Advances in Mining Heterogeneous Healthcare Data

Fenglong Ma, Muchao Ye, Junyu Luo, Cao Xiao & Jimeng Sun

KDD2021

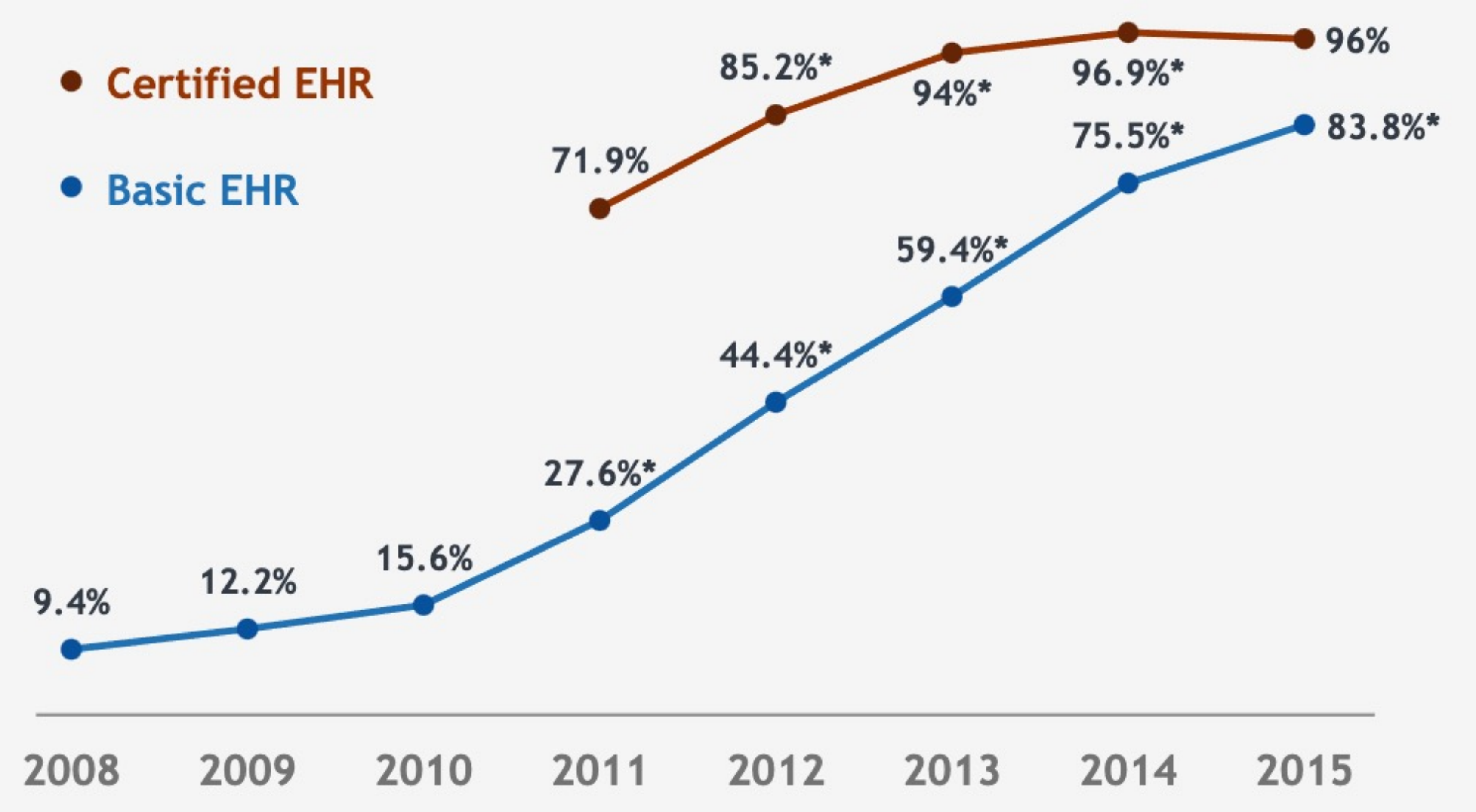
Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

Electronic Health Record (EHR)

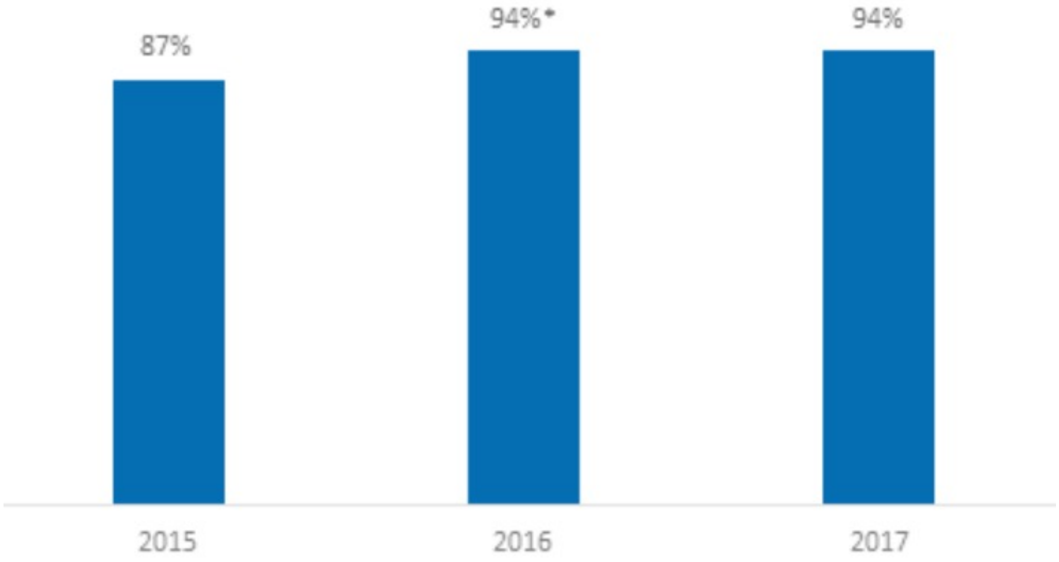
- A longitudinal record of patient health information generated by one or several encounters in any healthcare providing setting.

Adoption of Electronic Health Record Systems among U.S. Hospitals: 2008-2015

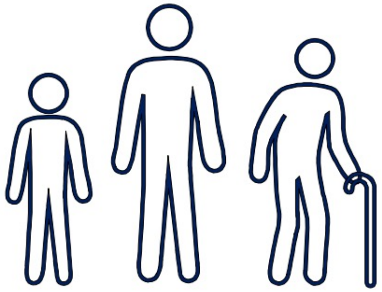


Hospitals' Use of Electronic Health Records Data, 2015-2017

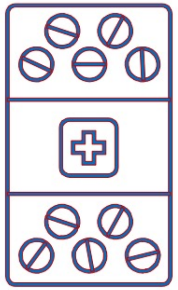
- As of 2017, 94 percent of hospitals used their EHR data to perform hospital processes that inform clinical practice.
- EHR data is most commonly used by hospitals to support quality improvement (82 percent), monitor patient safety (81 percent), and measure organization performance (77 percent).



Multiple Data Modalities in the EHR Systems



Demographics



Medications

Clinical Notes and Reports

Continuous Monitoring Data

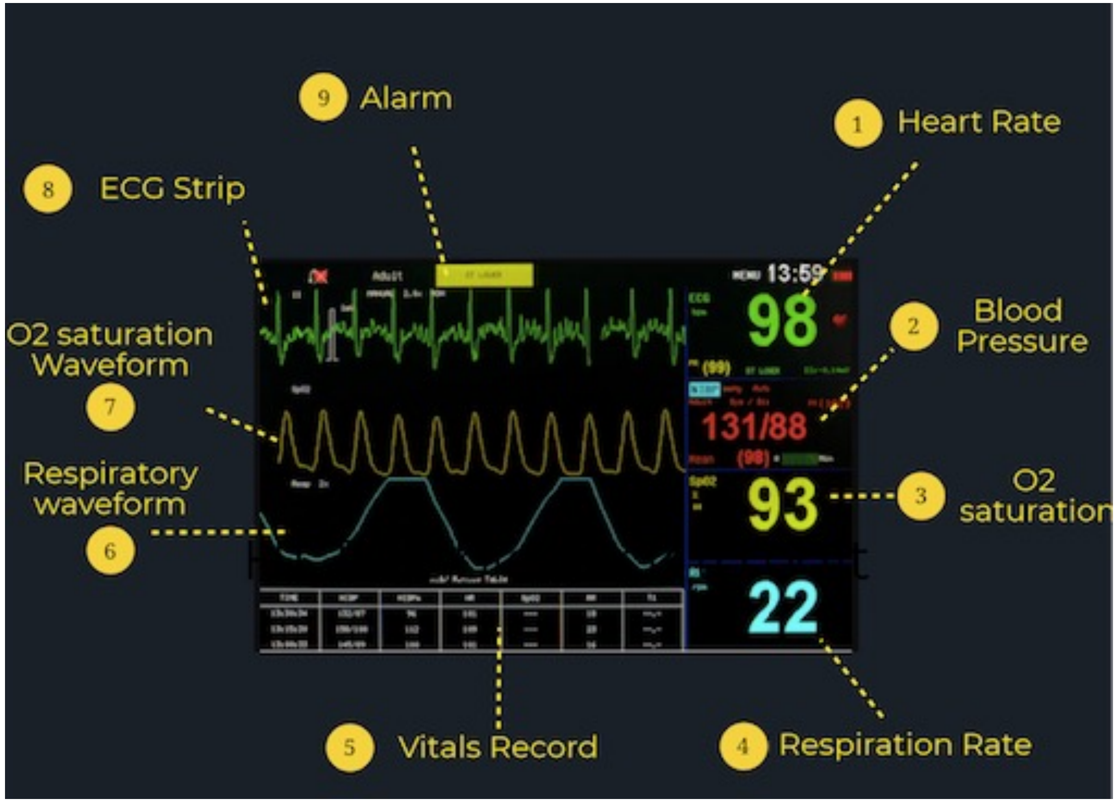
Multi-typed Medical Codes

Medical Images

Types of Data

- Demographics
  - Age, sex, socio-economic status, insurance type, language, religion, living situation, family structure, location, work, …

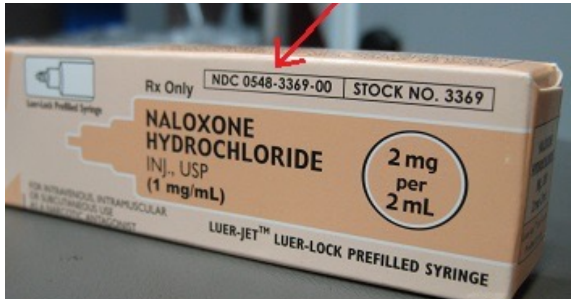
- Continuous Monitoring Data
  - Heart rate, pulse, respiration rate, body temperature, …



https://canadiem.org/how-to-read-patient-monitors/

Types of Data

- Medications
  - Prescriptions, over-the-counter drugs, illegal drugs, alcohol, …
  - Coding system
    - National Drug Code (NDC)
      - Each of sources provides NDC codes in a different format.
    - RxNorm
      - A standardized nomenclature for clinical drugs, is produced by the National Library of Medicine.



NDC	Item Description	rxnorm	rxnormName
00093820401	CIMETID TAB 400MG TEVA 100@	197507	Cimetidine 400 MG Oral Tablet
00781323909	CIPROFLOX I.V.BG2CMG1CMLSAN24@	1665210	100 ML Ciprofloxacin 2 MG/ML Injection
00781323909	CIPROFLOX I.V.BG2CMG1CMLSAN24@	1665210	100 ML Ciprofloxacin 2 MG/ML Injection
16714065301	CIPROFLOX TAB 750MG 50 NSTAR@	197512	Ciprofloxacin 750 MG Oral Tablet
68084006901	CIPROFLOX TB 250MG UD AHP 100@	197511	Ciprofloxacin 250 MG Oral Tablet
00703574811	CISPLATIN AQ 1MG/ML TEV 100ML@	309311	Cisplatin 1 MG/ML Injectable Solution
00039001810	CLAFORAN VIAL 1GM 10	1656316	Cefotaxime 1000 MG Injection
00039001910	CLAFORAN VIAL 2GM 10	1656320	Cefotaxime 2000 MG Injection
62037077760	CLARITHR ER TAB 500MG WAT 60	359385	24 HR Clarithromycin 500 MG Extended Rel

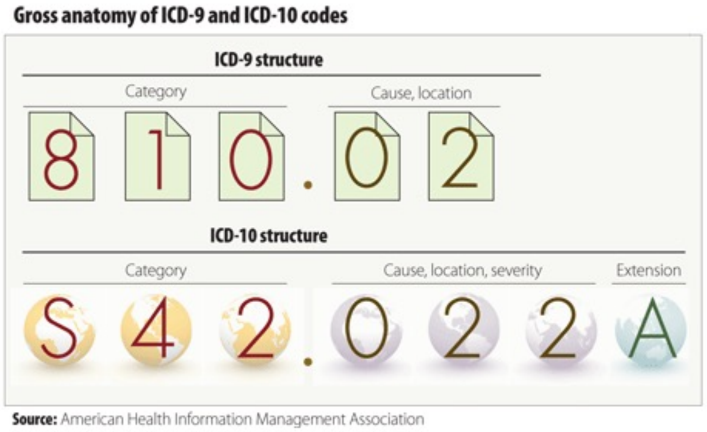
Types of Data

- Laboratory Results
  - Components of blood, urine, stool, saliva, spinal fluid (CSF), ascitic fluid, joint fluid, bone marrow, lung, …
  - Coding System
    - LOINC: Logical Observation Identifiers Names and Codes

Code value	Description
LOINC*	
1558-6	Fasting glucose [Mass/volume, mg/dL] in Serum or Plasma
14771	Fasting glucose [Moles/volume, mmol/L] in Serum or Plasma
1518-0	Glucose [Mass/volume, mg/dL] in Serum or Plasma --2 hr post 75 g glucose PO
14995-5	Glucose [Moles/volume, mmol/L] in Serum or Plasma --2 hr post 75 g glucose PO
2857-1	Prostate specific Ag [Mass/volume, ng/mL] in Serum or Plasma
35741-8	Prostate specific Ag [Mass/volume, µg/L] in Serum or Plasma by Detection limit $<=0.01$ ng/mL
19195-7	Prostate specific Ag [Units/volume, IU/L] in Serum or Plasma
33667-7	Prostate specific Ag protein bound [Mass/volume, ng/mL] in Serum or Plasma
10886-0	Prostate Specific Ag Free [Mass/volume, ng/mL] in Serum or Plasma

Types of Data

- Billing
  - Diagnoses (ICD-{9, 10})
    - International Classification of Diseases
    - The World Health Organization (WHO) currently develops and maintains the list for use by Member States.
  - Procedures (CPT and ICD)
    - CPT (Current Procedural Terminology) codes describe procedures performed
    - The American Medical Association administers and maintains the CPT list.



