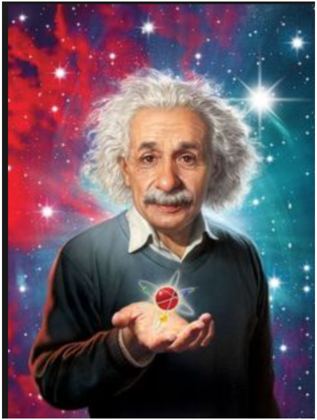
Deep Learning Models for Time Series Data Analysis with Applications to Health Care

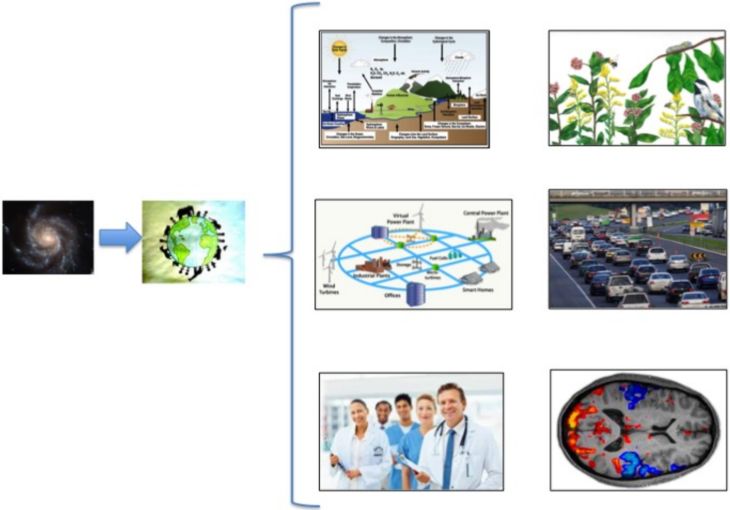
Yan Liu

Computer Science Department
University of Southern California
Email: yanliu@usc.edu



A human being is a part of a whole, called by us "universe", a part limited in time and space.

Large-scale Time Series Data Arise in Many Disciplines



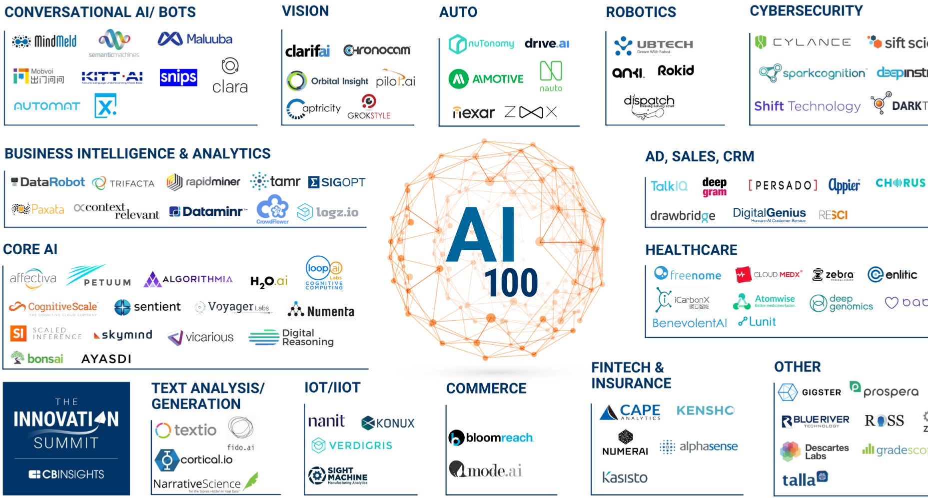
Machine Learning from Large-scale Time Series Observations

Developing scalable and effective solutions by leveraging recent progresses across disciplines

