Automatic extraction of swallows from patients with dysphagia using non-invasive neck sensors and deep learning

Yassin Khalifa[1], Cara Donohue[2], James L. Coyle[2] and Ervin Sejdic´[1]
[1]Electrical and Computer Engineering, University of Pittsburgh
[2]Department of Communication Science and Disorders, University of Pittsburgh

Automatic extraction of swallows from patients with dysphagia using non-invasive neck sensors and deep learning

Yassin Khalifa[1], Cara Donohue[2], James L. Coyle[2] and Ervin Sejdic´[1]
[1]Electrical and Computer Engineering, University of Pittsburgh
[2]Department of Communication Science and Disorders, University of Pittsburgh

Problem

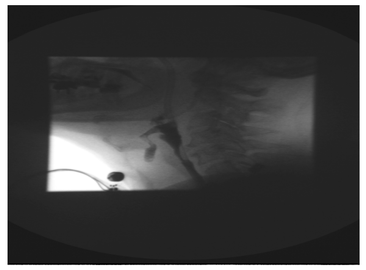
- Swallowing dysfunction (dysphagia) can occur due to a variety of underlying pathophysiologies and can result in impairments in swallowing safety (i.e., food and liquids entering the airway) and efficiency (pharyngeal residue).
- Dysphagia can lead to multiple negative consequences including aspiration pneumonia, malnutrition/dehydration, increased health care costs, and decreased quality of life.
- Several invasive imaging methods including videofluoroscopic swallow studies (VFSSs), fast pharyngeal CT/MRI, and fiberoptic endoscopic evaluation of swallowing can generate images and signals that can be used to detect the presence and the underlying pathophysiology of dysphagia; however, they are invasive, expensive, and are not feasible or easily accessible in all clinical settings or with all patients.
- High-resolution cervical auscultation (HRCA) is a sensor-based technology that uses advanced signal processing and machine learning techniques to characterize swallow function.
- A necessary first step for swallow function characterization in HRCA, is automatic swallow extraction which completes the non-invasive and subjective HRCA-based swallowing assessment.

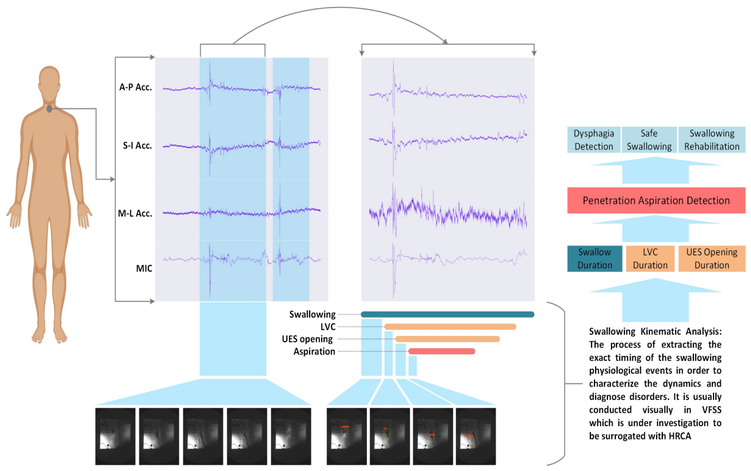
Videofluoroscopy Swallowing Study (VFSS): Gold Standard for swallow function evaluation
Pros:

- Evaluation of the entire swallowing process.
- Relatively noninvasive.

Cons:

- High x-ray dose.
- Not available in all clinics.
- An expert is always needed to interpret results.





