



PetuumMed: algorithms and system for EHR- based medical decision support

Eric Xing
Petuum Inc.
Carnegie Mellon University

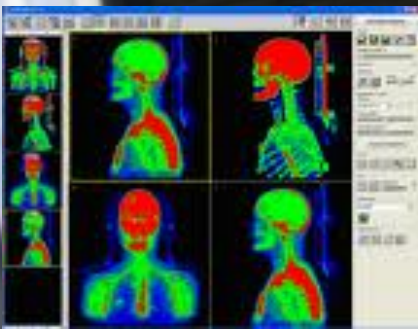
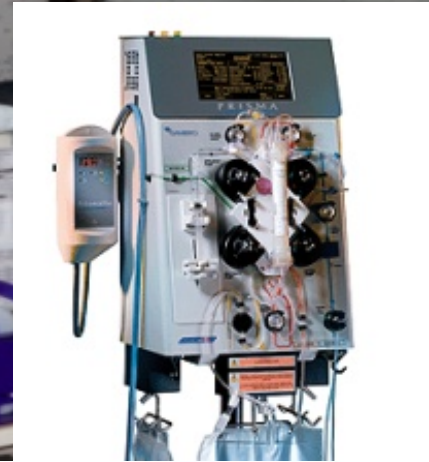
These Data Are ...

- Time Constrained
- Mission Critical
- Complex
- Teamwork
- Expensive

The Data Deluge



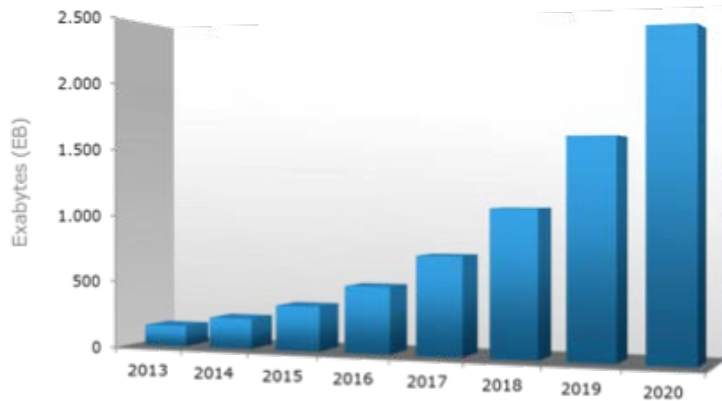
This is where evidence and information start





Data Deluge in Healthcare

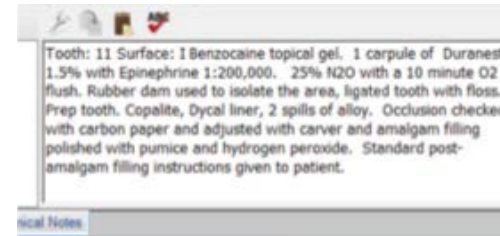
Volume



- 153 exabytes (one exabyte = one billion gigabytes) were produced in 2013
- An estimated 2,314 exabytes will be produced in 2020
- An overall rate of increase is at least 48 percent annually

Complexity

Notes



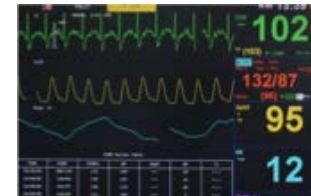
Image



Lab values

Test Name	Result	Normal Range	Units
Hemoglobin	12	12.0 - 16.0	g/dL
HCT	35	37.0-47.0	%
MCV	88	82-100	fL
MCH	89	27-31	pg
MCHC	89	31.9-36.0	g/dL
RBCWV	12	11.5-14.5	%
RBCWV2	88	85-98	%
WBC	8.7	4.5-11	10 ⁹ /dL
NEUT	88	40-70	%
LYMPH	88	20-40	%
MONO	8	1-10	%
EOSIN	2	0-5	%
PLT	296	150-450	10 ⁹ /dL
ESR	2	0-20	mm/hr

Vital signs



Test



Genomics



Billing



Literature

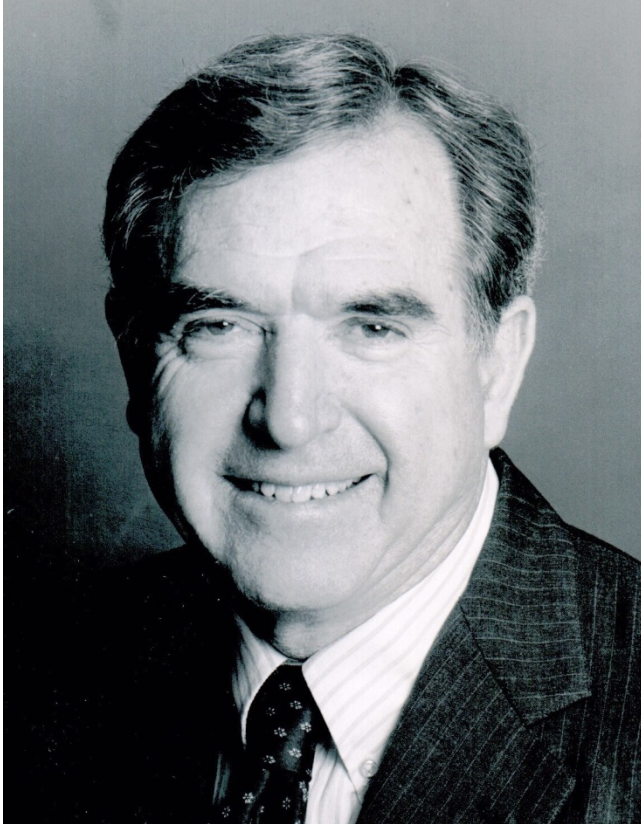


Social media



A conceptual image featuring a man in a dark suit and tie standing in a small, dark boat. The boat is floating on a vast, undulating sea of binary code (0s and 1s). The background consists of a blue sky with scattered white clouds. The overall scene is a metaphor for navigating through a large volume of data.

**In this sea of data, what
should a doctor do ?**



“Many people think that doctors make their recommendations from a basis of scientific certainty, that the facts are very clear and there’s only one way to diagnose or treat an illness. In reality, that’s not always the case. Many things are a matter of conjecture, tradition, convenience, habit.”

Arnold Relman (1923-2014)

Former Editor-in-Chief, New England Journal of Medicine



Pain Points in Healthcare

Quality

- 250,000 Americans die each year from medical errors (the third leading cause of death in the US).
- 12 million Americans are misdiagnosed each year.
- Preventable medication errors impact more than 7 million patients and cost almost \$21 billion annually.
- 15 to 25 percent of patients are readmitted within 30 days and readmissions are costly (e.g., \$41.3 billion in 2011).

Efficiency

- Patients wait on average 6 hours in emergency rooms. Nearly 400,000 patients wait 24 hours or more.
- Physicians spend only 27 percent of their office day on direct clinical face time with patients.
- The U.S. healthcare system wastes \$750 billion annually due to unnecessary services, inefficient care delivery, excess administrative costs, etc.



Machine Learning for Healthcare

Clinical Data

Machine Learning

Actionable Insights

Note
S

Image
S

Lab values

Vital signs
102
95
12

Test
S

Genomics

Billing
HEALTH INSURANCE

Literature
PubMed
National Library of Medicine

Social media
patientslikeme
Live better, together.



Components

Patterns

$\alpha \rightarrow \theta \rightarrow z \rightarrow x_N \rightarrow \beta_K$

Input image → Convolution (feature maps) → Mapping → Fully connected layer

$\theta_i \rightarrow e_{ij} \rightarrow \beta_{kl}$

Network diagram showing layers of nodes and connections.

Visualizations of patterns: a network graph and a cluster of colored nodes.



Extracted information

Recommended diagnosis and treatment

Suggested ICD codes

Detected lung nodule

Predicted mortality rate

Detected arrhythmia

History: aortic stenosis(4), atrial fibrillation(3), aortic valve replacement(2), mitral regurgitation(2), mitral regurgitation(2), coronary artery bypass grafting(1), cardiac catheterization(1), heart failure with reduced ejection fraction(1), vent(1), ar(1), mitral valve prolapse(1), congestive heart failure(1), coronary artery disease(1)

Comorbidities: diabetes mellitus(3), pvd(2), hypertension(1), pulmonary hypertension(1)

Symptoms: edema(2), pleural effusion(1), swn(1)

Diagnosis
42% Pneumonia

Medications
92% Levofloxacin
90% Penicillin Vk
88% Vancomycin

94% I50.9 heart failure, unspecified
88% J18.9 pneumonia, unspecified organism
67% N17.9 acute kidney failure, unspecified

24 Hours Mortality: 89%

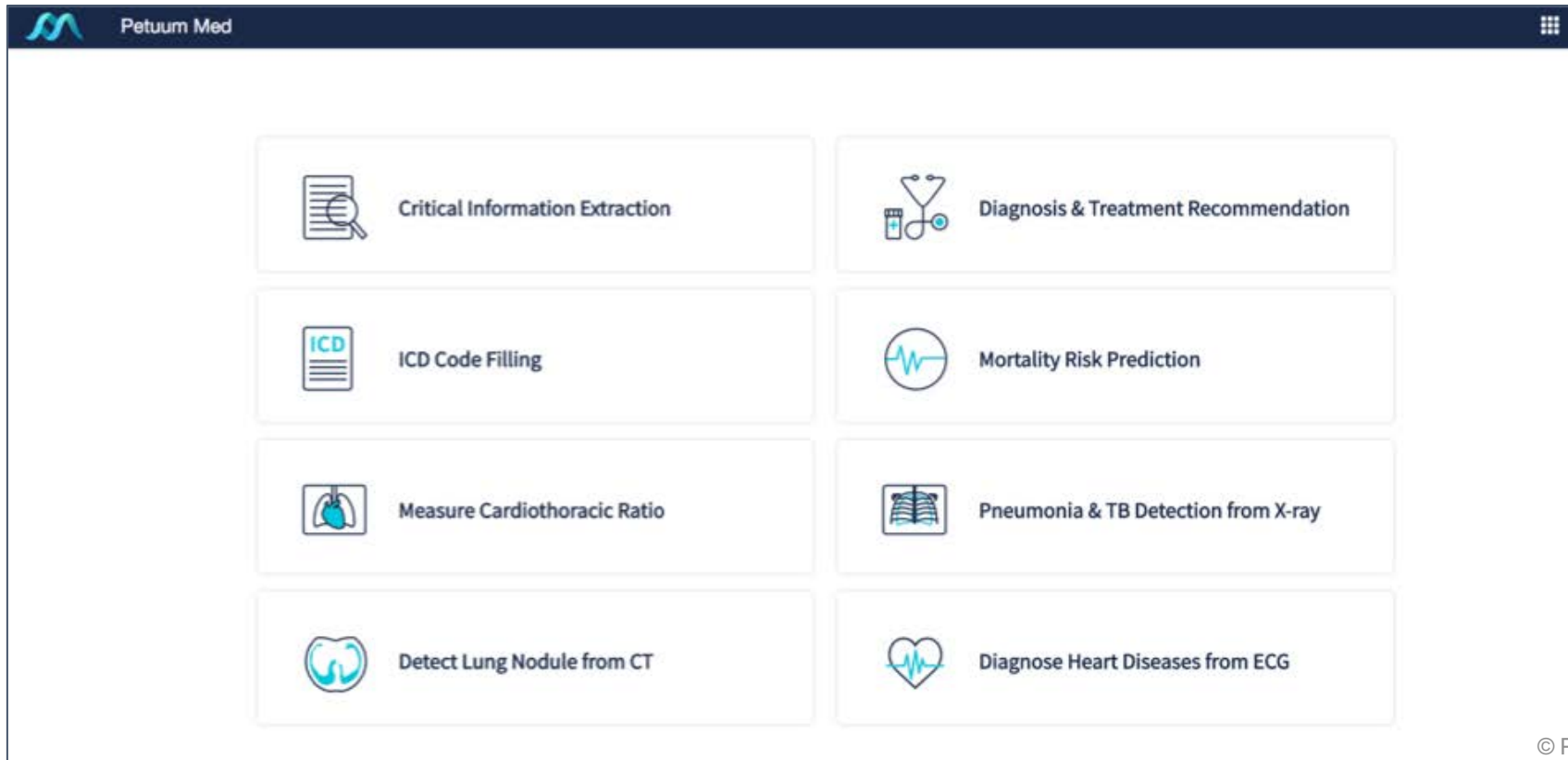
48 Hours Mortality: 37%

Right Bundle Branch Block



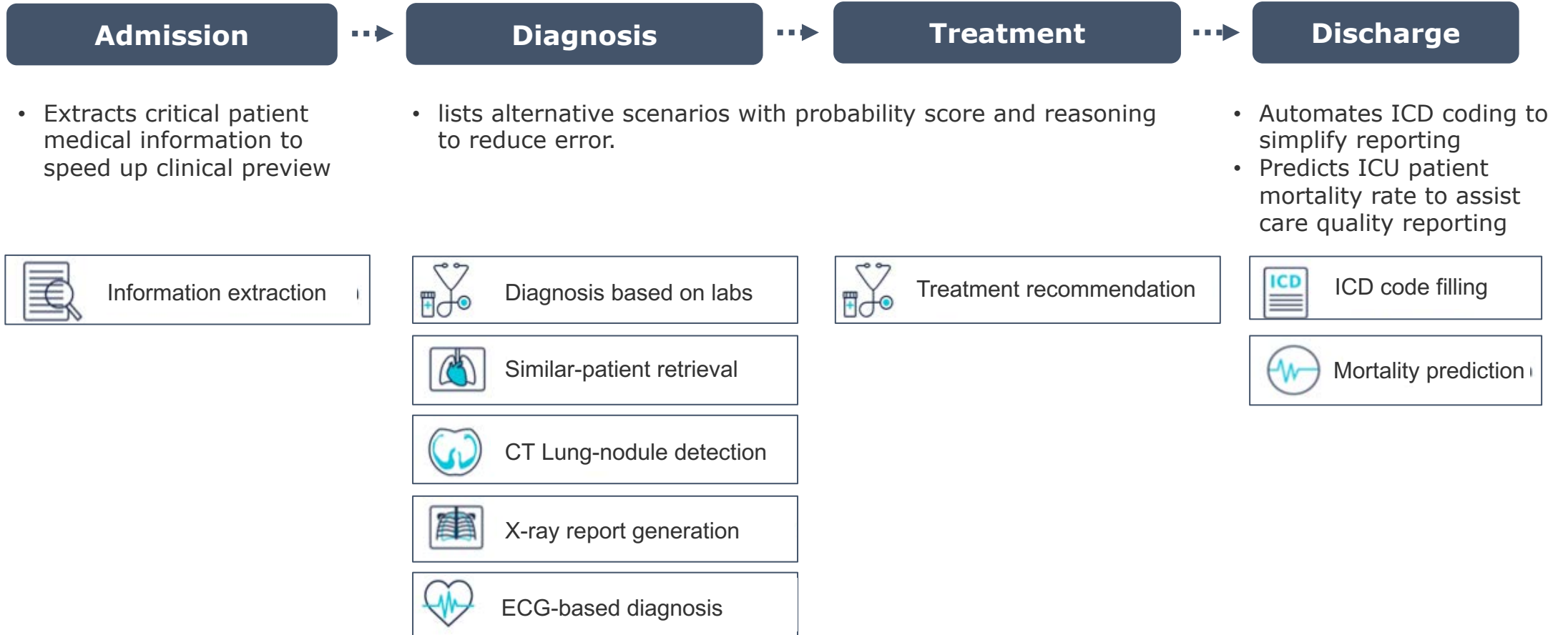
Petuum Healthcare Solutions

- **PetuumMed** is a clinical decision support platform powered by artificial intelligence and machine learning (AI/ML). It distills insights from massive and heterogeneous patient data, and empowers medical professionals make accurate and efficient decisions within the clinical flow





Supports clinical decisions at all points of care





Supports clinical decisions at all points of care



Admission

Diagnosis

Treatment

Discharge



- Extracts critical patient medical information to speed up clinical preview

- lists alternative scenarios with probability score and reasoning to reduce error.

- Automates ICD coding to simplify reporting
- Predicts ICU patient mortality rate to assist care quality reporting



Information extraction



- Automatic construction of disease-specific knowledge graph
- Medical synonym matching
- Named entity extraction
- Relation extraction



Diagnosis based on labs



Similar-patient retrieval



CT Lung-nodule detection



X-ray report generation



ECG-based diagnosis



Treatment recommendation



ICD code filling



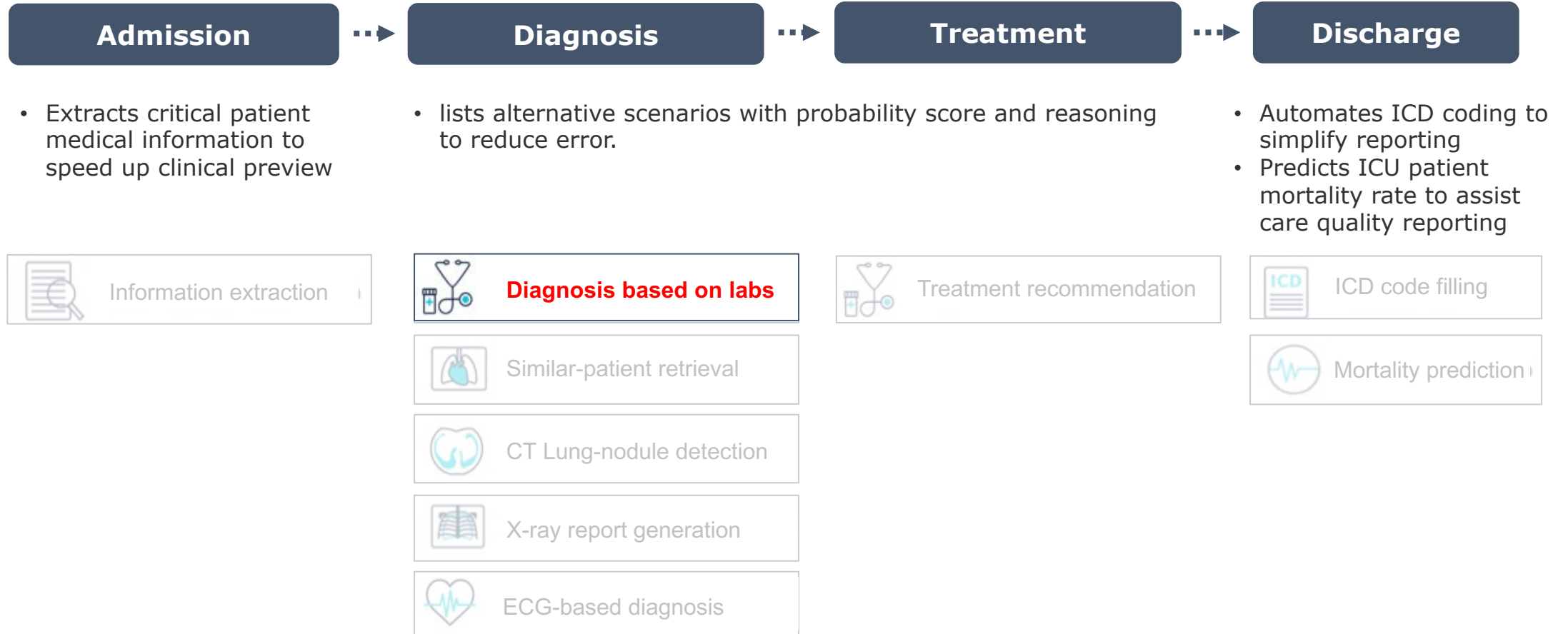
Mortality prediction



- Attentional matching
- Hierarchical classification
- Domain adaptation
- Isotonic constraints