- Temporal dependence discovery [KDD 2007, KDD 2009 (a,b), ISMB 2009, AAAI 2010, SDM 2012, ICML 2012, SDM 2013, KDD 2014, ICML 2015]
- Time series and spatial time series models [ICML 2010, CSB 2010, KDD 2013, NIPS 2014, ICML 2015, ICML 2016, NIPS 2016]
- Time series anomaly detection [SDM 2011, ICDM 2012, KDD 2014]
- Time series representation learning [AMIA workshop 2014, KDD 2015, AMIA 2015, AMIA 2016, ICLR 2017]
- Time series hashing [ICDM 2014]
- Time series clustering [ICML 2015]

Celebration for Tenure

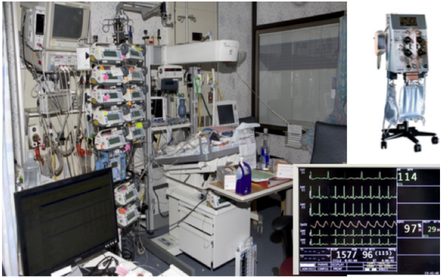
What is NEXT?

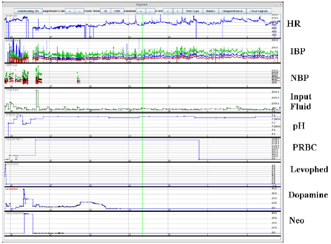

CONVERSATIONAL AI/ BOTS
MindMeld
semanticmachines
Maluuba
Mobvoi 出门问问
KITT·AI
snips
clara
AUTOMAT
x.
VISION
clarifai
Chronocam
Orbital Insight
pilot.ai
aptricity
GROKSTYLE
AUTO
nuTonomy
drive.ai
AIMOTIVE
nauto
nexar
ZOOX
ROBOTICS
UBTECH
anki
Rokid
dispatch
CYBERSECURITY
CYLANCE
sift science
sparkcognition
deepinstinct
Shift Technology
DARKTRACE
BUSINESS INTELLIGENCE & ANALYTICS
DataRobot
TRIFACTA
rapidminer
tamr
SIGOPT
Paxata
context relevant
Dataminr
CrowdFlower
logz.io
AI
100
AD, SALES, CRM
TalkIQ
deep gram
[PERSADO]
Appier
CHORUS
drawbridge
DigitalGenius
RESCI
CORE AI
affectiva
PETUUM
ALGORITHMIA
H2O.ai
loop ai Labs COGNITIVE COMPUTING
CognitiveScale
sentient
Voyager Labs
Numenta
SI SCALED INFERENCE
skymind
vicarious
Digital Reasoning
bonsai
AYASDI
HEALTHCARE
freenome
CLOUD MEDX
zebra
enlitic
iCarbonX
Atomwise
deep genomics
BenevolentAI
Lunit
THE INNOVATION SUMMIT
CBINSIGHTS
TEXT ANALYSIS/ GENERATION
textio
fido.ai
cortical.io
NarrativeScience
IOT/IIOT
nanit
KONUX
VERDIGRIS
SIGHT MACHINE
COMMERCE
bloomreach
mode.ai
FINTECH & INSURANCE
CAPE ANALYTICS
KENSHO
NUMERAI
alphasense
Kasisto
OTHER
GIGSTER
prospera
BLUE RIVER TECHNOLOGY
ROSS
Descartes Labs
gradescope
talla

Time Series in Critical Care Unit (ICU)

Critical care is among the most important areas of medicine.

- >5 million patients admitted to US ICUs annually.[1]
- Cost: \$81.7 billion in US in 2005: 13.4% hospital costs, ∼1% GDP.[1]
- Mortality rates up to 30%, depending on condition, care, age.[1]
- Long-term impact: physical impairment, pain, depression.





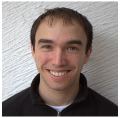
[1]Society of Critical Care Medicine website, Statistics page.

