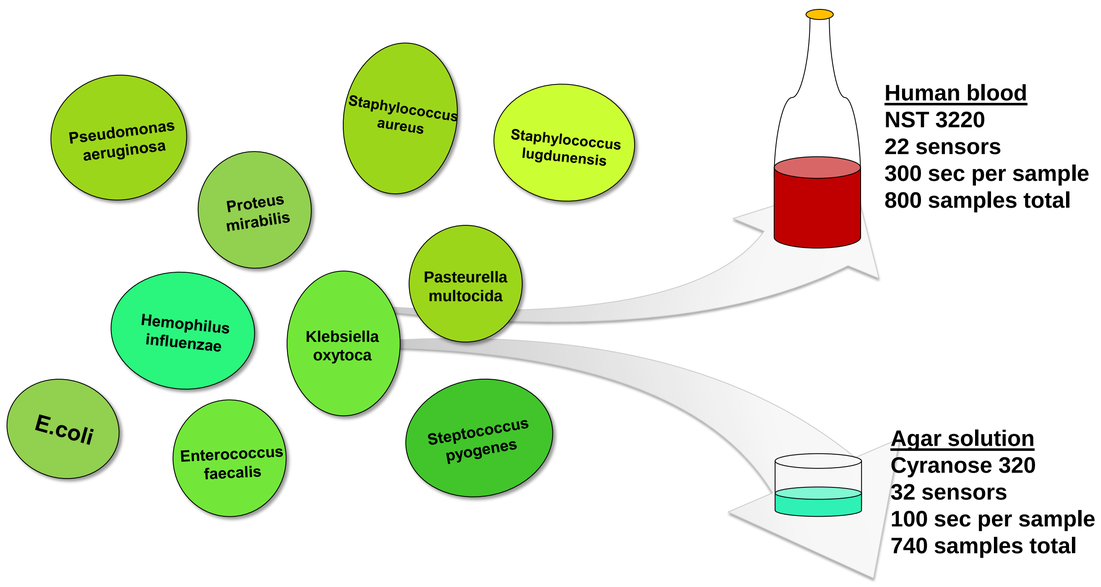
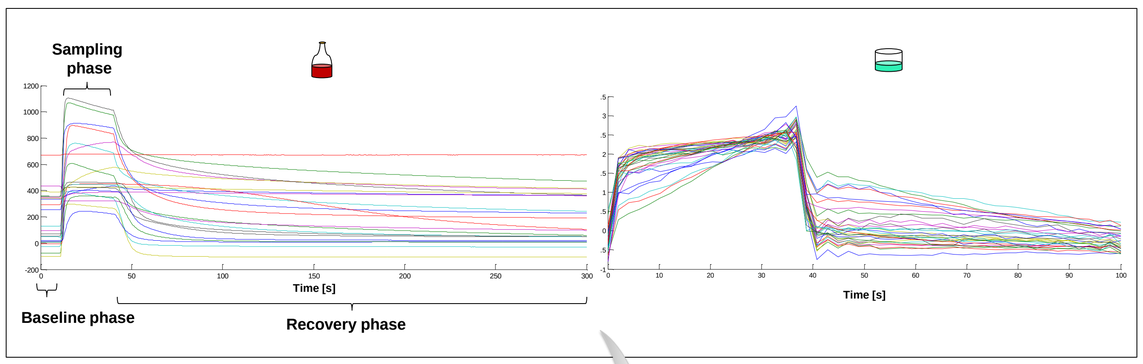
Unsupervised feature learning for electronic nose data applied to Bacteria Identification in Blood

