# S92: Interpretable Deep Models for ICU Outcome Prediction

**Zhengping Che**[1], Sanjay Purushotham[1], Robinder Khemani[2], and Yan Liu[1]

[1]University of Southern California
[2]Children's Hospital Los Angeles

Nov 16, 2016

# Disclosure

The speaker and authors have no relationships with financial or commercial interests related to this work.

# Outline

- Background
  - Our footprint and other related work
  - New challenges we addressed
- Deep model for ICU prediction
  - Deep models for multi-modal ICU data
  - Quantitative evaluations
- Interpretable deep model
  - Interpretable mimic learning framework
  - Model evaluations and interpretations
- Conclusion

# Time series data is ubiquitous in health care

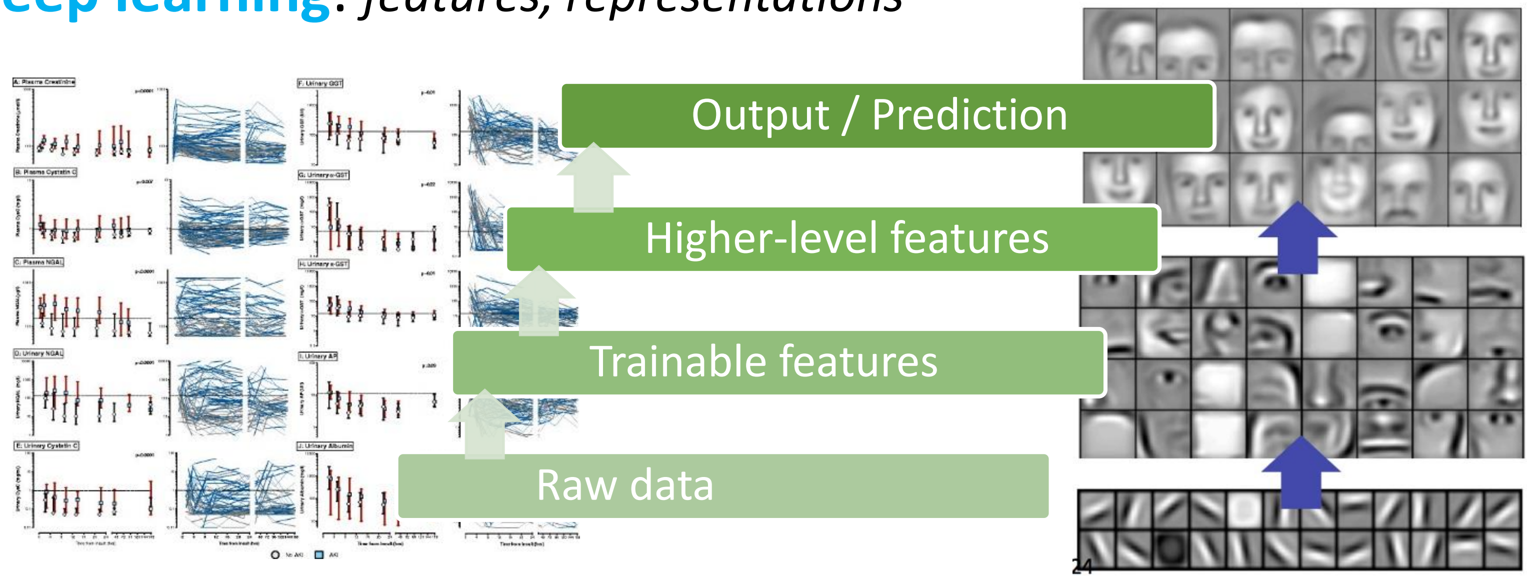
**Electronic Medical Records (EMR) from hospital ICUs**

**Personalized health data from mobile devices**

- Opportunities to improve health care quality
- Urgent needs of powerful and interpretable data-driven models

# Why deep learning (deep models)?

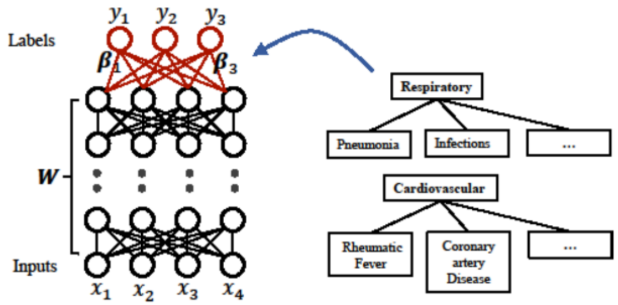
- **Health care**: *phenotypes, biomarkers*
- **Deep learning**: *features, representations*

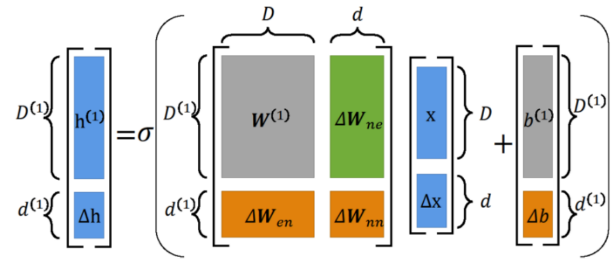



- ~~*Features from human knowledge…*~~ ***Features automatically learnt from data!***
- *Measurable!*
- *Useful!*
- *Interpretable?!*

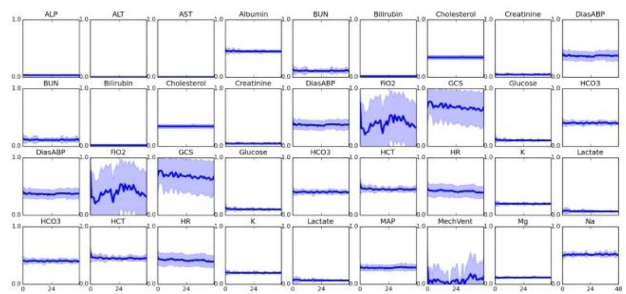
# Our roadmap

- Deep Computational Phenotyping [SIGKDD 2015]




- Causal Phenotype Discovery via Deep Networks [AMIA 2015]

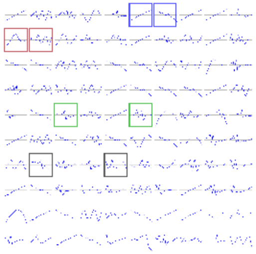


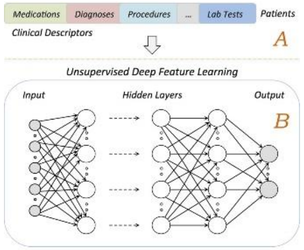

- Interpretable Deep Models for ICU Outcome Prediction [AMIA 2016] (this work)

# Related deep models for health care

- **Stacked Auto-encoder (SDA)**
- **Computational phenotyping**

[Lasko et al., 2013, Miotto et al., 2016]