CPT Code	CPT Code Description	Reimbursement*
CPT Code 99453	Initial set up and patient education on use of equipment.	$21.00 (one-time fee)
CPT Code 99454	Supply of devices, collection, transmission, and report/summary services to the clinician	$69.00
CPT Code 99457	Remote physiologic monitoring services by clinical staff/MD/QHCP for first 20 minutes of RPM services.	$54.00
CPT Code 99458	Remote physiologic monitoring services by clinical staff/MD/QHCP that exceeds first 20 minutes of RPM services	$43.00 (estimation)

*Based on current CMS Physician fee schedules

Types of Data

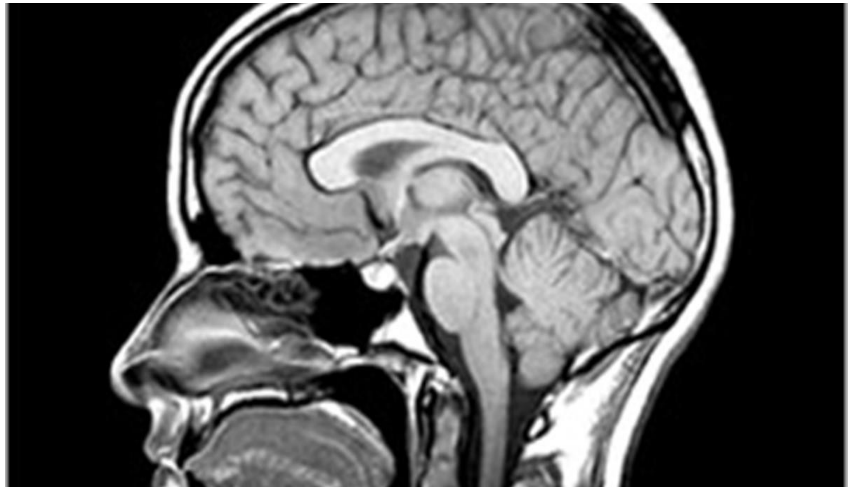
- Clinical Notes
  - Discharge summary
  - Attending and/or Resident
  - Nurse
  - Specialist
    - Radiology, Pathology, ECG, Nutrition, Respiratory, Social work, …
  - Consultant
  - Referring physician
  - Emergency Department

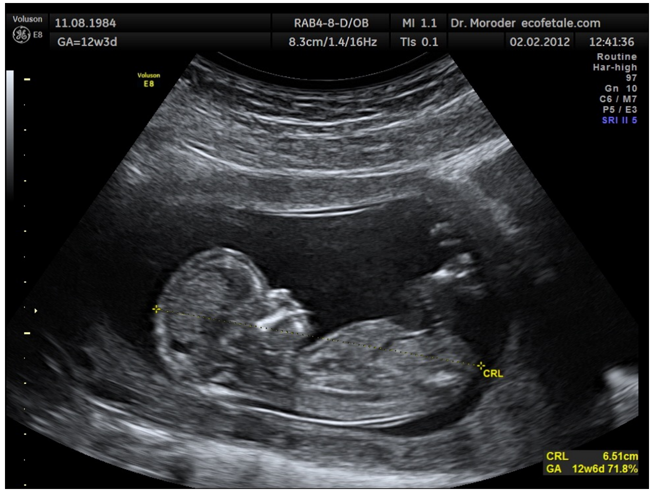
Admission Date :
⟨ deidentified ⟩
Discharge Date :
⟨ deidentified ⟩
Date of Birth :
⟨ deidentified ⟩ Sex :
F
Service :
SURGERY
Allergies :
Patient recorded as having No Known Allergies to Drugs
Attending :
⟨ deidentified ⟩
Chief Complaint :
Dyspnea
Major Surgical or Invasive Procedure :
Mitral Valve Repair
History of Present Illness :
Ms. ⟨ deidentified ⟩ is a 53 year old female who presents after a large bleed rhythmically lag to 2 dose but the patient was brought to the Emergency Department where he underwent craniotomy with stenting of right foot under the LUL COPD and transferred to the OSH on ⟨ deidentified ⟩ .
The patient will need a pigtail catheter to keep the sitter daily .

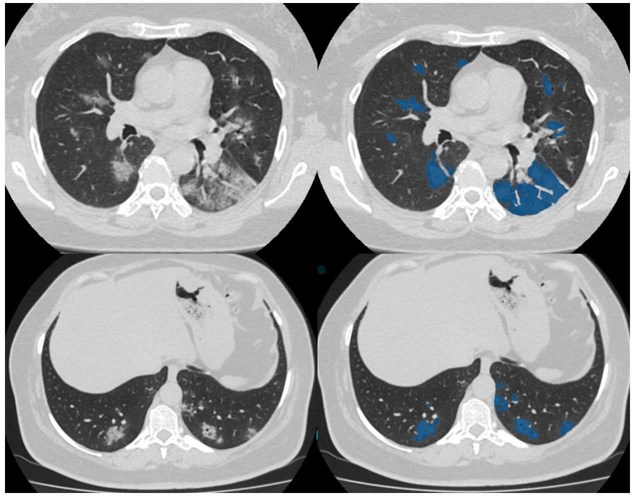
KDD

Types of Data

- Medical Images
  - X-ray
  - Ultrasound
  - CT
  - MRI
  - PET
  - Retinal
  - Endoscopy
  - Photographs

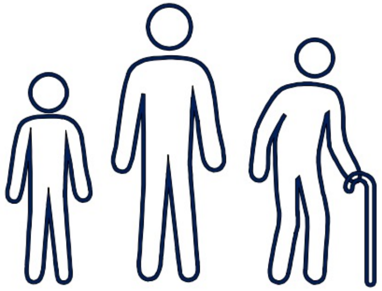




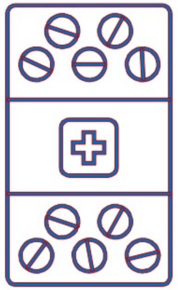


Multiple Data Modalities in the EHR Systems

Structured



Demographics



Medications

Continuous
Monitoring Data

Multi-typed
Medical Codes

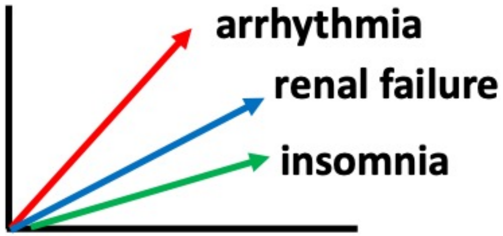
Unstructured

Clinical Notes
and Reports

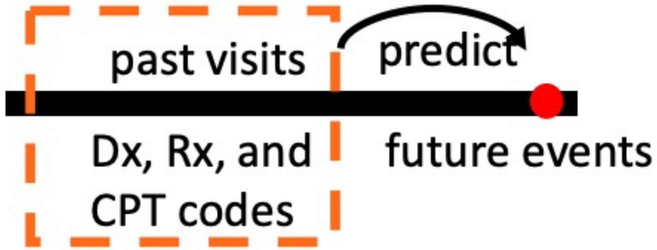
Medical
Images

Analytics Tasks using EHR Data

EHR Data



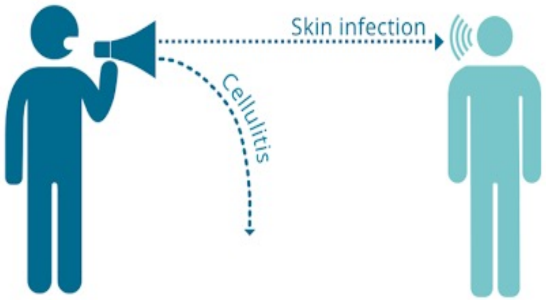
Phenotyping



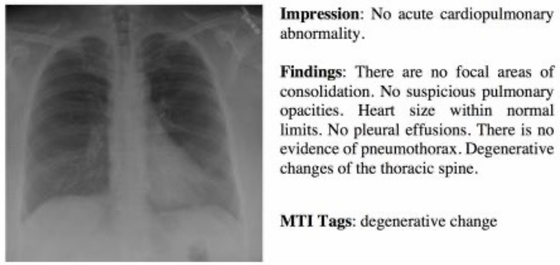
Risk Prediction

Medication Recommendation

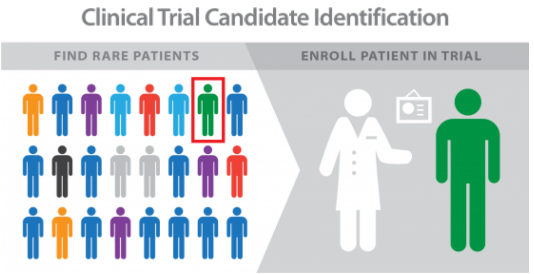
Disease Classification



Understandable Medical Language Translation



Medical Report Generation



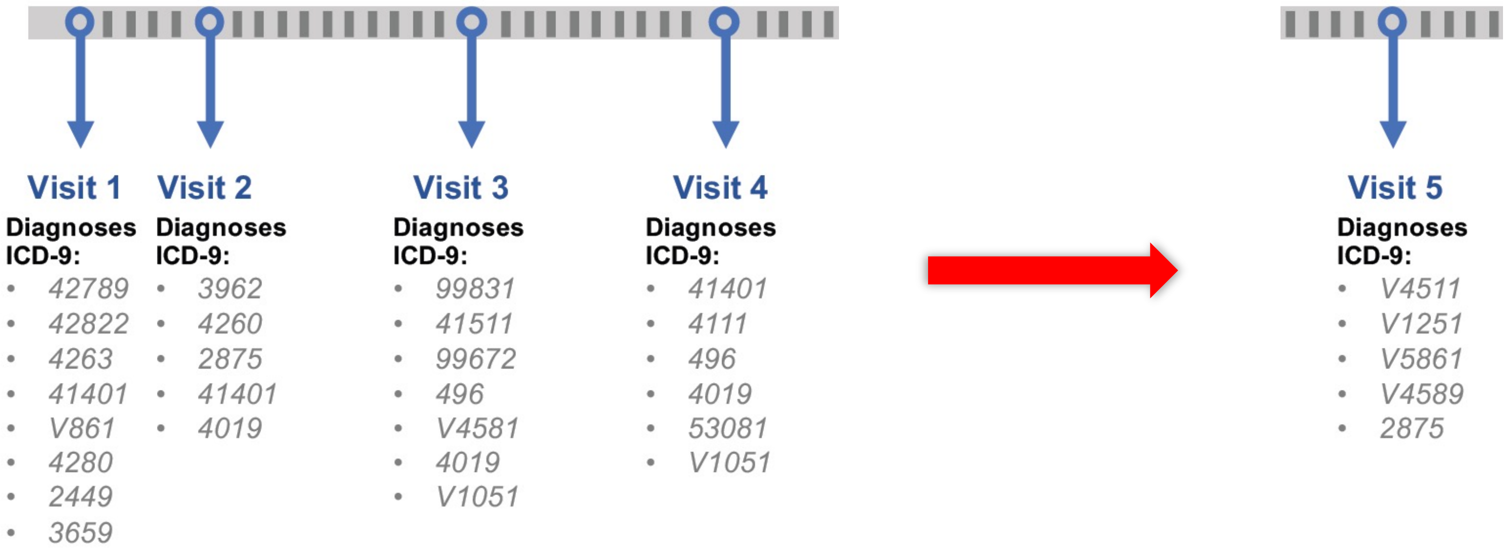
Clinical Trial Mining

Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- **Part I: Mining structured health data**
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

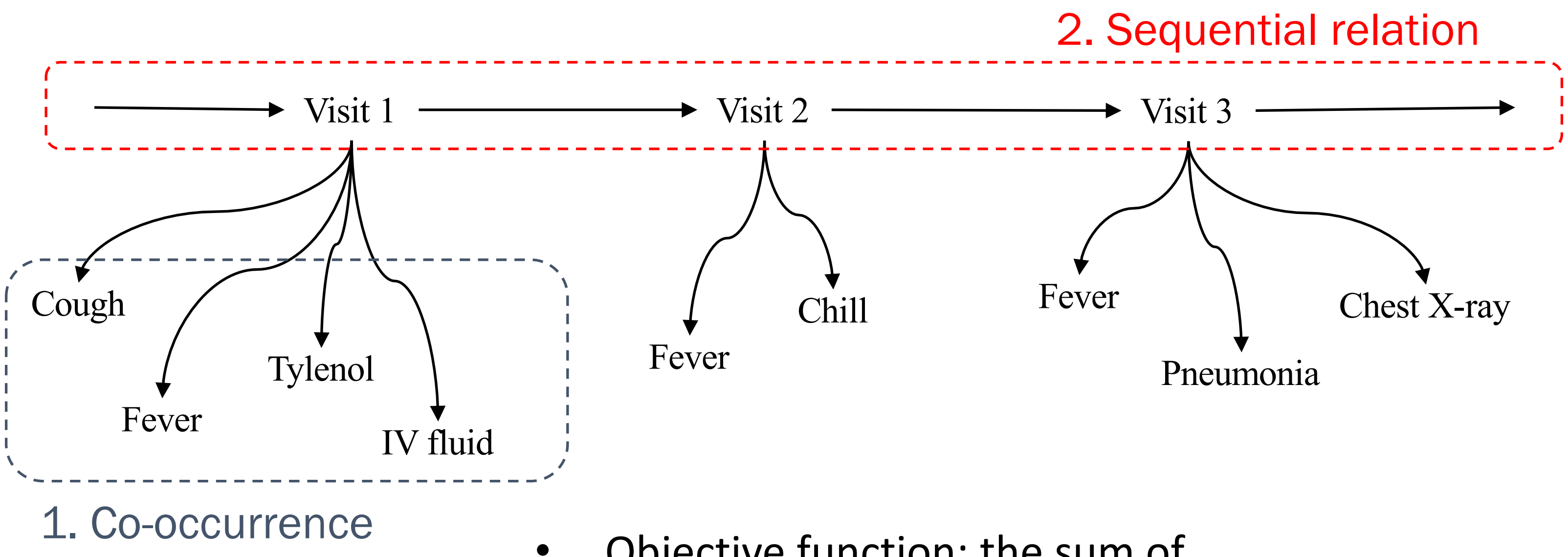
Phenotyping

- Goal: Learning medical concept representations from EHR data
- Approach: Predicting the next visit information according to all the previous visits