Deep Learning for Smart ICU

Collaborators:

Children's Hospital LOS ANGELES

MAYO CLINIC



David Sontag (MIT)

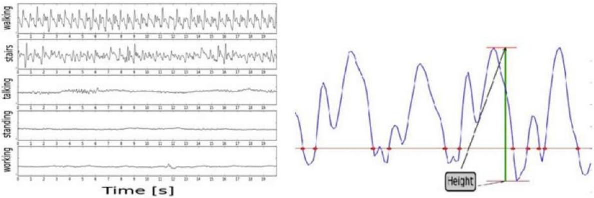
Kyunghyun Cho (NYU)

Tasks:

- Mortality prediction
- Ventilator free days
- Disease code

Deep Learning for Better Care of Diabetes Patients

Wearable devices provide large scale time series data regarding human activities, vital signs, environments, and real-time blood sugar levels.



(a) (b)

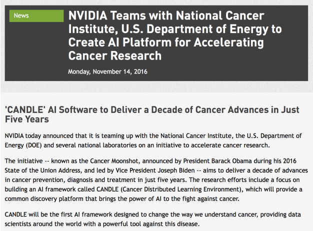
Collaborators: SAMSUNG Keck School of Medicine of USC

Tasks:

- Blood sugar hike prediction
- Intervention strategies

Deep Learning for Cancer Research

Cancer Moonshot projects:



News

NVIDIA Teams with National Cancer Institute, U.S. Department of Energy to Create AI Platform for Accelerating Cancer Research

Monday, November 14, 2016

'CANDLE' AI Software to Deliver a Decade of Cancer Advances in Just Five Years

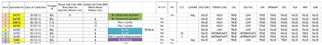
NVIDIA today announced that it is teaming up with the National Cancer Institute, the U.S. Department of Energy (DOE) and several national laboratories on an initiative to accelerate cancer research.

The initiative -- known as the Cancer Moonshot, announced by President Barack Obama during his 2016 State of the Union Address, and led by Vice President Joseph Biden -- aims to deliver a decade of advances in cancer prevention, diagnosis and treatment in just five years. The research efforts include a focus on building an AI framework called CANDLE (Cancer Distributed Learning Environment), which will provide a common discovery platform that brings the power of AI to the fight against cancer.

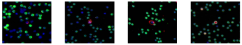
CANDLE will be the first AI framework designed to change the way we understand cancer, providing data scientists around the world with a powerful tool against this disease.

Time series data:

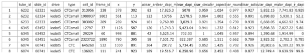
Clinical Data



Cancer cell microscopy data



Features extracted from cancer cell images



Collaborator:

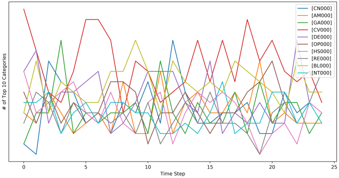
Tasks:

- Overall survival prediction for cancer patients
- Survival prediction after recurrence

Deep Learning for Opioid Addiction and Adverse Effect Analysis

Opioid use study on datasets from the Rochester Epidemiology Project (REP)[2] with more than 140k people

- To extract and understand risk factors and indicators for adverse opioid and opioid-related events
- To predict new opioid users and dependence and recognize misuse on opioid analgesics
- To provide health care providers with better suggestions on pain medication prescriptions



Collaborators:

[2] http://rochesterproject.org/

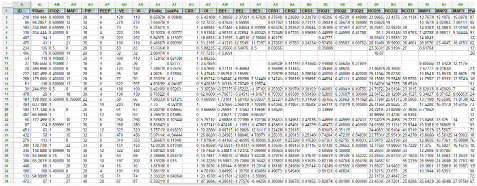
Deep Learning for Smart ICU - Dataset and Tasks

Children's Hospital Los Angeles (CHLA)
398 patients stay > 3 days
Static features (age, weight, etc.): 27 variables
Temporal features (Blood gas, ventilator signals,injury markers, etc.): 21 variables
MIMIC III Dataset
19714 patients stay for 2 days
All temporal features (input fluids, output fluids, lab tests, prescription): 99 variables
PhysioNet Challenge Part of MIMIC II dataset

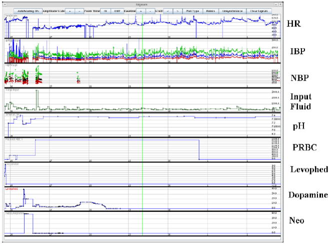
Task Prediction task (mortality, ventilator free days, and disease code), computational phenotyping, anomaly detection

Example of Health Care Data

Example 1:



Example 2:



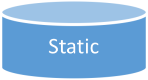
How are health care data different from the data from existing applications of deep learning?