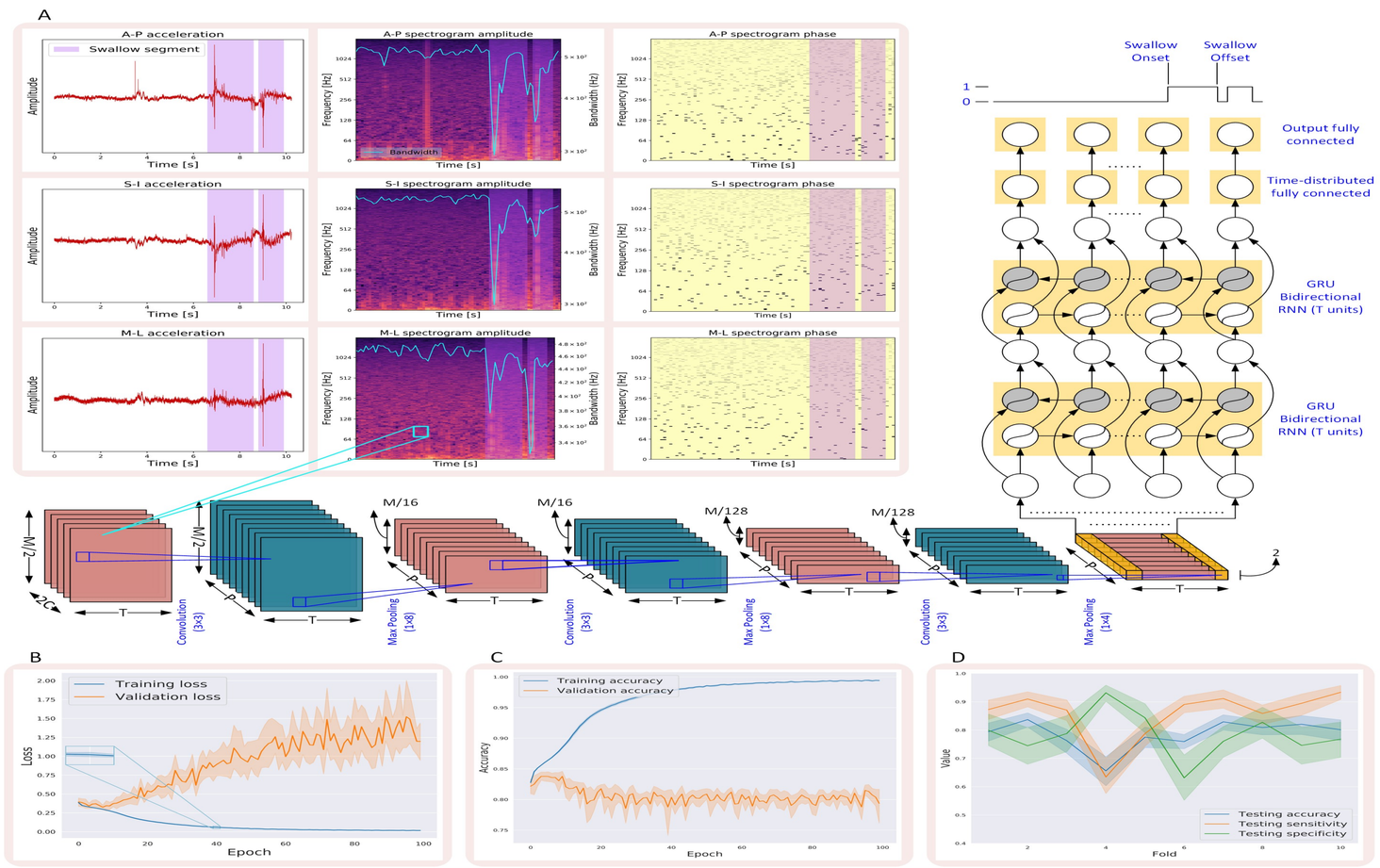
Objective

We investigate the ability of swallow segment extraction in HRCA signals using deep learning and power spectral estimation. The proposed deep learning framework combines convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to achieve high temporal swallow segment extraction efficiency through the sequential modeling power of RNNs.