Med2Vec

- Two-layered representation learning

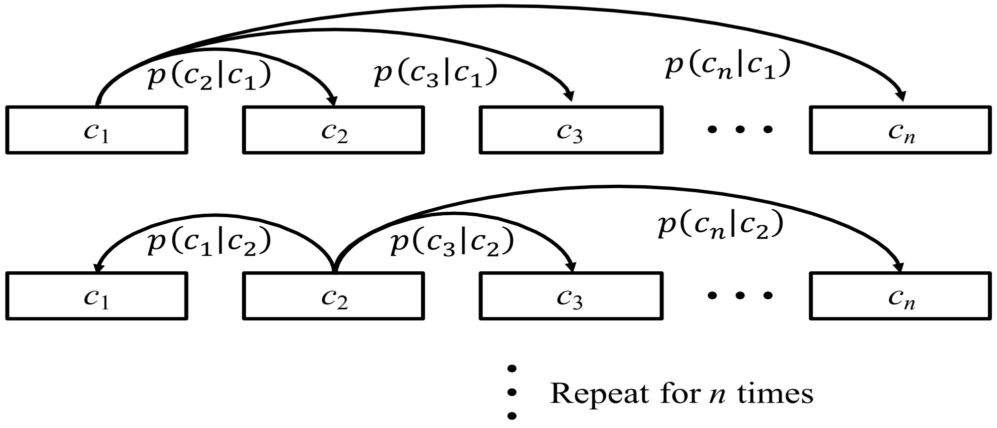


- Objective function: the sum of
  1. Negative intra-visit Skip-gram
     - Because Skip-gram objective function is to be maximized
  2. Inter-visit multi-label classification loss

Intra-visit Skip-gram

- Model all pairs of medical codes in a visit

- Visit contains codes $\{c_1, c_2, c_3, \dots c_n\}$
- c_i: i-th code among the code vocabulary C
- $p(c_i | c_j)$: Skip-gram probability (see below)
- Each code $c_1, c_2, c_3, \dots c_n$ is used as the "input"
- Learn $\boldsymbol{W}_c$, the code representation



$$\max_{\boldsymbol{W}'_c} \frac{1}{T} \sum_{t=1}^{T} \sum_{i:c_i \in V_t} \sum_{j:c_j \in V_t, j \neq i} \log p(c_j|c_i),$$

where
$$p(c_j|c_i) = \frac{\exp\left(\boldsymbol{W}'_c[:, j]^\top \boldsymbol{W}'_c[:, i]\right)}{\sum_{k=1}^{|\mathcal{C}|} \exp\left(\boldsymbol{W}'_c[:, k]^\top \boldsymbol{W}'_c[:, i]\right)}$$

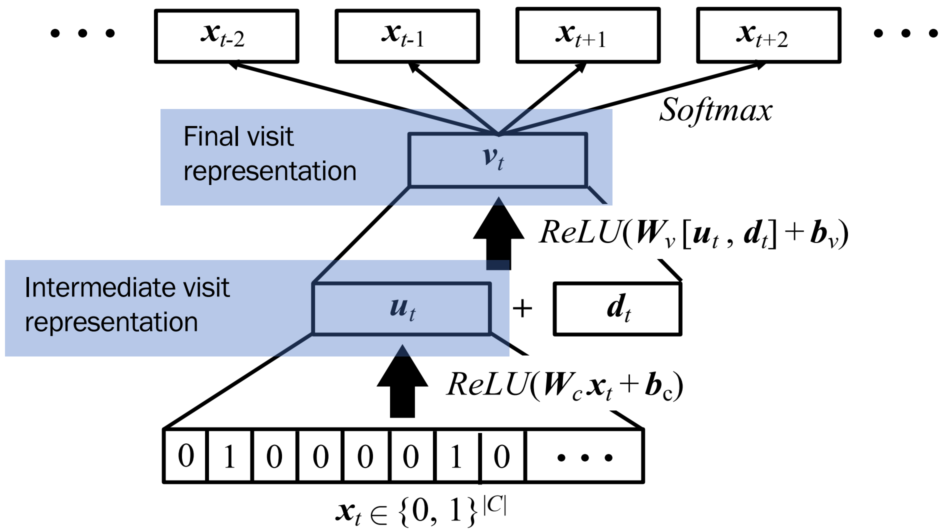
$$\boldsymbol{W}'_c = ReLU(\boldsymbol{W}_c) \in \mathbb{R}^{m \times |\mathcal{C}|}$$

- V_t: t-th visit
- c_i, c_j: codes in the visit V_t
- $[:, j]$: j-th column of the matrix

❖ Choi et al. *Med2Vec: Multi-layer Representation Learning for Medical Concepts*. KDD 2016.

Inter-visit Multi-label Classification Loss

- Model relations between nearby visits



- x_t: one-hot coded Dx, Rx, Pr at time t
- u_t: intermediate visit representation
- d_t: patient demographic information
- v_t: final visit representation
- W_c, W_v, b_c, b_v: weights to learn
- $|C|$: number of unique medical codes

$$\min_{\boldsymbol{W}_s, \boldsymbol{b}_s} \frac{1}{T} \sum_{t=1}^{T} \sum_{-w \le i \le w, i \ne 0} -\boldsymbol{x}_{t+i}^{\top} \log \hat{\boldsymbol{y}}_t - (\boldsymbol{1} - \boldsymbol{x}_{t+i})^{\top} \log(\boldsymbol{1} - \hat{\boldsymbol{y}}_t),$$

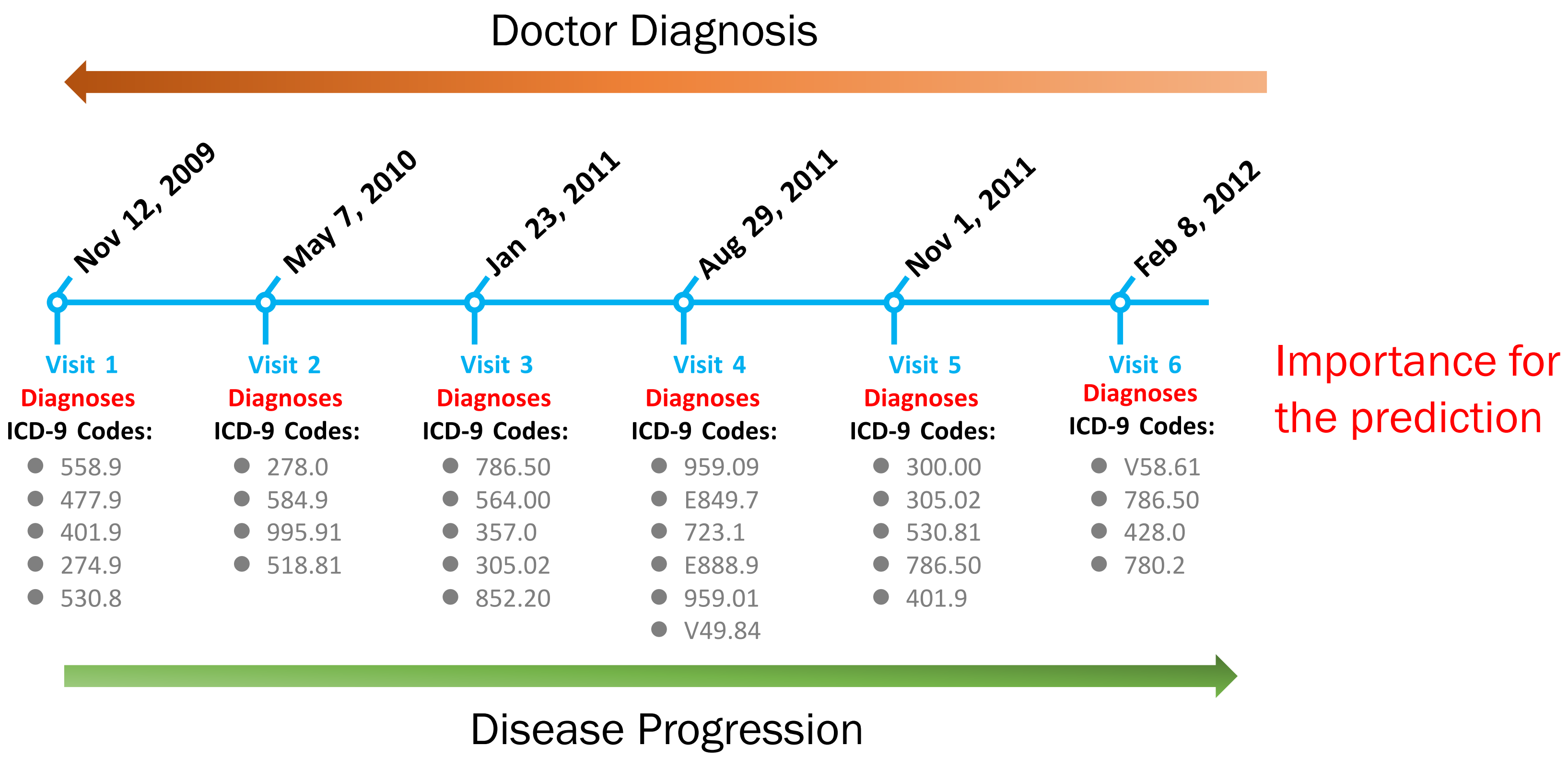
where
$$\hat{\boldsymbol{y}}_t = \frac{\exp(\boldsymbol{W}_s \boldsymbol{v}_t + \boldsymbol{b}_s)}{\sum_{j=1}^{|\mathcal{C}|} \exp(\boldsymbol{W}_s[j,:]\boldsymbol{v}_t + \boldsymbol{b}_s[j])}$$

- T: length of the visit record
- w: context visit window
- $[j,:]$: j-th row of the matrix
- $[j]$: j-th element of the vector
- W_s, b_s: weights for the Softmax

❖ Choi et al. *Med2Vec: Multi-layer Representation Learning for Medical Concepts*. KDD 2016.

Dipole

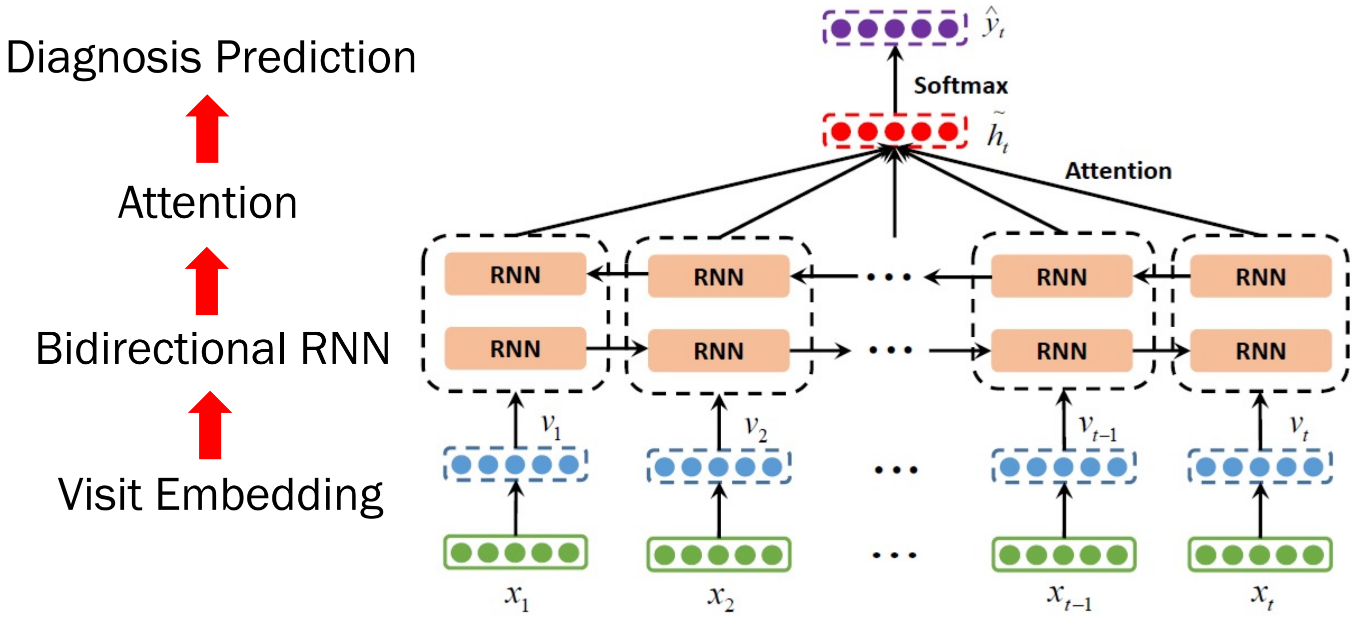
- Imitate doctors' diagnosis procedure + disease progression



❖ Ma et al. *Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks*. KDD 2017.

Dipole

- Motivations:
  - Bidirectional Recurrent Neural Networks (BRNN) to imitate both the procedure of doctor diagnosis and disease progression.
  - The importance of different visits for the final prediction should vary – Attention Mechanism!



❖ Ma et al. *Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks*. KDD 2017.