Supports clinical decisions at all points of care





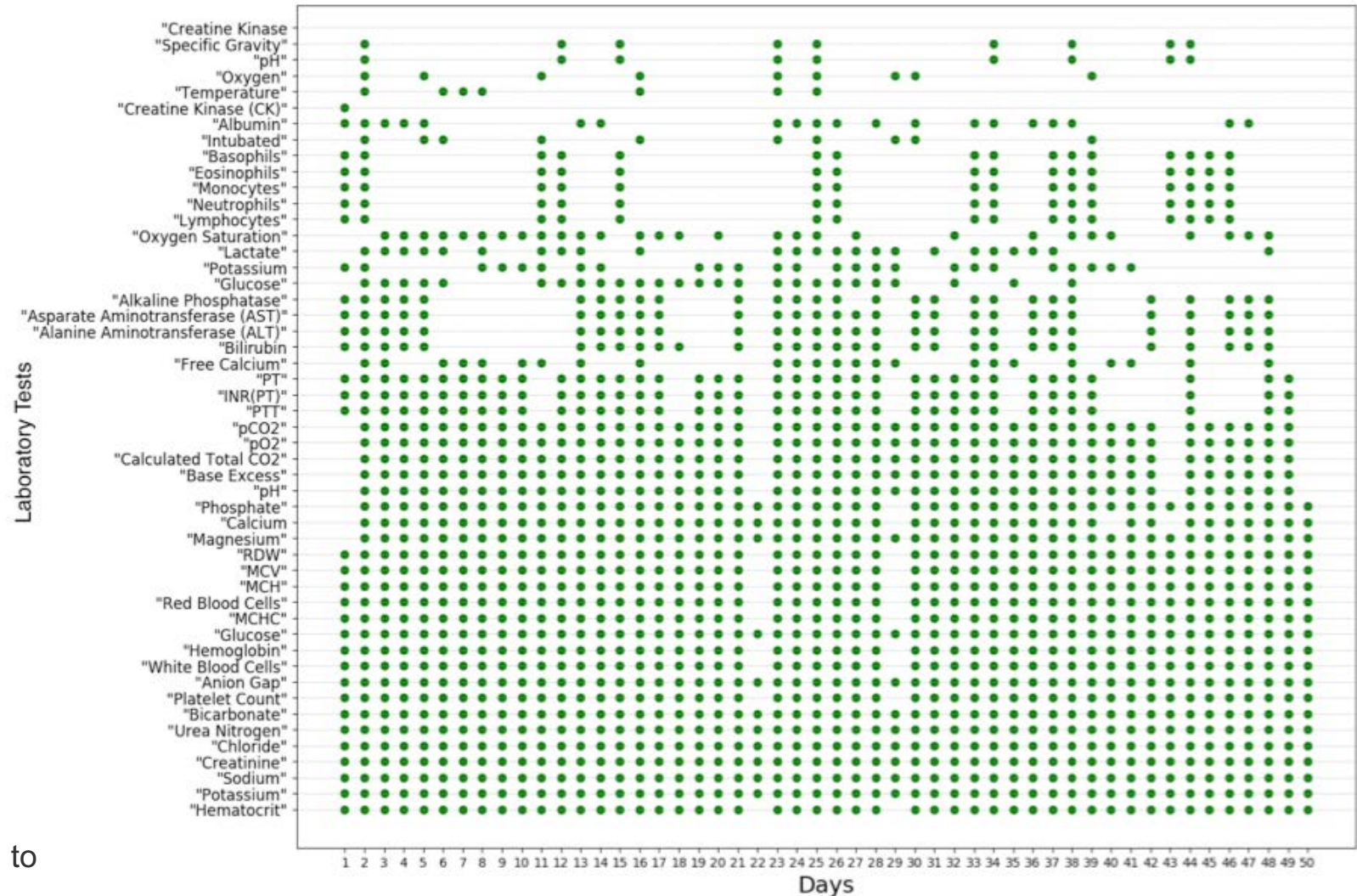
Background

Challenges:

- Missing values
- Temporal and multivariate structure

Problems of existing approaches

- Missing-value imputation methods (e.g., average filling, forward filling) are:
 - Linear, hence are less-expressive;
 - Heuristic, lacking a principled foundation.
- Missing-value imputation, time-series representation learning, and diagnosis prediction are performed separately, failing to take their inter-relations into account. For example, the imputed missing values may not be specifically good for the prediction task.

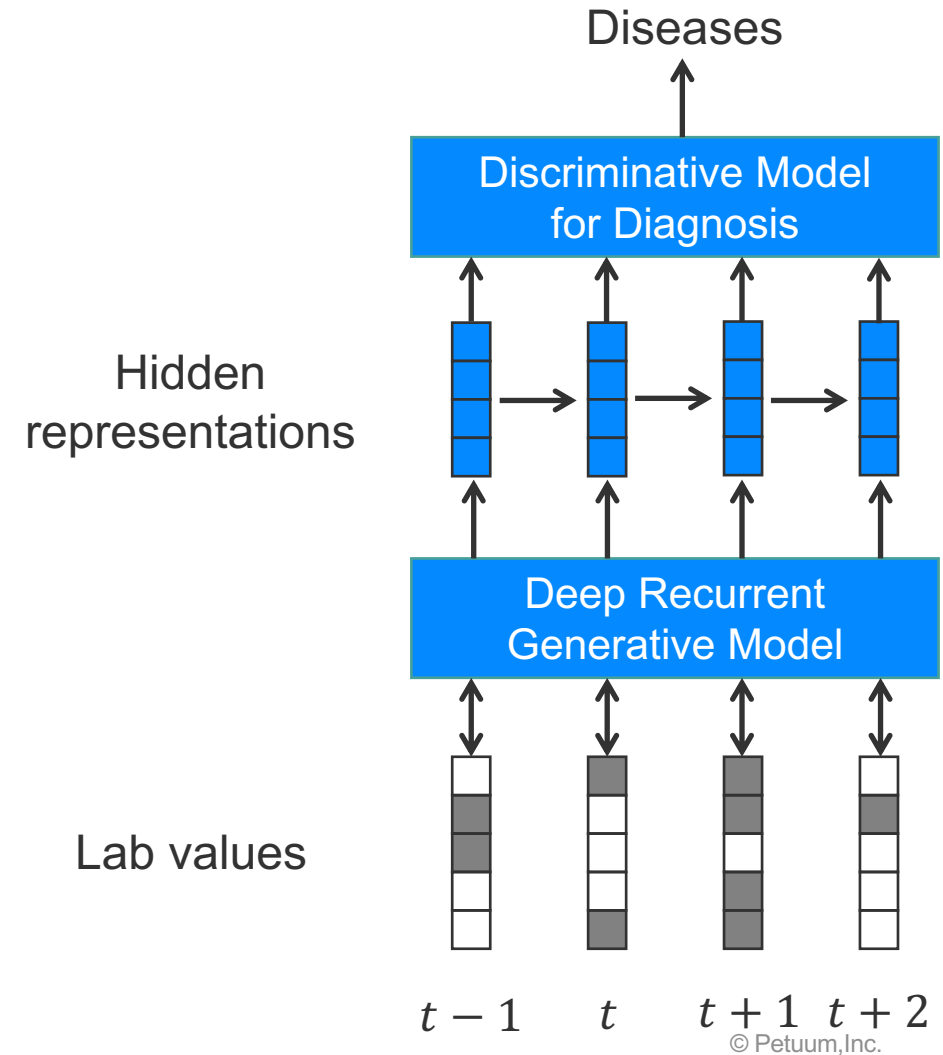


An example of a patient's laboratory test records. The green dot means there is a value, otherwise it is missing.



Diagnosis Based on Lab Values

- Solutions
 - Imputing missing values based on deep generative models. The approach is “indirectly supervised” and is able to capture nonlinear patterns in missing values.
 - Capturing time-series structure using recurrent neural networks.
 - Generative modeling and discriminative prediction (for diagnosis) are performed jointly, which is able to explore the correlations among sub-tasks. Missing-value-imputation is tailored to diagnosis prediction.






Imputation as Generation: a Probabilistic Approach

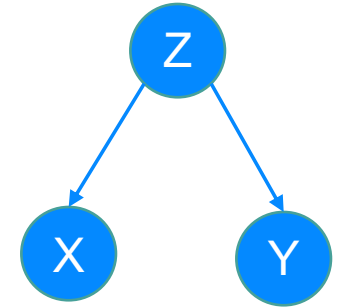
- Given the observed values \mathcal{X} , impute the missing values \mathcal{Y} by inferring $p(\mathcal{Y}|\mathcal{X})$

$$\begin{aligned}
 & p(\mathcal{Y}|\mathcal{X}) \\
 &= \int_{\mathbf{z}} p(\mathcal{Y}, \mathbf{z}|\mathcal{X}) \\
 &= \int_{\mathbf{z}} \frac{p(\mathcal{X}, \mathcal{Y}, \mathbf{z})}{p(\mathcal{X})} \\
 &= \int_{\mathbf{z}} \frac{p(\mathcal{X}, \mathcal{Y}|\mathbf{z})p(\mathbf{z})}{p(\mathcal{X})} \\
 &= \int_{\mathbf{z}} \frac{p(\mathcal{Y}|\mathbf{z})p(\mathcal{X}|\mathbf{z})p(\mathbf{z})}{p(\mathcal{X})} \\
 &= \int_{\mathbf{z}} \frac{p(\mathcal{Y}|\mathbf{z})p(\mathcal{X}, \mathbf{z})}{p(\mathcal{X})} \\
 &= \int_{\mathbf{z}} p(\mathcal{Y}|\mathbf{z})p(\mathbf{z}|\mathcal{X}) \\
 &= \mathbb{E}_{p(\mathbf{z}|\mathcal{X})}[p(\mathcal{Y}|\mathbf{z})] \\
 &\approx p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])
 \end{aligned}$$

Fixed-point Iteration



$$\begin{aligned}
 & p(\mathbf{z}|\mathcal{X}) \\
 &= \int_{\mathcal{Y}} p(\mathbf{z}, \mathcal{Y}|\mathcal{X}) \\
 &= \int_{\mathcal{Y}} \frac{p(\mathbf{z}, \mathcal{Y}, \mathcal{X})}{p(\mathcal{X})} \\
 &= \int_{\mathcal{Y}} \frac{p(\mathbf{z}|\mathcal{X}, \mathcal{Y})p(\mathcal{X}, \mathcal{Y})}{p(\mathcal{X})} \\
 &= \int_{\mathcal{Y}} p(\mathbf{z}|\mathcal{X}, \mathcal{Y})p(\mathcal{Y}|\mathcal{X}) \\
 &= \mathbb{E}_{p(\mathcal{Y}|\mathcal{X})}[p(\mathbf{z}|\mathcal{X}, \mathcal{Y})] \\
 &\approx p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])
 \end{aligned}$$



where z is a latent variable



Imputation as Generation: a Probabilistic Approach (Cont'd)

- Learning the parameters in $p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])$ and $p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])$ by maximizing the likelihood of observed data

$$\begin{aligned}
 & \log p(\mathcal{X}) \\
 &= \mathbb{E}_{p(\mathcal{Y}|\mathcal{X})}[\log p(\mathcal{X}, \mathcal{Y})] - \mathbb{E}_{p(\mathcal{Y}|\mathcal{X})}[\log p(\mathcal{Y}|\mathcal{X})] \\
 &\approx \log p(\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}]) - \mathbb{E}_{p(\mathcal{Y}|\mathcal{X})}[\log p(\mathcal{Y}|\mathcal{X})] \\
 &= \mathbb{E}_{p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])}[\log p(\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}], \mathbf{z})] - \mathbb{E}_{p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])}[\log p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])] - \mathbb{E}_{p(\mathcal{Y}|\mathcal{X})}[\log p(\mathcal{Y}|\mathcal{X})] \\
 &= \mathbb{E}_{p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])}[\log p(\mathcal{X}|\mathbf{z})p(\mathbb{E}[\mathcal{Y}|\mathcal{X}]|\mathbf{z})p(\mathbf{z})] - \mathbb{E}_{p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])}[\log p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])] - \mathbb{E}_{p(\mathcal{Y}|\mathcal{X})}[\log p(\mathcal{Y}|\mathcal{X})]
 \end{aligned}$$

Share the same distribution as $p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])$

- Iteratively perform learning and inference

Fixing the parameters in $p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])$ and $p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])$ infer $p(\mathcal{Y}|\mathcal{X})$

$$p(\mathcal{Y}|\mathcal{X}) \approx p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])$$

$$p(\mathbf{z}|\mathcal{X}) \approx p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])$$

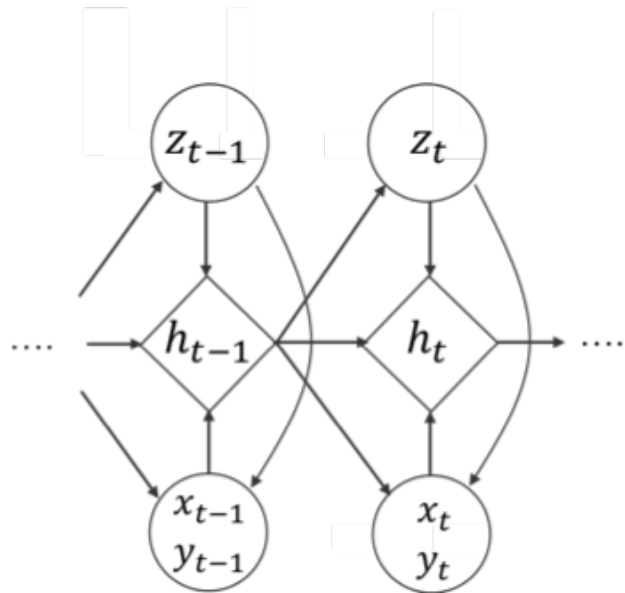

Fixing $p(\mathcal{Y}|\mathcal{X})$ learn the parameters in $p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])$ and $p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])$ by maximizing

$$\begin{aligned}
 & \mathbb{E}_{p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])}[\log p(\mathcal{X}|\mathbf{z})p(\mathbb{E}[\mathcal{Y}|\mathcal{X}]|\mathbf{z})p(\mathbf{z})] \\
 & - \mathbb{E}_{p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])}[\log p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])]
 \end{aligned}$$



Imputation as Generation: a Probabilistic Approach (Cont'd)

- Parameterize $p(\mathcal{Y}|\mathbb{E}[\mathbf{z}|\mathcal{X}])$ and $p(\mathbf{z}|\mathcal{X}, \mathbb{E}[\mathcal{Y}|\mathcal{X}])$ for time-series missing-value data
- Recurrent neural network with latent variables (Chung et al., 2015)



$$p(z_t|h_{t-1}) = \mathcal{N}(\xi_t, \text{diag}(\zeta_t^2)), \text{ where } [\xi_t, \zeta_t] = \phi(h_{t-1}).$$

$$p(x_t, y_t|z_t, h_{t-1}) = \mathcal{N}(\mu_t, \text{diag}(\sigma_t^2)), \text{ where } [\mu_t, \sigma_t] = \varphi(z_t, h_{t-1})$$

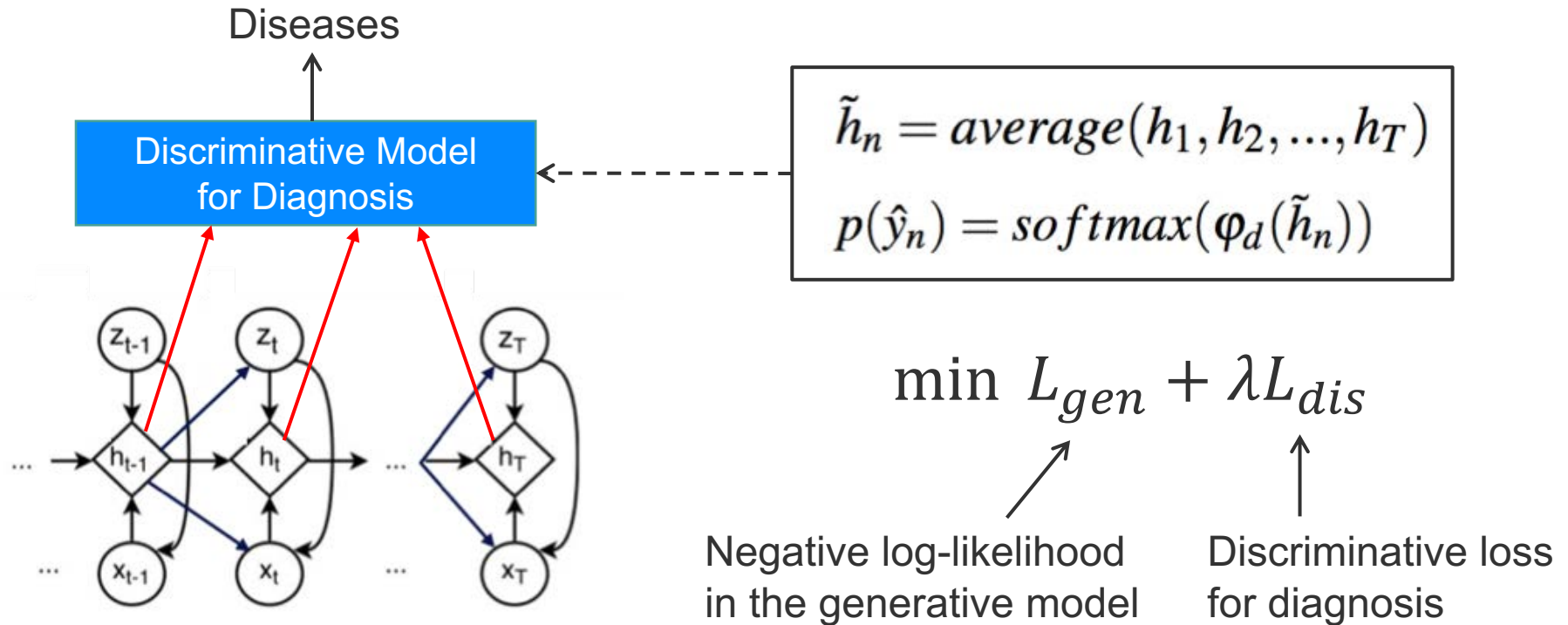
$$h_t = \text{LSTM}(x_t, \mathbb{E}[y_t|x_t], z_t, h_{t-1})$$

$$p(z_t|x_t, \mathbb{E}[y_t|x_t]) = \mathcal{N}(\pi_t, \text{diag}(\tau_t^2)), \text{ where } [\pi_t, \tau_t] = \omega(x_t, \mathbb{E}[y_t|x_t])$$