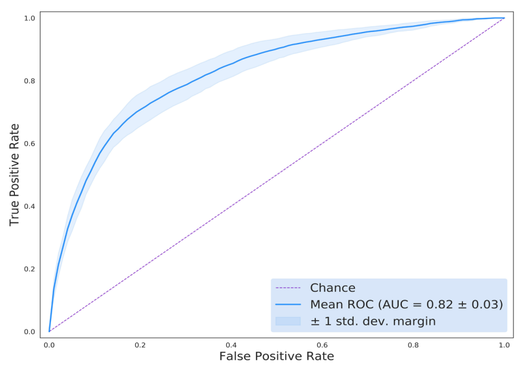
Dataset and Evaluation

- We analyzed HRCA signals of 3144 swallows from 248 patients (142 males, 106 females, age: 63.8±13.7) with suspected dysphagia who underwent VFSSs concurrently with HRCA recordings. Human judges rated the onset and offset of swallows by viewing VFSS images.
- A 10-fold cross validation scheme was used to evaluate the performance of the deep learning framework over the dataset. The swallow segment detected by the network was validated against the ground truth provided by the judges using accuracy, sensitivity, specificity and the overlap ratio between the detected segments and ground truth.
- Area under the reciever operating characteristic (ROC) curves was also used as a quality indicator for the detection performance of the framework across the 10 folds.

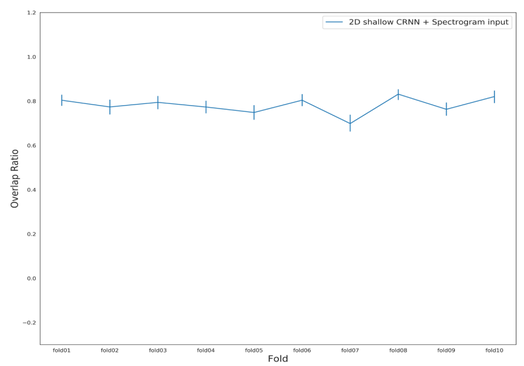
Methods



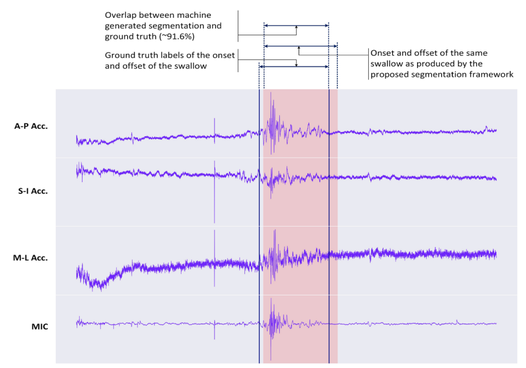
Results



ROC curves with the AUC values for the results of each fold of the 10 folds performed for evaluation.



The overlap ratio between the detected swallow segments and the ground truth segments marked by the judges for the 10 folds.



An example for the agreement (overlap) between the detected swallow segment by the network and the ground truth segment for one swallow (91.6% overlap).