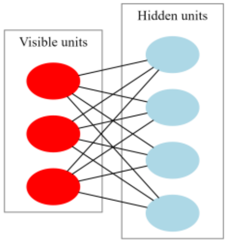


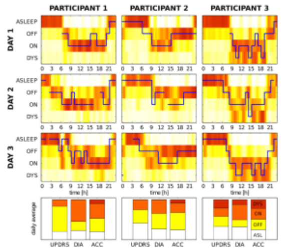

- **Deep neural networks (DNNs)**

  Restricted Boltzmann machine (RBM)
  Multi-layer perceptron (MLP)

- **Condition prediction**

[Dabek, Caban, 2015; Hammerla et al., 2015]



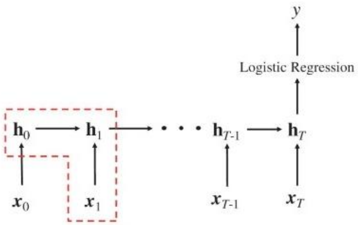




- **Recurrent neural networks (RNNs)**

  Long short-term memory (LSTM)
  Gated recurrent unit (GRU)

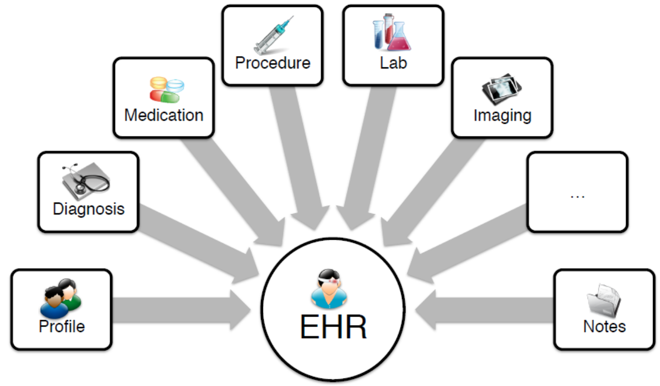
- **Diagnosis/event prediction**

[Lipton et al., 2015; Choi et al., 2015]

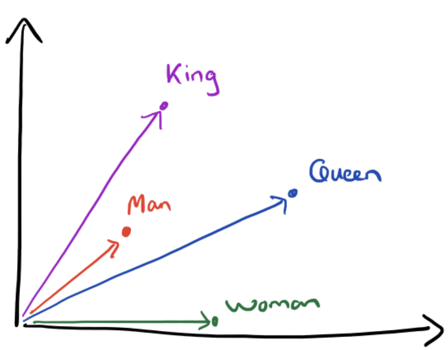



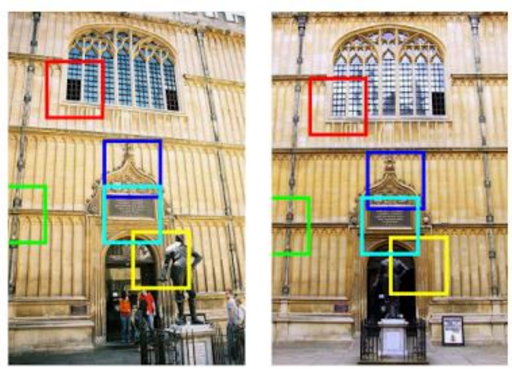
# Challenges still remain

- Handle multi-modal data in healthcare




- Provide fundamental and essential interpretability



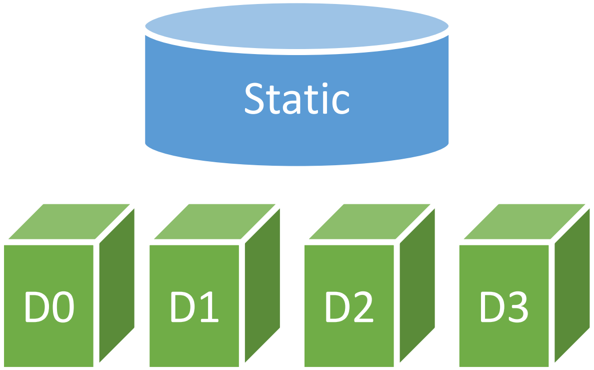



| | Risk Factor | Points |
|---|---|---|
| C | Recent Congestive HF | 1 |
| H | Hypertension | 1 |
| A | Age $\geqslant$ 75 | 1 |
| D | Diabetes | 1 |
| $S_2$ | Prior Stroke/TIA | 2 |

# Case study on ICU outcome prediction

- Acute hypoxemic respiratory failure
  - 398 patients at Children's Hospital Los Angeles (CHLA)
  - All patients stay >3 days

- Data
  - Static variables
    - 27 variables
    - Age, weight, etc.
  - Temporal variables
    - 21 variables x 4 days
    - Blood gas, ventilator signals, injury markers, etc.



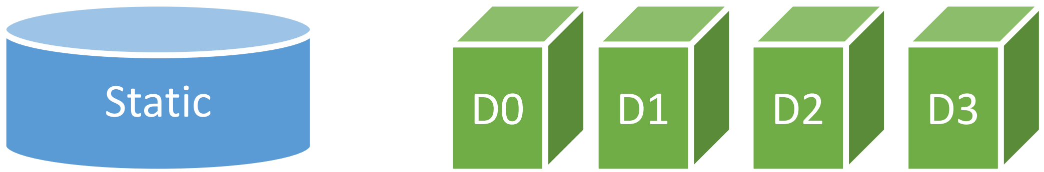

- Prediction tasks
  - Mortality
  - Ventilator free days

# Outline

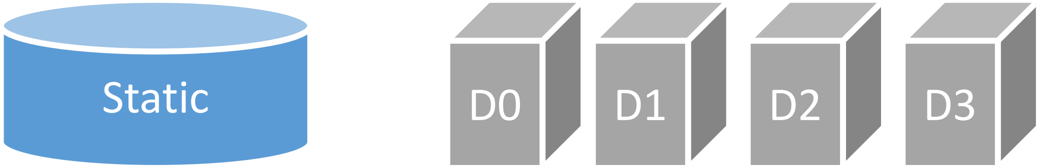
- Background
  - Our footprint and other related work
  - New challenges we addressed
- Deep model for ICU prediction
  - Deep models for multi-modal ICU data
  - Quantitative evaluations
- Interpretable deep model
  - Interpretable mimic learning framework
  - Model evaluations and interpretations
- Conclusion

# Our deep learning models

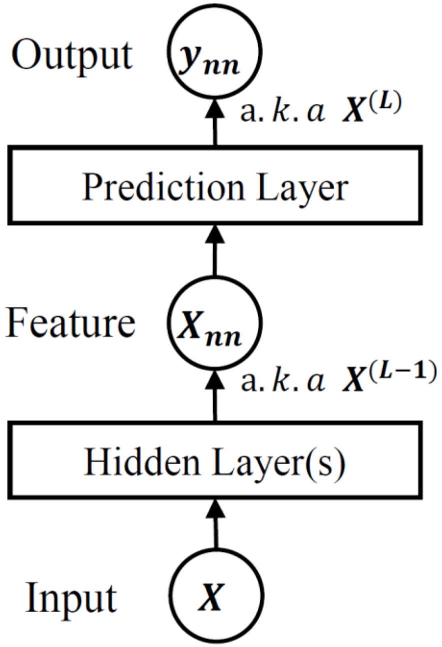
- Handle both static and temporal variables?