Combine Missing-Value-Imputation with Diagnosis

- Jointly perform generative and discriminative learning





MIMIC-III Dataset

- Electronic health records of patients in the intensive care units
- Processing
 - Time stamp is based on days
 - Maximum and minimum sequence length is 100 and 2.
 - Primary diagnosis is used as labels.
- Statistics

Original

# Hospital admissions	58,000
# Unique patients	46,520
# Diseases	2,833
# Lab tests	27.85 million
# Unique lab test items	635

After selection

# Hospital admissions	30,931
# Diseases	50 (most frequent ones)
# Unique lab test items	50 (most frequent ones)



Performance of Diagnosis

Diagnosis performances of different models

Model	Micro-F1	Macro-F1	Macro-F1-w	Micro-AUC	Macro-AUC	Macro-AUC-w
NN	0.376 ± 0.004	0.221 ± 0.003	0.347 ± 0.005	0.939 ± 0.001	0.905 ± 0.001	0.913 ± 0.001
AE+NN	0.366 ± 0.004	0.219 ± 0.002	0.344 ± 0.002	0.938 ± 0.001	0.903 ± 0.002	0.912 ± 0.001
VAE+NN	0.374 ± 0.003	0.226 ± 0.005	0.352 ± 0.004	0.941 ± 0.000	0.908 ± 0.001	0.916 ± 0.001
RNN+NN	0.395 ± 0.004	0.248 ± 0.003	0.373 ± 0.003	0.945 ± 0.003	0.918 ± 0.004	0.923 ± 0.003
VRNN+NN	0.426 ± 0.002	0.291 ± 0.006	0.407 ± 0.002	0.958 ± 0.000	0.937 ± 0.000	0.938 ± 0.001
VRNN+NN (early)	0.422 ± 0.005	0.285 ± 0.006	0.403 ± 0.004	0.957 ± 0.001	0.935 ± 0.001	0.937 ± 0.001
Performance of features derived from different models (with a simple NN classifier)						
$E(z_n)$ (VAE)	0.363 ± 0.004	0.195 ± 0.004	0.326 ± 0.003	0.936 ± 0.001	0.896 ± 0.003	0.906 ± 0.002
$E(z_n)$ (VAE+NN)	0.380 ± 0.004	0.228 ± 0.004	0.353 ± 0.002	0.943 ± 0.001	0.911 ± 0.003	0.918 ± 0.002
\tilde{h}_n (VRNN)	0.406 ± 0.003	0.261 ± 0.003	0.381 ± 0.003	0.953 ± 0.000	0.928 ± 0.001	0.930 ± 0.001
\tilde{h}_n (VRNN+NN)	0.427 ± 0.003	0.297 ± 0.004	0.410 ± 0.003	0.958 ± 0.001	0.936 ± 0.001	0.937 ± 0.000
Performance with different missing value imputation methods						
RNN+NN(zero)	0.395 ± 0.005	0.248 ± 0.003	0.374 ± 0.002	0.945 ± 0.003	0.918 ± 0.004	0.923 ± 0.003
RNN+NN(last&next)	0.385 ± 0.002	0.233 ± 0.003	0.360 ± 0.002	0.941 ± 0.001	0.912 ± 0.002	0.918 ± 0.001
RNN+NN(row mean)	0.393 ± 0.003	0.243 ± 0.005	0.369 ± 0.001	0.945 ± 0.002	0.917 ± 0.003	0.923 ± 0.002
RNN+NN(NOCB)	0.384 ± 0.003	0.231 ± 0.002	0.359 ± 0.001	0.941 ± 0.002	0.911 ± 0.003	0.917 ± 0.002
VRNN+NN	0.426 ± 0.002	0.291 ± 0.006	0.407 ± 0.002	0.958 ± 0.000	0.937 ± 0.001	0.938 ± 0.001



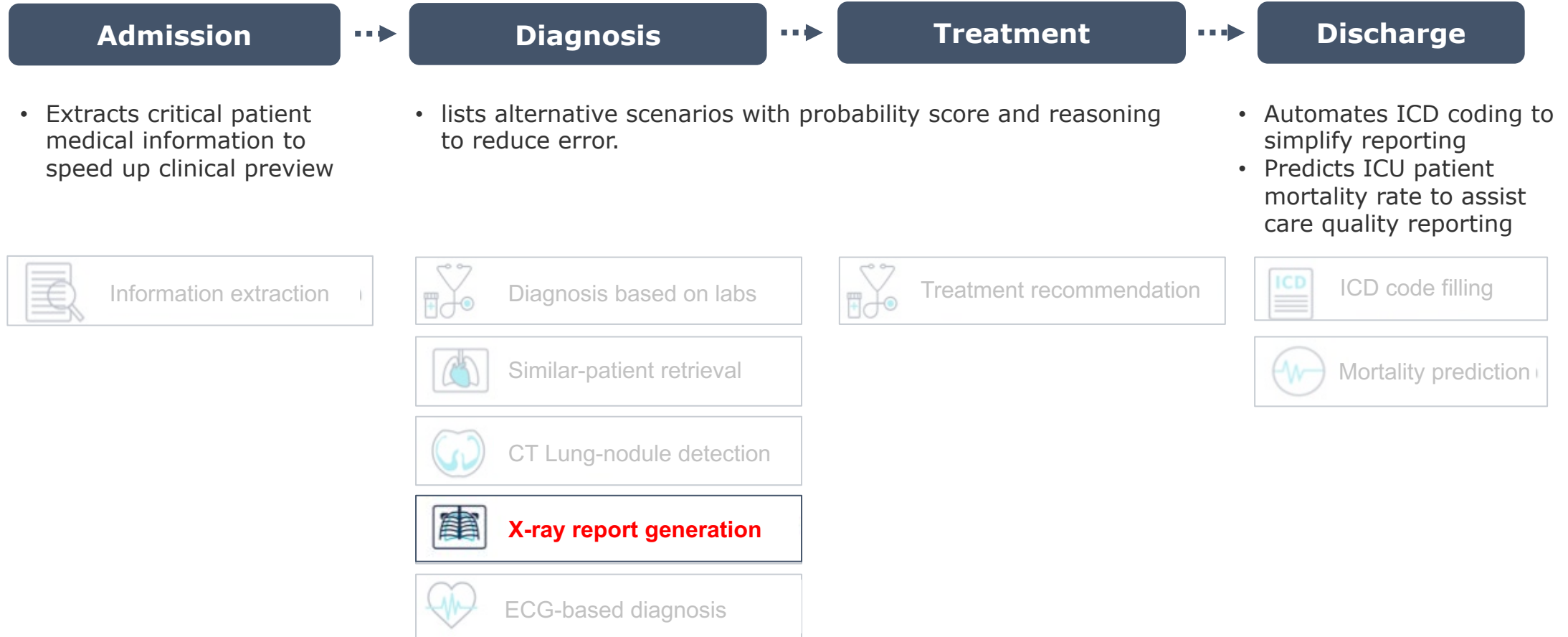
Performance of Missing Value Imputation

Different imputation methods

Imputation Methods	Imputation Error
Zero	0.909 ± 0.112
Last&next	0.434 ± 0.110
Row mean	0.541 ± 0.114
NOCB	0.547 ± 0.112
VRNN+NN	0.370 ± 0.110



Supports clinical decisions at all points of care





Chest X-ray Report

- A chest x-ray report consists of multiple sections of information.
 - **Findings:** the radiology observations and findings regarding each area of the body examined in the imaging study
 - **Impression:** the radiologist combines the findings, patient clinical history and indication for the imaging study and provides a diagnosis



Findings:

There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

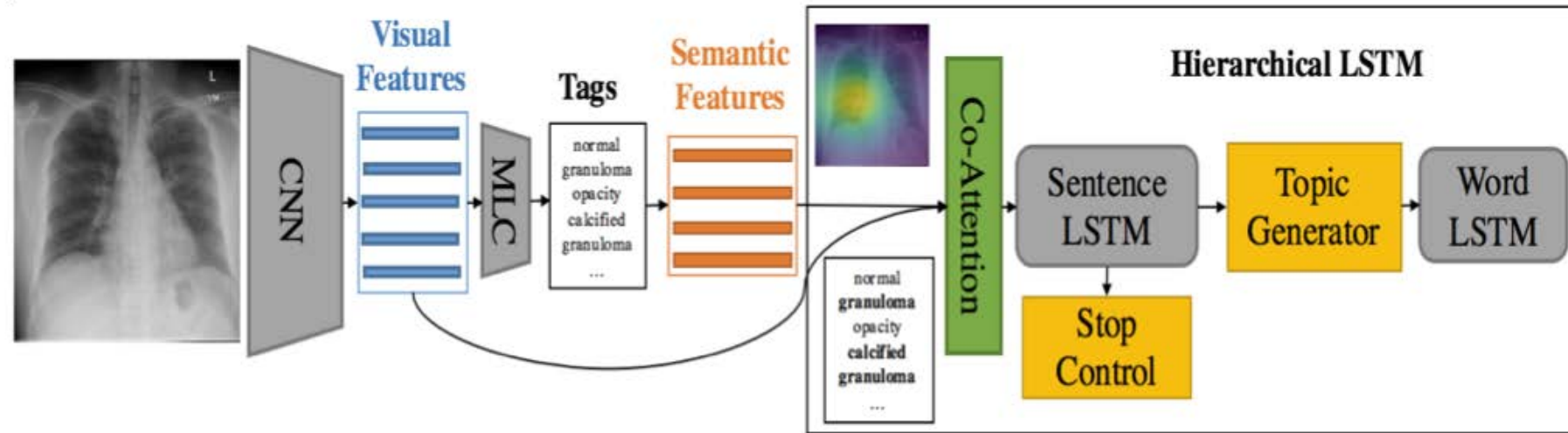
Impression:

No acute cardiopulmonary abnormality.

- Abnormal regions in medical images are difficult to identify.
- The reports are typically long, containing multiple sentences.
- Each sentence discusses a specific topic. How to localize the image regions and tags that are relevant to this topic?



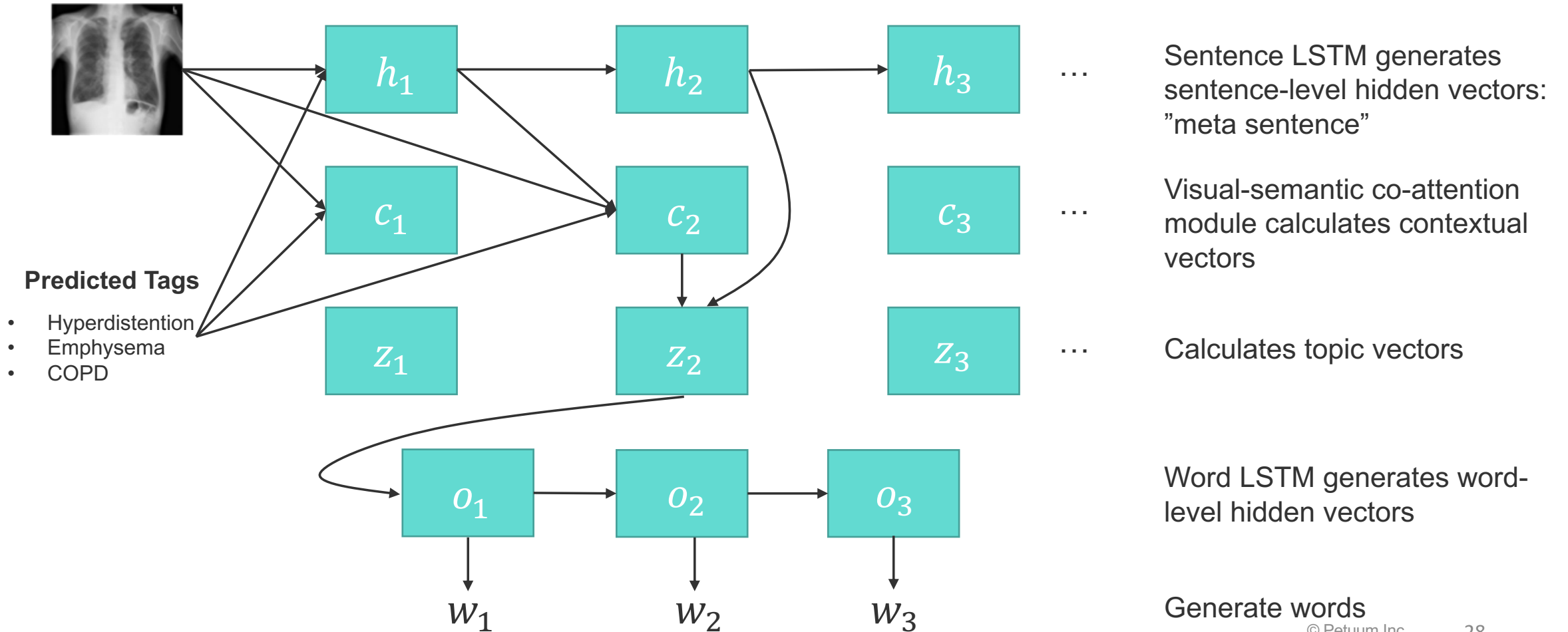
Model Architecture



- Semi-supervised learning for lesion tag prediction
- Hierarchical LSTM for long-paragraph generation
- Visual-semantic co-attention to localize the relevant image regions and tags for each sentence to be generated

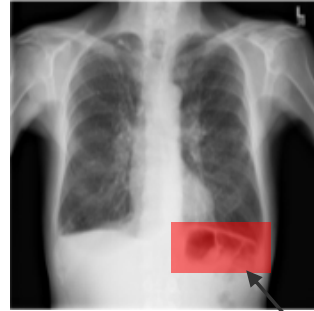


Hierarchical LSTM for Paragraph Generation





Visual-Semantic Co-Attention



- Hyperdistention
- Emphysema
- COPD
- **Cicatrix**

Visual attention

$$a_{tn}^{(v)} \propto \exp(\mathbf{U}^{(v)} \tanh(\mathbf{V}^{(v)} \mathbf{x}_n + \mathbf{W}^{(v)} \mathbf{h}_{t-1}))$$

Semantic attention

$$a_{tm}^{(s)} \propto \exp(\mathbf{U}^{(s)} \tanh(\mathbf{V}^{(s)} \mathbf{y}_m + \mathbf{W}^{(s)} \mathbf{h}_{t-1}))$$

... There is chronic pleural-parenchymal scarring within the lung bases. No lobar consolidation is seen. ...

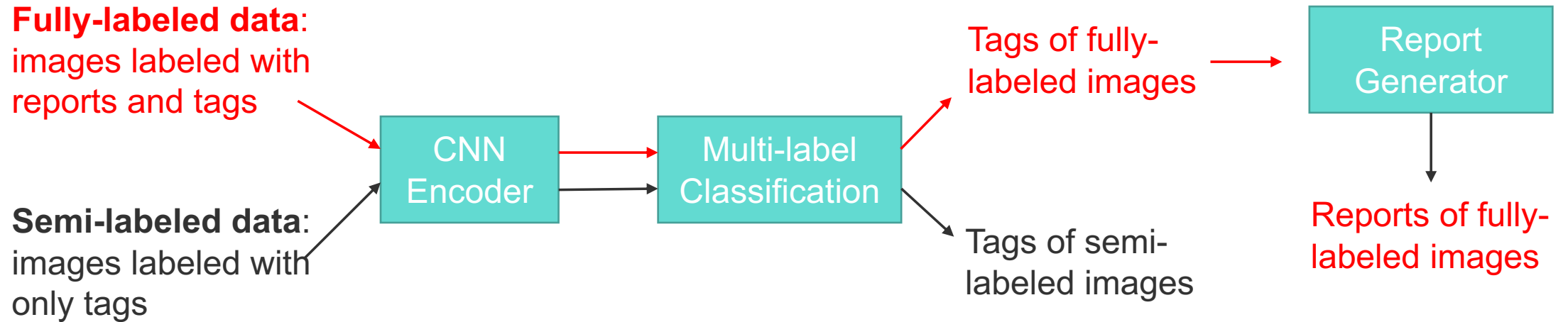
Co-attention

$$\mathbf{r}_t^{(v)} = \sum_{n=1}^N a_{tn}^{(v)} \mathbf{x}_n \quad \mathbf{r}_t^{(s)} = \sum_{m=1}^M a_{tm}^{(s)} \mathbf{y}_m$$

$$\mathbf{c}_t = \mathbf{Q}[\mathbf{r}_t^{(v)}; \mathbf{r}_t^{(s)}]$$



Semi-supervised Learning for Lesion Tagging



- The amount of semi-labeled data is large. Their tags can help train a better lesion-tagging model.



Experiments

- Fully-labeled dataset
 - Indiana University Chest X-ray Collection
 - 7,470 pairs of images and reports
 - Each image has a set of tags obtained from Medical Text Indexer
 - 1915 unique words and 572 unique tags
 - On average, each image is associated with 2.2 tags, 5.7 sentences and each sentence contains 6.5 words.
- Semi-labeled dataset
 - NIH Chest X-ray images
 - 108,948 frontal-view X-ray images
 - 14 lesion tags
- Evaluation metrics
 - Language quality: BLEU, METEOR, ROUGE and CIDER
 - Clinical correctness: F1



Evaluation of Language Quality

	Methods	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE	CIDER
Baselines	CNN-RNN [1]	0.316	0.211	0.140	0.095	0.159	0.267	0.111
	LRCN [2]	0.369	0.229	0.149	0.099	0.155	0.278	0.190
	Soft ATT [3]	0.399	0.251	0.168	0.118	0.167	0.323	0.302
	ATT-RK [4]	0.369	0.226	0.151	0.108	0.171	0.323	0.155
Our methods	No-Attention	0.505	0.383	0.290	0.224	0.200	0.420	0.259
	Semantic-Only	0.504	0.371	0.291	0.230	0.207	0.418	0.286
	Visual-Only	0.507	0.373	0.297	0.238	0.211	0.426	0.300
	Co-Attention	0.517	0.386	0.306	0.247	0.217	0.447	0.327

- Baselines are the state-of-the-art image captioning methods.



Evaluation of Clinical Correctness

- Whether the presence/absence of 16 major lesions is correctly predicted: atelectasis, calcinosis, consolidation, cardiomegaly, edema, effusion, emphysema, fibrosis, granuloma, hernia, infiltration, mass, nodule, pneumonia, pneumothorax, pleural thickening
- Labeling process: manually read the report and check whether a lesion exists
 - Example

Normal cardiomediastinal silhouette. Interval improvement in lung volumes bilaterally. Improved aeration of the right and left lung bases. Bilateral small **pleural effusions** and left base **atelectatic change**, with interval improvement. Visualized XXXX of the chest XXXX are within normal limits.



