2018). Neuere Deep-Learning-Systeme zeigen selbst bei der Videoverarbeitung in Echtzeit eine überlegene Latenzzeit von weniger als $0,3$ Sekunden und reduzierte Fehlraten für Polypen/Adenome (Alagappan, 2018), die normalerweise zwischen $6-22\%$ (Ahn et al, 2012) liegen und damit vergleichbar sind mit den Raten von Fehleinschätzungen, die zu einer reduzierten IRR bei FEES führen (Hey et al, 2015; Colodny, 2002; Butler et al, 2015). Diese vielversprechenden Ergebnisse lassen die Idee aufkommen, ein ähnliches Instrument für die FEES zu entwickeln.

Methodik. Proof-of-Concept-Studie anhand einer retrospektiven Datenanalyse von FEES-Videos (Ethikvotum der LÄK RLP: 2021-16141-retrospektiv). Durch die Annotation (by color und/oder sparse labeling) von 100 FEES-Videos (50 mit und 50 ohne Aspiration) soll eine KI antrainiert werden, Aspirationen zu erkennen (Annotationstool SATORI von Fraunhofer MEVIS). Validierungsdaten sollen genutzt werden, um die Modelleleistung während der des Trainings zu bewerten (Erkennung von Overfitting, Early Stopping).

Ergebnisse und Ausblick. Die Genauigkeit der Aspirations-Detektionsleistung der KI soll im Nachgang auf zuvor separierten Testdaten mittels der Metriken Precision und Recall bestimmt werden. Gemachte Erfahrungen und Empfehlungen für weiterführende Forschung werden dargestellt.

Alagappan, M., Brown, J., Mori, Y., & Berzin, T. M. (2018). Artificial intelligence in gastrointestinal endoscopy: The future is almost here. *World journal of gastrointestinal endoscopy*, 10(10), 239–249. DOI: 10.4253/wjge.v10.i10.239.

Ahn, S. B., Han, D. S., Bae, J. H., Byun, T. J., Kim, J. P., & Eun, C. S. (2012). The Miss Rate for Colorectal Adenoma Determined by Quality-Adjusted, Back-to-Back Colonoscopies. *Gut and liver*, 6(1), 64–70. DOI: 10.5009/gnl.2012.6.1.64

Butler, S. G., Markley, L., Sanders, B., & Stuart, A. (2015). Reliability of the penetration aspiration scale with flexible endoscopic evaluation of swallowing. *The Annals of otology, rhinology, and laryngology*, 124(6), 480–483. DOI: 10.1177/0003489414566267

Colodny N. (2002). Interjudge and intrajudge reliabilities in fiberoptic endoscopic evaluation of swallowing (fees) using the penetration-aspiration scale: a replication study. *Dysphagia*, 17(4), 308–315. DOI: 10.1007/s00455-002-0073-4

Hey, C., Pluschinski, P., Pajunk, R., Almahameed, A., Girth, L., Sader, R., Stöver, T., & Zaretsky, Y. (2015). Penetration-Aspiration: Is Their Detection in FEES® Reliable Without Video Recording?. *Dysphagia*, 30(4), 418–422. DOI: 10.1007/s00455-015-9616-33.

Steele, C. M., Mukherjee, R., Kortelainen, J. M., Pölönen, H., Jedwab, M., Brady, S. L., Theimer, K. B., Langmore, S., Riquelme, L. F., Swigert, N. B., Bath, P. M., Goldstein, L. B., Hughes, R. L., Leifer, D., Lees, K. R., Meretoja, A., & Muehleman, N. (2019). Development of a Non-invasive Device for Swallow Screening in Patients at Risk of Oropharyngeal Dysphagia: Results from a Prospective Exploratory Study. *Dysphagia*, 34(5), 698–707. DOI: 10.1007/s00455-018-09974-5