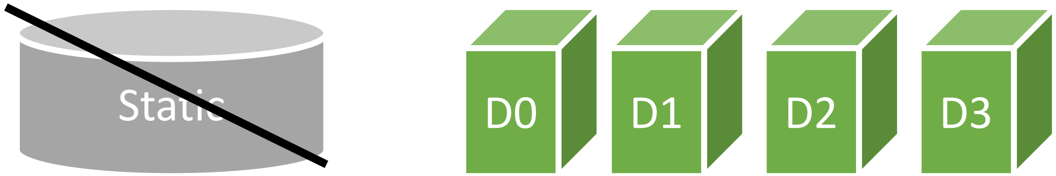
# Our deep learning model - DNN



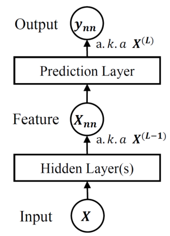

- Static + *flattened* temporal features
  - **DNN** (deep feed-forward neural net)



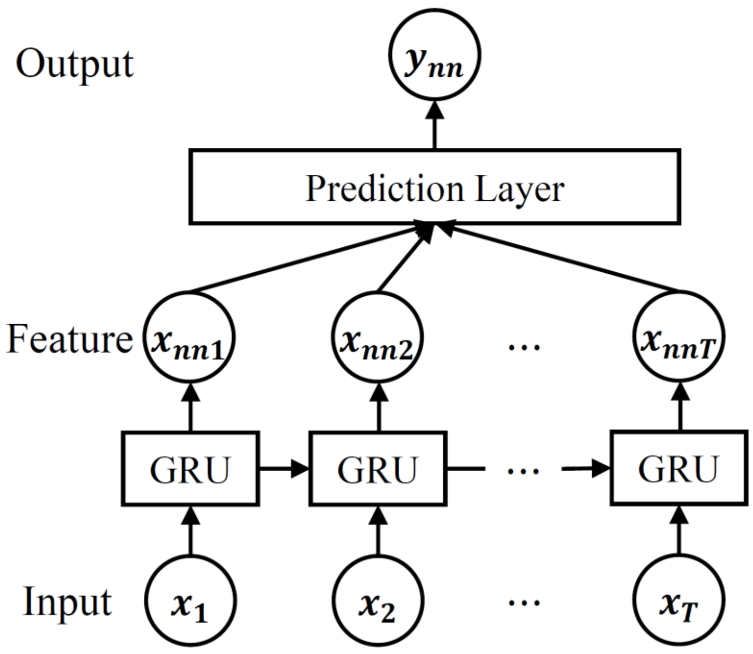
# Our deep learning model - GRU



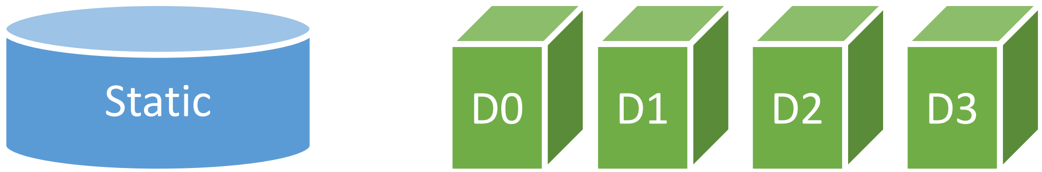

- *Static + (flattened) temporal features*
  - *DNN*



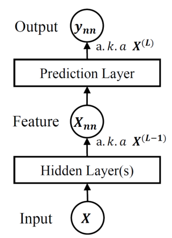

- Temporal features only
  - **GRU** (Gated Recurrent Unit)