Has effusion and atelectasis.
No other lesions.



Evaluation of Clinical Correctness (Cont'd)

- F1 scores

Methods	Our methods			
	No-Attention	Visual-Only	Semantic-Only	Co-Attention
Macro-F1	0.49	0.51	0.75	0.79



Success Cases



Ground Truth

No active disease. The heart and lungs have in the interval. Both lungs are clear and expanded. Heart and mediastinum normal.



No evidence of active disease. The lungs are clear. There is no focal airspace consolidation. No pleural effusion or pneumothorax. Heart size and mediastinal contour are within normal limits. There are multilevel degenerative changes of the spine.

Ours-CoAttention

No active disease. The heart and lungs have in the interval. Lungs are clear and expanded. Cardiomeastinal silhouette is within normal limits. No pleural effusion or pneumothorax is seen. No pleural effusion. No cavitary or pneumothorax.

No acute cardiopulmonary findings. Heart size is not enlarged. No focal airspace consolidation suspicious pulmonary opacity large pleural effusion or pneumothorax. No focal areas of consolidation. Degenerative changes of the spine. This is moderate exam of the hydropneumothorax. Lungs are clear. There is no focal airspace consolidation pleural effusion or pneumothorax.

Ours-no-Attention

The lungs are clear bilaterally. The are grossly normal. No focal lung consolidation. No acute bony abnormality. cm nodule within the right lower lobe on the lateral view. No pneumothorax or pleural effusion. No acute bony abnormality. The heart is not enlarged. The lungs are clear. No acute bony abnormality .

The lungs are clear bilaterally. The are grossly normal. No pleural effusion. The heart is normal in size and contour. The lungs are clear. There are no acute bony findings.

Soft Attention

No acute cardiopulmonary abnormality. The lungs are clear bilaterally. Specifically no evidence of focal airspace consolidation pleural effusion or pneumothorax. Cardio mediastinal silhouette is unremarkable. Visualized osseous structures of the thorax are without acute abnormality .

No acute cardiopulmonary abnormality. The lungs are clear bilaterally. There is no pleural effusion or pneumothorax. The heart and mediastinum are normal. There is no focal air space opacity to suggest a pneumonia.

- The underlined sentences are the descriptions of detected abnormalities.



Failure Cases

Ground Truth



No acute cardiopulmonary abnormality. Normal heart size mediastinal contours. Eventration of the right hemidiaphragm. No focal airspace consolidation. No pleural effusion or pneumothorax.



No acute cardiopulmonary abnormality. Heart size appears within normal limits . Pulmonary vasculature appears within normal limits. Overlying the middle cardiac silhouette representing a hiatal hernia. No focal consolidation pleural effusion or pneumothorax. No acute bony abnormality.

Ours-CoAttention

No acute cardiopulmonary abnormality. Stable appearance of the thoracic aorta. The right lateral lower lobe is noted in the right lower right midlung. No large pleural effusion or focal airspace disease. Mild interstitial opacities. Atherosclerotic calcifications bony structures bilaterally. There is no pleural effusion or pneumothorax developed in the right lower lobe.

No active disease. The heart and lungs have in the interval. Nipple and lateral lucency in the lungs suggestive of focal airspace disease. The lungs are hyperexpanded consistent with emphysema in the left lower lobe. This is most at the upper lobes. This may indicate hypoventilated irregularities or effusions. The lungs are otherwise grossly clear. Resolution of by normal pleural effusion.

Ours-no-Attention

The lungs are clear bilaterally. The are grossly normal. No acute bony abnormality. The lungs are otherwise clear. No acute osseous abnormality. No acute osseous abnormality. The heart and mediastinum are normal. There is no focal air space opacity.

No acute cardiopulmonary abnormality. The lungs are clear bilaterally. The are grossly normal. No focal airspace consolidation. No pneumothorax or pleural effusion. Heart size and pulmonary vascularity within normal limits. There is no pneumothorax or pleural effusion.









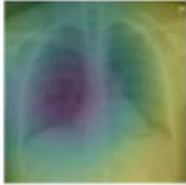





Soft Attention

No acute cardiopulmonary abnormality. The lungs are clear bilaterally. There is no focal airspace consolidation. No pleural effusion or pneumothorax. Heart size and pulmonary vascularity appear within normal limits.

No acute cardiopulmonary abnormality. There is no focal airspace consolidation. No pneumothorax or pleural effusion. No acute bony abnormality. Heart size is normal .

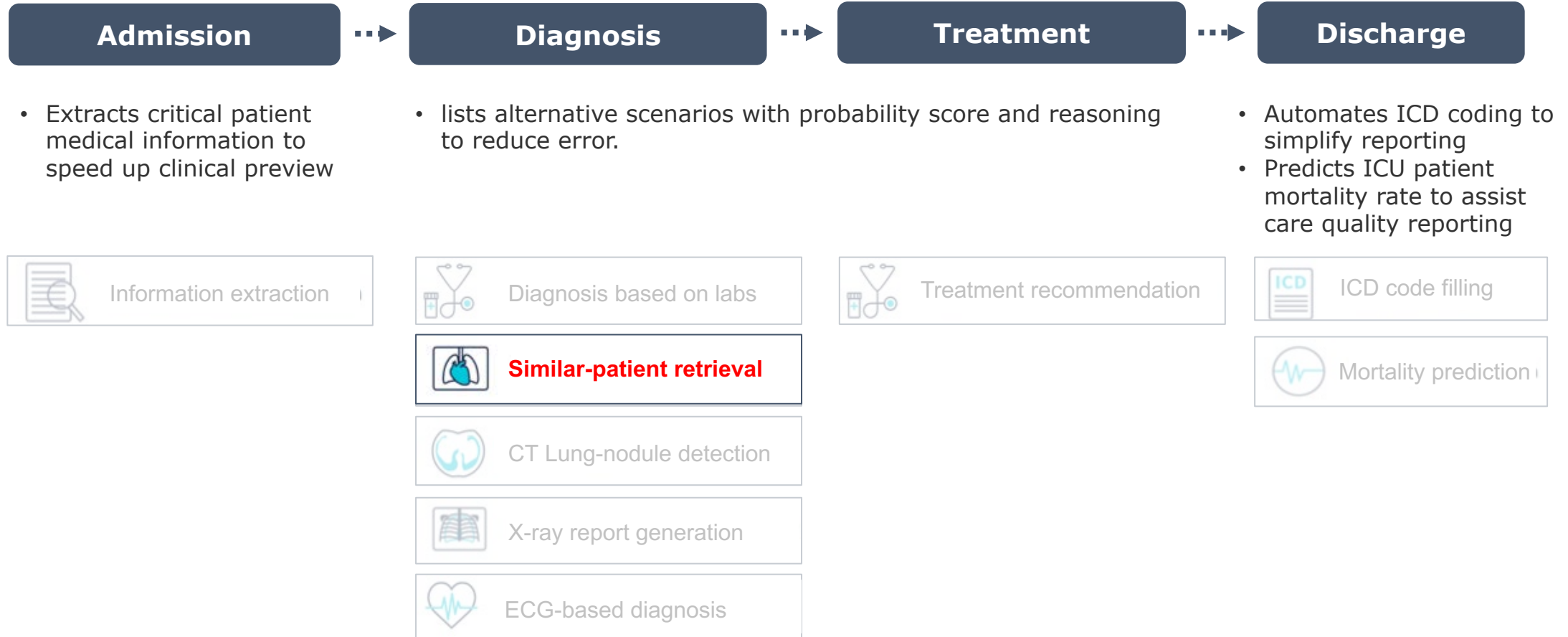


Visualization of Co-attention

						
degenerative change; obstruction	normal; degenerative change; nodule; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy	normal; degenerative change; nodule ; calcified granuloma; hyper expansion; granuloma ; pneumonia; scarring; sternotomy	normal; degenerative change; nodule ; calcified granuloma; hyper expansion; granulomatous disease; granuloma ; pneumonia; scarring; sternotomy	normal; degenerative change; nodule ; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy	normal; degenerative change; nodule ; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy	normal; degenerative change; nodule ; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy
No acute intrathoracic abnormality.	No acute intrathoracic abnormality.	No bony abnormality.	The cardio mediastinal silhouette is within normal limits for appearance.	No focal areas of pulmonary consolidation.	Breast motion.	There is an age indeterminate deformity of a mid-thoracic vertebral body.
No acute cardiopulmonary finding. The heart size and cardiopulmonary silhouette is normal. There is no focal airspace opacity pleural effusion or pneumothorax. The obstruction are intact with mild degenerative change in the thoracic spine.						
						
normal	normal; calcified granuloma; granulomatous disease; granuloma ; scarring; opacity; degenerative change; sternotomy; thoracic aorta; nodule	normal; calcified granuloma; granulomatous disease; granuloma; scarring; opacity; degenerative change ; sternotomy; thoracic aorta; nodule	normal; calcified granuloma; granulomatous disease; granuloma; scarring; opacity; degenerative change ; sternotomy; thoracic aorta; nodule	normal; calcified granuloma; granulomatous disease; granuloma; scarring; opacity; degenerative change ; sternotomy ; thoracic aorta; nodule	normal; calcified granuloma; granulomatous disease; granuloma; scarring; opacity; degenerative change ; sternotomy ; thoracic aorta; nodule	normal; calcified granuloma; granulomatous disease ; granuloma; scarring; opacity; degenerative change; sternotomy ; thoracic aorta; nodule
	Right upper lobe infiltrate.	Lungs are clear .	Stable heart size and aortic contours.	No acute displaced rib fractures.	No focal airspace opacities or consolidation.	No visualized of pneumothorax.
No acute cardiopulmonary abnormality identified. The examination consists of frontal and lateral radiographs of the chest. The cardio mediastinal contours are within normal limits. Pulmonary vascularity is within normal limits. No focal consolidation pleural effusion or pneumothorax identified. The visualized osseous structures and upper abdomen are unremarkable.						

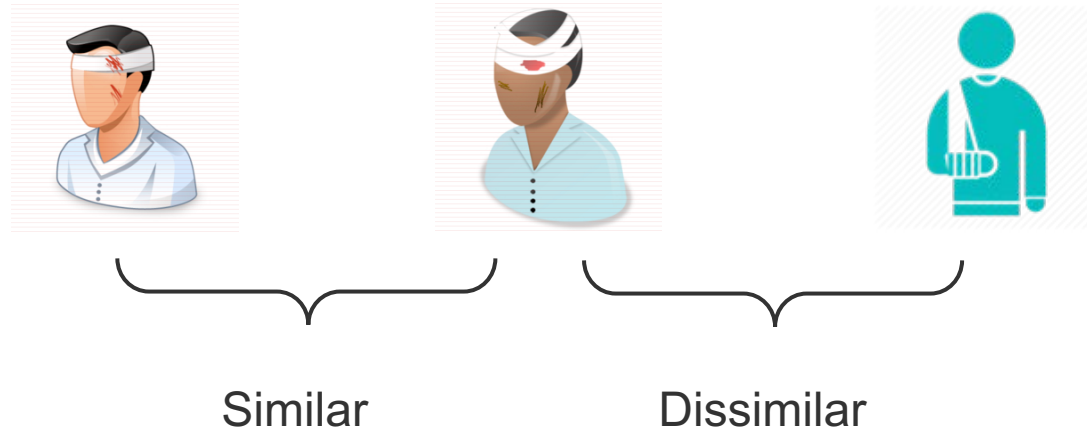


Supports clinical decisions at all points of care





Patient Similarity





Patient Similarity for Personalized Diagnosis

Query Patient



What disease?

Head Injury

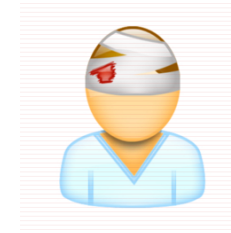
Similar Patients in Database



Head Injury



Head Injury



Head Injury





Patient Similarity for Personalized Treatment

Query Patient



What medication?

Medication F

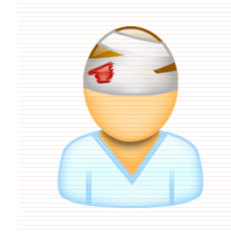
Similar Patients in Database



Medication A
Medication F
Medication H



Medication F
Medication M

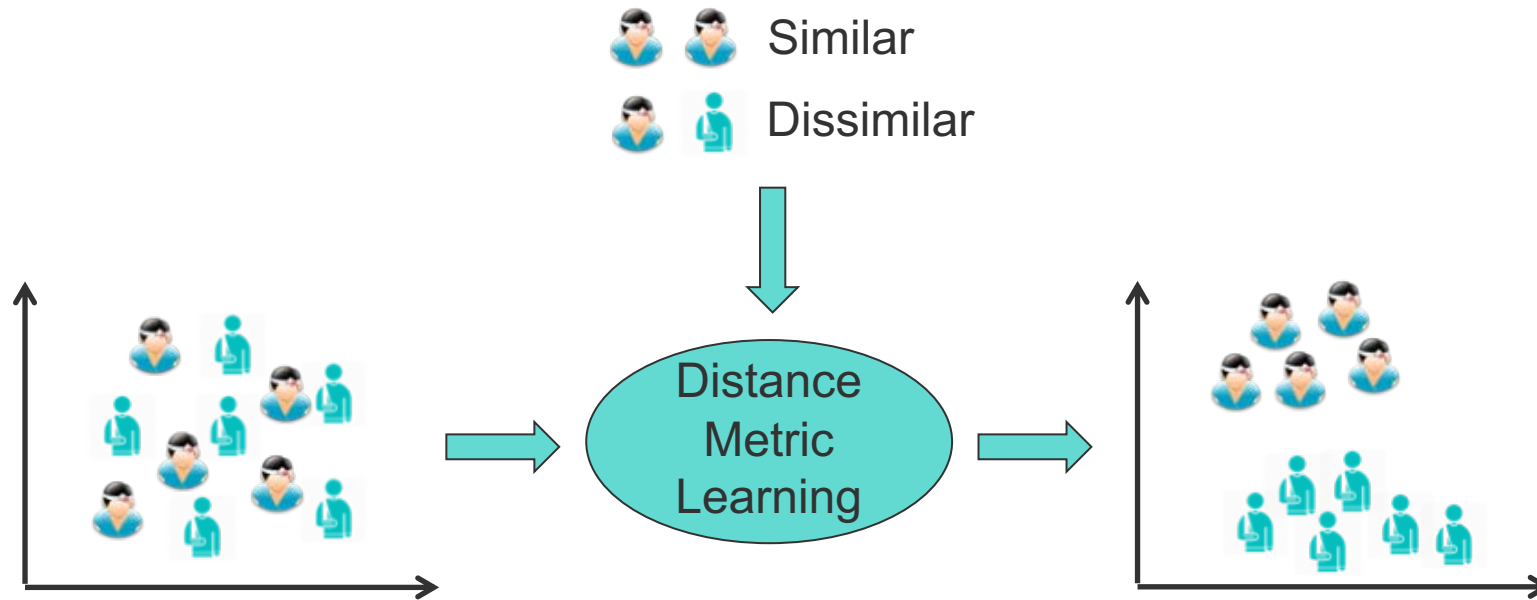


Medication D
Medication E
Medication F



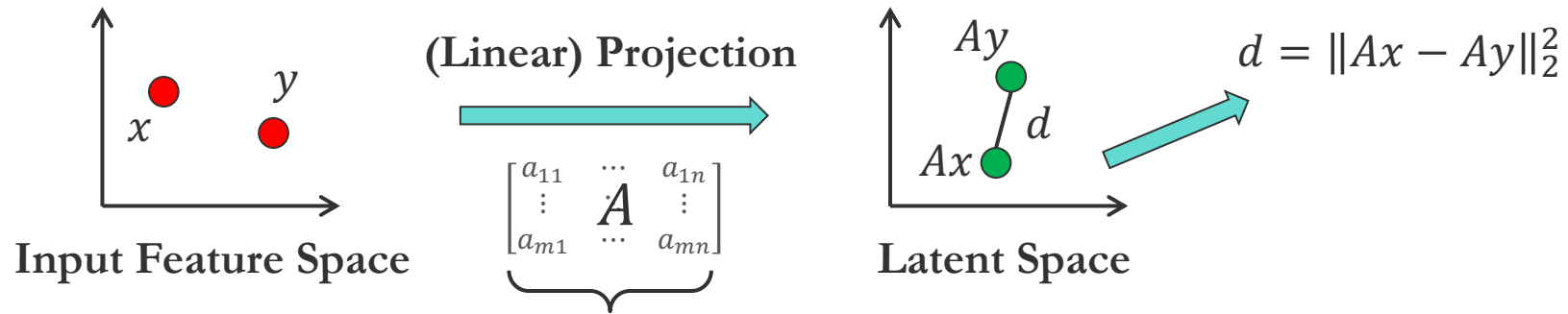


Distance Metric Learning (DML)





Distance Metric Learning (Cont'd)



