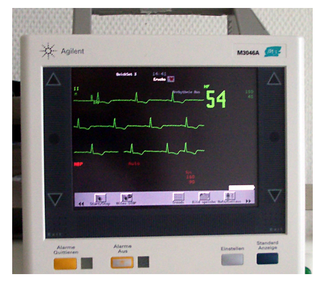
Feature Learning in Clinical Time Series using Deep Learning

Biswajit Paria[1], David Kale[2], and Yan Liu[2]
[1]Indian Institute of Technology Kharagpur; [2]University of Southern California
biswajitsc@iitkgp.ac.in, {dkale, yanliu.cs}@usc.edu



Critical Care Medicine

Among the most important areas of medicine.

- >5 million US patients admitted to annually.[a]
- US cost in 2005: $81.7 bn, ∼1% GDP.[a]
- Mortality up to 30%, depending on condition.[a]
- Impact: impairment, pain, depression.

Modeling critical illness: *a grand challenge*.

- *Big Data:* 100's of data points/patient/hour.
- Diverse conditions: respiratory failure, sepsis, etc.
- Symptoms complex, vary across patients, overlap between conditions.

[a] *Society of Critical Care Medicine*: `http://www.sccm.org/Communications/Pages/CriticalCareStats.aspx`

Phenotype Discovery as Feature Learning

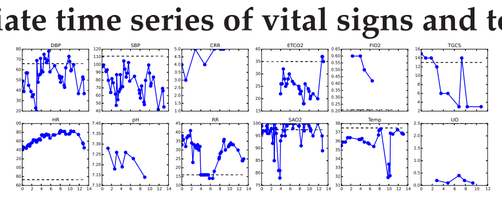
Medicine: *phenotypes, biomarkers*

- Measurable attributes of patient/disease.
- Independent of other biomarkers.
- Separate patients into meaningful groups.
- Improve outcome prediction, risk assessment.
- Clinically plausible, interpretable.

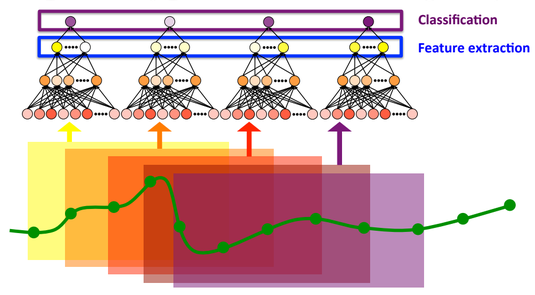
Machine learning: *features, representations*

- Measurable properties of objects.
- Independent, disentangle factors of variation.
- Form natural clusters.
- Useful for discriminative, predictive tasks.
- Interpretable, provide insight into problem.

Multivariate time series of vital signs and test results



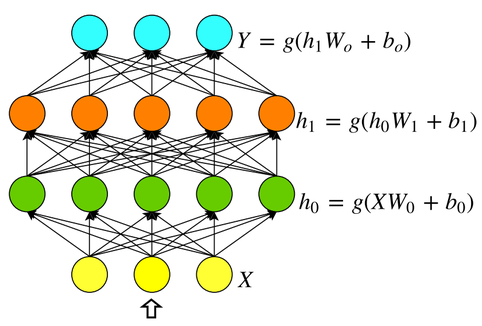
Temporal phenotype detection using *deep learning*

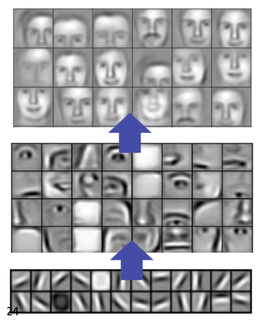


Feature Learning using Deep Learning

Stacked Denoising Autoencoders

- can perform complex feature transformations.
- can learn high level features.
- can capture complex interactions between input features.





Experimental Setup

Deep Learning Setting

- **Unsupervised Pre-Training:** SDAs for complex feature transformation.
- **Supervised Finetuning** - Gradient Descent using Back Propagation.

Neural Network Parameters

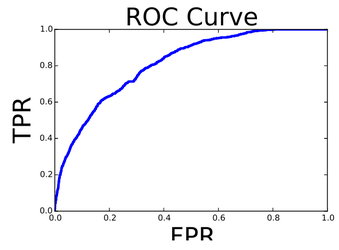
- **Layers:** In → 500 nodes → 100 nodes → 100 nodes → Out.
- **Regularizers:** L1 cost on highest layer weigths to impose sparsity on learnt features.
- **Stopping Criterion:** Early stopping based on validation error.

