SEGNET-BASED DEEP REPRESENTATION LEARNING FOR DYSPHAGIA CLASSIFICATION

Siddharth Subramani, Achuth Rao M V*, Anwesha Roy*, Prasanna Suresh Hegde†, Prasanta Kumar Ghosh**

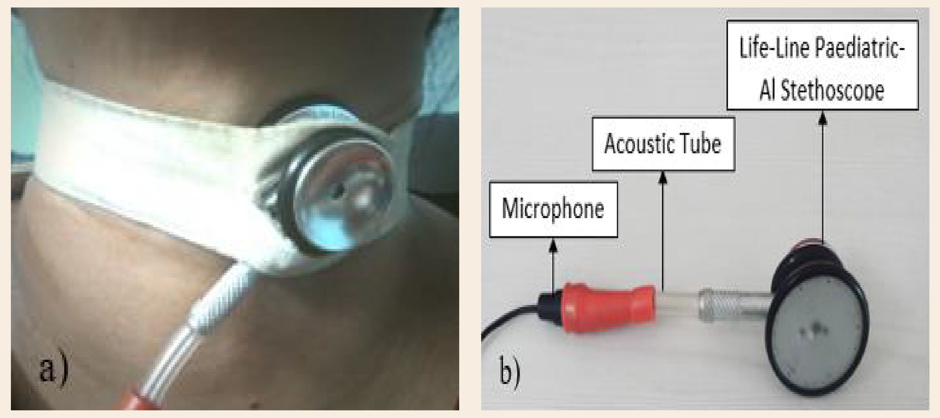
*Department of Electrical Engineering, Indian Institute of Science, Bangalore-560 012, India, †Dept. of Head and Neck Surgery, Health Care Global Enterprises Ltd Bangalore-560 002, India

Introduction

- **Dysphagia** - difficulty in swallowing food
- Dysphagia can potentially cause aspiration, dehydration & malnutrition
- Swallowing can be studied non-invasively through **Cervical Auscultation** (CA)
  - CA captures vibrations from the surface of the throat when swallowing food bolus
  - ... and gives reliable representations just as golden standards such as VideoFluoroScopy[1]
- **Objective**:
  - Learn acoustic representations to classify swallows as healthy or dysphagic