Row vectors of A are called “**components**”, each corresponding to one dimension of the latent space

$$\min_A \sum_{(x,y) \in S} \|Ax - Ay\|^2 \quad (\text{Xing et al., 2002})$$

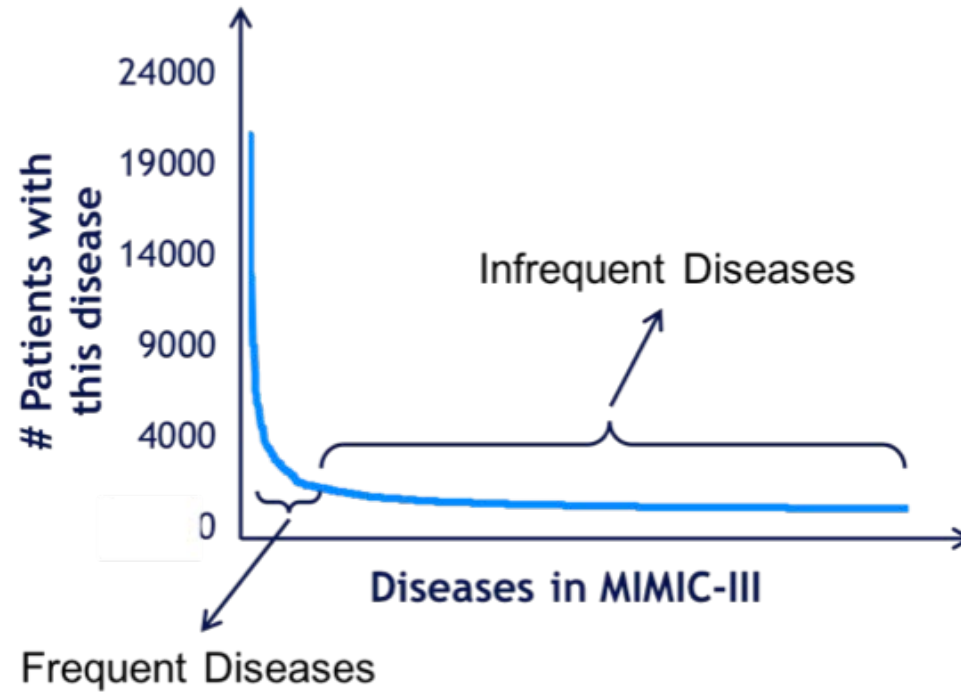
↑
Similar Pairs

$$s. t. \quad \forall (x, y) \in D, \|Ax - Ay\|^2 \geq 1$$

↑
Dissimilar Pairs

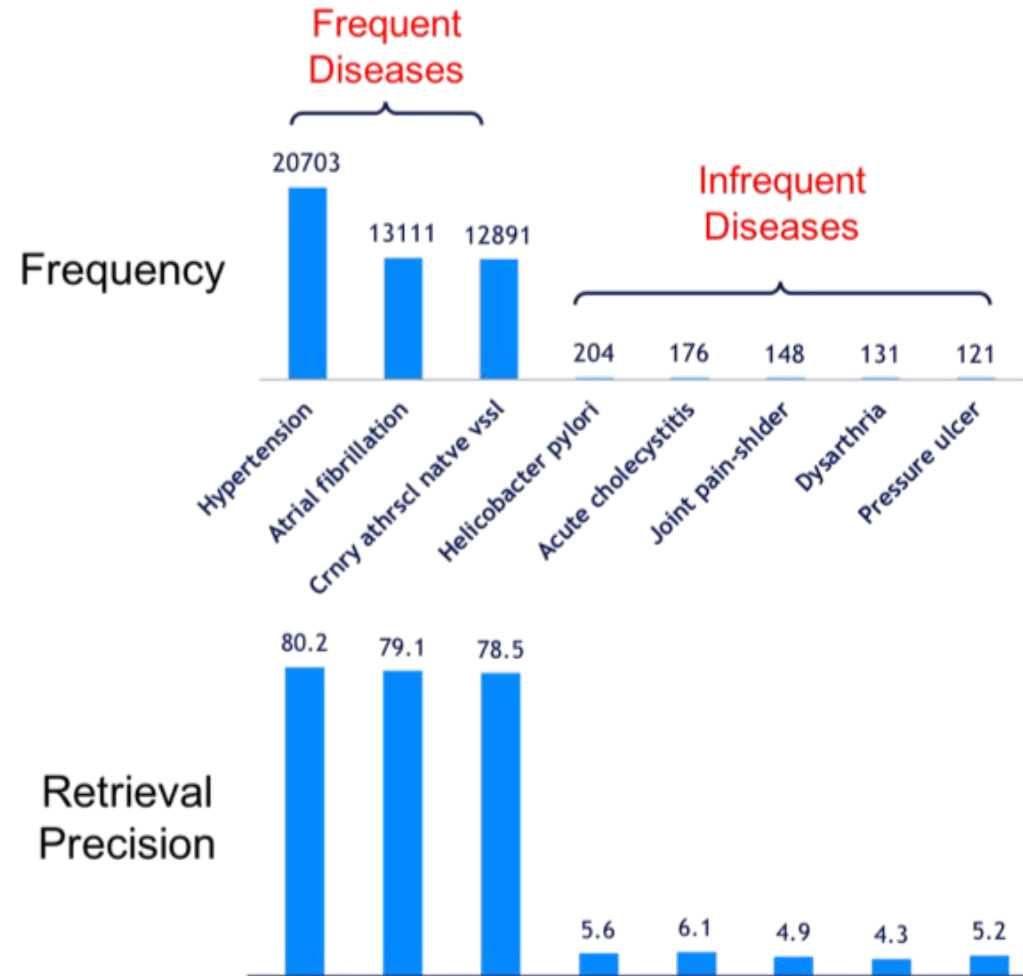


Power Law Distribution of Diseases





DML is Biased to Frequent Diseases



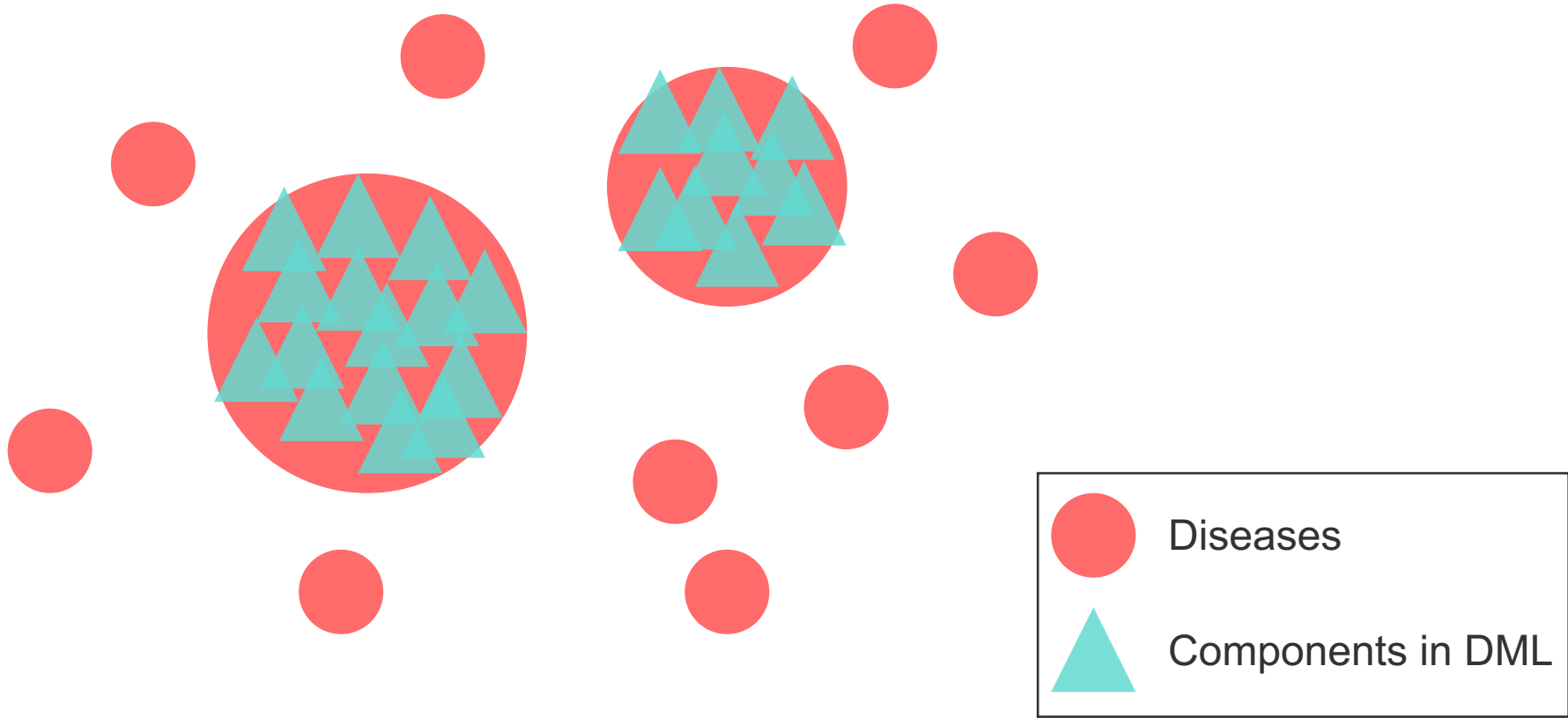


Infrequent Diseases Are Not Neglectable

- Many infrequent diseases are life-threatening
 - “Flail chest is a life-threatening medical condition that occurs when a segment of the rib cage breaks due to trauma and becomes detached from the rest of the chest wall.”
 - Flail chest is infrequent, occurring only in 0.1% MIMIC-III patients
- The total amount of infrequent disease is very large
 - In MIMIC-III, the number of infrequent diseases (with < 200 patients) is 2458, accounting for 86.8% of all diseases
 - The number of patients with infrequent diseases are 36278, accounting for 78.0% of all patients

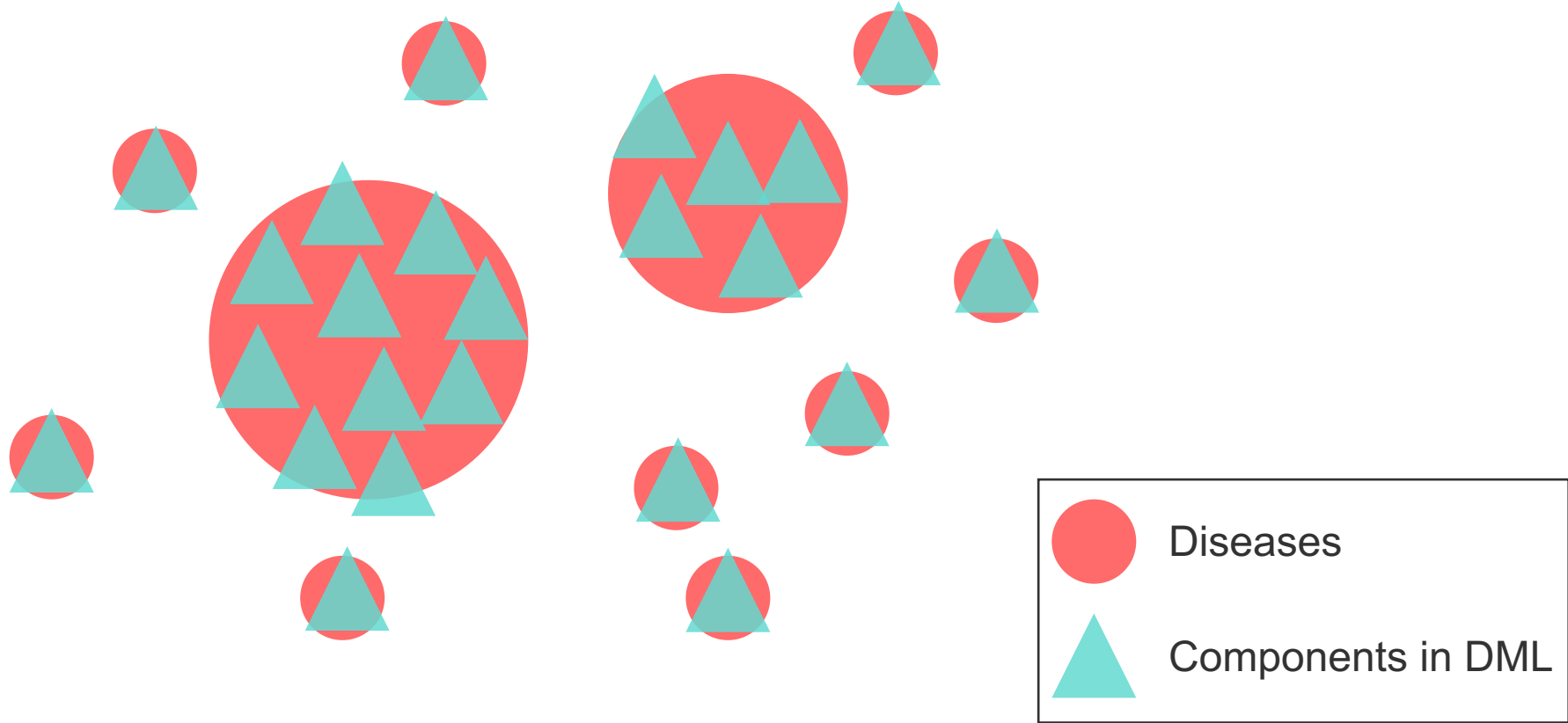


Why DML is Biased to Frequent Diseases





“Diversify” Components to Capture Infrequent Diseases





Encourage Near-Orthogonality among Components

- Vector a_i and a_j are near-orthogonal if $a_i \cdot a_j \approx 0$, $\|a_i\|_2 \approx 1$, $\|a_j\|_2 \approx 1$
- A set of vectors are near-orthogonal if the Gram matrix is near to the identity matrix

$$\begin{bmatrix} a_1 \cdot a_1 & a_1 \cdot a_2 & a_1 \cdot a_3 \\ a_2 \cdot a_1 & a_2 \cdot a_2 & a_2 \cdot a_3 \\ a_3 \cdot a_1 & a_3 \cdot a_2 & a_3 \cdot a_3 \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- How to measure the nearness between two matrices?



Diversity-Promoting DML

$$\min_A \sum_{(x,y) \in S} \|Ax - Ay\|^2 + \lambda R(A)$$

$$s. t. \quad \forall (x, y) \in D, \|Ax - Ay\|^2 \geq 1$$

- Bregman divergence between Gram matrix G and identity matrix I :

$$R(A) = D_\phi(AA^T, I)$$

- Under von Neumann divergence

$$R_{vn}(A) = \text{tr}((AA^T)\log(AA^T) - AA^T) + k$$

- Under Log-Det divergence

$$R_{ld}(A) = \text{tr}(AA^T) - \log\det(AA^T) - k$$



Summary of Algorithm

- Convex relaxation
 - Using semidefinite programming relaxation to convexify the data-dependent loss
 - Using the properties of eigenvalues to convexify the regularizers
- Proximal gradient descent
 - Eliminating constraints on “dissimilar” pairs
 - Deriving proximal operators of the von Neumann and LogDet regularizers



Experiments: Feature Extraction

- Demographics
 - Gender, age
- Clinical notes
 - 5000-dimensional bag-of-words representation
 - 300-dimensional Word2Vec representation
- Lab tests
 - 635 test items
 - Zero-order, first-order, second-order time series features
- Dimensionality reduction
 - Reduce the feature dimension from 7207 to 1000 using principle component analysis

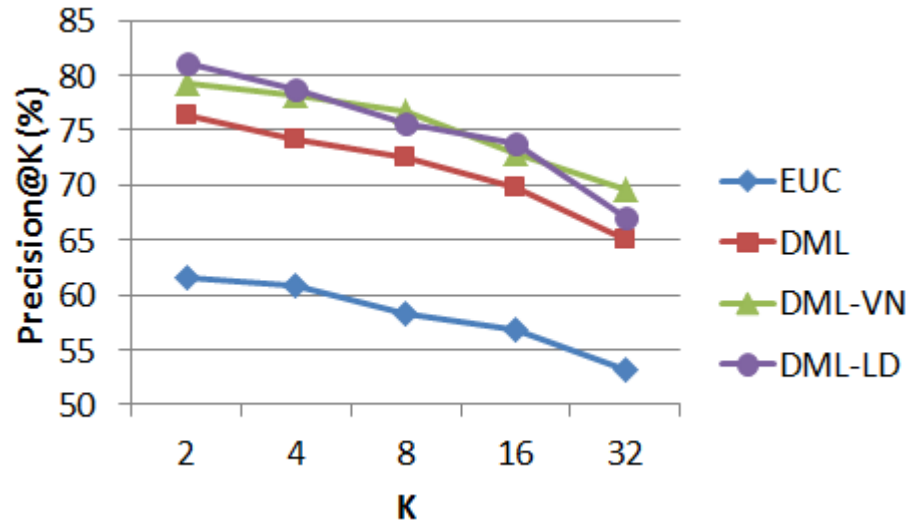


Experimental Setup

- Determine whether two patients are similar or dissimilar based on their diagnostic results (a set of diseases)
 - Given the disease set S_i and S_j of patient i and j , measure their overlap score $o(S_i, S_j)$
 - Determine similarity/dissimilarity label: if $o(S_i, S_j)$ is greater than a threshold c , then patient i and j are regarded as similar; otherwise, dissimilar
 - Training data: 0.1 million similar pairs and 0.1 million dissimilar pairs
- Application of the learned distance metric: similar-patient retrieval
 - Given the query patient, compute its distance with patients in the database using the learned distance metric
 - Use precision@K to evaluate the retrieval performance: among the top K retrieved patients, how many are similar to the query patient



Similar-Patient Retrieval Results

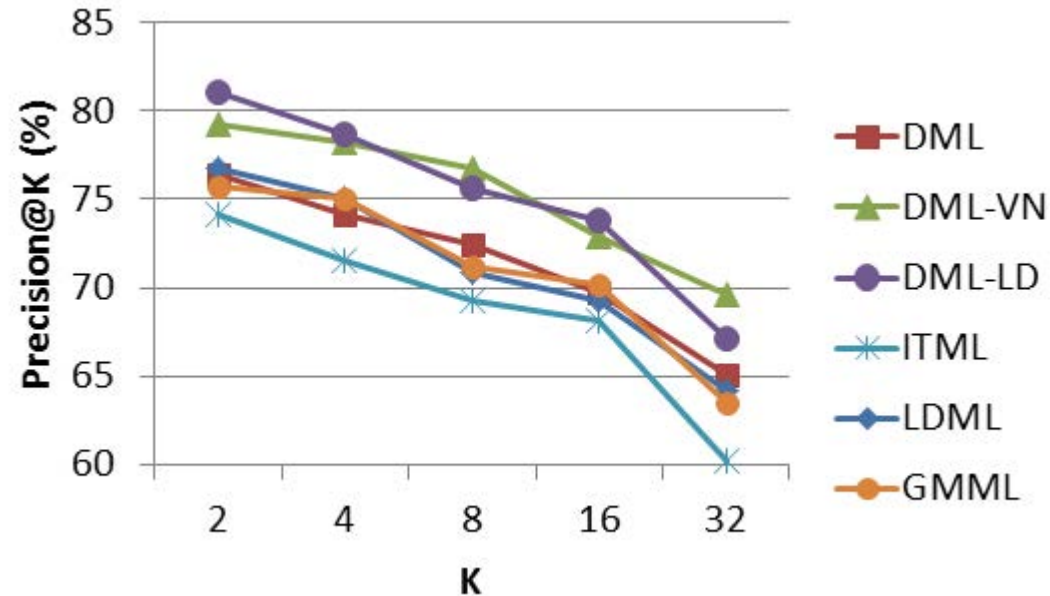


- EUC: Euclidean distance
- DML: Distance metric learning
- DML-VN: DML with von Neumann divergence regularization
- DML-LD: DML with Log-Det divergence regularization

- Compared with the vanilla Euclidean distance, DML dramatically improves retrieval precision.
- Adding diversity-promoting regularization significantly improves the performance of DML.