Interpretation for Code Representations (Diabetes Dataset)

$$\mathbf{v}_i = \mathrm{ReLU}(\boxed{\mathbf{W}_v}\mathbf{x}_i + \mathbf{b}_v) \qquad \mathbf{x}_i \in \{0,1\}^{|C|}$$

Latent Space

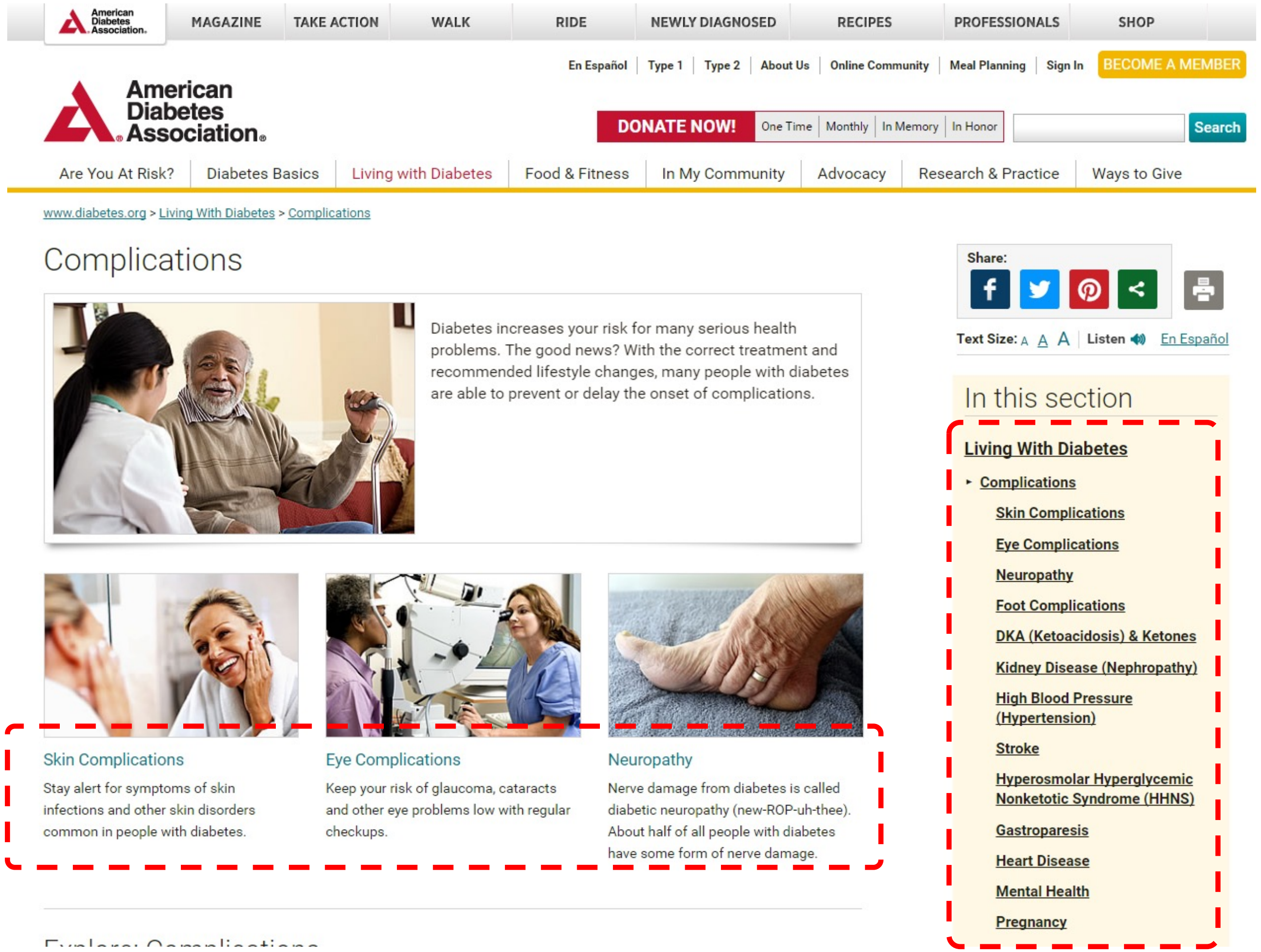
Medical Code

$\mathbf{W}_v$

Eye Complications & Alzheimer's Disease	Neuropathy	Heart Diseases
Coordinate 10	**Coordinate 38**	**Coordinate 77**
Glaucoma (365) Fracture of one or more tarsal and metatarsal bones (825) Dementias (290) Psoriasis and similar disorders (696) Mild mental retardation (317) Cataract (366) Injury, other and unspecified (959) Rheumatoid arthritis and other inflammatory polyarthropathies (714) Thyrotoxicosis with or without goiter(242) Blindness and low vision (369)	Hereditary and idiopathic peripheral neuropathy (356) Other disorders of soft tissues (729) Dermatophytosis (110) Other disorders of urethra and urinary track (599) Mononeuritis of lower limb (355) Diabetes mellitus (250) Mononeuritis of upper limb and mononeuritis multiplex (354) Sprains and strains of sacroiliac region (846) Osteoarthrosis and allied disorders (715) Other and unspecified disorders of back (724)	Cardiac dysrhythmias (427) Chronic pulmonary heart disease (416) Special screening for malignant neoplasms (V76) Angina pectoris (413) Other hernia of abdominal cavity without mention of obstruction (553) Cardiomyopathy (425) Ill-defined descriptions and complications of heart disease (429) Diabetes mellitus (250) Acute pulmonary heart disease (415) Gastrointestinal hemorrhage (578)
Coordinate 79	**Coordinate 141**	**Coordinate 142**
Neurotic disorders (300) Other current conditions in the mother classifiable elsewhere (648) Symptoms concerning nutrition metabolism and development (783) Obesity and other hyperalimentation (278) Diseases of esophagus (530) Other organic psychotic conditions (chronic) (294) Schizophrenic disorders (295) Asthma (493) Chronic liver disease and cirrhosis (571) Spondylosis and allied disorders (721)	Viral hepatitis (070) Other cellulitis and abscess (682) Other personal history presenting hazards to health (V15) Cellulitis and abscess of finger and toe (681) Bacterial infection in conditions classified elsewhere (041) Episodic mood disorders (296) Chronic ulcer of skin (707) Mononeuritis of upper limb and mononeuritis multiplex (354) Other diseases due to viruses and Chlamydiae (078) Diabetes mellitus (250)	Essential hypertension (401) Hypertensive renal disease (403) Hypertensive heart disease (402) Chronic renal failure (585) Other disorders of kidney and ureter (593) Other psychosocial circumstances (V62) Secondary hypertension (405) Nonspecific abnormal results of function studies (794) Calculus of kidney and ureter (592) Other organic psychotic conditions (chronic) (294)
Mental Health	Skin Complications	High Blood Pressure

❖ Ma et al. *Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks*. KDD 2017.

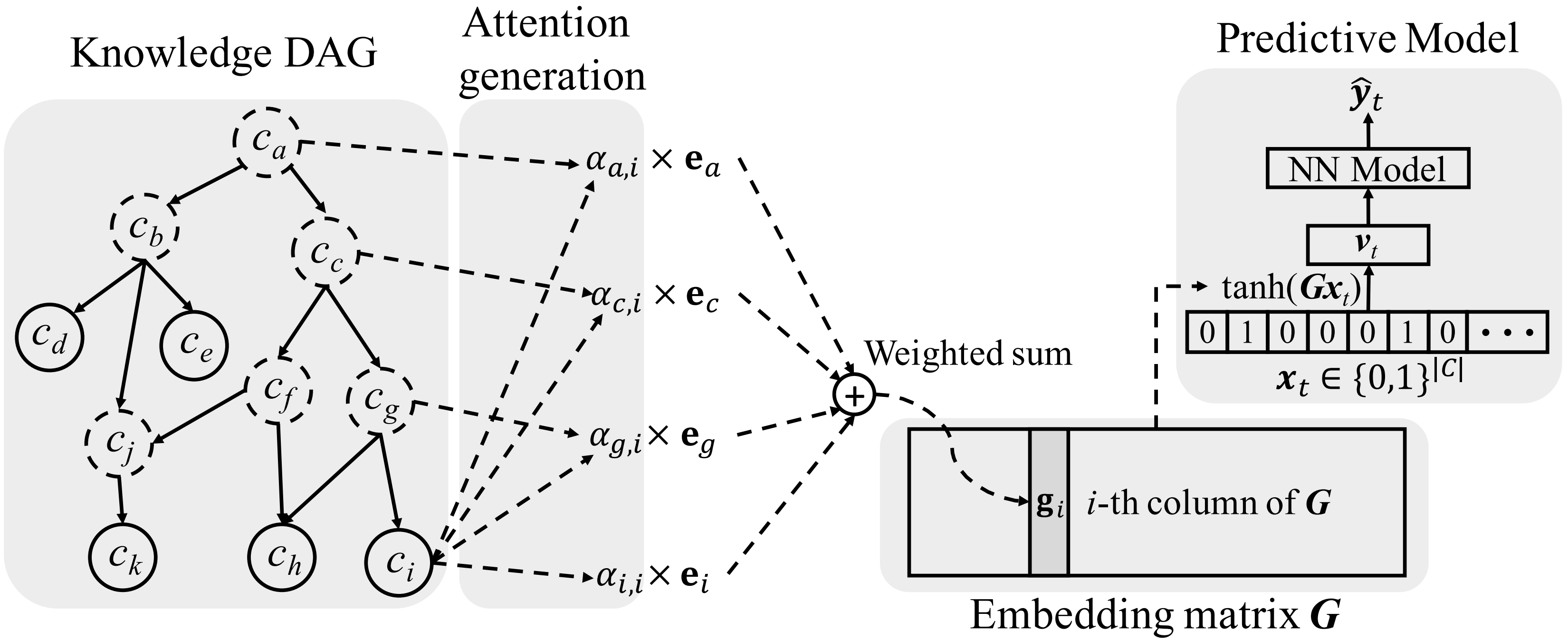
Interpretation for Code Representations



❖ Ma et al. *Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks*. KDD 2017.

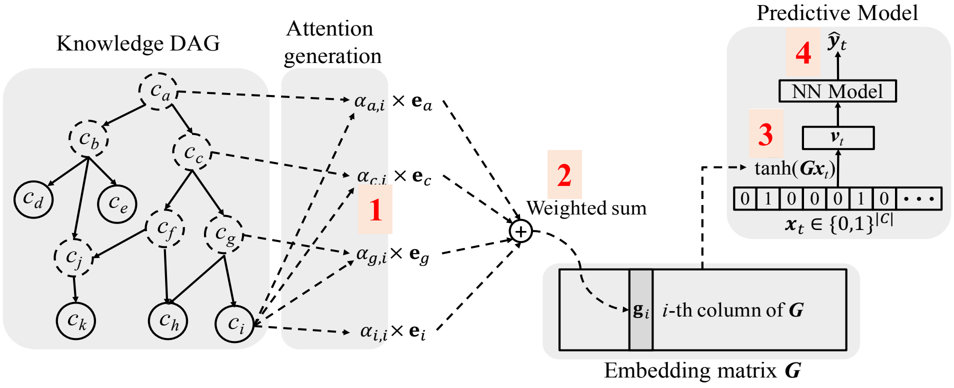
GRAM

- Generate a medical code representation vector by combining the representation vectors of its ancestors using the attention mechanism



❖ Choi et al. *GRAM: Learn representations of medical codes leveraging medical ontologies*. KDD 2017.

GRAM algorithm

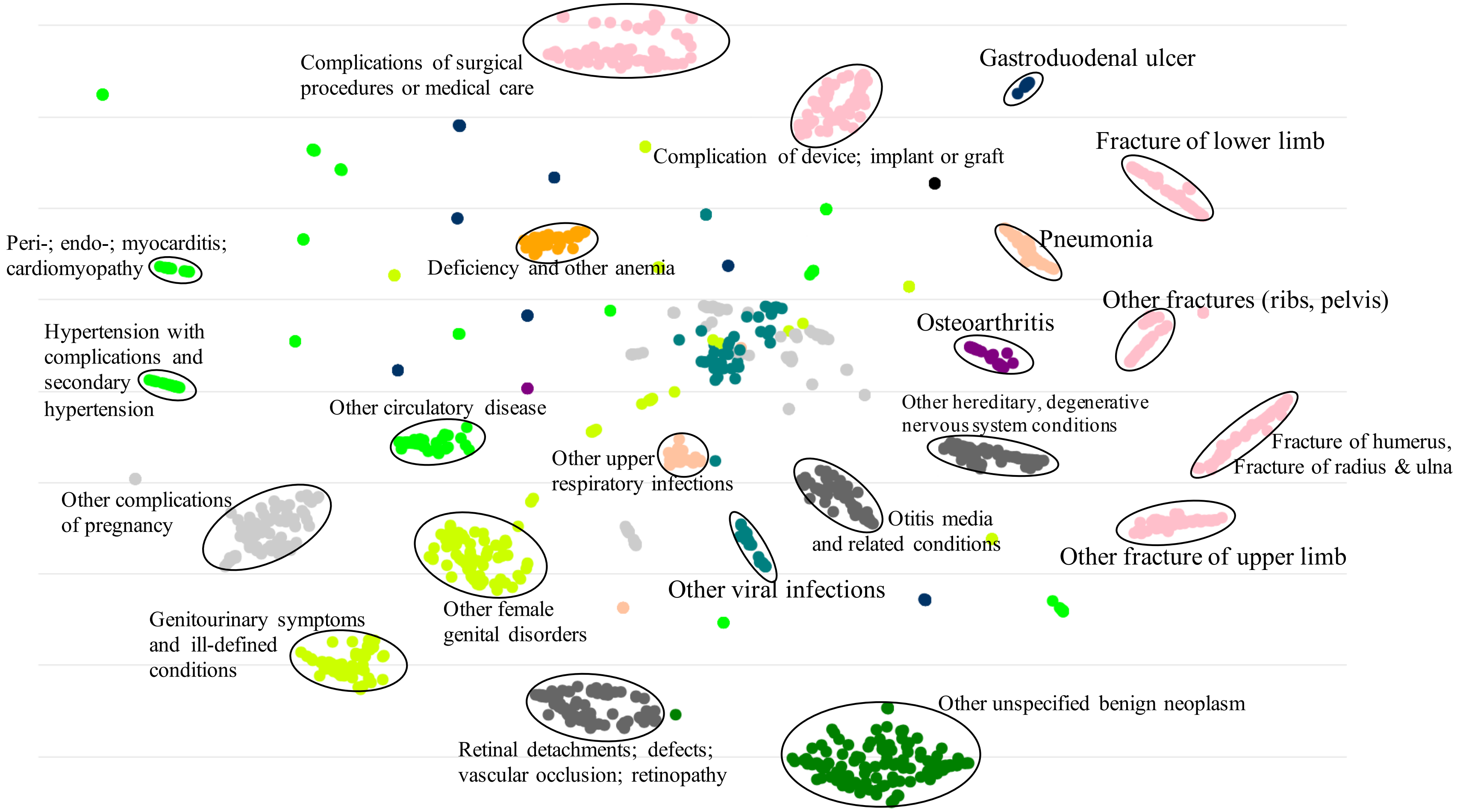


1	$\alpha_{ij} = \frac{\exp(f(\mathbf{e}_i, \mathbf{e}_j))}{\sum_{k \in \mathcal{A}(i)} \exp(f(\mathbf{e}_i, \mathbf{e}_k))}$ where $f(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{u}_a^\top \tanh(\mathbf{W}_a \begin{bmatrix} \mathbf{e}_i \\ \mathbf{e}_j \end{bmatrix} + \mathbf{b}_a)$	
	Attention weights are generated for all pairs of basic embeddings $\mathbf{e}_i$ and its ancestors $\mathbf{e}_j$.	
2	$\mathbf{g}_i = \sum_{j \in \mathcal{A}(i)} \alpha_{ij} \mathbf{e}_j,$	
	Final representation $\mathbf{g}_i$ is the weighted sum of attention weights and basic embeddings.	
3	$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t = \tanh(\mathrm{G}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t])$	
	Sequence of visit representations are obtained using the Embedding matrix $\boldsymbol{G}$.	
4	$\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_t = \mathrm{RNN}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t, \theta_r),$ $\widehat{\mathbf{y}}_t = \widehat{\mathbf{x}}_{t+1} = \mathrm{Softmax}(\mathbf{W}\mathbf{h}_t + \mathbf{b}),$	
	Performing sequential diagnoses prediction, outcomes are generated by RNN and Softmax.	