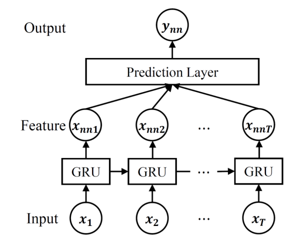
# Our deep learning model - DNN + GRU



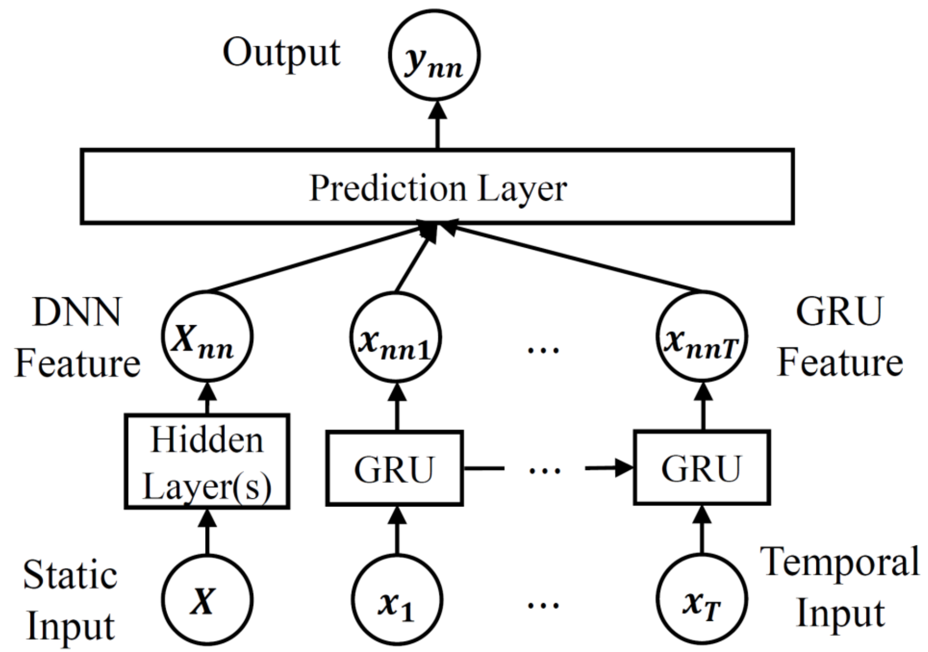

- *Static + (flattened) temporal features*
  - *DNN*




- *Temporal features only*
  - *GRU*

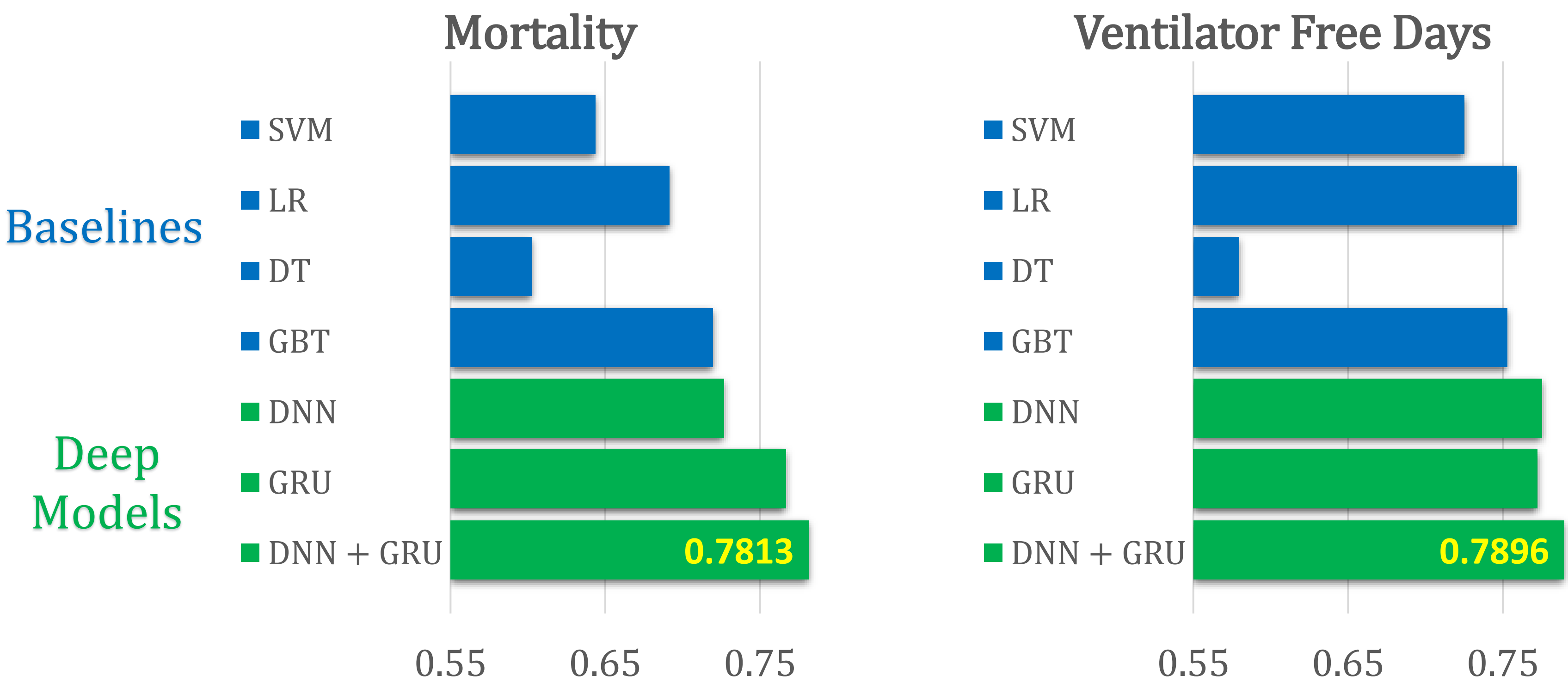



- Static + temporal features
  - **DNN + GRU** (combination)



# Quantitative results

**Prediction on patients with acute hypoxemic respiratory failure (AUROC)**

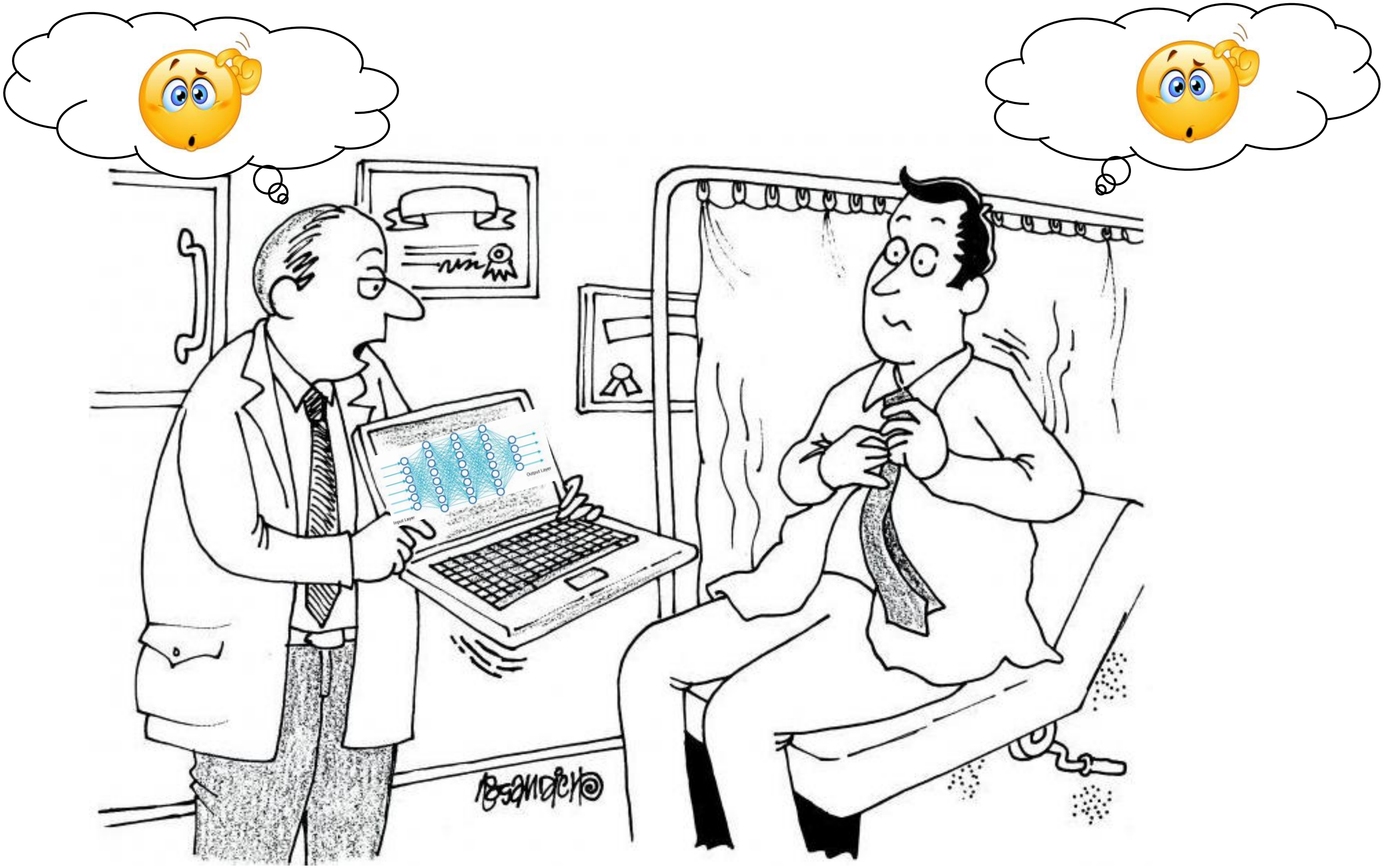



SVM: support vector machine; LR: logistic regression;
DT: decision tree; GBT: gradient boosting tree.
Results are based on 5-fold cross-validation.