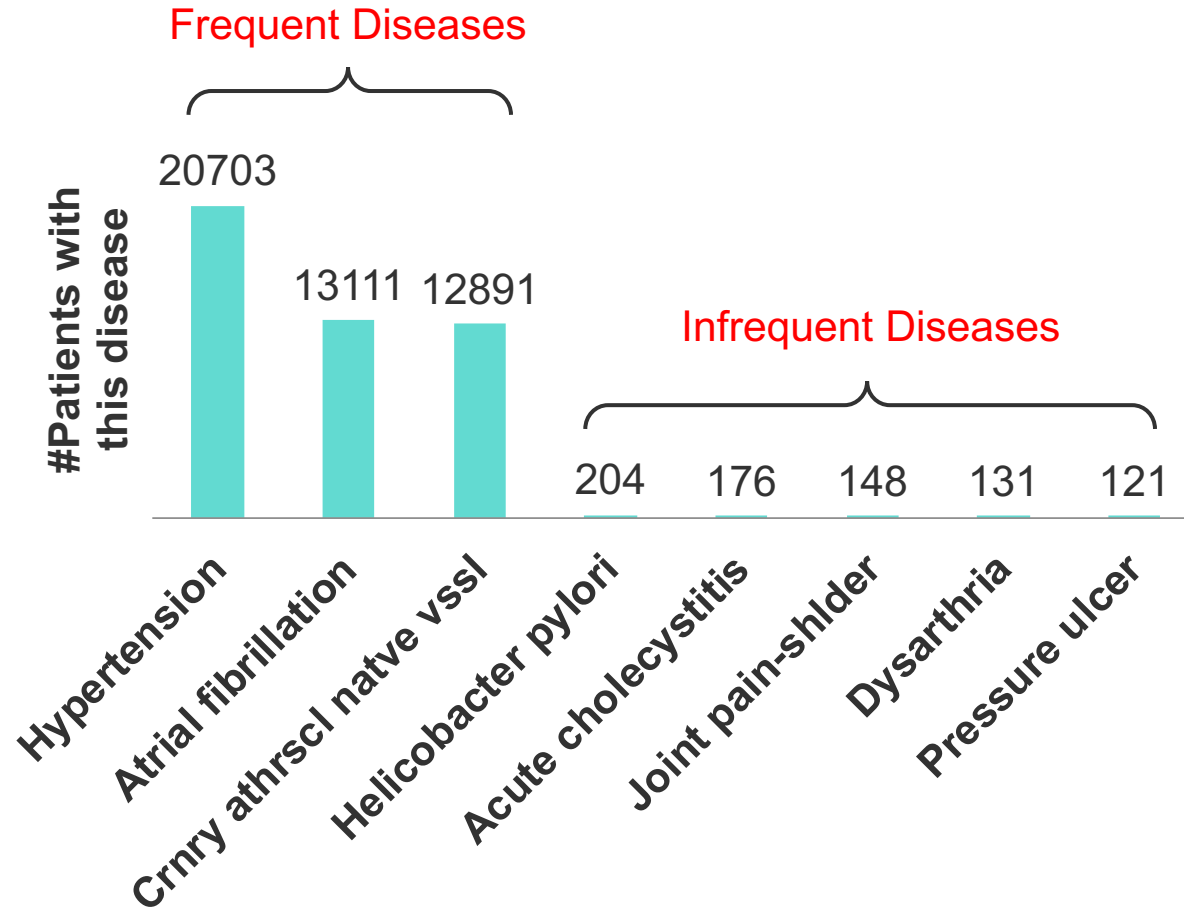
Comparison with More Baseline Methods



- ITML: Information theoretical metric learning (David et al., 2007)
- LDML: Logistic discriminant metric learning (Guillaumin et al., 2009)
- GMML: Geometric mean metric learning (Zadeh et al., 2016)

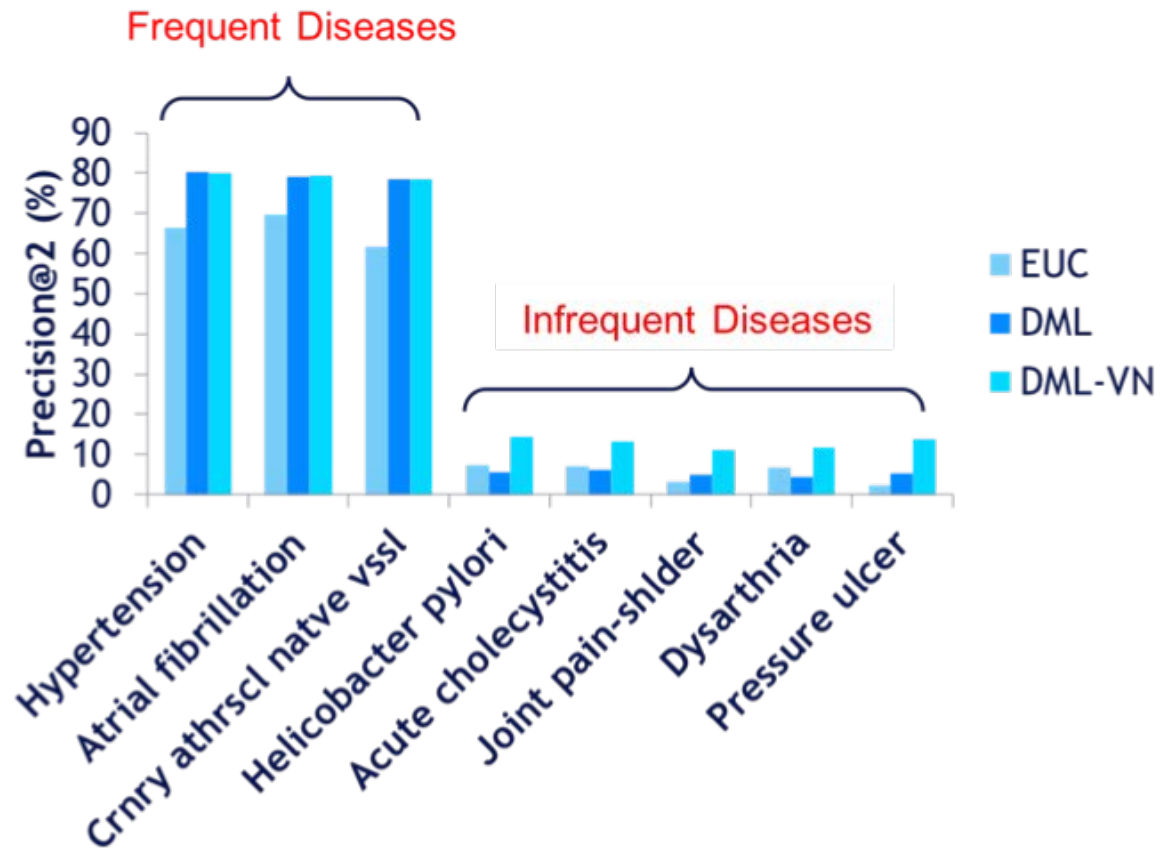


Performance on Infrequent Diseases





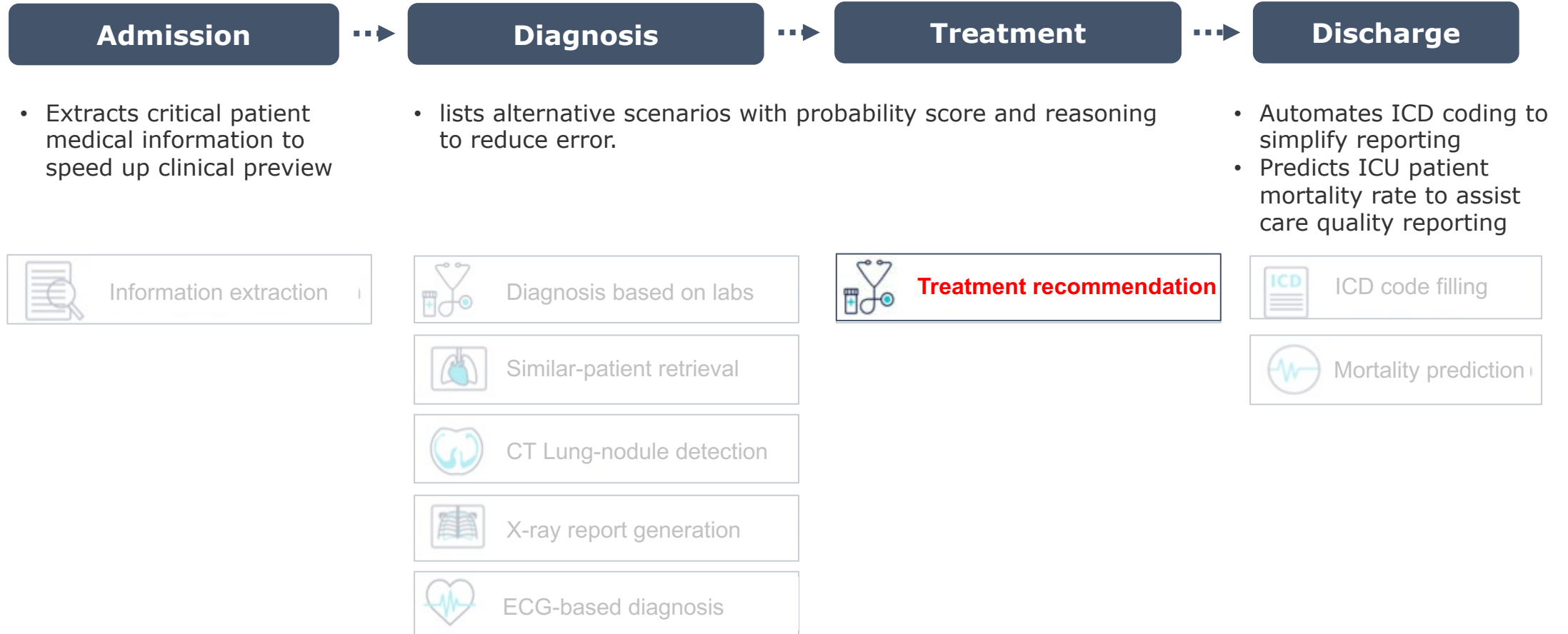
Performance on Infrequent Diseases (Cont'd)



- DML-VN (promoting diversity) significantly improves the precision on infrequent diseases without sacrificing the performance on frequent diseases



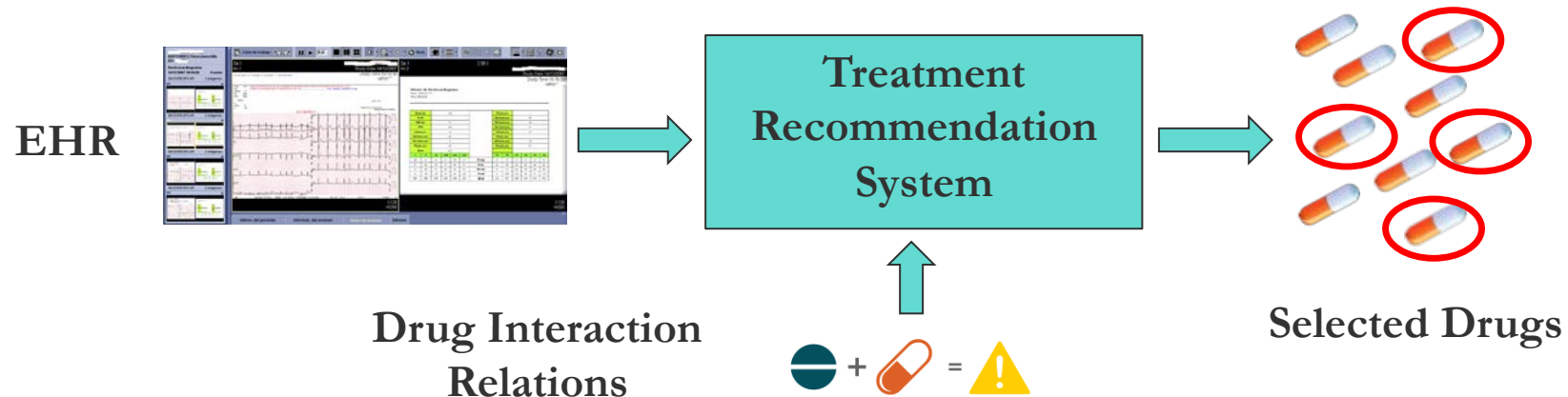
Supports clinical decisions at all points of care





Treatment Recommendation

- Based on the EHR of a patient, select a subset of drugs that
 - Can effectively treat the conditions/ diseases of the patient
 - Discourage adverse interactions
 - Encourage beneficial interactions





Multiple Diagnosis for Each Patient

PATIENT_ID	DISEASE CODE
109	33829
109	78900
109	79092
112	53100
112	41071
112	2859
112	41401
112	725
113	1915
113	3314
113	53081
114	41401
114	4111
114	48283
114	2859
114	2720
114	3051
115	1940
115	1977
115	2553
115	4240
115	5845


Each patient is diagnosed with multiple diseases/conditions





Multiple Drug Prescriptions for Each Patient

Each patient is prescribed with multiple drugs

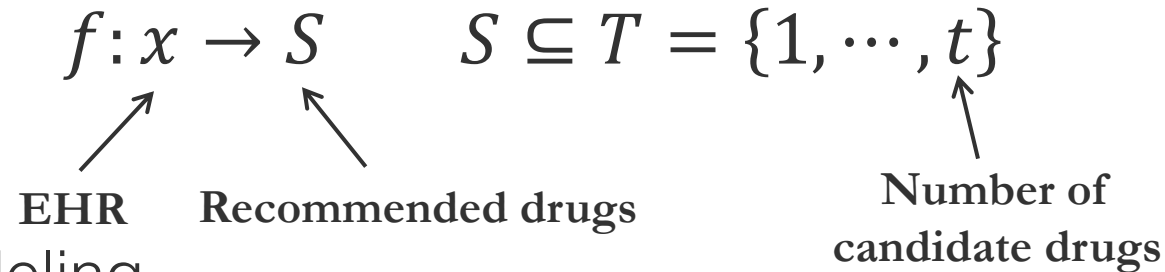


PATIENT_ID	DRUG	DOSE_VAL_RX	DOSE_UNIT_RX	PROD_STRENGTH	STARTDATE	ENDDATE
6	Tacrolimus	2	mg	1mg Capsule	6/11/2175 0:00	6/12/2175 0:00
6	Warfarin	5	mg	5mg Tablet	6/11/2175 0:00	6/12/2175 0:00
6	Heparin Sodium	25,000	UNIT	25,000 unit Premix Bag	6/11/2175 0:00	6/12/2175 0:00
6	D5W	250	ml	HEPARIN BASE	6/11/2175 0:00	6/12/2175 0:00
6	Furosemide	20	mg	20mg Tablet	6/11/2175 0:00	6/12/2175 0:00
6	Warfarin	1	dose	Check with MD for Dose	6/11/2175 0:00	6/15/2175 0:00
6	Heparin Sodium	25,000	UNIT	25,000 unit Premix Bag	6/12/2175 0:00	6/12/2175 0:00
6	D5W	250	ml	HEPARIN BASE	6/12/2175 0:00	6/12/2175 0:00
6	Heparin Sodium	25,000	UNIT	25,000 unit Premix Bag	6/12/2175 0:00	6/13/2175 0:00
6	Warfarin	2	mg	2mg Tab	6/12/2175 0:00	6/13/2175 0:00
6	D5W	250	ml	HEPARIN BASE	6/12/2175 0:00	6/13/2175 0:00
6	Tacrolimus	5	mg	5mg Capsule	6/12/2175 0:00	6/13/2175 0:00
6	Tacrolimus	2	mg	1mg Capsule	6/12/2175 0:00	6/13/2175 0:00
13	Heparin Sodium	25000	UNIT	25,000 unit Premix Bag	1/8/2167 0:00	1/9/2167 0:00
13	Nitroglycerin	100	mg	100MG/250 PM	1/8/2167 0:00	1/9/2167 0:00
13	Docusate Sodium	100	mg	100MG CAP	1/8/2167 0:00	1/9/2167 0:00
13	Atropine Sulfate	0.5	mg	1MG/10ML SYRINGE	1/8/2167 0:00	1/9/2167 0:00
13	Zolpidem Tartrate	5	mg	5mg Tab	1/8/2167 0:00	1/12/2167 0:00
13	Midazolam HCl	2	mg	2MG/2ML VIAL	1/9/2167 0:00	1/9/2167 0:00
13	Nitroglycerin SL	0.3	mg	0.3MG SL TAB	1/9/2167 0:00	1/9/2167 0:00
13	Lorazepam	1	mg	2MG/ML SYR	1/9/2167 0:00	1/9/2167 0:00
13	Magnesium Sulfate	2	gm	1gm/2ml vial	1/9/2167 0:00	1/9/2167 0:00



Treatment Recommendation as a Combinatorial Subset Selection Problem

- Formulation: a combinatorial subset selection problem with relational constraints



- Probabilistic modeling
 - Modeling: define conditional probability $p(S|x)$
 - Training: maximize the conditional likelihood
 - Inference: $f(x) = \operatorname{argmin}_S p(S|x)$



Determinantal Point Process (DPP)

- The DPP defines a probability distribution over subsets, with tractable partition function (Kulesza and Taskar, 2012)
- Given a set of items $\{y_i\}_{i=1}^t$, DPP defines the probability over a subset $S \subseteq T = \{1, \dots, t\}$ as

$$p(S|x) = \frac{\det(L_S)}{\det(L + I)}$$

- L is a $t \times t$ matrix where $L_{ij} = k(y_i, y_j)$
- L_S is a sub-matrix of L indexed by elements in S
- The partition function $\det(L + I)$ is computable in polynomial time



Treatment Recommendation based on Conditional DPP

- Use DPP to define $p(S|x)$
- Model the correlation between drugs
 - Represent drugs
 - Define kernel function over the representations of drugs
- Model the dependency between drug and EHR
 - Define a dependency score function
 - Build the score function into drug-drug kernel function



Drug Representation

- Many web articles contain rich information of drugs

Losartan ⓘ

Generic Name: losartan (loe SAR tan)
Brand Names: Cozaar

[Overview](#) [Side Effects](#) [Dosage](#) [Interactions](#) [Patient Education](#)

What is losartan?
Losartan (Cozaar) belongs to a group of drugs called angiotensin II receptor antagonists. It keeps [blood vessels](#) from narrowing, which lowers blood pressure and improves blood flow.

Losartan is used to treat high [blood pressure](#) (hypertension). It is also used to lower the risk of stroke in certain people with heart disease.

Losartan is used to slow long-term kidney damage in people with type 2 diabetes who also have high blood pressure.

Losartan may also be used for purposes not listed in this medication guide.

Important information
Do not use if you are pregnant. Stop using and tell your doctor right away if you are pregnant or planning to get pregnant, as it can cause injury or death to the unborn baby if you take the medication while pregnant. Use effective birth control.

How should I take losartan?

Take losartan exactly as prescribed by your doctor. Do not take more than the recommended dose or take it more often than recommended. Do not stop taking losartan without talking to your doctor. You may occasionally change your dose to make sure you are getting the most benefit from your medication or to avoid side effects or for longer than recommended.

You may take losartan with or without food.

Call your doctor if you have ongoing vomiting or diarrhea, as you may easily become dehydrated while taking this medication. This may lead to a serious electrolyte imbalance.

Your [blood pressure](#) will need to be checked often.

It may take 3 to 6 weeks of using losartan before you notice the full effect of the medication. Talk with your doctor about the best way to use the medication as directed.

If you are being treated for high blood pressure, keep your blood pressure under control. High blood pressure often has no symptoms. You may need to use medication to keep your blood pressure under control.

Store at room temperature away from moisture, heat, and light.

See also: [Dosage Information \(in more detail\)](#)

Losartan Side Effects

[Overview](#) [Side Effects](#) [Dosage](#) [Interactions](#)

In Summary

Commonly reported side effects of losartan include: chest pain, dizziness, weakness, and fatigue. **Other side effects include:** hypotension, hyperkalemia, and renal impairment. For a comprehensive list of adverse effects, see the full list of side effects.

