European Society for Swallowing Disorders
ESSD
November 30, 2023, Toulouse, France

JG|U UNIVERSITÄTSmedizin. MAINZ
Interprofessional Study Center of Motion Research

AI-based aspiration detection in Flexible Endoscopic Evaluation of Swallowing: An explainable AI approach

Jürgen Konradi[1], Milla Zajber[1],Ulrich Betz[1], Philipp Drees[2], Annika Gerken[3], Hans Meine[3]
[1]Institute for Physical Therapy, Prevention and Rehabilitation, University Medical Center Mainz, Germany, [2]Department of Orthopaedics and Traumatology, University Medical Center Mainz, Germany, [3]Fraunhofer Institute for Digital Medicine MEVIS, Germany

Background and Aim

Flexible Endoscopic Evaluation of Swallowing (FEES) is considered a gold standard in dysphagia diagnosis, but is prone to human error. This is demonstrated by an improvable interrater reliability (~.8) for aspiration detection [1]. An AI-based tool could support this. Apart from one further approach [2], only non-endoscopic machine learning approaches exist, which mostly have black box properties, because they do not explain their model output (identification of an aspiration) transparently. This makes it difficult to trust such model decisions. Our goal was to introduce an explainable artificial intelligence (XAI) approach to detect aspiration in FEES [3].

Method

Proof-of-Concept study based on a retrospective data analysis of FEES videos. By annotating a total of 1330 image frames (2895 were labeled as not showing the glottis) in 92 videos (50 with aspirations) (see Table 1), an AI (CNN, U-Net) was trained to segment relevant anatomical structures (vocal cords and glottis) as well as to detect and visually highlight aspirations (see Fig. 1).

Table 1. Distribution of annotated frames across the different data sets.

AI data subset	Segmented frames	Frames with aspiration	Frames not showing glottis
training	1029	424	2220
validation	103	17	186
test	199	63	489

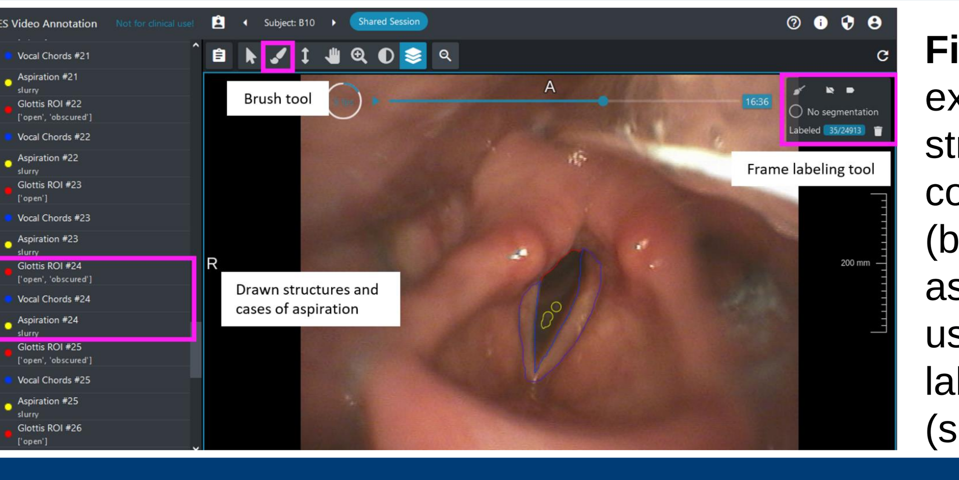

Figure 1: Annotation example with drawn structures of vocal cords (red), glottis (blue) and cases of aspiration (yellow) using the frame-labeling tool (screenshot).

Results

During the training, the curves of the learning progress for glottis and vocal cord segmentation (Jaccard scores 1 and 2) rise steeply from the beginning, unlike the aspiration detection task (Jaccard score 4), where the AI does not learn until about 19,000 iterations. After the rise of Jaccard score 4 (aspiration detection) only the training loss curve but not the validation loss curve progresses to decline. The best model performance (highest mean Jaccard score) is reached at 32,000 iterations (see Fig. 2), building the basis for the test run. During the tests, the AI successfully detected the glottis and vocal cords, but was not yet able to reach an expert level in the detection of aspirations (F1 score = .632) (see Fig. 3 and Table 2). Selected video frames in Figure 4 demonstrate this heterogeneity of results. As a means for post-hoc interpretation of the model outcome by the examiner, we implemented a concept of identifying meaningful frames in sequences. Therefore, an automated video analysis applies the CNN to an entire video to create a new video in which all AI-based segmentations and detections of aspirations are drawn into all frames of the video sequence (Fig. 5), serving as a first visual aid for key frames.