- Privacy, privacy!
- Heterogeneity
- Lots lots of missing data
- Big small data
- *Worst of all: doctors do not believe anything they cannot understand no matter how cool and how deep they are!!*

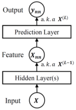
Road Map

- Heterogeneity
  Deep computational phenotyping [SIGKDD 2015, AMIA 2015]
- Missing data
  Gated recurrent neural networks for missing data [aXriv 2016]
- Big small data
  Variational recurrent adversarial deep domain adaptation [ICLR 2017]
- Interpretation
  Interpretable deep models for ICU outcome prediction [AMIA 2016]

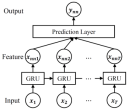
Deep learning model: DNN + GRU



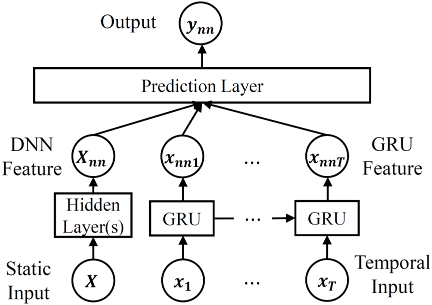
- *Static + (flattened) temporal features*
  - *DNN*



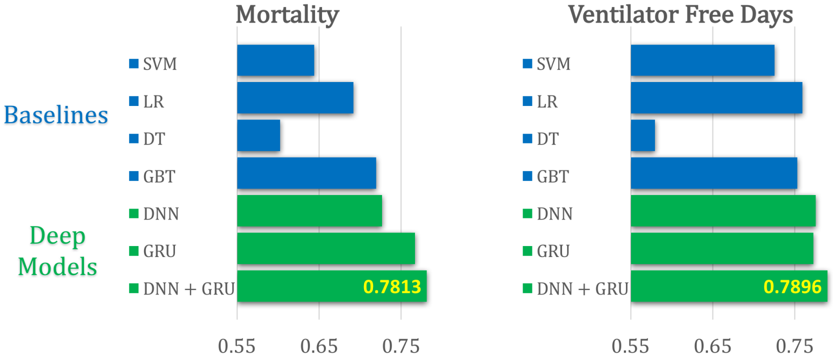
- *Temporal features only*
  - *GRU*



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



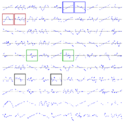
Experiment Results

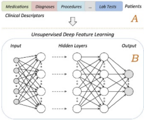


SVM: support vector machine; LR: logistic regression;
DT: decision tree; GBT: gradient boosting tree.
Results are based on 5-fold cross-validation.

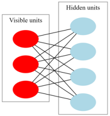
Related Work

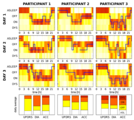
Stacked Auto-encoder (SDA)
Computational phenotyping [Lasko et al., 2013, Miotto et al., 2016]



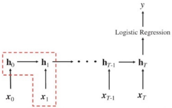


Deep neural networks (DNNs)
Restricted Boltzmann machine (RBM)
Multi-layer perceptron (MLP)
Condition prediction [Dabek, Caban, 2015; Hammerla et al., 2015]





Recurrent neural networks (RNNs)
Long short-term memory (LSTM) Gated recurrent unit (GRU)
Diagnosis/event prediction [Lipton et al., 2015; Choi et al., 2015]



Road Map

- Heterogeneity
  Deep computational phenotyping [SIGKDD 2015, AMIA 2015]
- Missing data
  Gated recurrent neural networks for missing data [aXriv 2016]
- Big small data
  Variational recurrent adversarial deep domain adaptation [ICLR 2017]
- Interpretation
  Interpretable deep models for ICU outcome prediction [AMIA 2016]

Motivation

Limited amount of data across age groups

- Studies have shown age is a factor for survival in a medical ICU [Critical Care Med. 1983]
- Pediatricians catch phrase - Children are not little adults.
  - However, medical care for children is based on adults [American Journal of Respiratory and Critical Care Medicine, 2010]

Target	Model Trained on Adult	Model trained on Children
Children	0.56	0.70

- Training models for each age group not ideal
  - Small target dataset
  - Difficult to get labels

Question: How do we adapt models from Adults (source domain) to Children (target domain)?