For the Consumer

Applies to losartan: oral tablet

In addition to its needed effects, some unwanted effects may be caused by losartan. In the event that any of these side effects do occur, they may require medical attention.

Major Side Effects

You should check with your doctor immediately if any of these side effects occur when taking losartan:

More common:

- Abdominal or stomach pain

What conditions/diseases the drug can treat

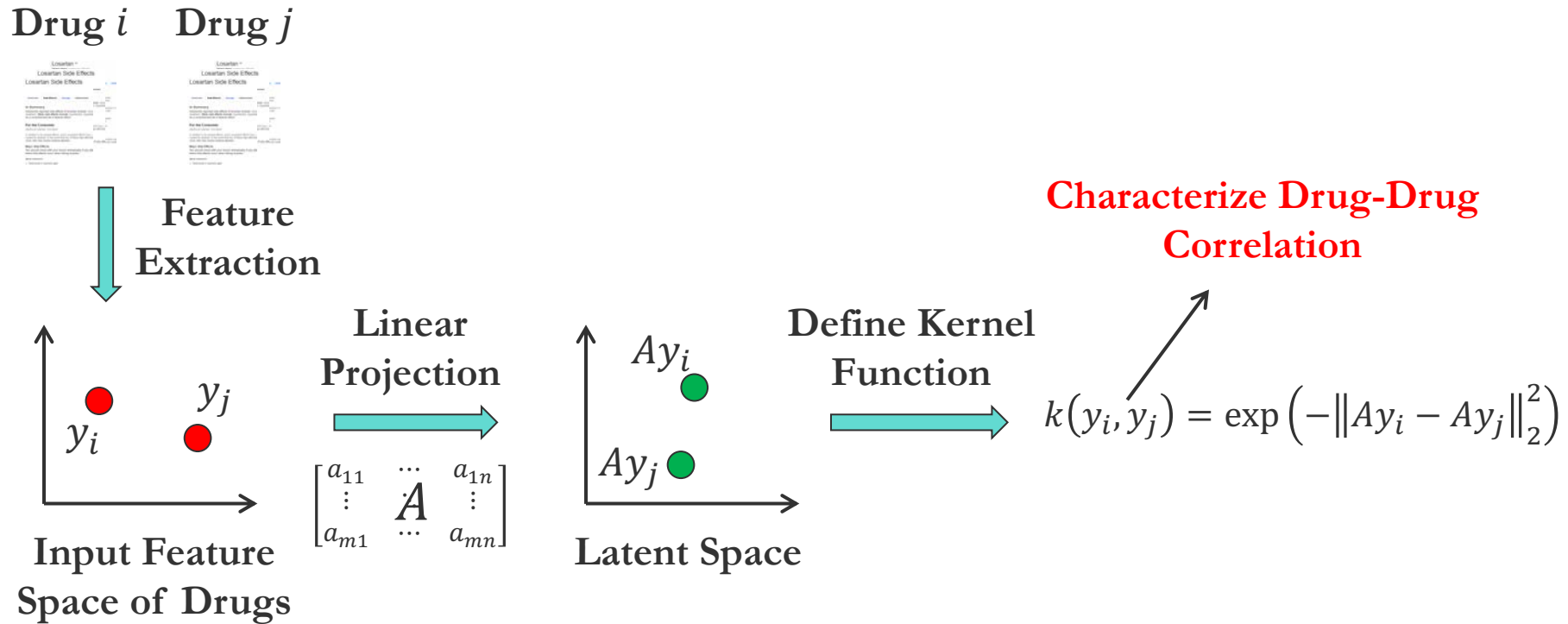
How to take the drug

Side Effect

- Use these articles to represent drugs

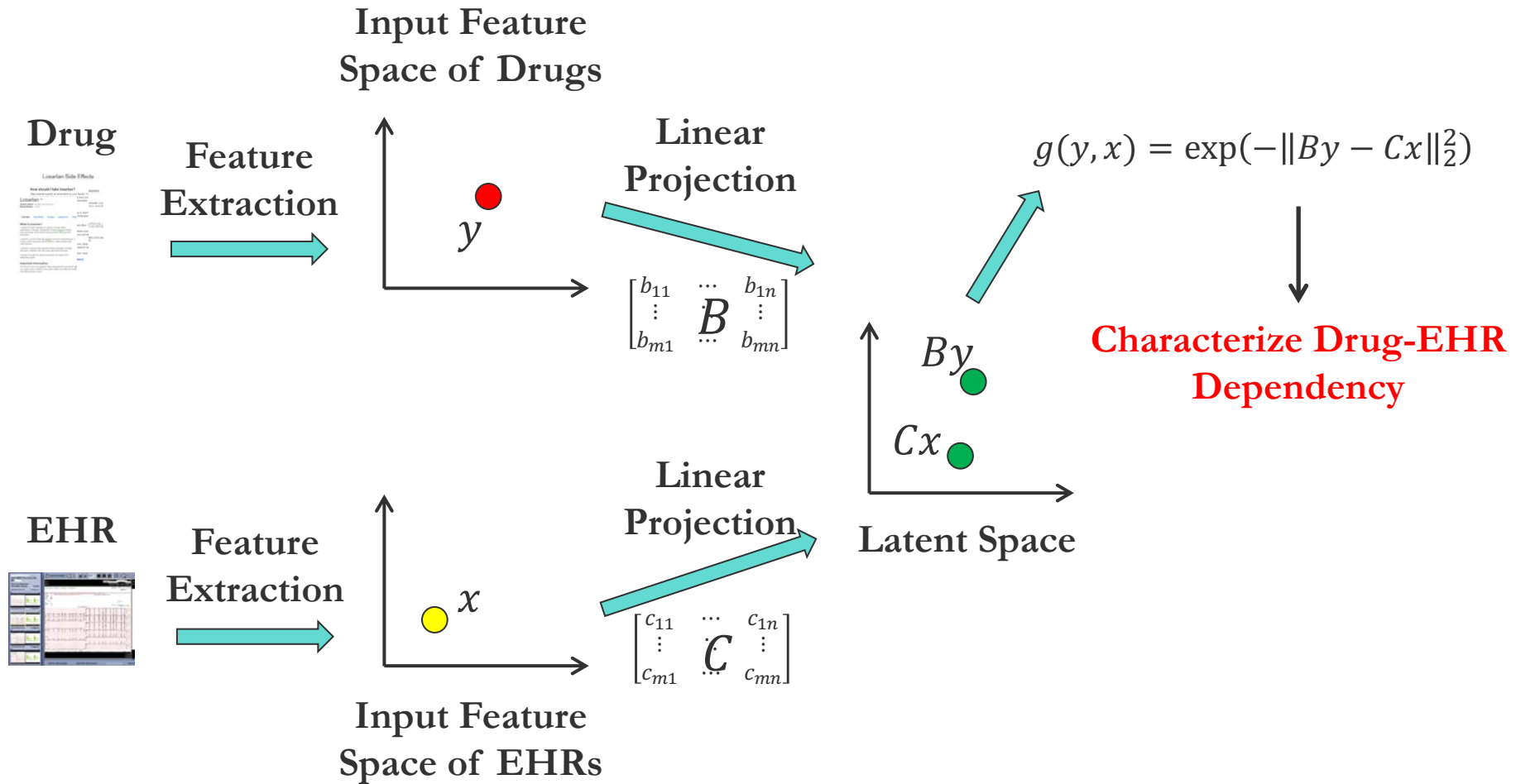


Characterize Correlation between Drugs



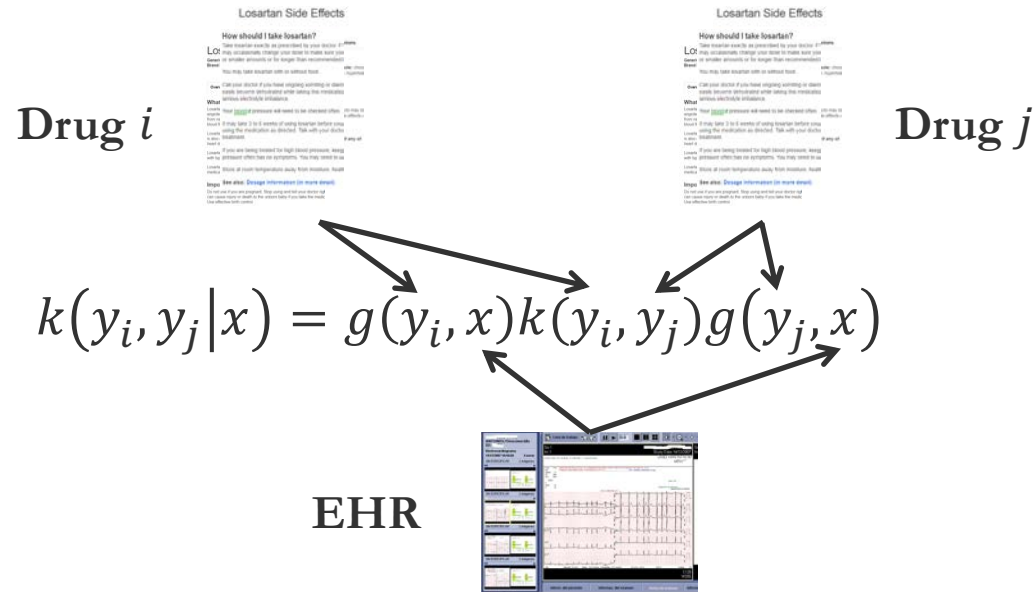


Characterize Dependency between Drug and EHR





Simultaneously Characterize Drug-Drug Correlation and Drug-EHR Dependency



A conditional DPP:
$$p(S|x) = \frac{\det(L_S(x; A, B, C))}{\det(L(x; A, B, C) + I)}$$

where
$$L_{ij}(x) = g(y_i, x)k(y_i, y_j)g(y_j, x) = \exp\left(-\|By_i - Cx\|_2^2 - \|Ay_i - Ay_j\|_2^2 - \|By_j - Cx\|_2^2\right)$$



Parameter Learning

- Maximum Likelihood Estimator
 - N patients, each has an EHR x and a set of prescribed drugs S_i

$$\max_{A,B,C} L(\{(x_i, S_i)\}_{i=1}^N)$$

$$L(\{(x_i, S_i)\}_{i=1}^N) = \prod_{i=1}^N p(S_i|x_i) = \prod_{i=1}^N \frac{\det(L_{S_i}(x_i; A, B, C))}{\det(L(x_i; A, B, C) + I)}$$



Rich Interactions between Drugs

- When used together, one drug can affect the activity of another in different ways:
 - **Antagonism** (adverse interaction): two drugs may lead to a decrease in the beneficial effects and bring in adverse effects
 - E.g. *Antacids* can prevent *antibiotics* from being absorbed into the blood stream
 - Antagonistic drugs are discouraged to be used together
 - **Synergy** (beneficial interaction): two drugs work together to cause a positive effect greater than the sum of its parts
 - E.g. When mixed with *ibuprofen*, *codeine* is more effective in relieving pain
 - Synergic drugs are preferred to be used together

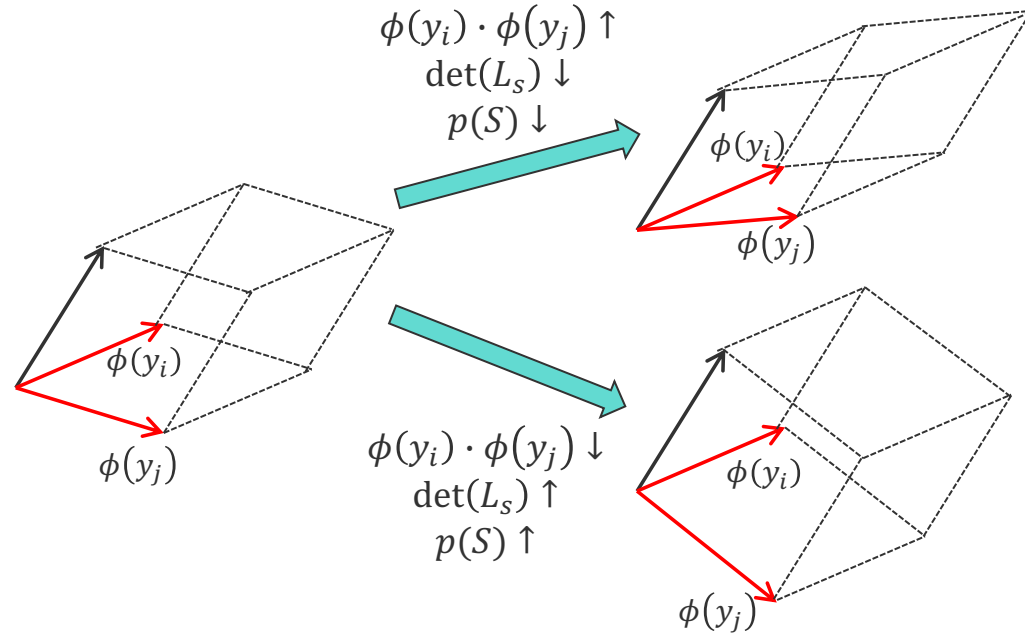


En-/Dis-couraging Co-selection

$$p(S) \propto \det(L_S) = \begin{vmatrix} \phi(y_1) \cdot \phi(y_1) & \cdots & \phi(y_1) \cdot \phi(y_S) \\ \vdots & \ddots & \vdots \\ \phi(y_S) \cdot \phi(y_1) & \cdots & \phi(y_S) \cdot \phi(y_S) \end{vmatrix}$$



Volume of the parallelepiped formed by $\{\phi(y_i)\}_{i=1}^{|S|}$



To discourage the co-selection of i and j , increase $\phi(y_i) \cdot \phi(y_j)$

To encourage the co-selection of i and j , decrease $\phi(y_i) \cdot \phi(y_j)$

$$\max_{A,B,C} L(\{(x_i, S_i)\}_{i=1}^N) + \lambda \left(\sum_{(i,j) \in AR} k(y_i, y_j) + \sum_{(i,j) \in SR} -k(y_i, y_j) \right)$$

↑
↑
 Antagonism Relation Synergy Relation



Algorithm for Parameter Learning

- Stochastic gradient descent

- Log-likelihood $\log L(\{(x_i, S_i)\}_{i=1}^N) = \sum_{i=1}^N (\log \det(L_{S_i}(x_i; A, B, C)) - \log \det(L(x_i; A, B, C) + I))$
- Major bottleneck of computing gradient

$$\frac{\partial \log \det(L(x_i; A, B, C) + I)}{\partial L(x_i; A, B, C)} = (L(x_i; A, B, C) + I)^{-1} \longleftarrow \text{Matrix inverse, cubic complexity}$$

- Method of inducing points (Silverman, 1985)

- Introduce a set of inducing point $\{(u_k)\}_{k=1}^r$ where $(r \ll t)$
- Reparameterize the drug-drug kernel

$$k(y_i, y_j) = v_i^T U v_j$$

where $v_i \in \mathbb{R}^r$ and $v_{ik} = \exp(-\|Ay_i - Au_k\|_2^2)$; $U \in \mathbb{R}^{r \times r}$ and $U_{kl} = \exp(-\|Au_k - Au_l\|_2^2)$

- Compute the low-rank kernel matrix $L(x; A, B, C) = V^T U V$

where $V \in \mathbb{R}^{r \times t}$ and the i th column of V is $g(y_i, x)v_i$ with $g(y_i, x) = \exp(-\|By_i - Cx\|_2^2)$



Drug Recommendation: Mode Inference

- Given the input features x , the recommended drug set is:

$$S^* = \operatorname{argmax}_{S \subseteq T} \log \det(L_S(x))$$

- Exponentially many subsets; NP hard
- Approximate inference (Gillenwater et al., 2012)

- Continuous relaxation

$$F(x) = \log \sum_{S \subseteq T} \prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i) \det(L_S(x)) = \log \det(\operatorname{diag}(x)(L - I) + I) \quad \text{where } x \in [0,1]^t$$

- Find the optimal solution \mathbf{x}^* of $F(\mathbf{x})$ using gradient method

$$\frac{\partial F(x)}{\partial x_i} = \operatorname{tr}((\operatorname{diag}(x)(L - I) + I)^{-1}(L - I)_i)$$

- $(\operatorname{diag}(x)(L - I) + I)^{-1}$ can be efficiently computed using the Sherman–Morrison–Woodbury formula since L is low-rank
- $(L - I)_i$ denotes the matrix obtained by zeroing all except the i -th row of $L - I$
- Rounding: Drug i is selected if $\operatorname{round}(x_i) = 1$ and is not selected if $\operatorname{round}(x_i) = 0$



Summary of Algorithm

- Training
 - Using the inducing point method to perform low-rank re-parameterization of the kernel matrix
 - Using Sherman–Morrison–Woodbury formula to compute matrix inverse
 - Using stochastic gradient method to learn model parameters
- Inference
 - Relaxing the 0-1 programming problem into a continuous one
 - Finding the optimal solution of the relaxed problem
 - Rounding the continuous solution to $\{0,1\}$



Feature Extraction from MIMIC-III

- Diagnostic results
 - 2833 diseases
- Demographics
 - Gender, age
- Clinical notes
 - 5000-dimensional bag-of-words representation
 - 300-dimensional Word2Vec representation
- Lab tests
 - 635 test items
 - Zero, first, second – order time series features
- Total feature dimension: 7207
- 57461 EHRs, 40000 for training, 17461 for testing



Drug Representation and Relations

- 1687 drugs
- Drug representation
 - For each drug, crawl its “profile” articles from Drugs.com, including *overview, side effects, dosage, professional*
 - 5000-dimensional bag-of-words representation
 - Words are filtered using Unified Medical Language System (UMLS)
 - Weighted using TF-IDF
 - 300-dimensional Word2Vec representation
- Drug relation
 - 8106 antagonism relations obtained from Drugs.com
 - 1372 synergy relations obtained from Drugs.com



Summary of Experimental Settings

$$\max_{A,B,C} L(\{(x_i, S_i)\}_{i=1}^N) + \lambda \left(\sum_{(i,j) \in AR} k(y_i, y_j) + \sum_{(i,j) \in SR} -k(y_i, y_j) \right)$$

$$p(S|x) = \frac{|L_S(x; A, B, C)|}{|L(x; A, B, C) + I|} \quad L_{ij}(x) = \exp \left(-\frac{1}{\gamma_2} \|By_i - Cx\|_2^2 - \frac{1}{\gamma_1} \|Ay_i - Ay_j\|_2^2 - \frac{1}{\gamma_2} \|By_j - Cx\|_2^2 \right)$$

N	# Training data	40,000
AR	# Antagonism relations	8,106
SR	# Synergy relations	1,372
Dim(x)	Dim. of EHR features	10,040
Dim(y)	Dim. of drug features	5,300
Size(L)	Size of kernel matrix	1687 x 1687 (# Drugs)
Size(A)	Size of projection matrix	500 x 5300
Size(B)	Size of projection matrix	500 x 5300
Size(C)	Size of projection matrix	500 x 10040

→
Dim. of latent space



Recommendation Performance

Methods			F1(%)
Baselines	Logistic Regression		55.8
	Neural Network		57.2
	Structured Energy Network (Belanger and McCallum, 2016)		59.1
Ours	Using Drug Documents	Using Drug Interactions	
1	No	No	62.9
2	Yes	No	67.5
3	Yes	Yes	71.4

- Compare baselines and 1: Conditional DPP performs better than other subset selection methods
- Compare 1 and 2: it is very helpful to incorporate external medical articles to enrich the representation of drugs
- Compare 2 and 3: leveraging drug interaction relations can greatly improve the quality of recommendation



Example: Effect of Incorporating Medical Articles for Drug Representation



Patient 4655 in
MIMIC-III

Diagnosis Results

Recommended to



Lorazepam

Article of Lorazepam
from Drugs.com

Bipolar disorder

Lorazepam is used to
treat anxiety disorders

Matches

Lorazepam [Ⓢ]

Generic Name: lorazepam (oral) (lor A ze pam)
Brand Names: Ativan

Overview Side Effects Dosage Interactions P

What is lorazepam?