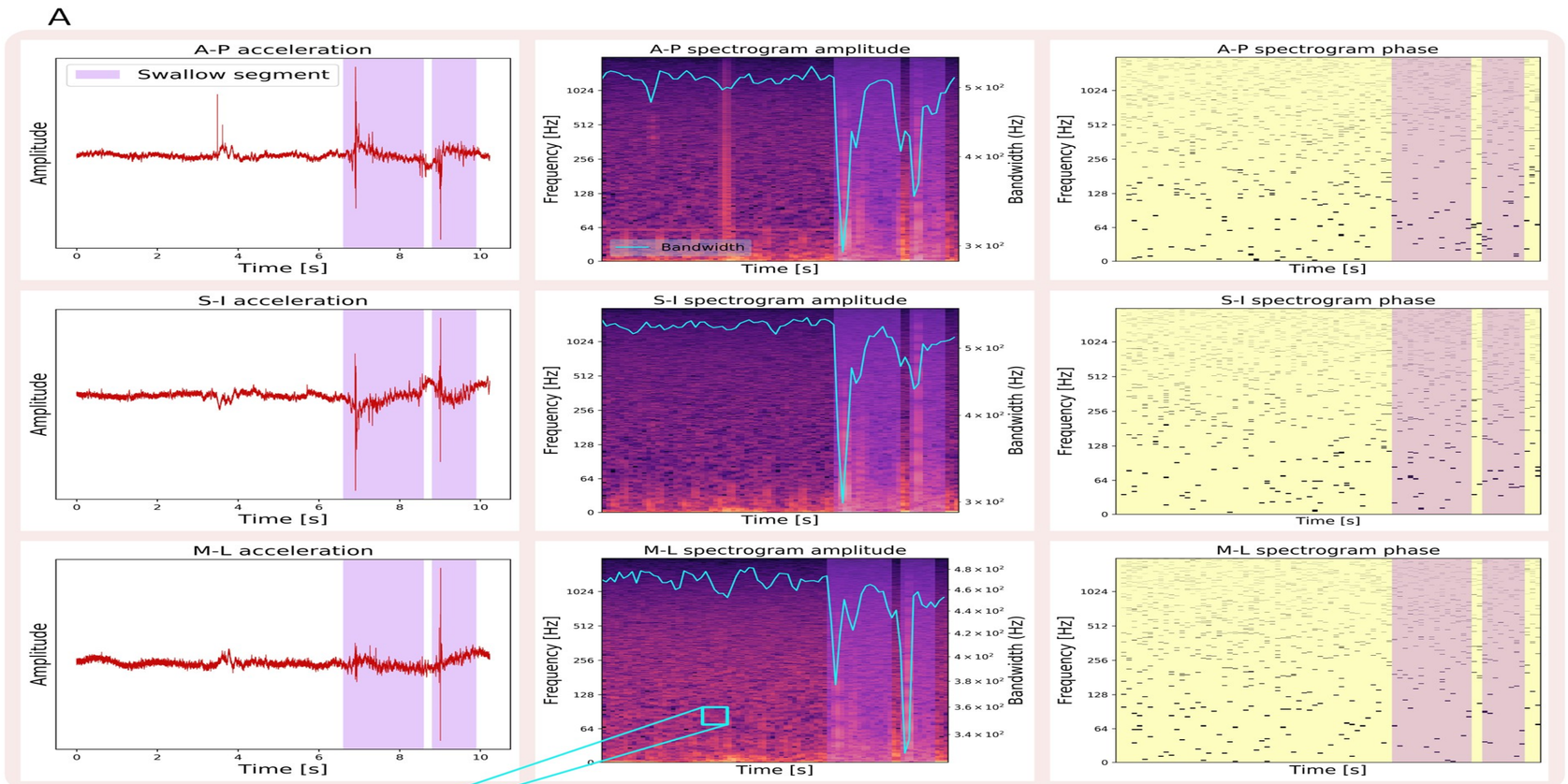
Conclusion

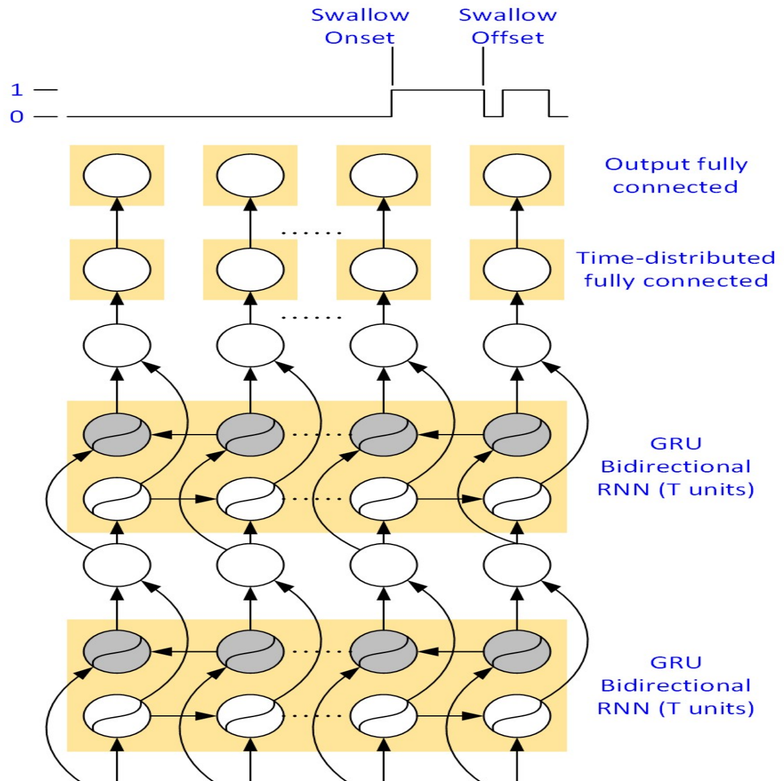
- The ability to autonomously detect swallow onset and swallow offset is vital for dysphagia clinical and research applications.
- Elimination of the time consuming manual extraction of swallows increases the efficiency and speed of HRCA-based swallow function assessment and research.
- The proposed deep learning framework achieved high detection accuracy and demonstrated robustness in swallow segment extraction.
- The combined use of CNNs and RNNs contributed to achieve good detection accuracy due to sequence modeling superioirity over traditional machine learning techniques.