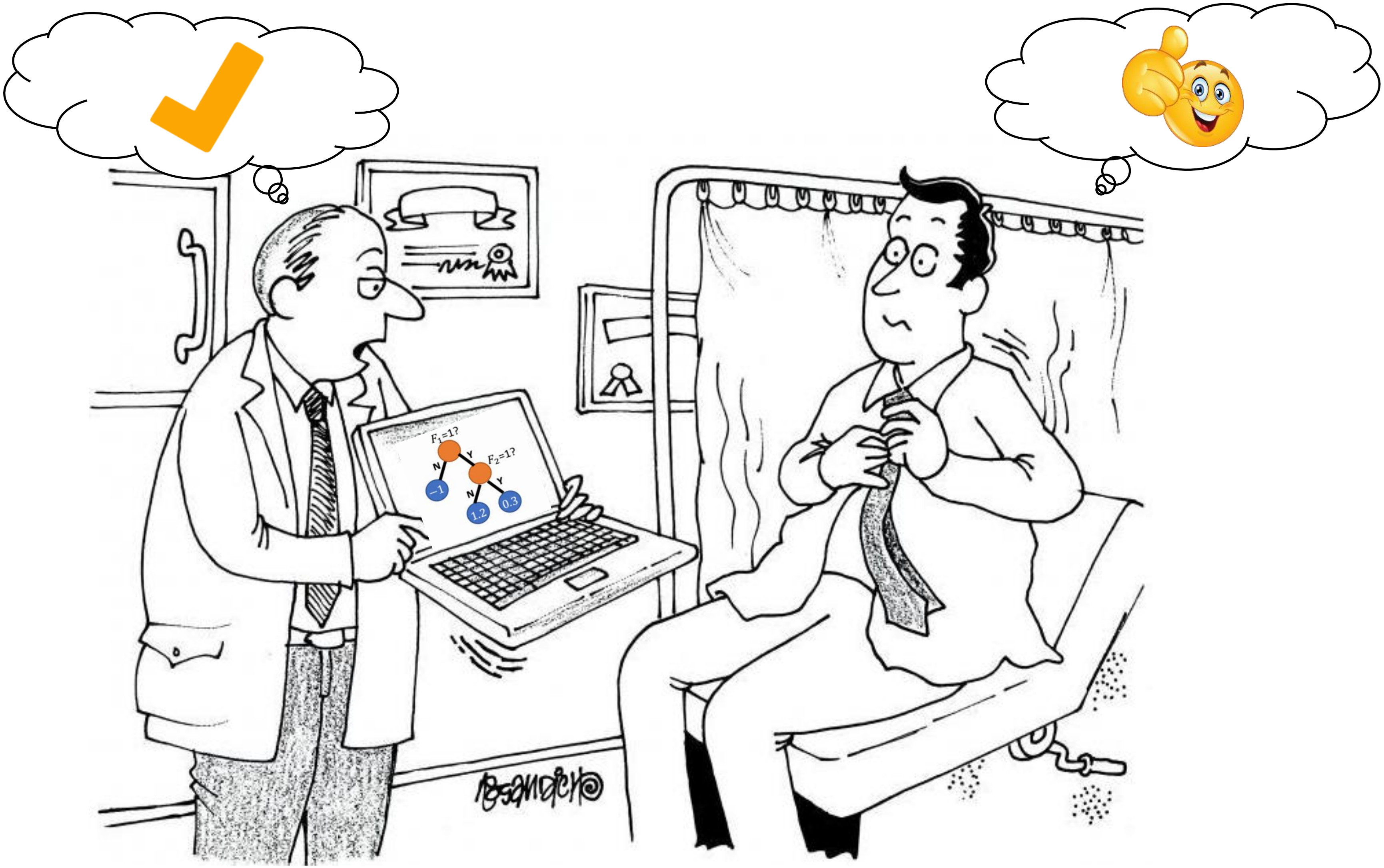
# Outline

- Background
  - Our footprint and other related work
  - New challenges we addressed
- Deep model for ICU prediction
  - Deep models for multi-modal ICU data
  - Quantitative evaluations
- Interpretable deep model
  - Interpretable mimic learning framework
  - Model evaluations and interpretations
- Conclusion

# Interpretability is necessary

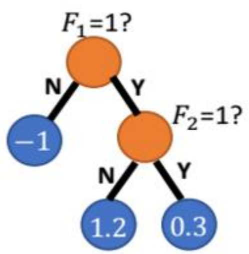


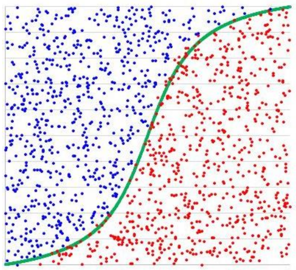
# Interpretability is necessary



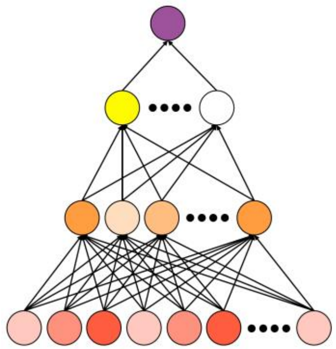
# Performance vs. Interpretability

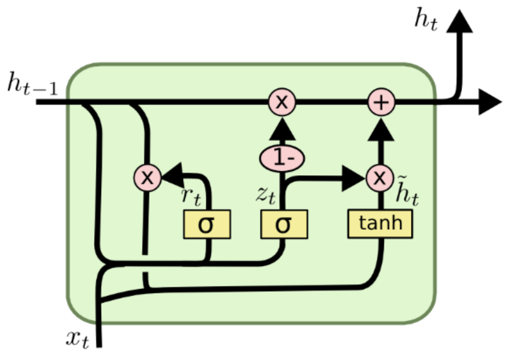
- Simple and commonly used models
  - Easy to interpret, mediocre performance