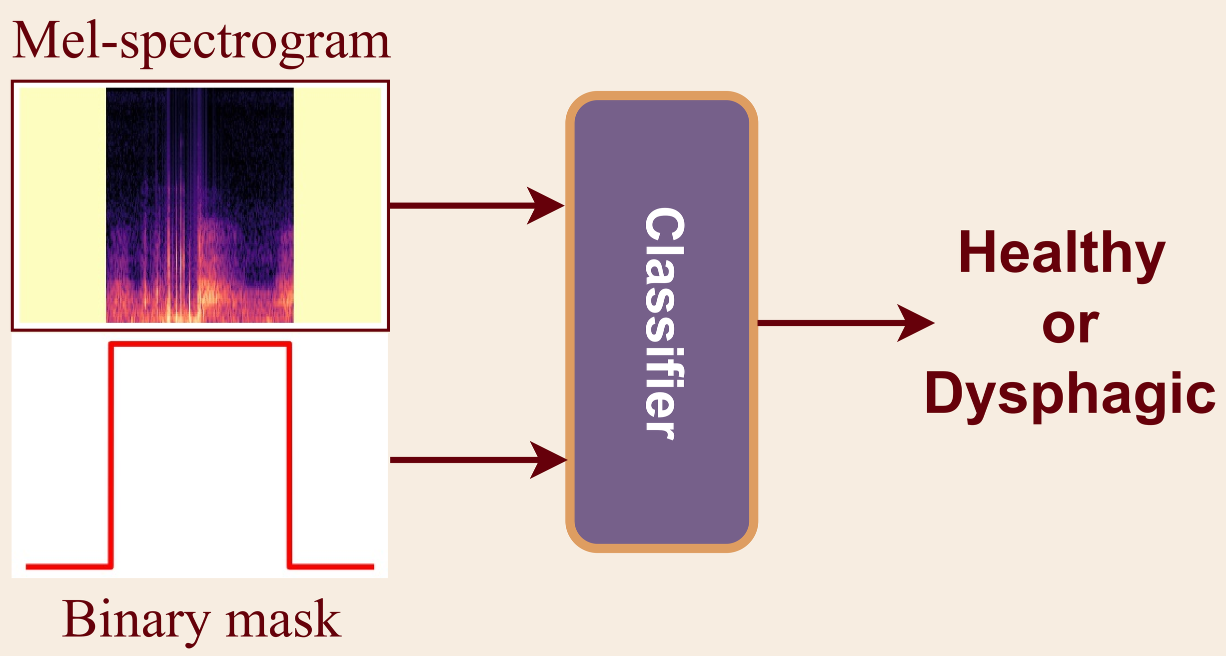
Dataset

- 24 subjects: 14 healthy controls, 10 patients



- **CA device setup**: Paediatric-Al Stethoscope connected to microphone (30Hz - 15000Hz) was patched on throat surface posterior to cricoid cartilage, around the trachea
- **Swallow tasks**: Dry, 5ml, 10ml and 15ml of water, 3-4 times per subject