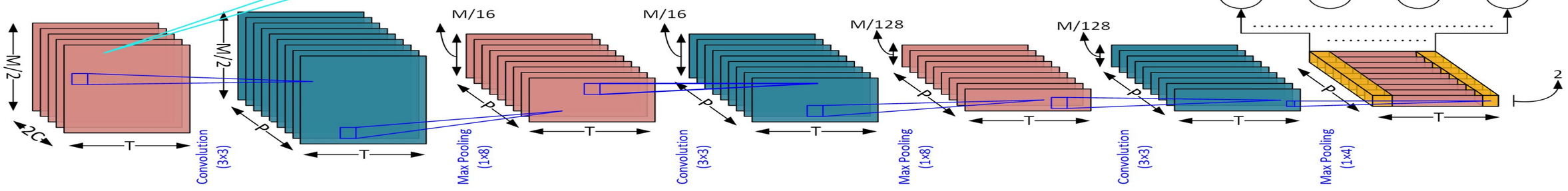
Acknowledgment

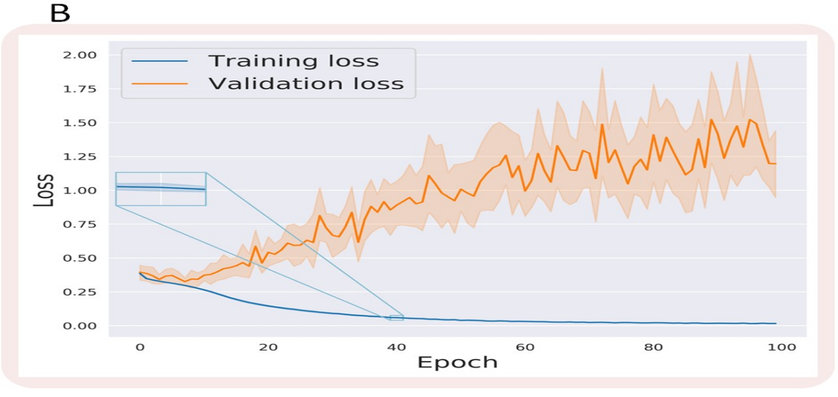
This study was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number R01HD092239, while the data was collected under award number R01HD074819. The computational resources utilized in this study were provided by Microsoft and its could service, Azure, through Microsoft's generous support to Pittsburgh CREATES. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding organizations.