- Deep learning solutions
  - Superior performance, hard to explain



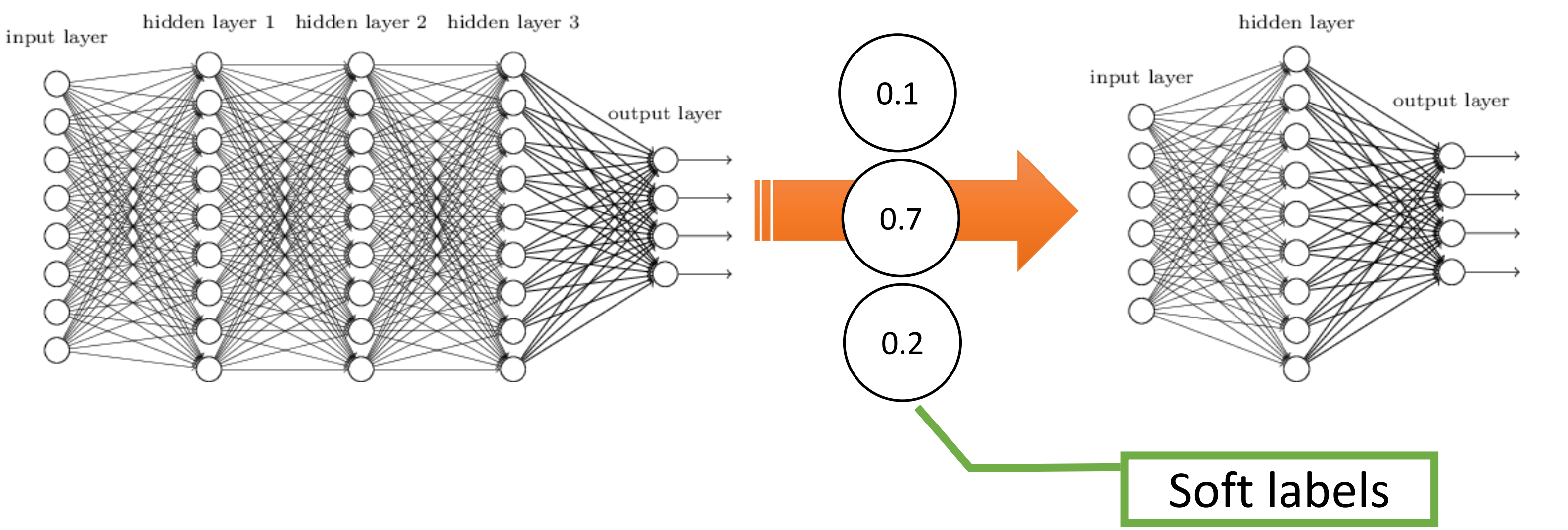



Achieving / Tradeoff
both / between
**interpretation**
and
**performance**

# Mimic learning (knowledge distillation)

[Ba, Caruana, 2014]; [Hinton, et al., 2015]

**Teacher (base) model** **Student (mimic) model**




## Method

- $\sum_{i=1}^{N} \left| y_{soft,i} - F_{mimic}(X_i) \right|^2$

## Explanation

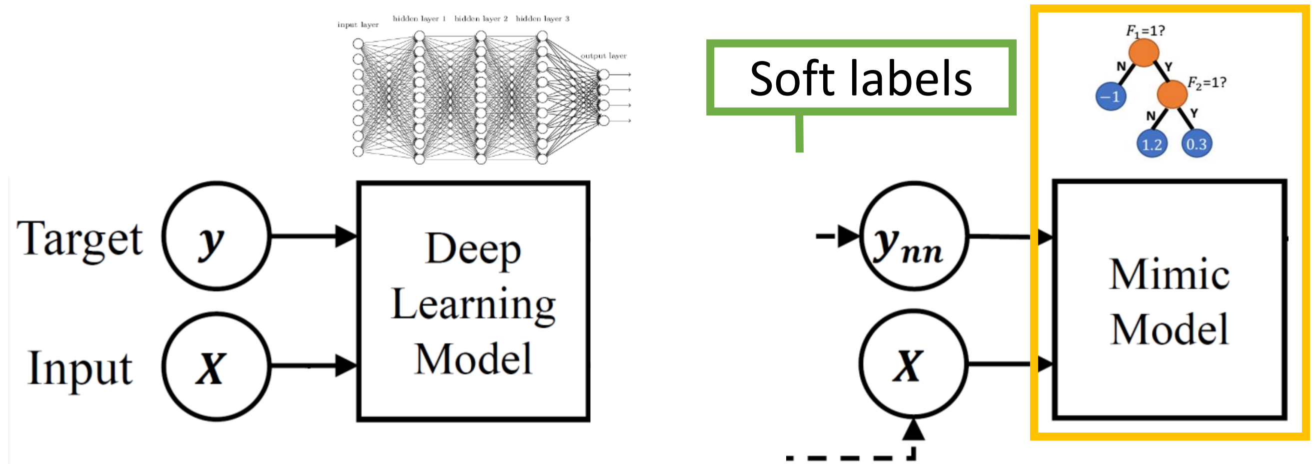
- More information from teacher models, reduced noises, implicit regularizations, etc.

# Interpretable mimic learning framework

## Main idea

Use Gradient Boosting Trees (GBT) to mimic deep learning models.