- **172** healthy & **118** dysphagic swallow recordings in total - Dataset **InD**

Swallow features and process outline



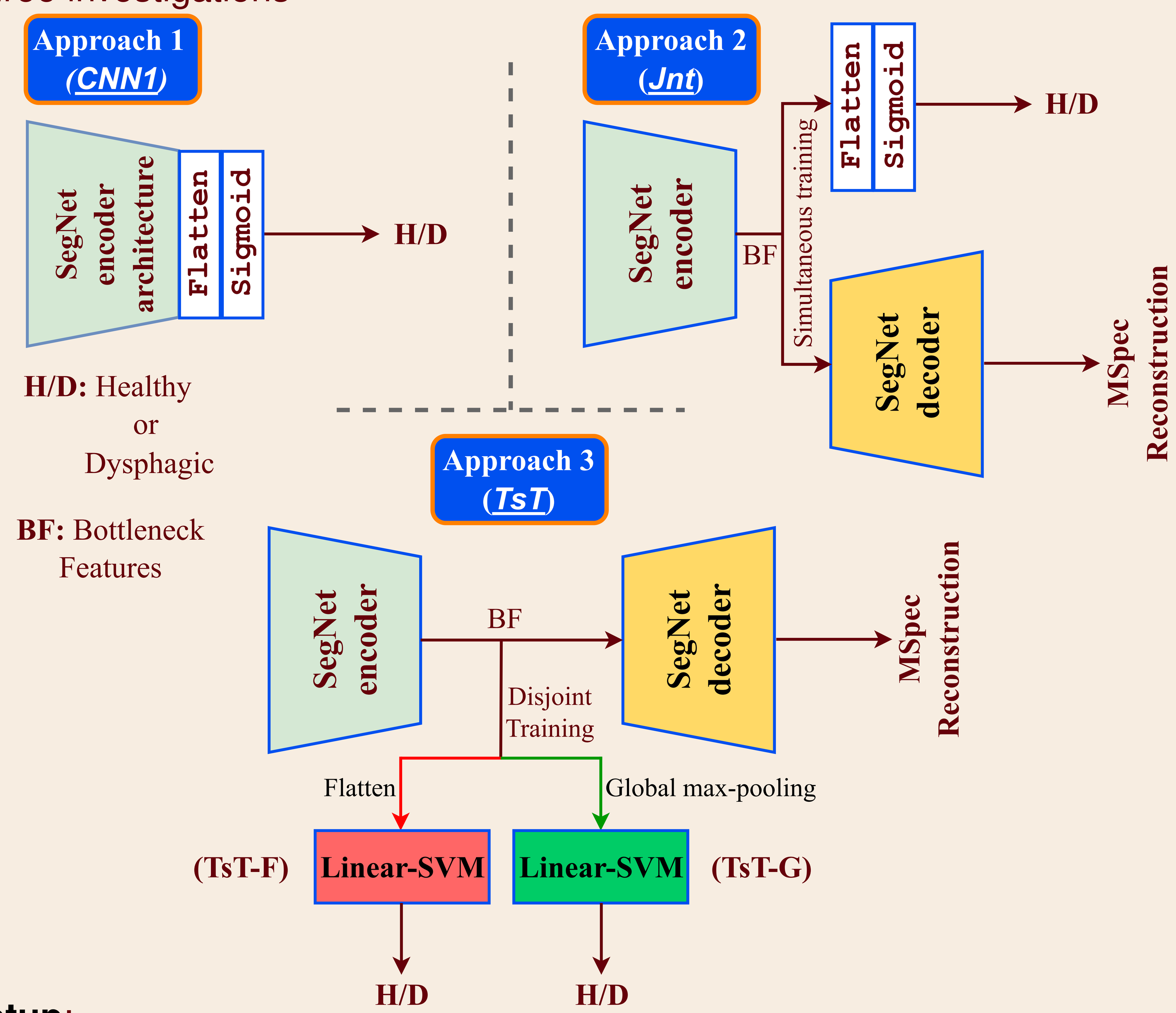
- **Input feature**: Mel-Spectrogram (MSpec) for every swallow signal was computed with 20ms long hamming winsow and 2ms long hop
  - **Pre-processing** - MSpecs (not signals) padded with zeros to equalize time length of all MSpecs