❖ Choi et al. *GRAM: Learn representations of medical codes leveraging medical ontologies*. KDD 2017.

GRAM learns representations well aligned with knowledge ontology

Scatterplot of GRAM representations

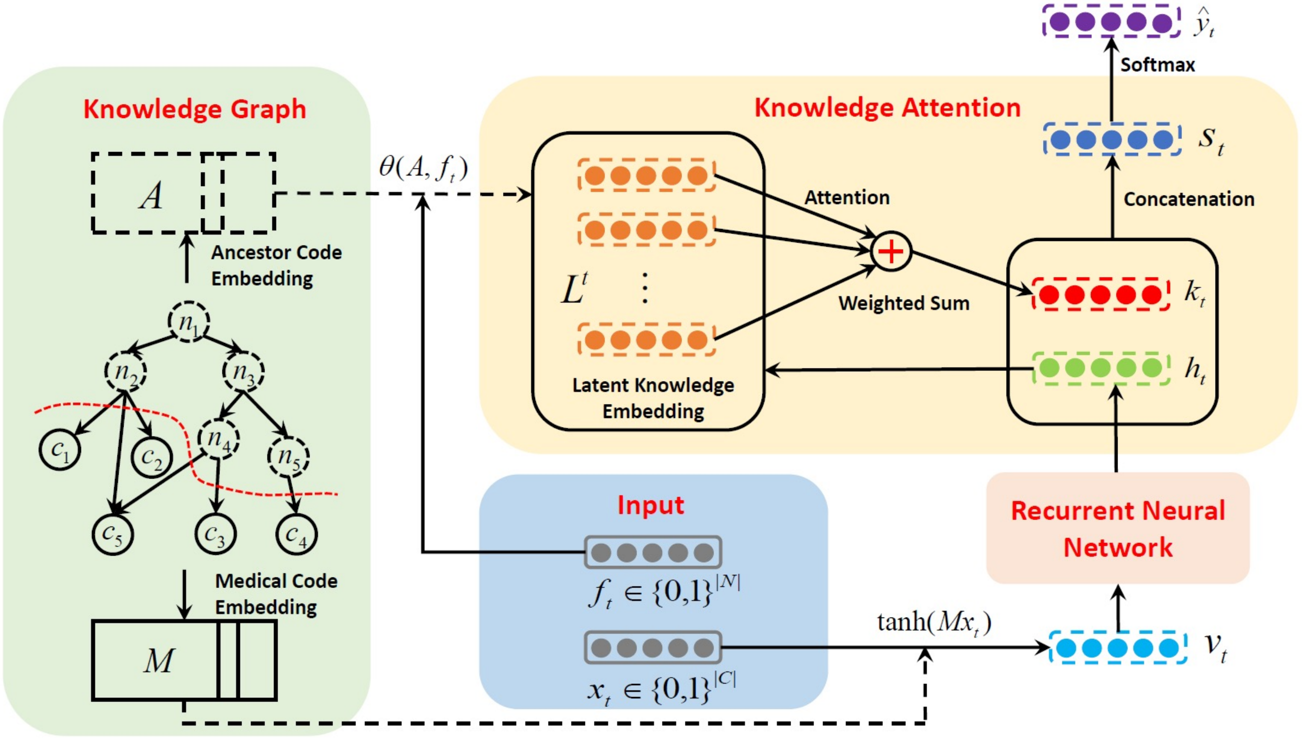


❖ Choi et al. *GRAM: Learn representations of medical codes leveraging medical ontologies*. KDD 2017.

KDD

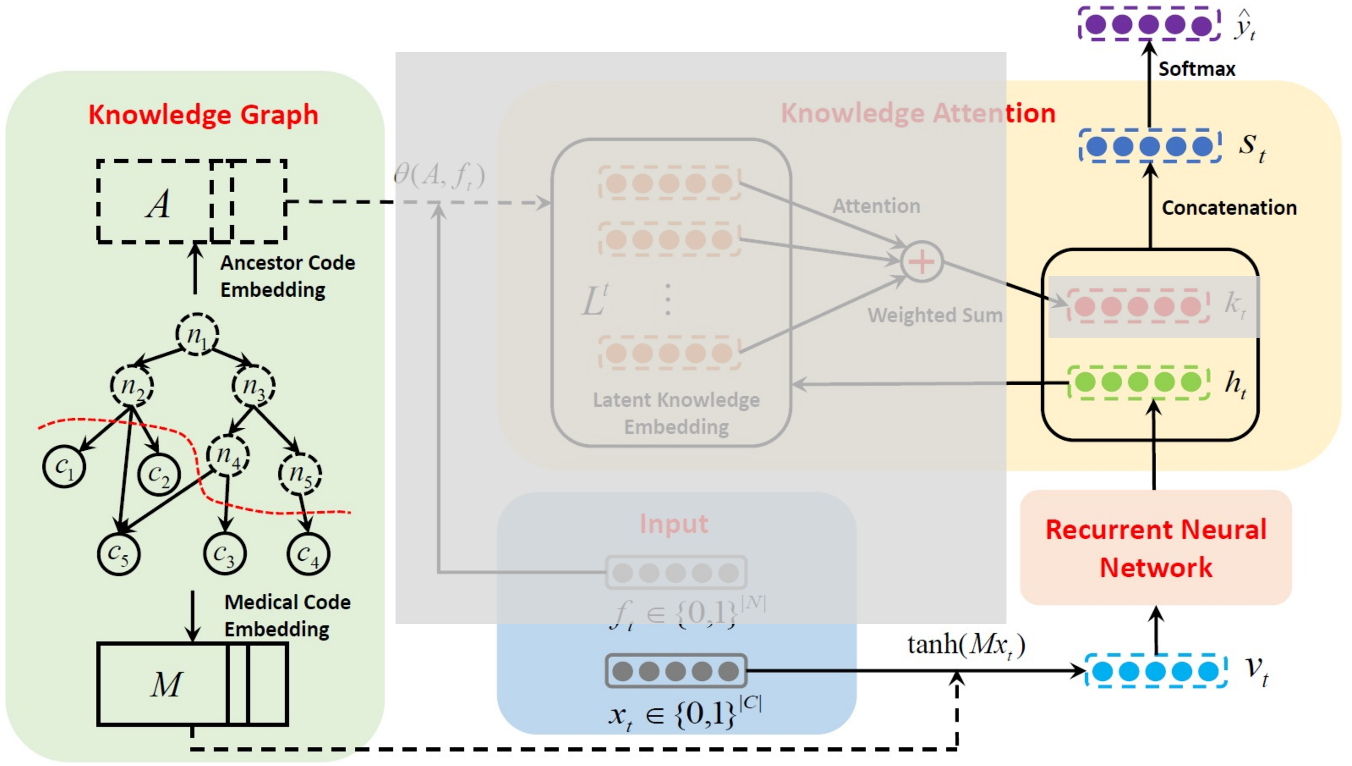
KAME

- Take high-level visit information as input.
- Propose a knowledge attention mechanism.
- Consider general knowledge when making prediction.



KAME vs GRAM

- KAME is the generalization of the state-of-the-art diagnosis prediction model GRAM.
- When removing the proposed knowledge-based attention component (i.e., deleting $\mathbf{k}_t$), then the proposed KAME is reduced to GRAM.



❖ Ma et al. *KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare*. CIKM 2018.

Performance Evaluation

Dataset	Model	Visit-Level Precision@k						Code-Level Accuracy@k					
		5	10	15	20	25	30	5	10	15	20	25	30
Medicaid	KAME	**0.6107**	**0.7475**	**0.8168**	**0.8606**	**0.8920**	**0.9154**	**0.5461**	**0.7037**	**0.7808**	**0.8305**	**0.8667**	**0.8940**
	GRAM	0.5832	0.7189	0.7902	0.8367	0.8717	0.8976	0.5279	0.6842	0.7630	0.8146	0.8528	0.8819
	Dipole	0.5943	0.7226	0.7892	0.8340	0.8680	0.8942	0.5406	0.6903	0.7637	0.8130	0.8503	0.8791
	RNN+	0.5964	0.7210	0.7919	0.8397	0.8746	0.9011	0.5402	0.6867	0.7642	0.8166	0.8550	0.8845
	RNN	0.5448	0.6737	0.7503	0.8036	0.8433	0.8740	0.4914	0.6370	0.7200	0.7782	0.8222	0.8564
Diabetes	KAME	**0.5881**	**0.7313**	**0.8054**	**0.8523**	**0.8859**	**0.9107**	**0.5147**	**0.6939**	**0.7779**	**0.8293**	**0.8666**	**0.8949**
	GRAM	0.5596	0.7048	0.7822	0.8326	0.8684	0.8962	0.4958	0.6776	0.7617	0.8158	0.8546	0.8848
	Dipole	0.5697	0.7015	0.7765	0.8267	0.8640	0.8921	0.5110	0.6771	0.7585	0.8120	0.8520	0.8824
	RNN+	0.5680	0.7007	0.7769	0.8279	0.8649	0.8943	0.5086	0.6740	0.7569	0.8118	0.8519	0.8838
	RNN	0.5515	0.6851	0.7639	0.8179	0.8575	0.8877	0.4984	0.6611	0.7459	0.8024	0.8445	0.8765
MIMIC-III	KAME	**0.7103**	**0.6568**	**0.6967**	**0.7562**	**0.8091**	**0.8470**	**0.3167**	**0.5100**	**0.6379**	**0.7240**	**0.7862**	**0.8303**
	GRAM	0.6998	0.6447	0.6847	0.7439	0.8007	0.8424	0.3123	0.5026	0.6296	0.7142	0.7798	0.8266
	Dipole	0.6220	0.5839	0.6310	0.6953	0.7556	0.8059	0.2774	0.4556	0.5801	0.6671	0.7354	0.7902
	RNN+	0.6158	0.5803	0.6243	0.6912	0.7542	0.8017	0.2760	0.4548	0.5751	0.6647	0.7350	0.7867
	RNN	0.6580	0.6186	0.6637	0.7254	0.7836	0.8272	0.2941	0.4836	0.6106	0.6961	0.7629	0.8119

- The performance of the proposed KAME is better than that of all the baselines on the three datasets.
- Fully utilizing medical knowledge graph is important!
- The proposed KAME achieves robust results on different datasets.

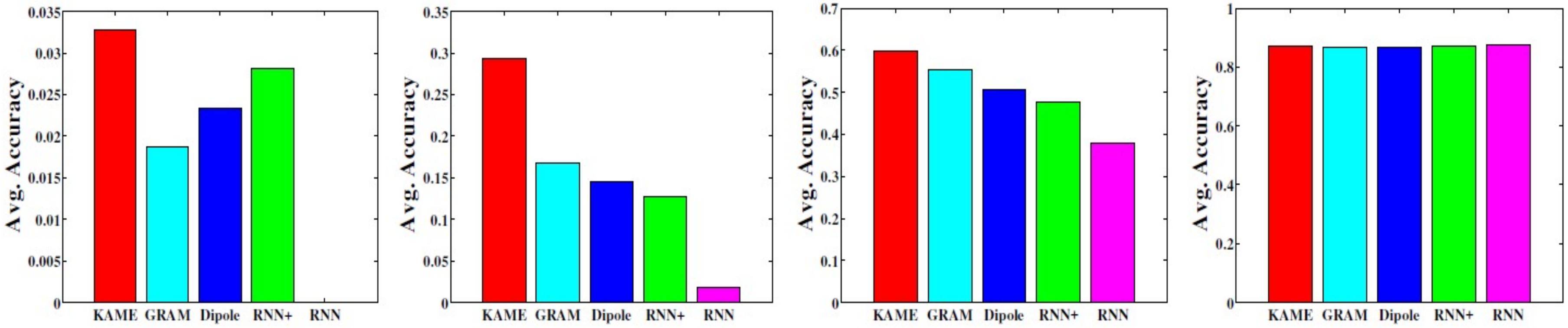
Data Sufficiency Evaluation

- Divide medical codes into four groups: 0-25, 25-50, 50-75 and 75-100, based on their frequency in the training set.
- The 0-25 group represents the most rare codes in the training set, while codes in the 75-100 group are the most common ones.
- Calculate the average accuracy of codes in each group on the testing set.

❖ Ma et al. *KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare*. CIKM 2018.