Problem Formulation

Problem: unsupervised domain adaptation for multivariate time series

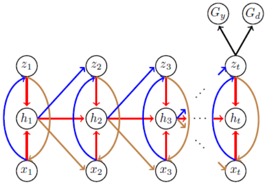
Case study: acute hypoxemic respiratory failure

Our Solution:

Deep learning model with Adversarial training and Variational methods

Domain invariant representation while transferring temporal dependencies

Variational Adversarial Deep Domain Adaptation (VADDA) [ICLR 2017]



VRNN Objective Function

$$\mathcal{L}_r(x_t^i;\theta_e,\theta_g) = E_{q_{\theta_e}(z^i_{\le T^i}|x^i_{\le T^i})}\sum_{t=1}^{T^i}(-D(q_{\theta_e}(z_t^i|x^i_{\le t},z^i_{<t})||p(z_t^i|x^i_{<t},z^i_{<t}))+\log p_{\theta_g}(x_t^i|z^i_{\le t},x^i_{<t}))$$

Source Classification Loss with regularizer

$$\min_{\theta_e,\theta_g,\theta_y}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^i}\mathcal{L}_r(\mathbf{x^i};\theta_e,\theta_g)+\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_y(\mathbf{x^i};\theta_y,\theta_e)+\lambda\mathcal{R}(\theta_e)$$

Domain Regularizer

$$\mathcal{R}(\theta_e) = \max_{\theta_d}\left[-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_d(\mathbf{x^i};\theta_d,\theta_e)-\frac{1}{n'}\sum_{i=n+1}^{N}\mathcal{L}_d(\mathbf{x^i};\theta_d,\theta_e)\right]$$

Overall Objective Function

$$E(\theta_e,\theta_g,\theta_y,\theta_d) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^i}\mathcal{L}_r(\mathbf{x^i};\theta_e,\theta_g)+\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_y(\mathbf{x^i};\theta_y)-\lambda(\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_d(\mathbf{x^i};\theta_d)+\frac{1}{n'}\sum_{i=n+1}^{N}\mathcal{L}_d(\mathbf{x^i};\theta_d)))$$

Experiments

Case Study: Acute Hypoxemic Respiratory Failure

- Datasets
  - Pediatric ICU: Child-AHRF
    - 398 patients at Children's Hospital Los Angeles (CHLA) Group 1: children (0-19 yrs)
  - MIMIC-III : Adult-AHRF
    - 5527 patients Group 2: working-age adult (20 to 45 yrs); Group 3: old working-age adult (46 to 65 yrs, Group 4: elderly (66 to 85 yrs); Group 5: old elderly (> 85 yrs)
- Data Temporal variables - 21 (Blood gas, ventilator signals, injury markers, etc.) for 4 days
- Prediction tasks - Mortality label
- Comparison
  - Non-domain adaptation: Logistic regression, Adaboost, Deep Neural Networks
  - Deep Domain adaptation: DANN (JMLR 2016), R-DANN, VFAE (ICLR 2016)

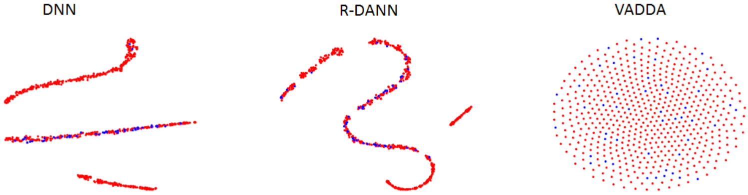
Preliminary Results

AUC Comparison for AHRF Mortality Prediction task with and without Domain Adaptation