Classification pipeline

- 1D-CNN (13 layers) SegNet-based[2] feature learner
- Linear-SVM classifier
- MSpecs multiplied with binary 1D-mask prior to convolution layers in SegNet

Proposed approaches & Experimental Setup

- Three investigations



- **Setup**:
  - 5 **groups** of 24 subjects with **two** patients and **three** controls in every group - 3 groups as training set, 1 group each for validation and test sets (with no common subjects); 5-fold CV
  - SegNet recontruction loss: MSE, validation accuracy based selection of SVM regularization parameter
  - Baseline scheme (BLS) comparison[3]: fourier transform and spectrogram features to train an RBF-SVM classifier

References

[1] Geovana de Paula Bolzan et al., "Contribution of the cervical auscultation in clinical assessment of the oropharyngeal dysphagia," Revista CEFAC, vol. 15, no. 2, pp. 455–465, 2013.

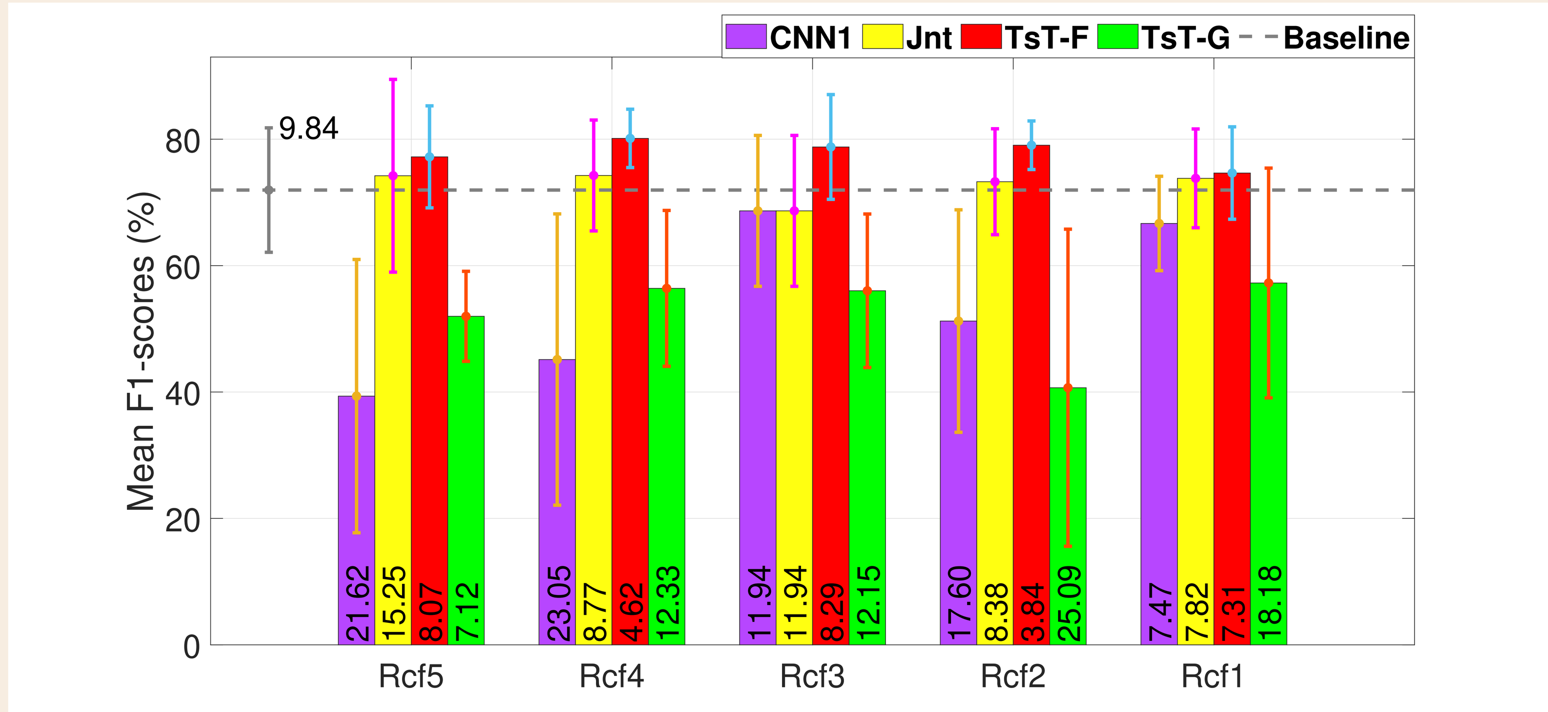
[2] Vijay Badrinarayanan et al., "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE trans. on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[3] Shigeyuki Miyagi et al., "Classifying dysphagic swallowing sounds with support vector machines," in Healthcare. Multidisciplinary Digital Publishing Institute, 2020, vol. 8, p. 103.

Results & Discussion

- Depth analysis: 5 "RCF*i*-trials" with bottleneck features from each of the 5 progressing maxpooling layers in SegNet encoder
  - Learning at different depths identifies time-scale resolution that maximizes performance
  - Each maxpooling layer has Receptive Field Size (RCF) of 3, 10, 26, 58, 122 respectively
- Best: TsT-F in RCF4-trial with a mean (across folds) test F1-score of **80.13%**

	Mean F1-score (%)	Mean Sensitivity (%)	Mean Specificity (%)	Mean SD (%)
CNN1	54.21	43.05	62.14	±16.33
Jnt	72.84	72.91	62.92	±10.43
TsT-F	77.96	78.29	66.98	±6.43
TsT-G	52.46	49.28	47.73	±14.97
Baseline F1-score = 71.95%, Sensitivity = 76.67%, Specificity = 42.92%				



- BLS suffered on **InD** (larger and imbalanced data)
- Jnt performance only competitive with BLS but still showed low F1-scores
- TsT-F outperformed all approaches, including TsT-G, with highest mean F1-score and lowest mean SD across all RCF*i*-trials

Conclusion

- Unsupervised feature learning followed by supervised training performed better than fully supervised training
- Performance variation at different depths hints the influence of swallow phase transition patterns
- **Future directions**
  - Adapt proposed method to identify severity of dysphagia
  - Explore time-scale analysis to outline spectral signatures pertaining to levels of severity

Acknowledgement: The authors thank the Pratiksha Trust and the Department of Science and Technology, Govt. of India for their support.

http://spire.ee.iisc.ac.in/spire/ ssd9709@gmail.com, achuthr@iisc.ac.in, anwesha97@yahoo.in, prasanna.h@hcgel.com, prasantg@iisc.ac.in