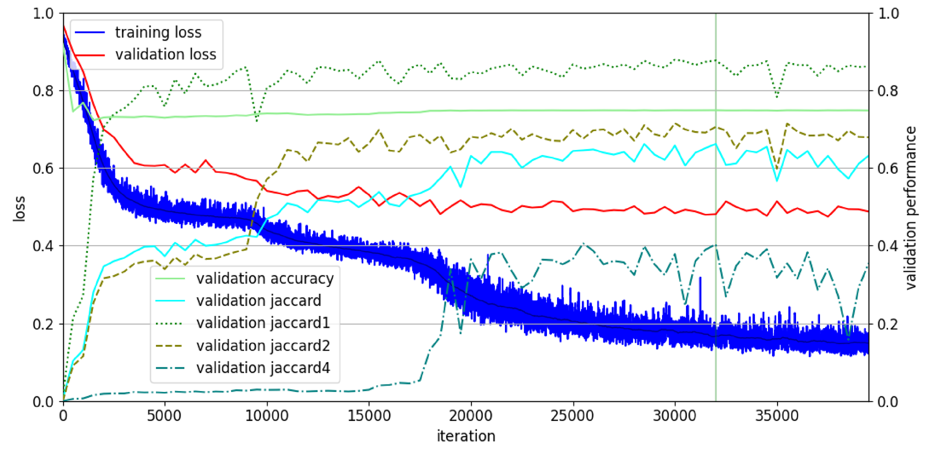

Figure 2: Loss curves of training (blue) and validation (red) as well as validation Jaccard scores (turquoise = mean of all; 1/dotted = segmentation of glottis; 2/dashed = segmentation of vocal cords; 4/dashed and dotted = detection of aspiration) show overlaps with the references. The vertical line shows the moment of the optimally working model.

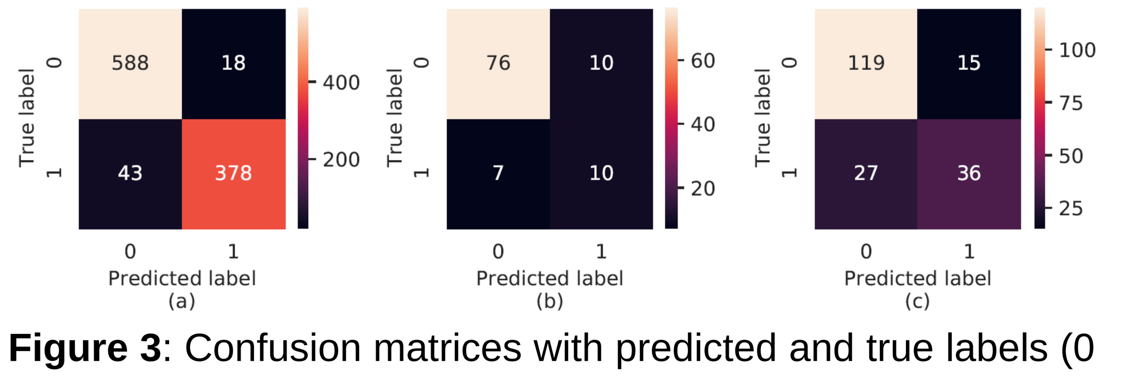

Figure 3: Confusion matrices with predicted and true labels (0 = negative / 1 = positive) for aspiration detection for training (a), validation (b), and testing (c) with heat-map scales for result interpretation (right in each case).

Table 2. Metrics for aspiration detection for all data sets.

Metrics	Training	Validation	Test
Precision	.955	.500	.706
Recall	.898	.588	.571
F1-Score	.925	.541	.632