Experimental Results

Prediction of Acute Respiratory Distress Syndrome (ARDS)

ARDS

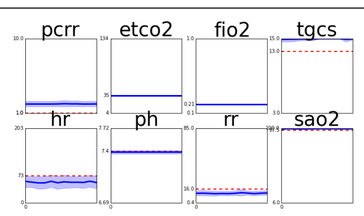
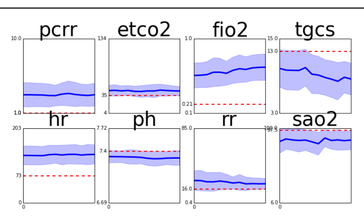
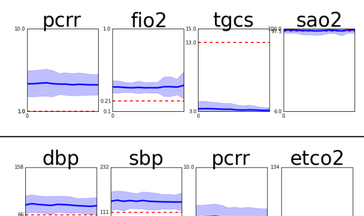
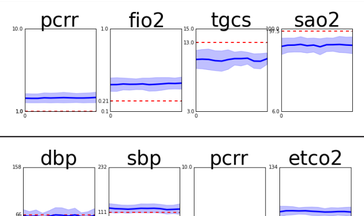
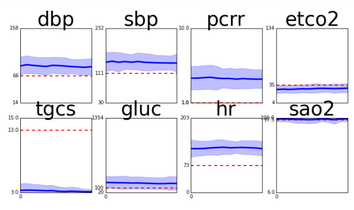
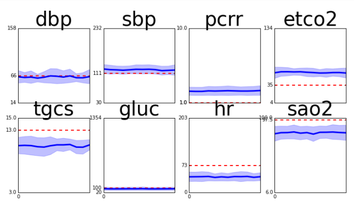
- not enough oxygen from the lungs and into the blood.
- fluid buildup in lungs.
- treated in an ICU.
- patients often deeply sedated.



Area under curve (AUC) = 81.10%

Analysis of learnt features

Top three features selected based on logistic layer weights for ARDS label.

	Feature Weight	Maximally Activating	Minimally Activating	Learnt variable trends
A	-1.7847			high pcrr, slightly higher etco2, high fio2, low tgcs, high hr, low ph, high rr, low sao2
B	1.4835			high pcrr, slighly high fio2, low tgcs, unstable sao2
C	-0.6502			normal and stable dbp & sbp, normal and stable pcrr, high etco2, slightly low tgcs, normal gluc, low hr, low sao2

Variables trends for learnt features

- High pcrr ⟹ Dehydration.
- Low tgcs ⟹ Unconscious.
- Low sao2 ⟹ Impaired oxygen intake.
- High fio2 ⟹ Oxygen masks.

Symptoms of ARDS

- Severe shortness of breath.
- Labored and unusually rapid breathing.
- Low blood pressure.

Learnt features and symptoms of ARDS look similar

Interpreting Learnt Features

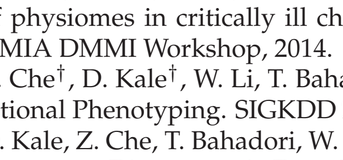
Goal: To interpret learnt features as traditional diagnostic rules.

Decision Trees (DT) can extract set of *if else* rules for learnt features.

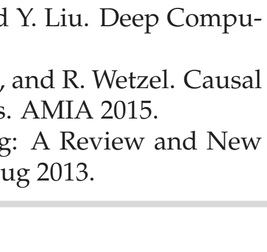
Gradient Boosted Decision Trees (GBDT) can extract set of **weighted Decision Trees** for learnt features.

Learning DTs for the learnt feature A

	Model	RMSE
I	DT - max leaf nodes = 50	0.6167
II	GBDT - depth = 2, est = 1	0.3285
III	GBDT - depth = 2, est = 10	0.2497
IV	GBDT - depth = 2, est = 100	0.1995



```
if (mean(tgcs) <= 0.89){
    if (mean(sao2) <= 0.96)
    {
        return -0.2307
    }else{
        return -0.0303
    }
}else{
    if(mean(temp) <= 0.61)
    {
        return 0.1664
    }else{
        return -0.0527
    }
}
```



Scatter plot for model IV

DT rules can approximate neural net features, are interpretable but complex. Promising direction for understanding deep learning.

Observations and Future Work

Observations

- Decision trees approximate the learnt features fairly well and are often interpretable.
- Multiple trends are learnt by a single feature.
- Redundant features pose a challenge to interpretation.

Future work

- Use *dropout*, etc., to impose further sparsity and disentangle features.
- Prevent duplication of features by imposing a *prior* (e.g., disease ontologies).
- Experiment with *recurrent*, *convolutional* nets.

Publications

- D. Kale, Z. Che, Y. Liu, and R. Wetzel. Computational discovery of physiomes in critically ill children using deep learning. 1st AMIA DMMI Workshop, 2014.
- Z. Che†, D. Kale†, W. Li, T. Bahadori, and Y. Liu. Deep Computational Phenotyping. SIGKDD 2015.
- D. Kale, Z. Che, T. Bahadori, W. Li, Y. Liu, and R. Wetzel. Causal Phenotype Discovery via Deep Networks. AMIA 2015.
- Y. Bengio, et al. Representation Learning: A Review and New Perspectives. TPAMI 35 (8): 1798-1828, Aug 2013.