## Pipeline

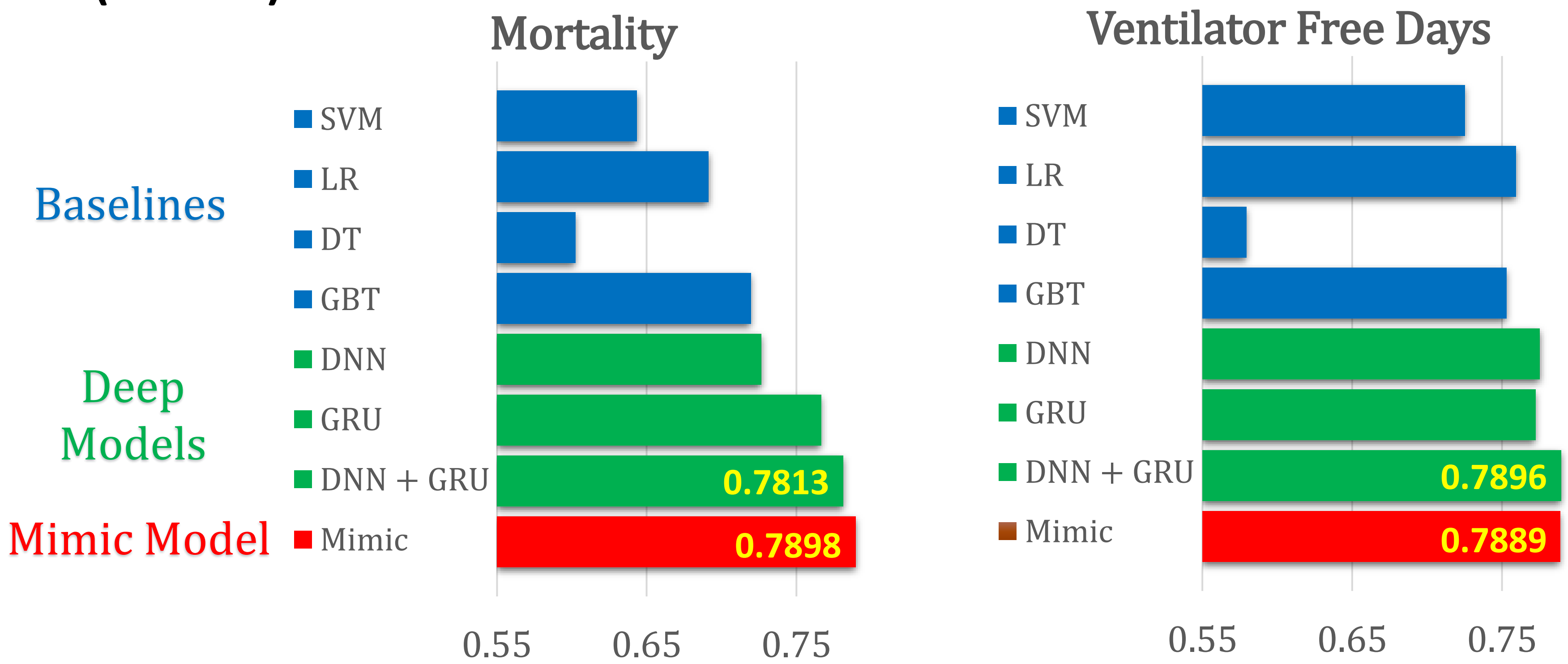



## Benefits

- Good performance
- Less overfitting
- Interpretations

# Quantitative results

**Prediction on patients with acute hypoxemic respiratory failure (AUROC)**



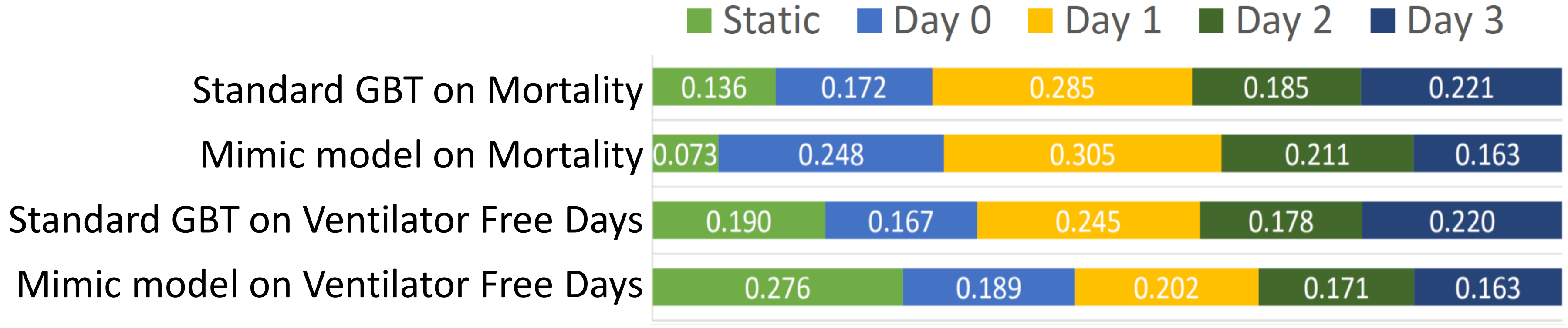

SVM: support vector machine; LR: logistic regression;
DT: decision tree; GBT: gradient boosting tree.
Results are based on 5-fold cross-validation.

# Interpretability: feature importance

## Most important features *(and importance scores)*

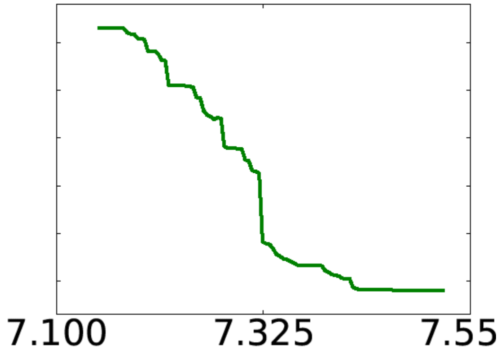
| Task | Mortality | | Ventilator Free Days | |
|---|---|---|---|---|
| **Model** | **Standard GBT** | **Mimic model** | **Standard GBT** | **Mimic model** |
| **Features** | PaO2-Day2 *(0.0539)* | BE-Day0 *(0.0433)* | MAP-Day1 *(0.0423)* | MAP-Day1 *(0.0384)* |
| | MAP-Day1 *(0.0510)* | deltaPF-Day1 *(0.0431)* | PH-Day3 *(0.0354)* | PIM2S *(0.0322)* |
| | BE-Day1 | PH-Day1 | MAP-Day2 | VE-Day0 |
| | FiO2-Day3 | PF-Day0 | MAP-Day3 | VI-Day0 |
| | PF-Day0 | MAP-Day1 | PRISM12 | PaO2-Day0 |

## Feature importance for variables on each day



# Interpretability: feature dependency

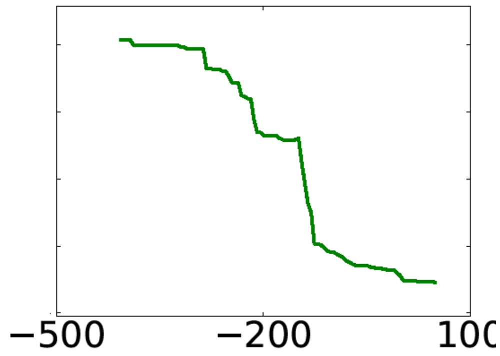
**How features are evaluated for mortality prediction in our model**