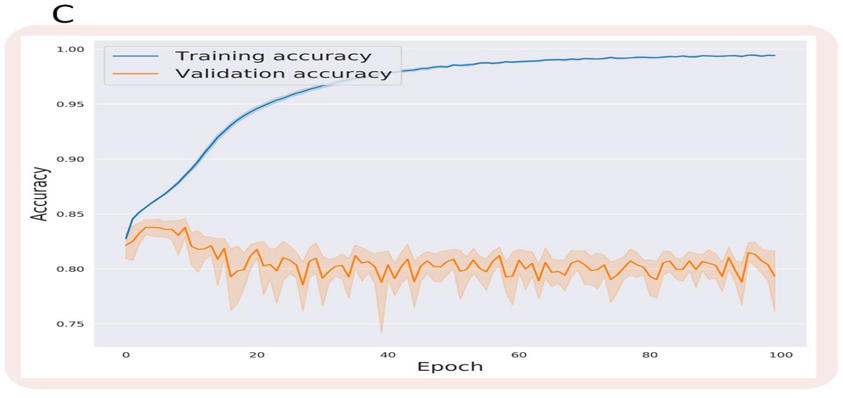
Methods

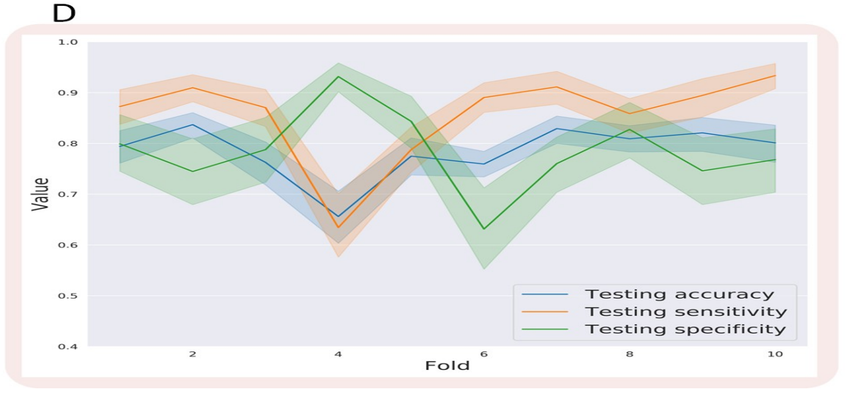

A
A-P acceleration
Swallow segment
Amplitude
Time [s]
A-P spectrogram amplitude
Frequency [Hz]
Bandwidth (Hz)
Bandwidth
Time [s]
A-P spectrogram phase
Frequency [Hz]
Time [s]
S-I acceleration
Amplitude
Time [s]
S-I spectrogram amplitude
Frequency [Hz]
Bandwidth (Hz)
Time [s]
S-I spectrogram phase
Frequency [Hz]
Time [s]
M-L acceleration
Amplitude
Time [s]
M-L spectrogram amplitude
Frequency [Hz]
Bandwidth (Hz)
Time [s]
M-L spectrogram phase
Frequency [Hz]
Time [s]


Swallow Onset
Swallow Offset
1
0
Output fully connected
Time-distributed fully connected
GRU Bidirectional RNN (T units)
GRU Bidirectional RNN (T units)