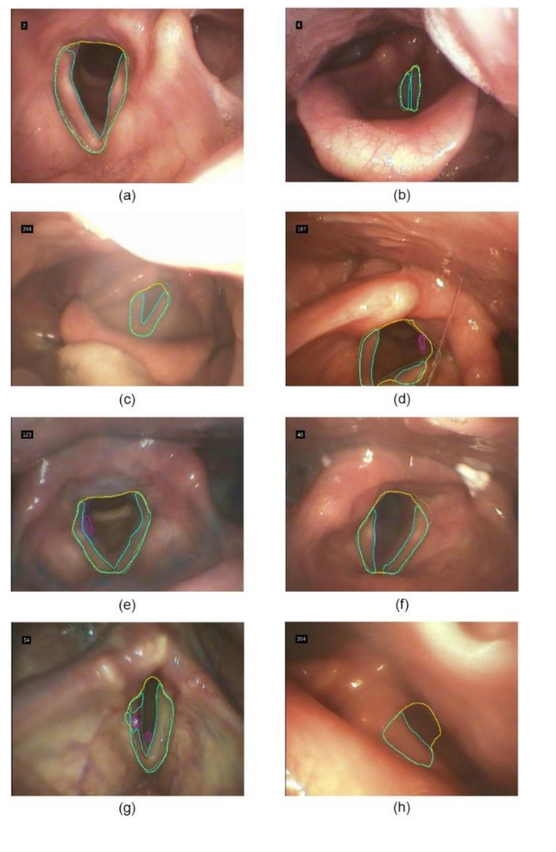

Figure 4: Examples of segmentation results in the test set across different videos: (a-c) high overlap between references (dotted) and AI-based (drawn through) segmentation in different states (open, closed) and light conditions, (d) segmentation errors of partially visible glottis close to the image edge, (e) correct detection of aspiration, (f) missed detection of aspiration, (g) false positive detection of aspiration, (h) false positive segmentation of glottis and vocal cords on frame without visible glottis.

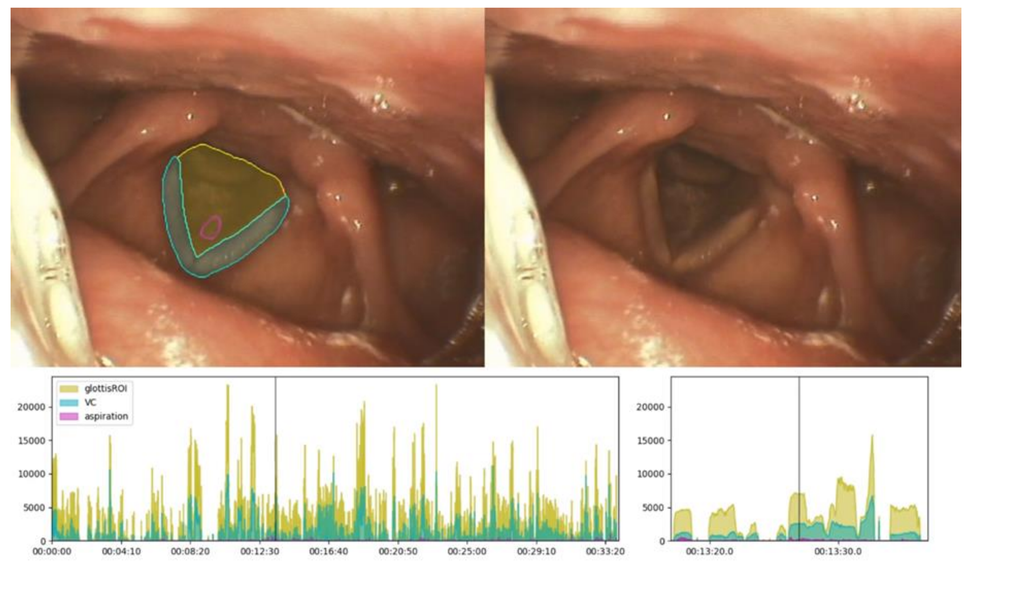

Figure 5: Visual aids to find meaningful frames for interpretation of model output. Screenshot of AI-based segmentation and detection of aspiration results (upper left) respective normal view (upper right). Timeline and timeline zoom with a curve displaying the number of pixels for the segmentation tasks and the detected aspiration (below). Vertical line indicates the point in time.

Contact

juergen.konradi@unimedizin-mainz.de
SCAN ME

Discussion and Conclusions

For the first time we have introduced an XAI that has been trained to detect aspiration in endoscopic swallowing videos. Although the recognition performance still needs to be optimized, our architecture leads to a final model that explains its evaluation by finding meaningful frames with relevant aspiration events and by highlighting the presumed bolus, thus making it applicable for clinical use. Further research will be necessary to enhance the model performance and to develop a real time tool for bedside use. After implementation of this tool in a FEES software, it will aid endoscopists to improve accuracy (thereby potentially saving lives), shorten the duration of the administration, and altogether safe costs as positive contributions for healthcare.

References

1. Curtis JA, Borders JC, Perry SE, Dakin AE, Seikaly ZN, Troche MS. Visual Analysis of Swallowing Efficiency and Safety (VASES): A Standardized Approach to Rating Pharyngeal Residue, Penetration, and Aspiration During FEES. Dysphagia. 2022;37(2):417-35.
2. Weng, W., et al., Expert-level aspiration and penetration detection during flexible endoscopic evaluation of swallowing with artificial intelligence-assisted diagnosis. Sci Rep, 2022. 12(1): 21689.
3. Konradi, J., et al., AI-Based Detection of Aspiration for Video-Endoscopy with Visual Aids in Meaningful Frames to Interpret the Model Outcome. Sensors, 2022. 22(23): 9468.