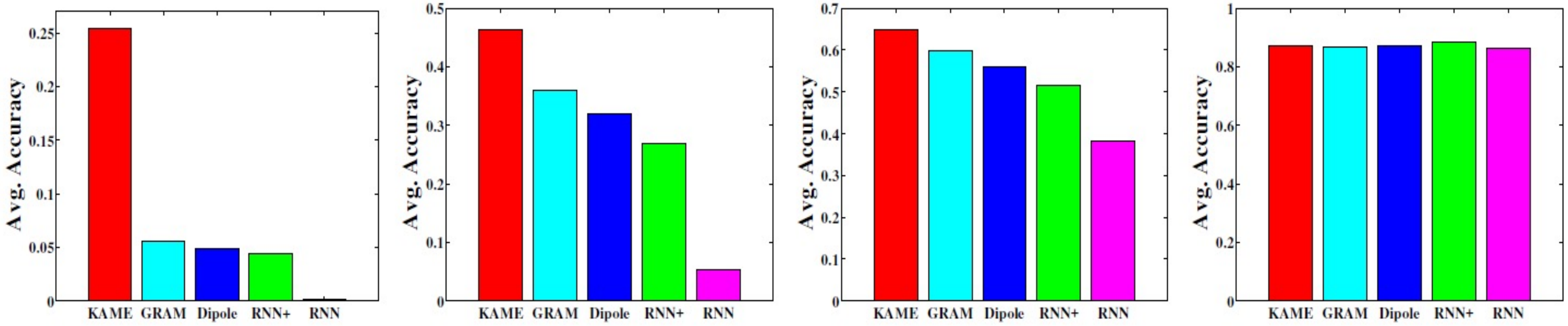
Data Sufficiency Evaluation

Diabetes



(a) 0-25 (b) 25-50 (c) 50-75 (d) 75-100

Medicaid



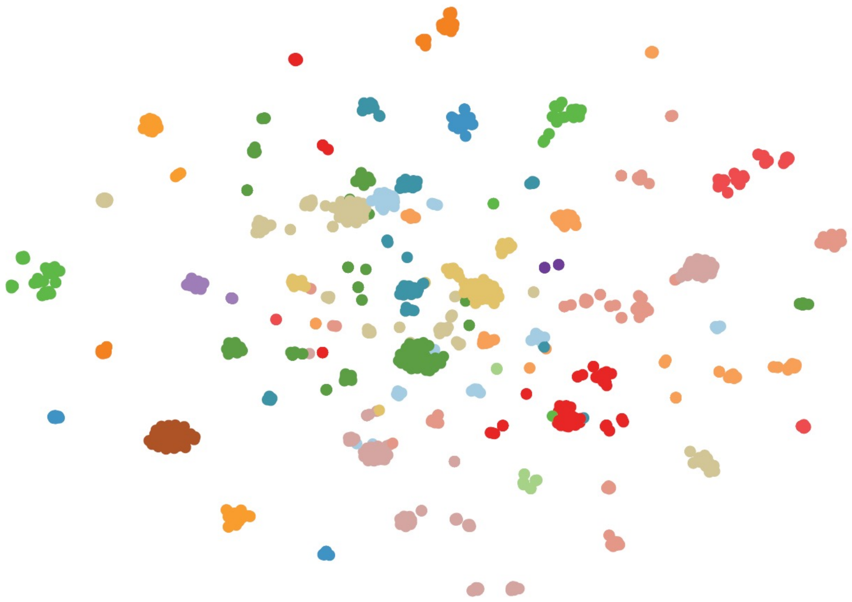
(a) 0-25 (b) 25-50 (c) 50-75 (d) 75-100

❖ Ma et al. ***KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare***. CIKM 2018.

Interpretability Analysis

- Interpretability of the learned medical code representations
  - Randomly select 2000 medical codes and then plot on a 2-D space with *t*-SNE using their learned embeddings.

- Each dot represents a diagnosis code. The colors of the dots represents the disease categories, i.e., cluster labels.
- Ideally, the dots with the same color should be in the same cluster, and there are margins among different clusters.



❖ Ma et al. *KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare*. CIKM 2018.

HAP

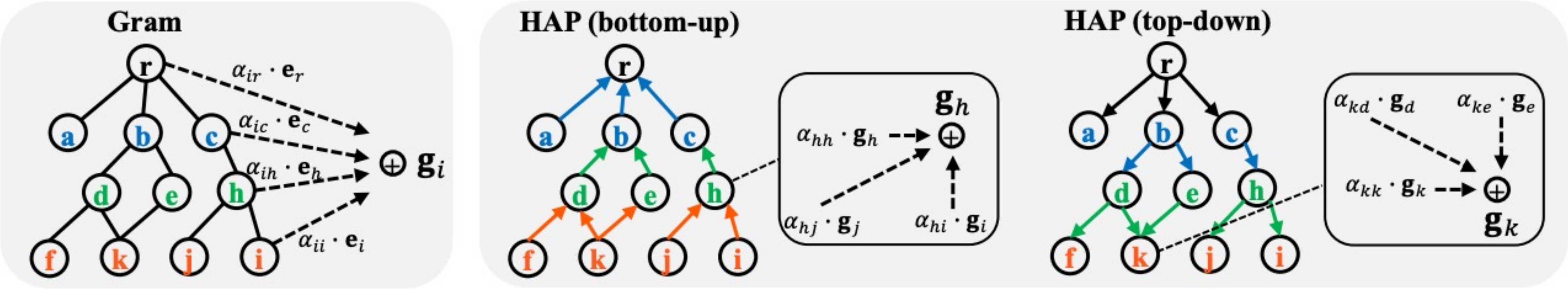
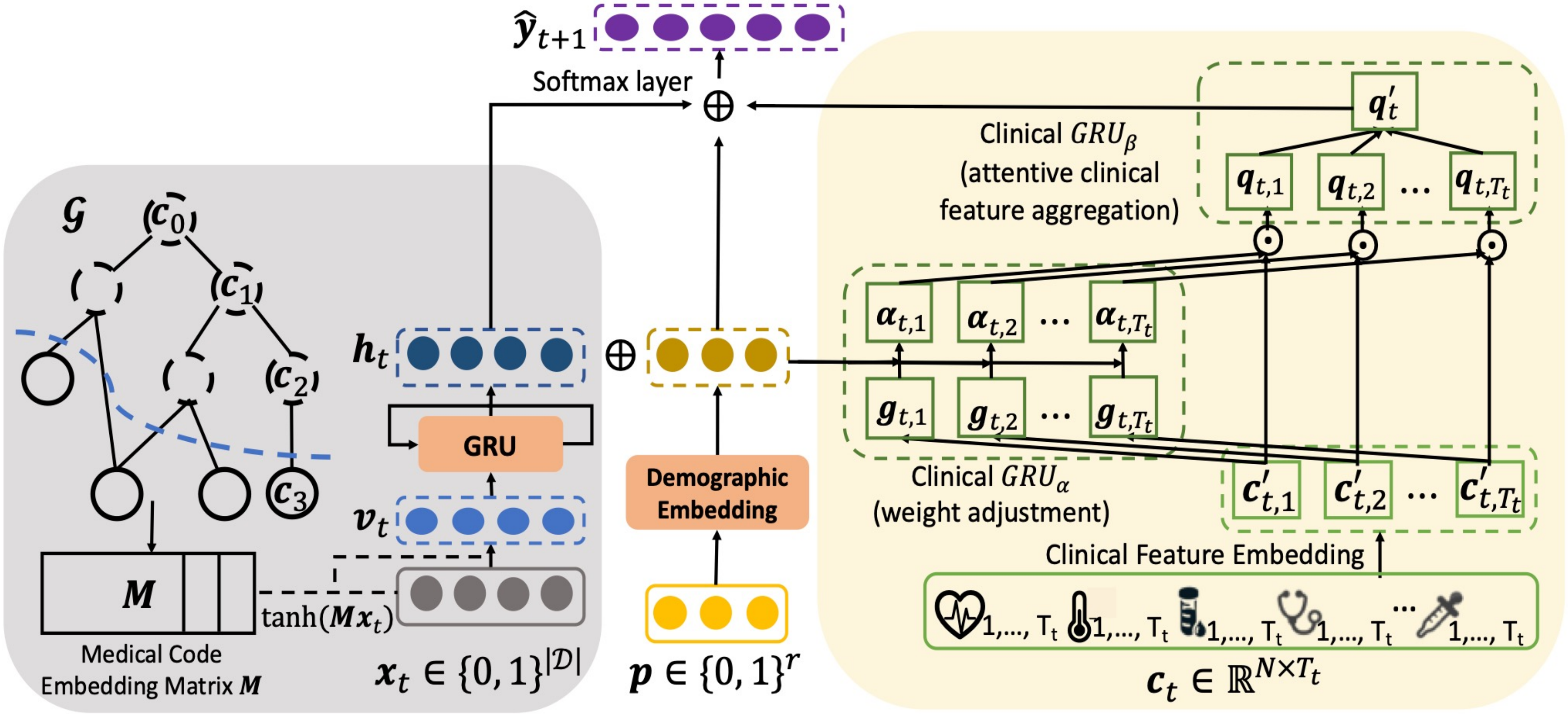


Figure 1: Comparison between Gram and HAP. Gram only considers a node's unordered ancestor set to compute its embedding. HAP hierarchically propagates information across the graph. In the bottom-up round, each parent aggregates information from its children. In the top-down round, each child aggregates information from its parents. The final embedding of each node effectively absorbs information from not only its ancestors, but the entire graph (ancestors, descendants, siblings and others).

❖ Zhang et al. *Hierarchical Attention Propagation for Healthcare Representation Learning*. KDD 2020.

Integrating Multimodal Electronic Health Records



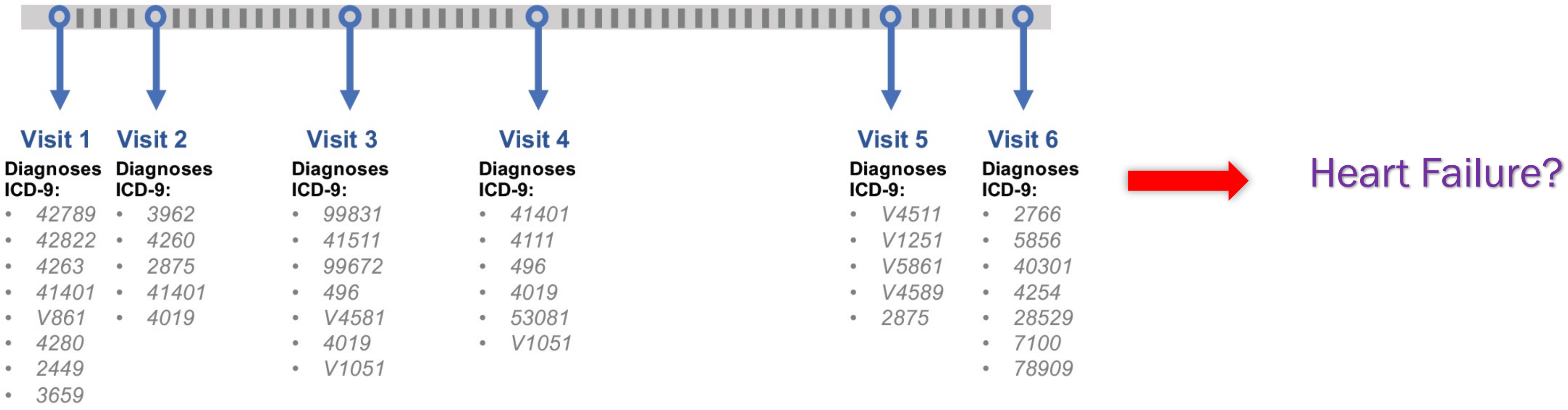
❖ Li et al. *Integrating Multimodal Electronic Health Records for Diagnosis Prediction*. AMIA 2021.

Outline

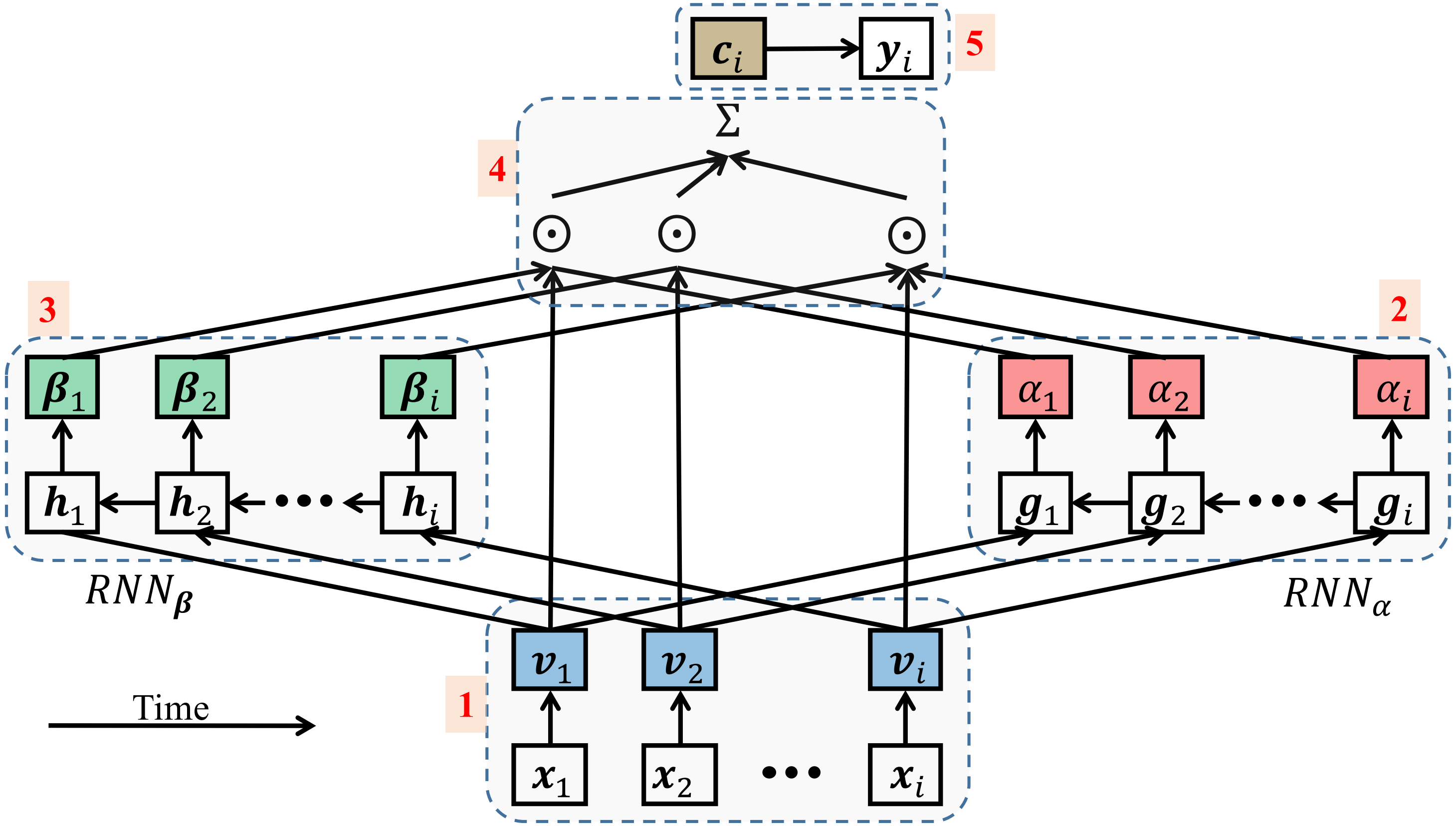
- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

Risk Prediction

- Predicting whether a patient will suffer a given disease/condition.