Lorazepam belongs to a group of drugs called benzodiazepines. It affects chemicals in the brain that may be unbalanced in people with anxiety.

Lorazepam is used to treat anxiety disorders

Lorazepam may also be used for purposes not listed in this medication guide.

- Without incorporating external articles, the system didn't recommend Lorazepam to this patient.



Example: Effect of Incorporating Drug Interaction Relations



Patient 3026 in
MIMIC-III

**Recommended
drugs** before
incorporating
interaction relations

Haloperidol
Ketoconazole

**Recommended
drugs** after
incorporating
interaction relations

Haloperidol

**Adverse interaction
between Haloperidol
and Ketoconazole**

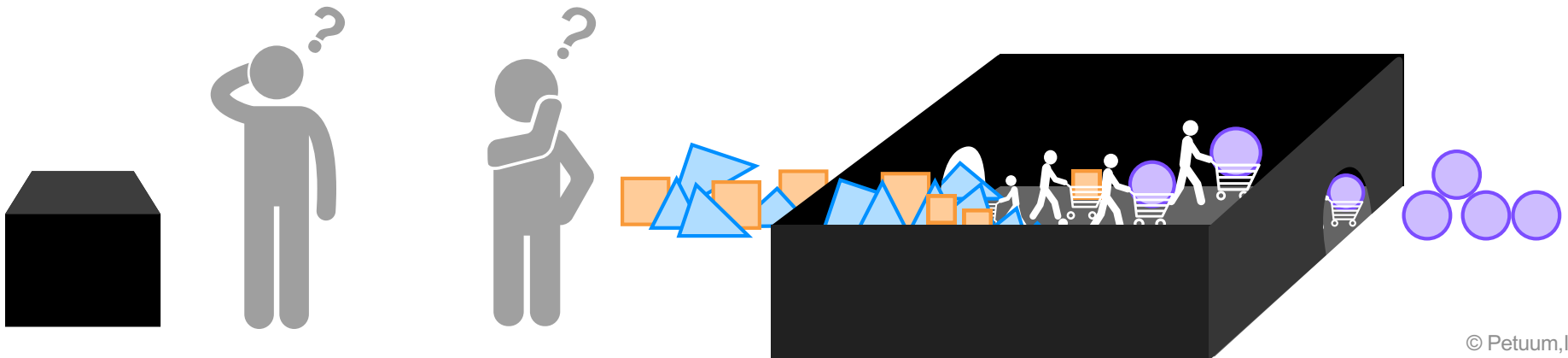
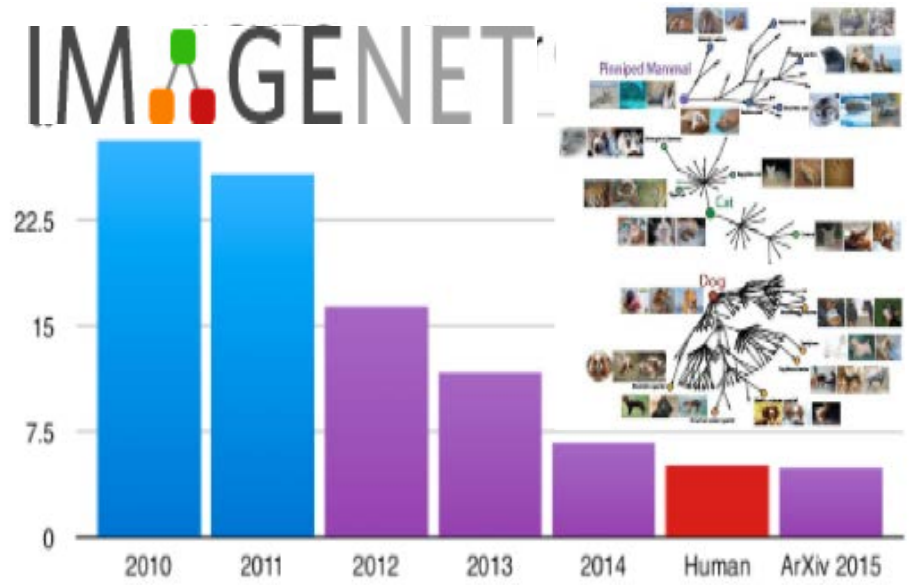
Using haloperidol together
with ketoconazole can
increase the risk of an
irregular heart rhythm that
may be serious and
potentially life-threatening



Toward New System for Automatic ML



AI as we see it now ...





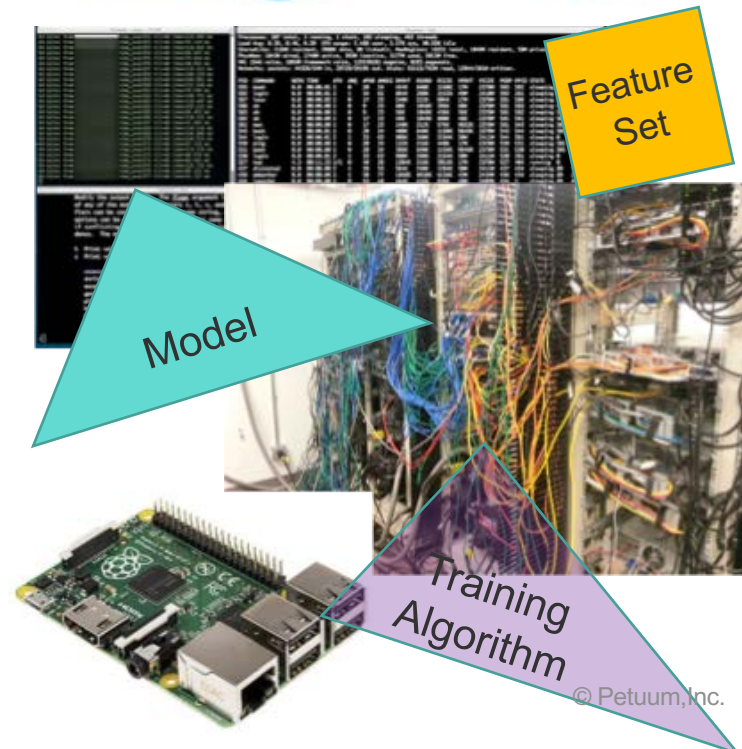
Petuum Simplifies Model/Algorithm Development

Tune/modify models
(e.g, hyperparameter, regularizer)

Experiment with
(changes/alternatives/tricks)

- Distributed training
- Model compiling
- Resource provisioning
- Fault recovery
- Model versioning
- Feature engineering
- Data cleaning
- Software Libraries
- Different Datacenter/IoT Hardware
- Connecting Devices
- ML models debugging
- Learning rate schedules
- Network protocols
- ...

DIY “tuning”
 (“tuning” AND all else)





Petuum Simplifies Model/Algorithm Development

Petuum dev
(nothing but, e.g., tuning)



ML Building Blocks

ML Engine

Data Machine

PetuumOS

Tune/modify models

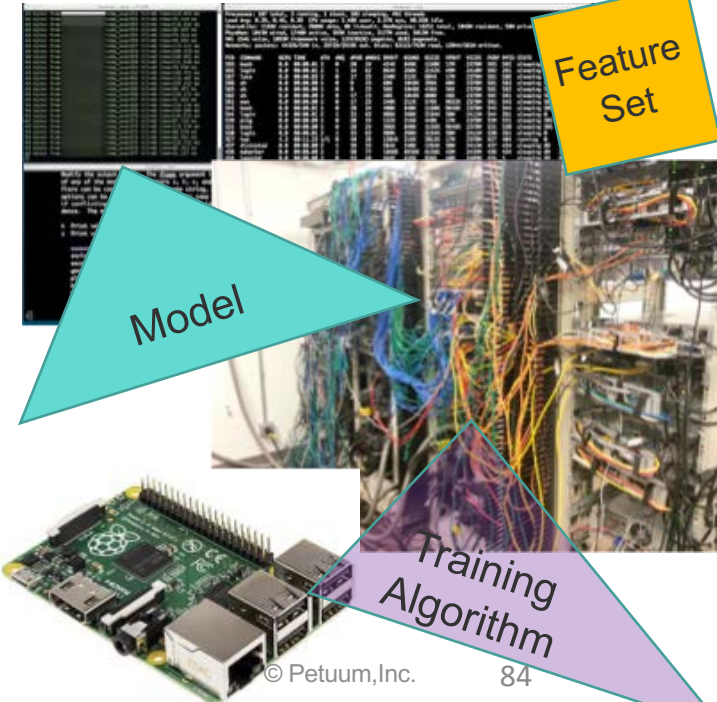
(e.g, hyperparameter, regularizer)

Experiment with

(changes/alternatives/tricks)

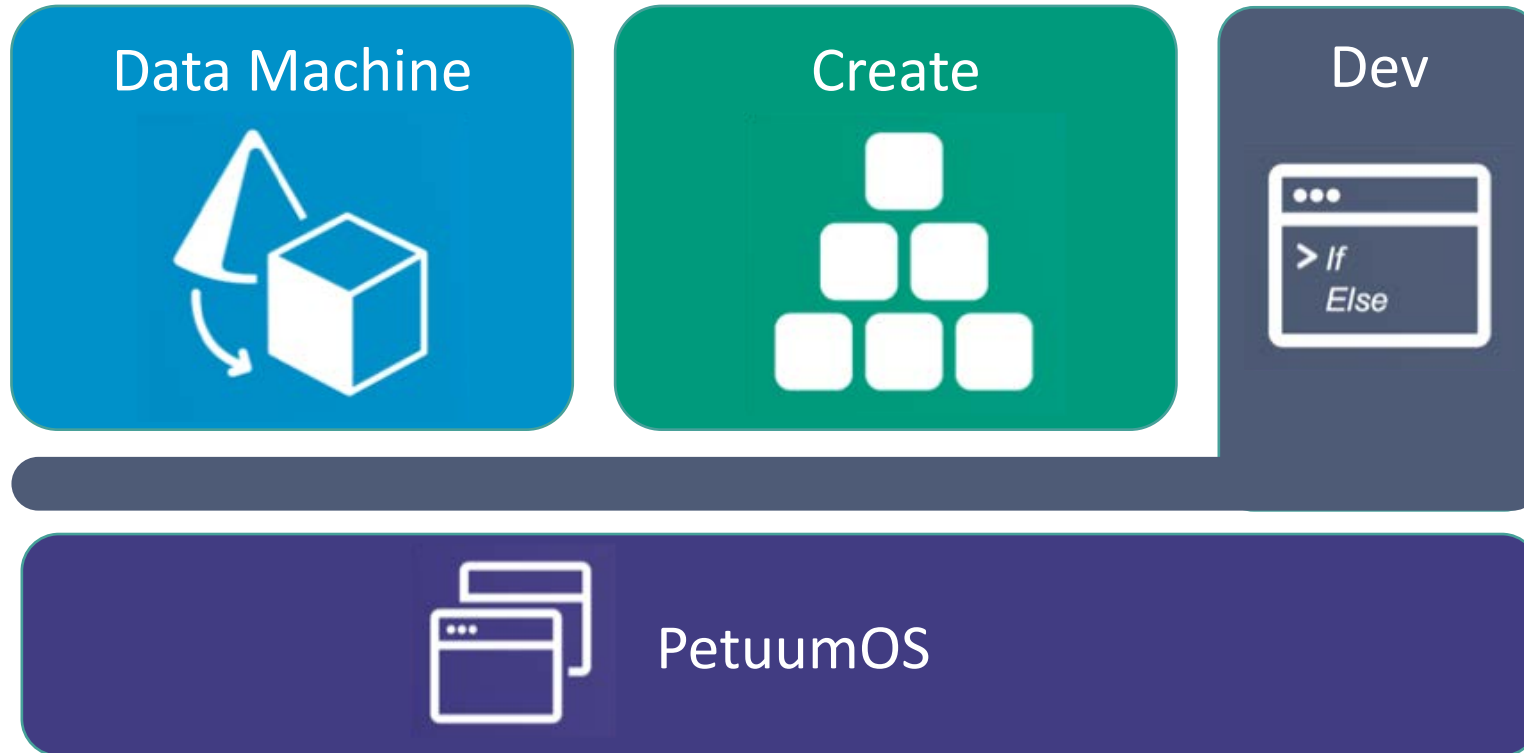
- Distributed training
- Model compiling
- Resource provisioning
- Fault recovery
- Model versioning
- Feature engineering
- Data cleaning
- Software Libraries
- Different Datacenter/IoT Hardware
- Connecting Devices
- ML models debugging
- Learning rate schedules
- Network protocols
- ...

DIY “tuning”
 (“tuning” AND all else)



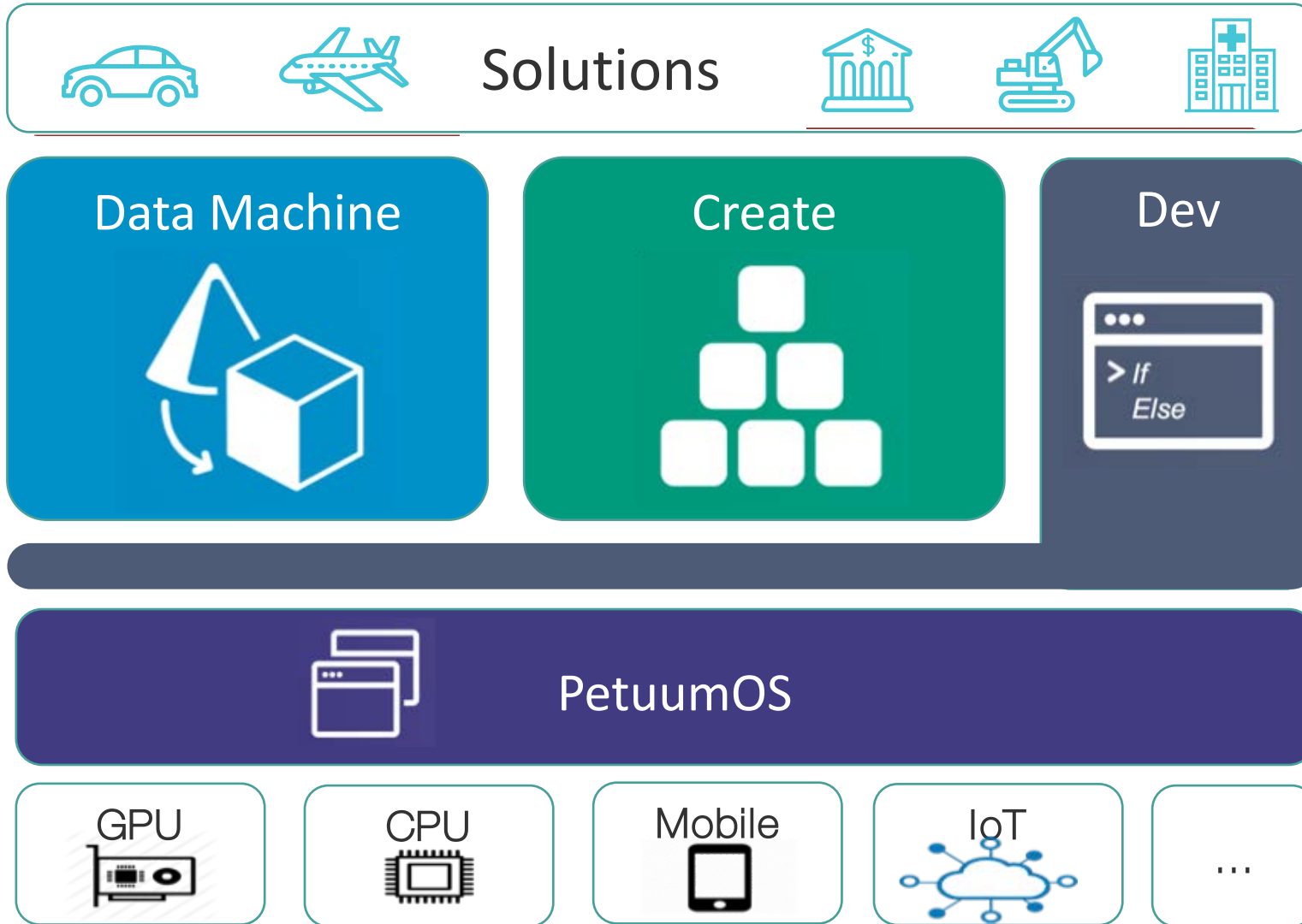


Petuum Symphony





Petuum Symphony





Stack Building Blocks to Create Complex Applications

- Implementing a state-of-the-art ML solutions as simple as stacking available components
 - Traditionally: thousands of lines of code
- Emphasize modularization and abstraction
- Clean and simple interfaces for users to assemble models
- Broad coverage of both common and cutting-edge models



Hundreds of thousands of lines of code



With the toolbox





PetuumMed: AI-aided Healthcare

- ❑ State-of-the-art healthcare AI embedded in existing clinical workflows
- ❑ Leverage all available clinical inputs: medical notes, images, vitals, lab tests, genomic data, and more
- ❑ Productivity: offers distilled patient info & second opinion at practitioners' fingertip
- ❑ Deploys & scales with Petuum Symphony



AI Meets Applications At Fingertips

– Experiencing & Using AI at Any Scale Like Never Before with Petuum UI



The background is a vibrant blue with a central graphic of a globe. From the globe, numerous thin white lines radiate outwards, creating a sunburst effect. Interspersed among these lines are various mathematical symbols and formulas in a light blue color, including $\pi = 3.14$, e^{-x} , $x=y$, 1001 , $(\cos(\pi-x))$, $a+b$, $2x+5y=-1$, $</>$, $(b^2-4ac)/2a$, $2\sin x$, 01010001 , \sum , $a_n \cos$, $f(x) = -2x^2$, $a=b$, \int , $\frac{3}{4}$, $(x \pm y) \pm x$, $(\cos(\pi-x))$, $\forall a.T$, N_e , $(2\pi)^k$, π , and $x \pm y \pm x$.

Thank You!