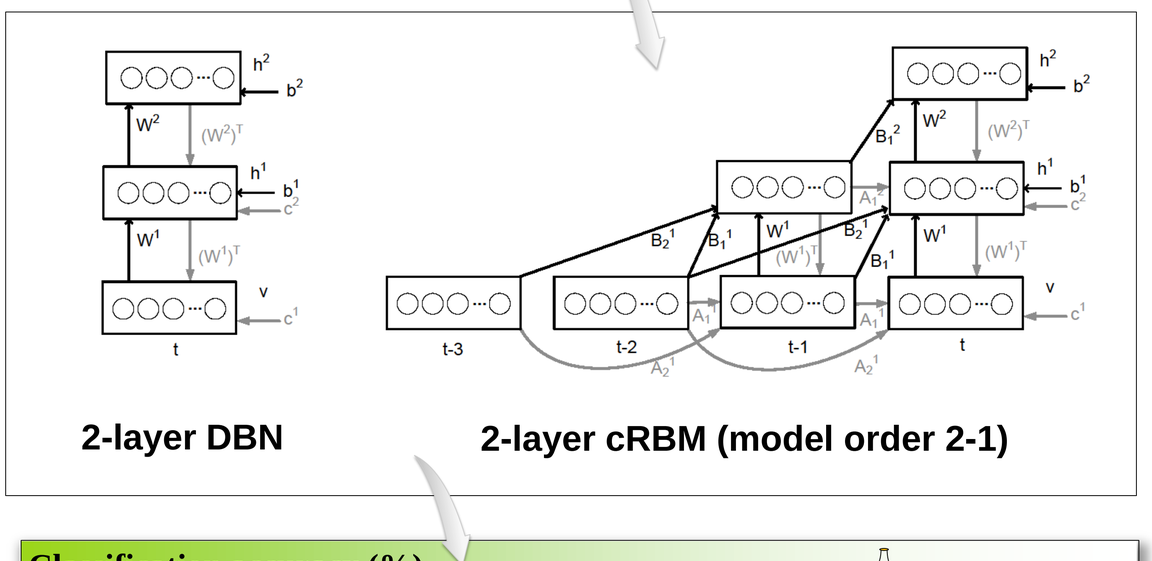
Martin Längkvist, Amy Loutfi

Applied Autonomous Sensor Systems, School of Science and Technology, Örebro University, Sweden

Quickly identifying the presence of bacteria has both health and economic benefits. A new, quick, and reliable method is to use an electronic nose. Data is unintuitive, difficult to interpret and has high redundancy. Is it possible to use deep learning methods to eliminate the need for hand-designed features?







Classification accuracy (%)		
Features + SVM [1]	93.7	84.0
1-layer DBN (20 sample window)	93.8	41.9
2-layer DBN (20 sample window)	**96.2**	47.3
2-layer DBN (full sample window)	43.8	44.6
1-layer cRBM (model order 5)	85.0	96.0
2-layer cRBM (model order 10-10)	75.0	90.0
2-layer cRBM (model order 5-5)	85.0	**96.0**

Introduction

This project aims to use an electronic nose in order to discriminate between different type of bacteria that is typically found in blood and can lead to blood poisoning (sepsis). 10 different bacteria was individually put in human blood and agar solution. Each sample was smelled by two electronic nose systems for a duration of days resulting in two big data sets. Number of sensors and sample length differed between the two systems.

Data

Electronic nose data represents multivariate time-series from chemical gas sensors exposed to a gas. Normalization was done by subtracting each sensor signal with the baseline value and dividing by maximum value. Data was divided into 80% training, 10% validation, and 10% testing.

Method

Raw sensor data was used in both a DBN and a cRBM architecture with 1 and 2 layers. A full window input and a 20 sample moving average window was used. Classification was done by attaching one set of softmax units to the top hidden layer units. The same set of softmax units where used for all window positions.

Results

A DBN with a 20 sample moving average window outperformed a feature-based approach on the bacteria in blood data set. A cRBM outperformed a feature-based approach on the bacteria in agar data set. A smaller window width for DBN and model order for cRBM gave better results than a full sample window or higher model order. Adding a second layer gave same or better results.

References

[1] M. Trincavelli, S. Coradeschi, A. Loutfi, B. Söderquist, P. Thunberg, Direct identification of bacteria in blood culture samples using an electronic nose, IEEE Trans Biomedical Engineering 57 (Issue 12), 2010

Acknowledgments

L. Barkman, B. Söderquist, and P. Thunberg at Örebro University Hospital. M. Trincavelli and S. Coradeschi at Örebro University. D. Wulsin and G. Taylor for code sharing. This work was funded by NovaMedTech.

website: aass.oru.se/~mlt

Universitetssjukhuset Örebro ÖREBRO LÄNS LANDSTING · ÖREBRO UNIVERSITY · aass · NovaMedTech