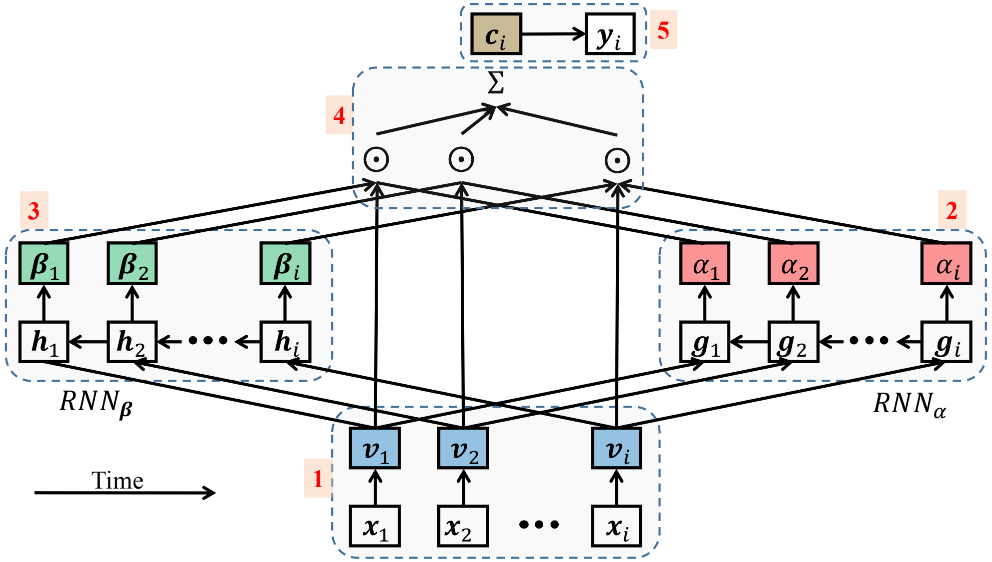


RETAIN: REverse Time AttentIoN model



❖ Choi et al. *RETAIN: An Interpretable Predictive Model for Healthcare Using Reverse Time Attention Mechanism*. NeurIPS 2016.

RETAIN



1	$\mathbf{v}_i = \mathbf{E}\mathbf{x}_i$
	Multi-hot representation of the visit is linearly projected by the embedding matrix **E**.
2	$\mathbf{g}_i, \mathbf{g}_{i-1}, \ldots, \mathbf{g}_1 = \mathrm{RNN}_\alpha(\mathbf{v}_i, \mathbf{v}_{i-1}, \ldots, \mathbf{v}_1)$, $\alpha_1, \alpha_2, \ldots, \alpha_i = \mathrm{Softmax}(\mathbf{w}_\alpha^\top [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_i] + b_\alpha)$
	RNN_α generates α_i, the scalar attention weight for the *i*-th visit. The visit representations v_i's are fed to the RNN_α in reverse order.
3	$\mathbf{h}_i, \mathbf{h}_{i-1}, \ldots, \mathbf{h}_1 = \mathrm{RNN}_\beta(\mathbf{v}_i, \mathbf{v}_{i-1}, \ldots, \mathbf{v}_1)$ $\boldsymbol{\beta}_j = \tanh\left(\mathbf{W}_\beta \mathbf{h}_j + \mathbf{b}_\beta\right) \quad \text{for} \quad j = 1, \ldots, i$
	RNN_β generates $\boldsymbol{\beta}_i$, the vector attention weight for the medical codes in the *i*-th visit. v_i's are fed to the RNN_β in reverse order as well.
4	$\mathbf{c}_i = \sum_{j=1}^{i} \alpha_j \boldsymbol{\beta}_j \odot \mathbf{v}_j$
	The attention weights α_i and $\boldsymbol{\beta}_i$ are combined with the visit representation v_i to obtain the context vector c_i.
5	$\widehat{\mathbf{y}}_i = \mathrm{Softmax}(\mathbf{W}\mathbf{c}_i + \mathbf{b})$
	Using the context vector c_i, we make the final prediction.

LSAN

- Motivation

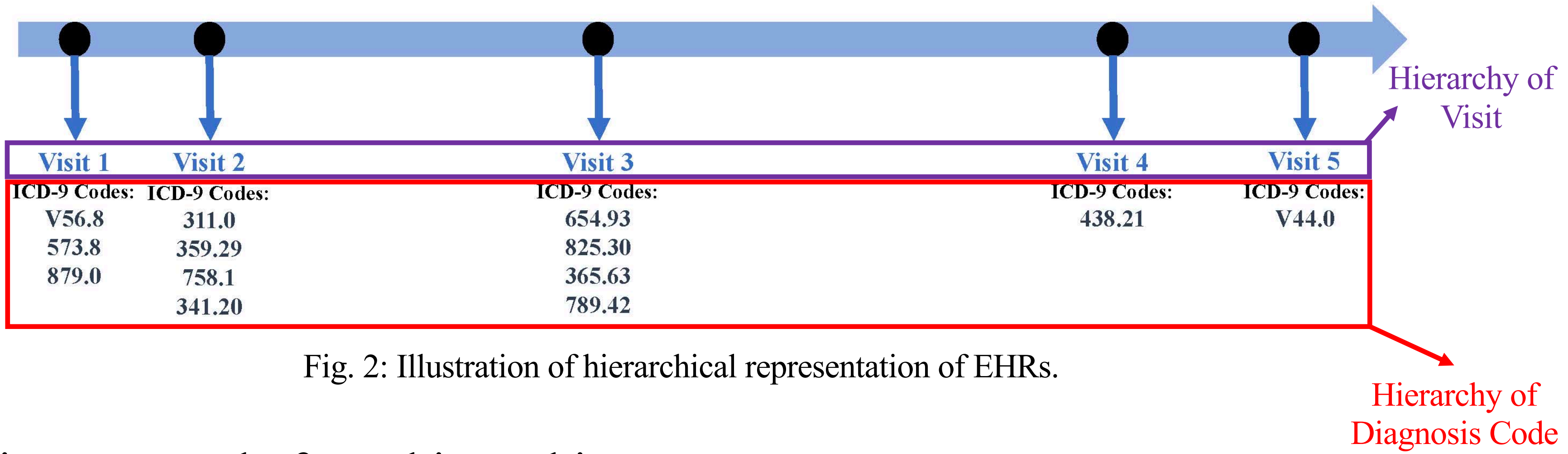


Fig. 2: Illustration of hierarchical representation of EHRs.

- EHR is composed of two hierarchies.
- In the hierarchy of diagnosis code, we should reduce the noise information to learn a better embedding for each visit.
- In the hierarchy of visit, we should pay attention to the correlations among visits.

❖ Ye et al. *LSAN: Modeling Long-term Dependencies and Short-term Correlations with Hierarchical Attention for Risk Prediction*. CIKM 2020.

Motivation

- Within each visit, there may exist diagnosis codes that are unrelated to the target task.
- Distinguishing the importance of diagnosis codes within each visit.
- In the hierarchy of visit, capturing the temporal patterns of disease changes is always important.
- Filtering out noise by extracting local temporal correlations among neighboring visits and utilizing the long-term dependencies information.

❖ Ye et al. *LSAN: Modeling Long-term Dependencies and Short-term Correlations with Hierarchical Attention for Risk Prediction*. CIKM 2020.

LSAN

- Modeling the Long-term dependencies and Short-term correlations with the utilization of a hierarchical Attention Network

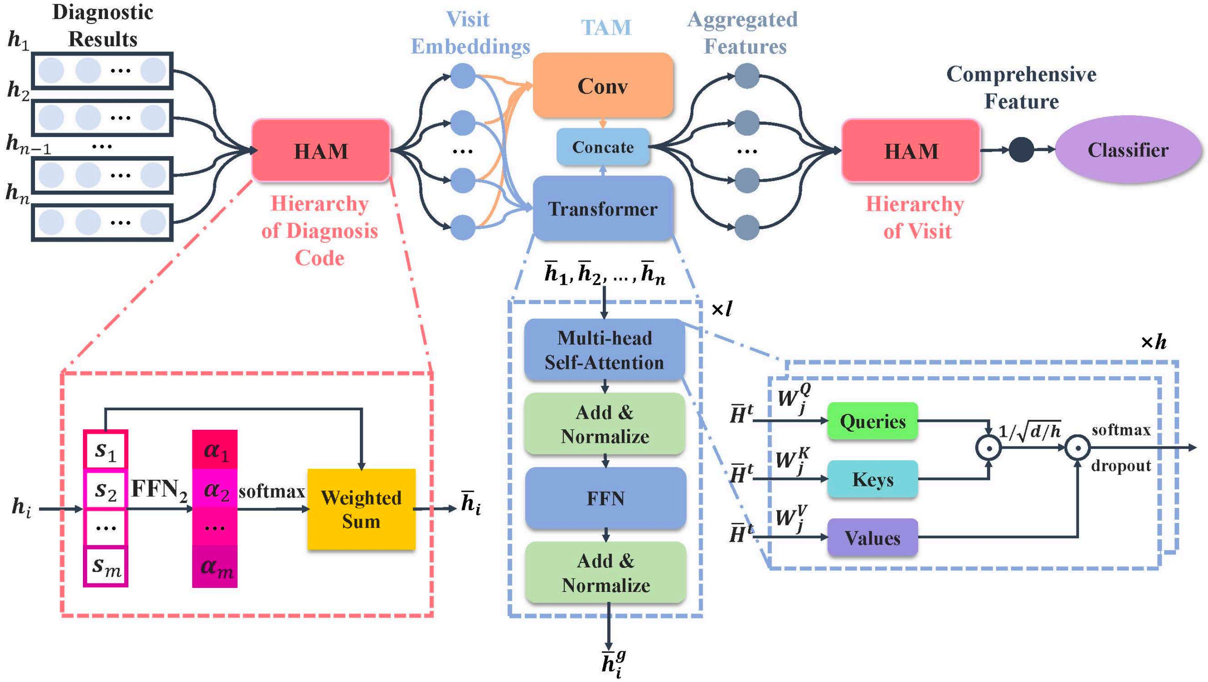


Fig. 3: The proposed model.

❖ Ye et al. *LSAN: Modeling Long-term Dependencies and Short-term Correlations with Hierarchical Attention for Risk Prediction*. CIKM 2020.

HAM

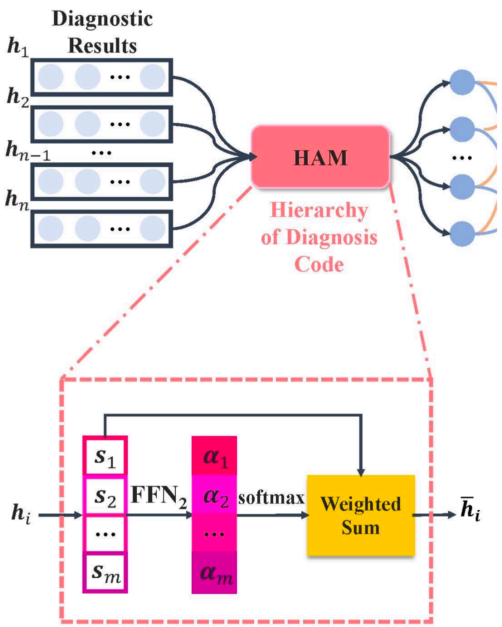


Fig. 4: HAM.

- HAM has a hierarchical attention mechanism in the hierarchies of diagnosis code and visit.
- In the hierarchy of diagnosis code, it gets a single dense diagnosis embedding for each visit by summing up the diagnosis code embeddings with code-level attention weights.
- In the hierarchy of visit, it attends the aggregated visit embeddings by their relevance to target disease and attains a comprehensive representation for risk prediction.

❖ Ye et al. *LSAN: Modeling Long-term Dependencies and Short-term Correlations with Hierarchical Attention for Risk Prediction*. CIKM 2020.