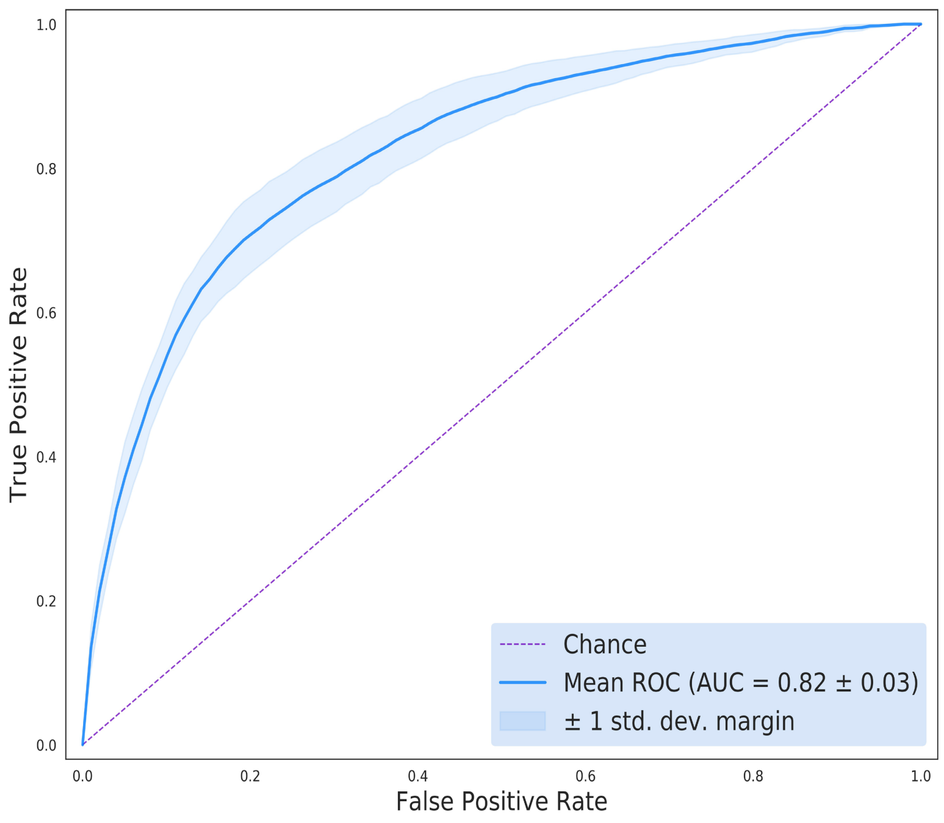

M/2
2C
T
Convolution (3x3)
M/2
P
T
M/16
Max Pooling (1x8)
M/16
Convolution (3x3)
M/128
Max Pooling (1x8)
M/128
Convolution (3x3)
Max Pooling (1x4)
2