Source-Target	LR	Adaboost	DNN	DANN	VFAE	R-DANN	VRDDA
3- 2	0.555	**0.562**	0.569	0.572	0.615	0.603	**0.654**
4- 2	0.624	**0.645**	0.569	0.589	0.635	0.584	**0.656**
5- 2	0.527	**0.554**	0.551	0.540	0.588	0.611	**0.616**
2- 3	**0.627**	0.621	0.550	0.563	0.585	0.708	**0.724**
4- 3	**0.681**	0.636	0.542	0.527	0.722	**0.821**	0.770
5- 3	0.655	**0.706**	0.503	0.518	0.608	0.769	**0.782**
2- 4	0.585	**0.591**	0.530	0.560	0.582	0.716	**0.777**
3- 4	**0.652**	0.629	0.531	0.527	0.697	**0.769**	0.764
5- 4	0.689	**0.699**	0.538	0.532	0.614	0.728	**0.738**
2- 5	**0.565**	0.543	0.549	0.526	0.555	0.659	**0.719**
3- 5	0.576	**0.587**	0.510	0.526	0.533	0.630	**0.721**
4- 5	**0.682**	0.587	0.575	0.548	0.712	0.747	**0.775**
5- 1	0.502	**0.573**	0.557	0.563	0.618	0.563	**0.639**
4- 1	**0.565**	0.533	0.572	0.542	**0.668**	0.577	0.636
3- 1	0.500	0.500	0.542	0.535	0.570	0.591	**0.631**
2- 1	**0.520**	0.500	0.534	0.559	0.578	0.630	**0.637**

VADDA mostly outperforms all domain adaptation and non-domain adaptation models

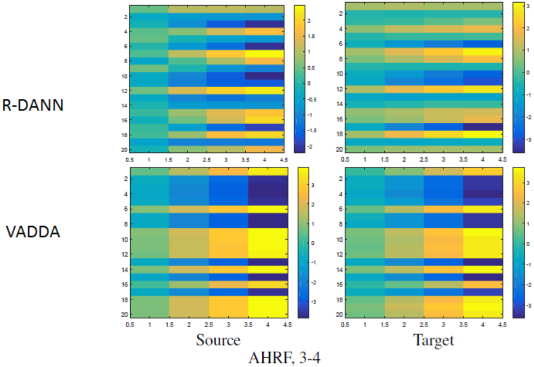
Domain-invariant representations



t-SNE projections for the latent representations for domain adaptation from Adult-AHRF to Child-AHRF

VADDA has better distribution mixing than DANN

Temporal dependencies Visualization



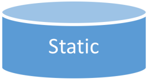
Memory cell state neuron activations of the R-DANN and VADDA

Activation patterns of VADDA are more consistent across time-steps than for R-DANN

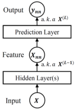
Road Map

- Heterogeneity
  Deep computational phenotyping [SIGKDD 2015, AMIA 2015]
- Missing data
  Gated recurrent neural networks for missing data [aXriv 2016]
- Big small data
  Variational recurrent adversarial deep domain adaptation [ICLR 2017]
- Interpretation
  Interpretable deep models for ICU outcome prediction [AMIA 2016]

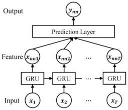
Deep learning model: DNN + GRU



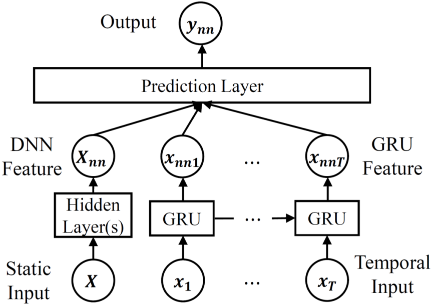
- *Static + (flattened) temporal features*
  - *DNN*