x-axis: PH value; y-axis: mortality risk

PH value in blood

- A very narrow normal range around 7.35-7.45




x-axis: delta-PF ratio; y-axis: mortality risk

Change of PaO2/FiO2 ratio

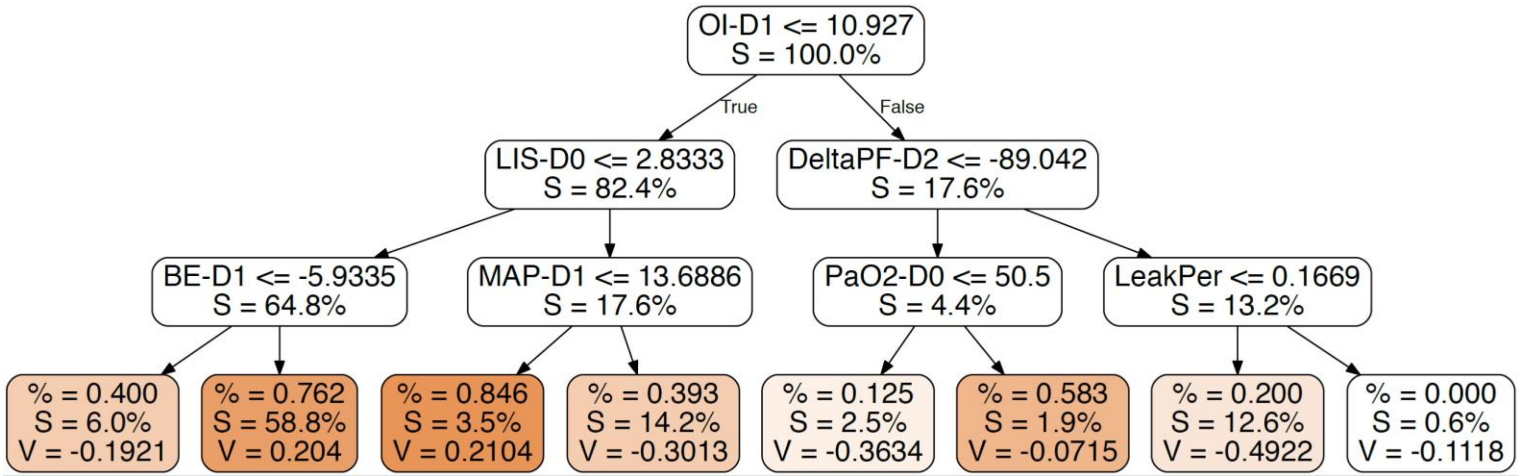
- Normal range: 400-500 mmHg
- < 200: necessary for the diagnosis of respiratory distress syndrome

# Interpretability: decision tree

**Most important trees for ventilator free days prediction**

Ventilator free days

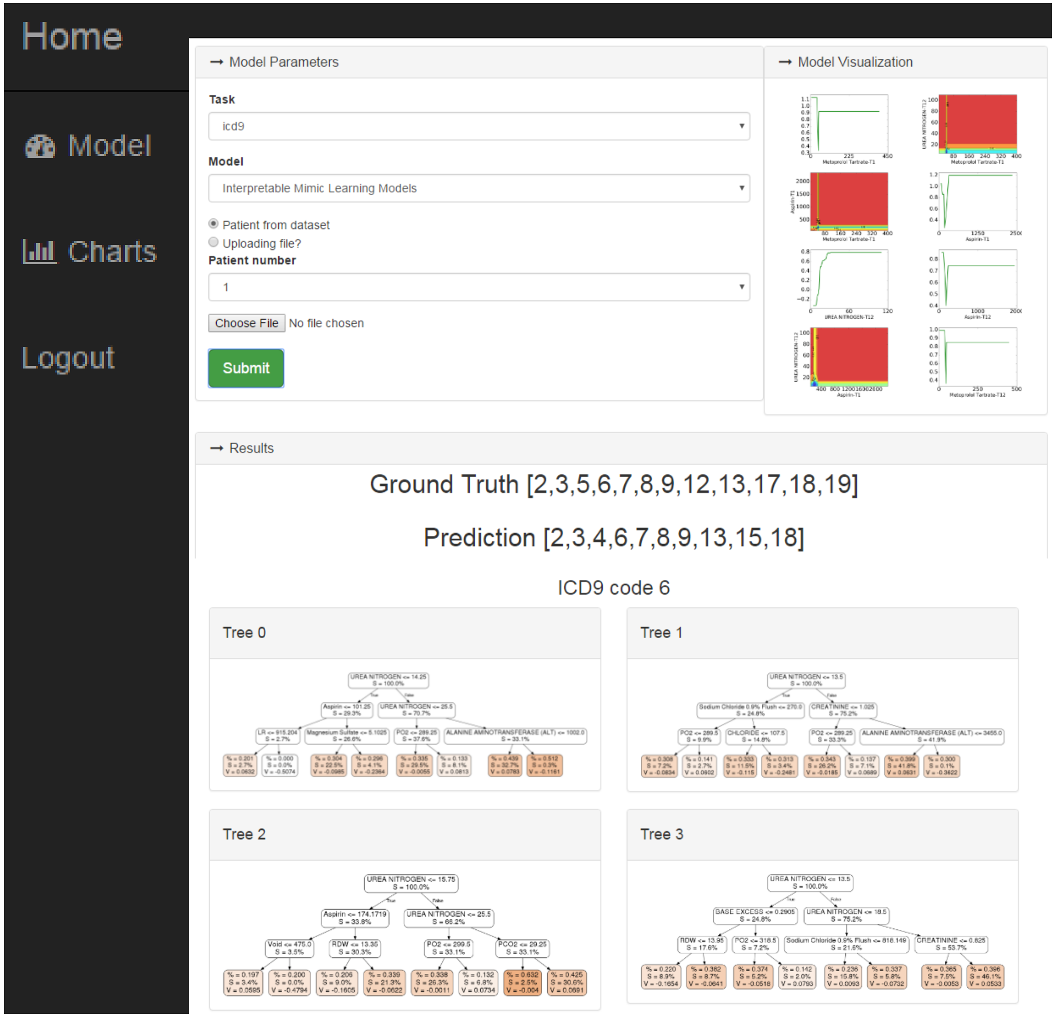
- Lung injury score
- Oxygenation index
- Change of PaO2/FiO2 ratio




% and color: class distribution; S: # of samples; V: prediction value

# Our system under development

- A health care benchmark platform for deep models
- Interpretable deep models and other ML models trained on MIMIC-III dataset
- Apply on public and user-provided EHR datasets
- Provide model interpretation, visualization and results
- Welcome to try and collaborate!



# Conclusion

- Summary
  - Good prediction performance from deep neural networks
  - Interpretability from simple and effective mimic methods
- Future work
  - Further clinical validation and investigation
  - Scalable system for more medical tasks and features
- Contact
  - Zhengping Che zche@usc.edu
  - USC Melady Lab http://www-bcf.usc.edu/~liu32/melady.html

**Thank you!**

# References


**[Lasko et al., 2013]** Lasko, Thomas A., Joshua C. Denny, and Mia A. Levy. "Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data." PloS one 8.6. 2013.

**[Miotto et al., 2016]** Miotto, Riccardo, et al. "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records." Scientific reports 6 (2016).

**[Dabek, Caban, 2015]** Dabek, Filip, and Jesus J. Caban. "A neural network based model for predicting psychological conditions." International Conference on Brain Informatics and Health. Springer International Publishing, 2015.

**[Hammerla et al., 2015]** Hammerla, Nils Yannick, et al. "PD Disease State Assessment in Naturalistic Environments Using Deep Learning." AAAI. 2015.

**[Lipton et al., 2015]** Lipton, Zachary C., et al. "Learning to Diagnose with LSTM Recurrent Neural Networks." arXiv preprint arXiv:1511.03677. 2015.

**[Choi et al., 2015]** Choi, Edward, Mohammad Taha Bahadori, and Jimeng Sun. "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks." arXiv preprint arXiv:1511.0594. 2015.

**[Ba, Caruana, 2014]** Ba, Jimmy, and Rich Caruana. "Do deep nets really need to be deep?." Advances in Neural Information Processing Systems. 2014.

**[Hinton et al., 2015]** Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531. 2015.