B
Training loss
Validation loss
Loss
Epoch

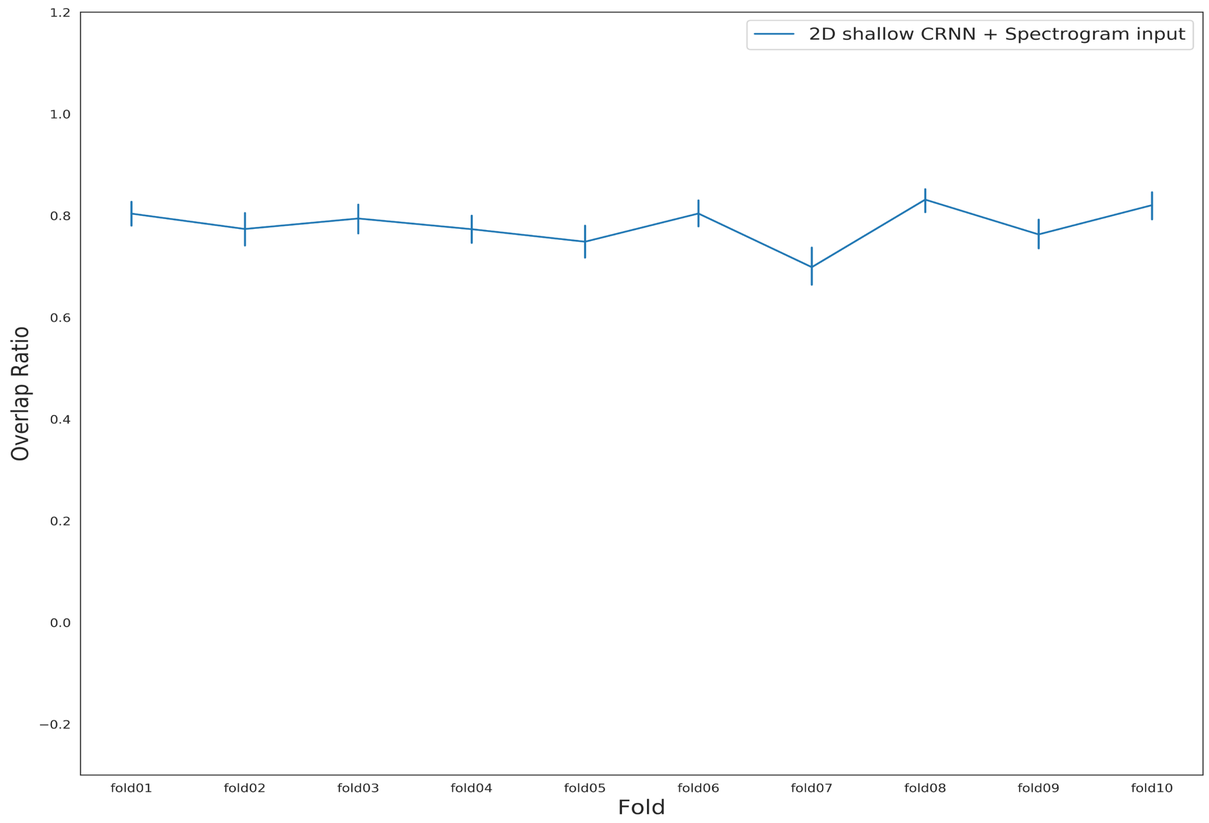

C
Training accuracy
Validation accuracy
Accuracy
Epoch


D
Testing accuracy
Testing sensitivity
Testing specificity
Value
Fold

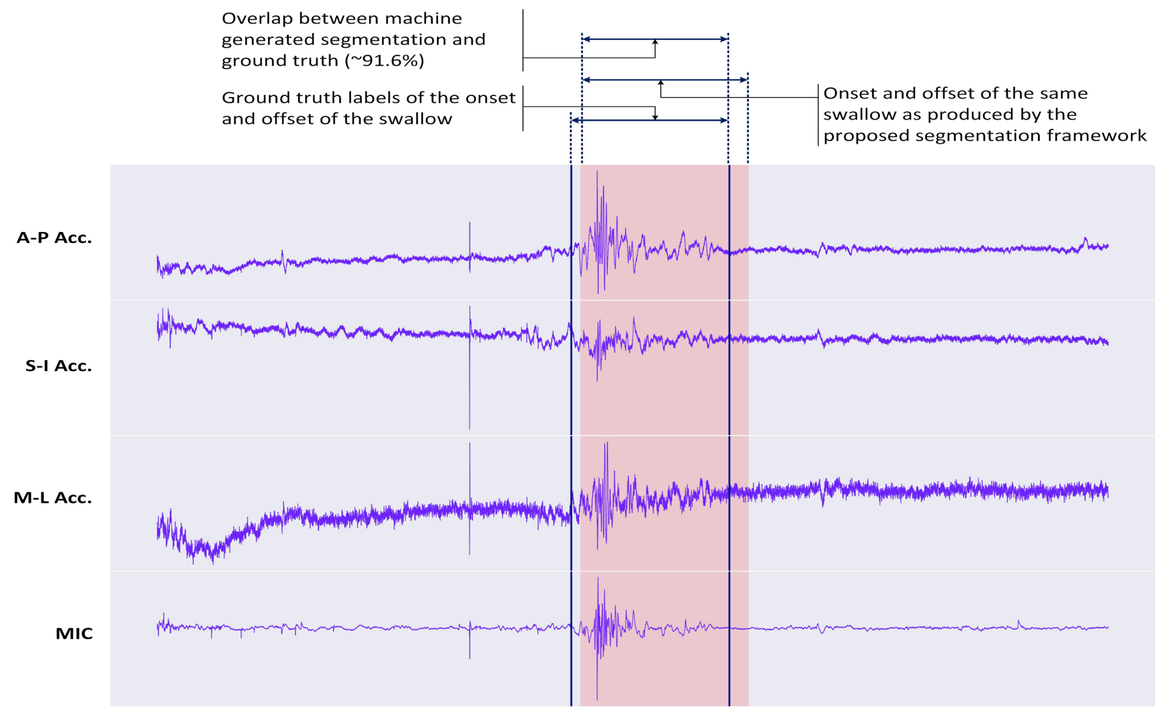
Results



ROC curves with the AUC values for the results of each fold of the 10 folds performed for evaluation.



The overlap ratio between the detected swallow segments and the ground truth segments marked by the judges for the 10 folds.



An example for the agreement (overlap) between the detected swallow segment by the network and the ground truth segment for one swallow (91.6% overlap).