- *Temporal features only*
  - *GRU*

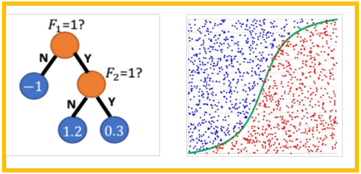


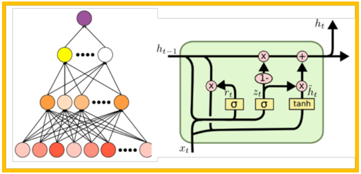
- Static + temporal features
  - **DNN + GRU** (combination)



Interpretable Model is Necessary

Interpretable predictive models are shown to result in faster adoptability among clinical staff and better quality of patient care.





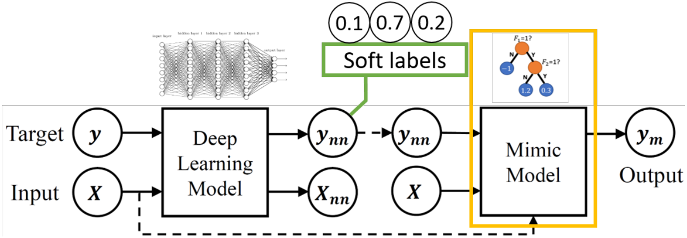
- Simple and commonly use models
- Easy to interpret, mediocre performance

- Deep learning solutions
- Superior performance, hard to explain

Can we learn interpretable models with robust prediction performance?

Interpretable Mimic Learning Framework

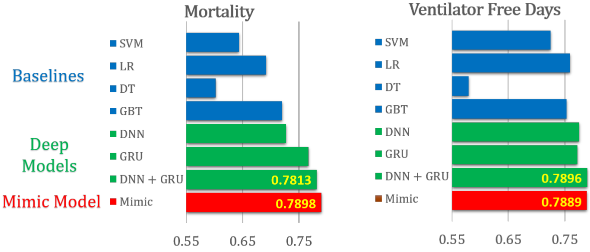
- Main ideas:
    - Borrow the ideas from knowledge distillation [Hinton, et al., 2015] and mimic learning [Ba, Caruana, 2014].
    - Use **Gradient Boosting Trees (GBTs)** to mimic deep learning models.
- Training Pipeline:



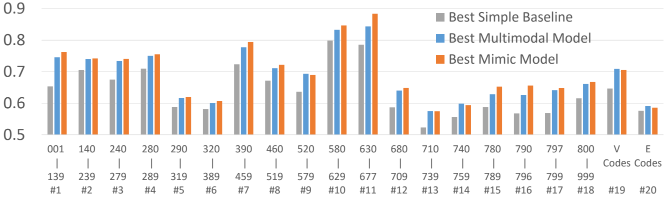
- Benefits: Good performance, less overfitting, interpretations.

Quantitative Evaluation

AUROC score of prediction on patients with acute hypoxemic respiratory failure.

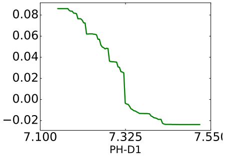


AUROC score of 20 ICD-9 diagnosis category prediction tasks on MIMIC-III dataset.



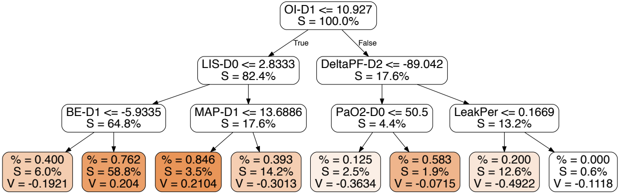
Model/Feature Interpretation

Partial dependency plot for mortality prediction on patients with acute hypoxemic respiratory failure.



- pH value in blood should stay in a normal range around 7.35-7.45.
- Our model predicts a higher mortality change when the patient pH value below 7.325.

Most Useful Decision Trees for ventilator free days prediction.



Useful features:

- Lung injury score
- Oxygenation index
- PF ratio change

AI for Health Care - in Hollywood Movie

AI for Health Care - in Practical World

Thank You!

Questions and Comments?