TAM

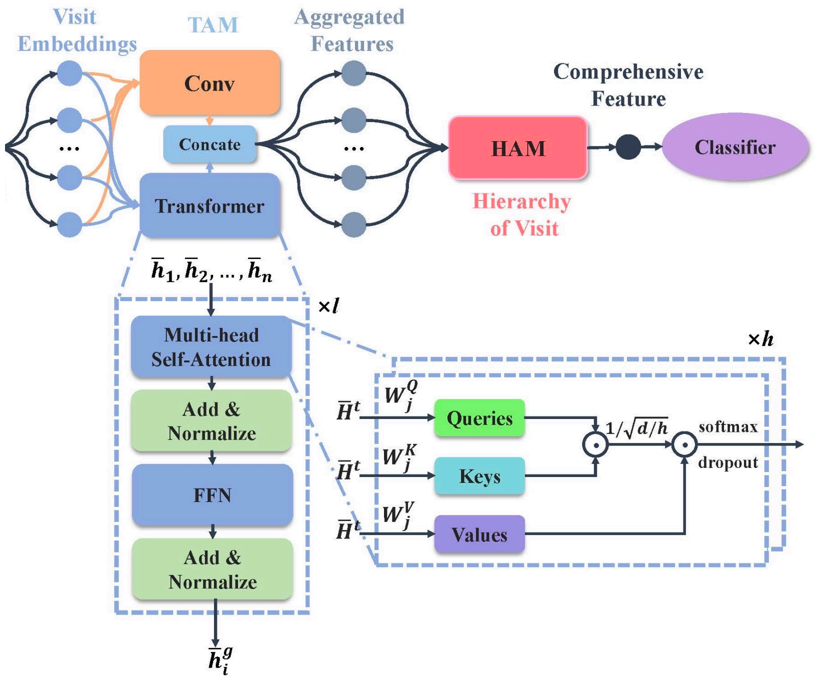
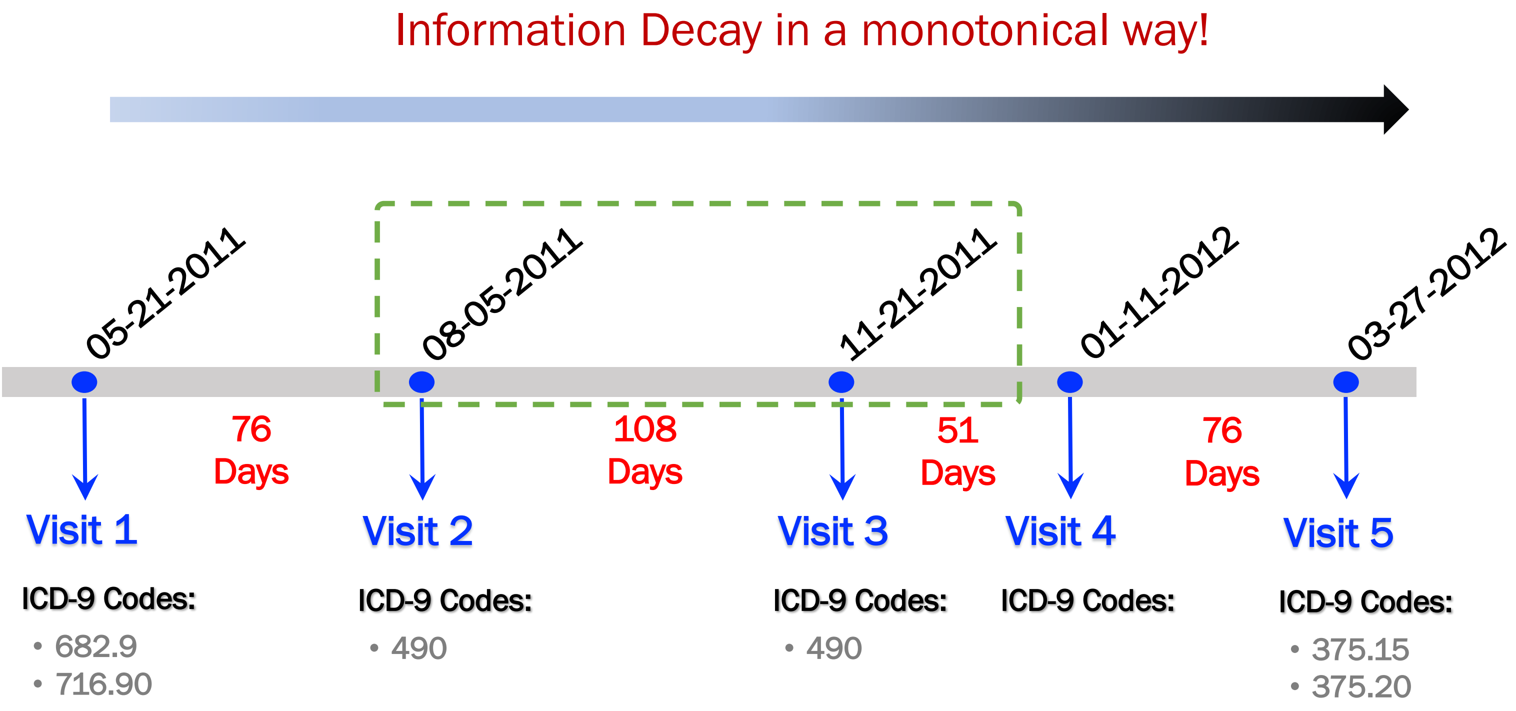


Fig. 5: TAM.

- TAM aggregates the visit embeddings with two kinds of temporal information from global and local temporal structures.
- When the features of all visit are put into TAM, it models long-term dependencies in the global structure by Transformer and short-term correlations in the local structure by a convolutional layer.

❖ Ye et al. *LSAN: Modeling Long-term Dependencies and Short-term Correlations with Hierarchical Attention for Risk Prediction*. CIKM 2020.

Importance of Time Information



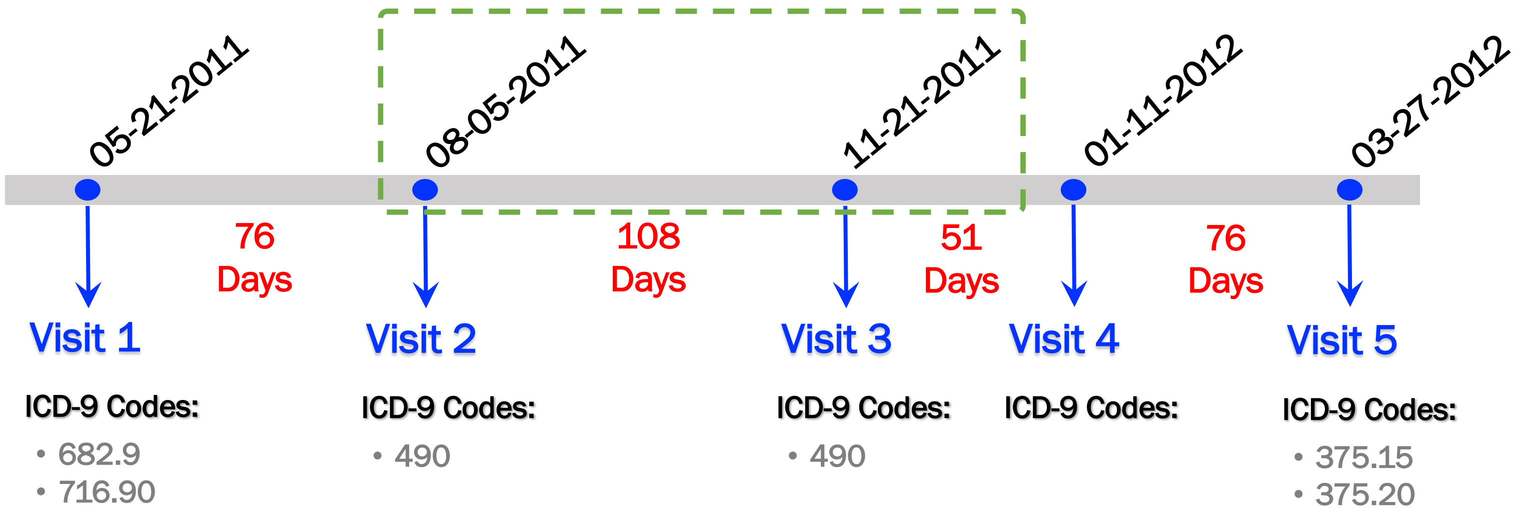
An example of a patient's visit information

❖Baytas et al. *Patient subtyping via time-aware LSTM networks*. KDD 2017.

❖Bai et al. *Interpretable Representation Learning for Healthcare via Capturing Disease Progression through Time*. KDD 2018.

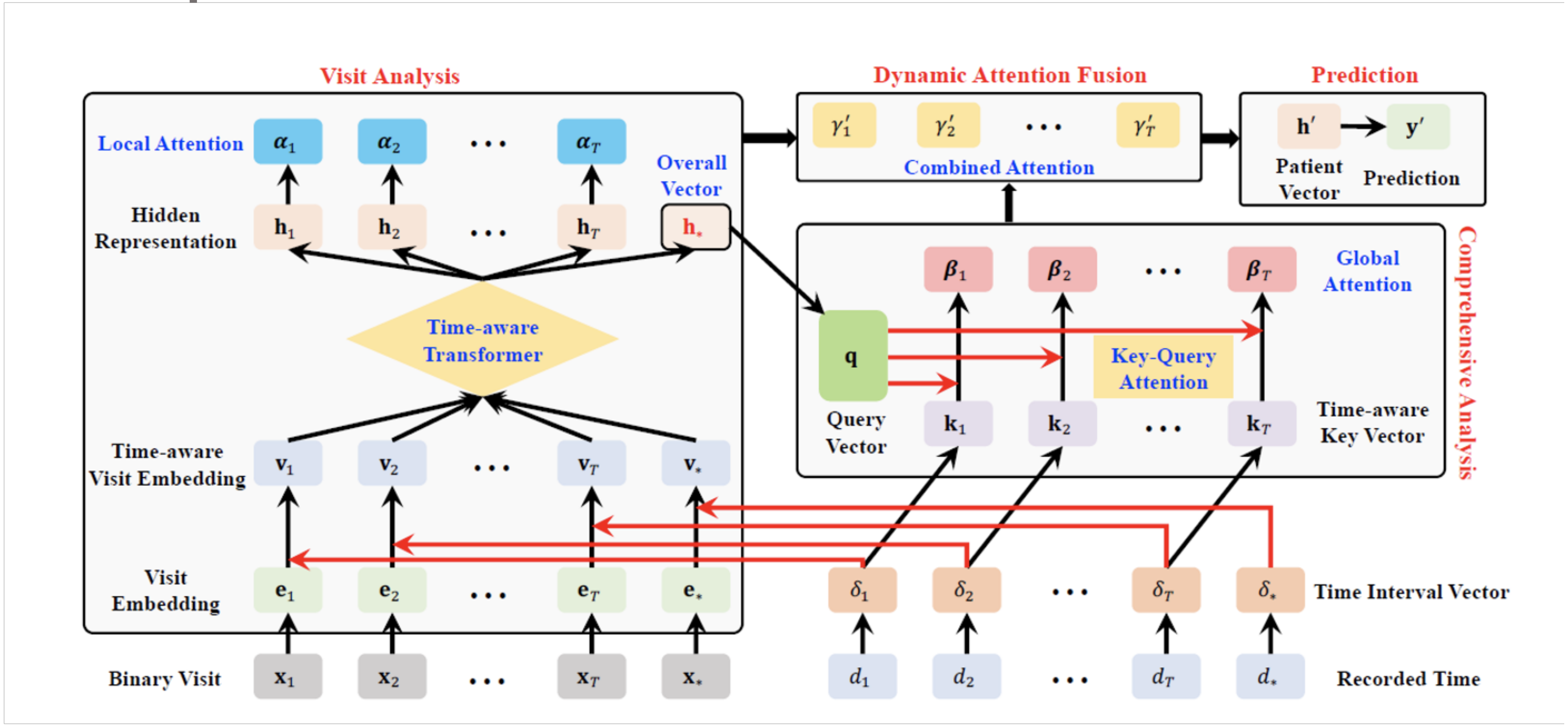
HiTANet: Hierarchical Time-aware Attention

- Motivations
  - The importance of historical patient information with respect to current health risk does not decay monotonically.
  - The importance of previous timestamps varies among patients.

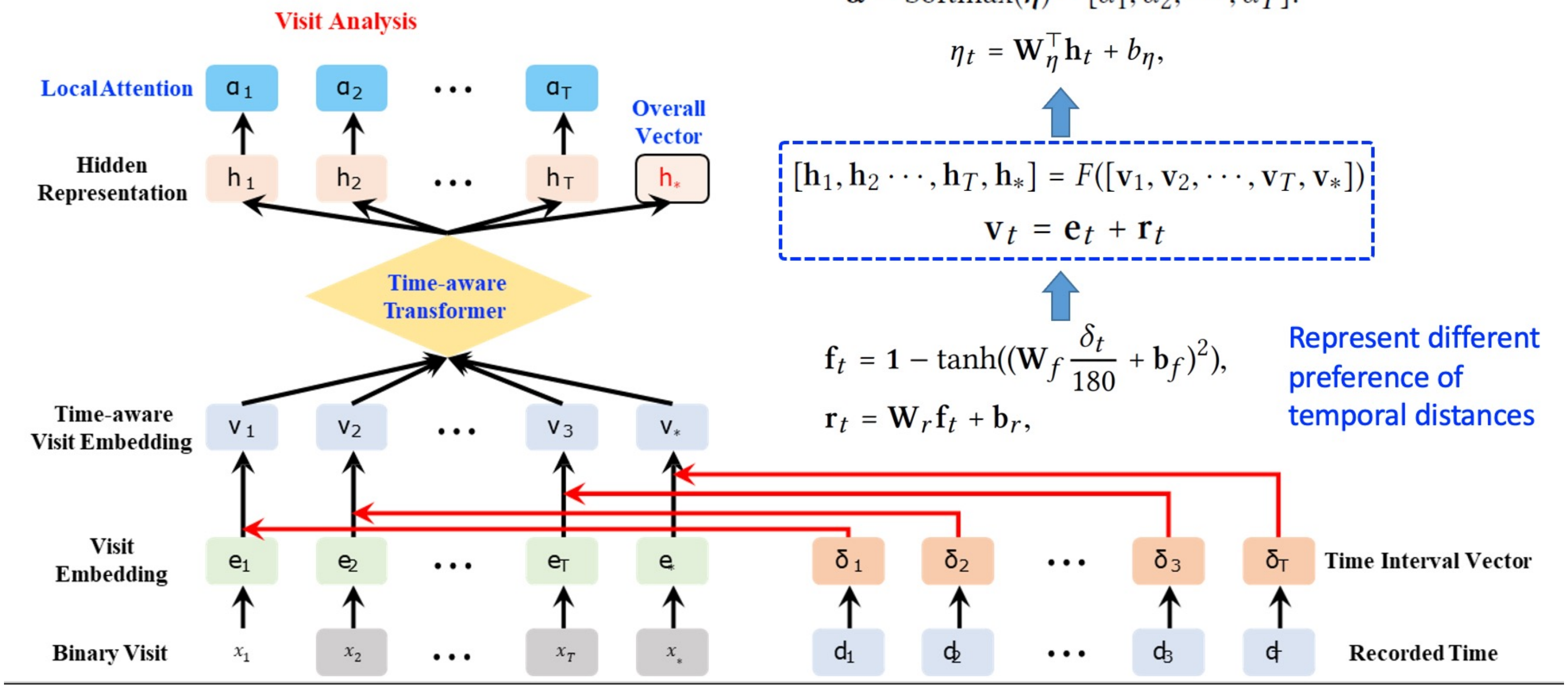


❖ Luo et al. *HiTANet: Hierarchical time-aware attention networks for risk prediction on electronic health records*. KDD 2020.

The Proposed HiTANet Model



Visit Analysis



$$\boldsymbol{\alpha} = \mathrm{Softmax}(\boldsymbol{\eta}) = [\alpha_1, \alpha_2, \cdots, \alpha_T].$$

$$\eta_t = \mathbf{W}_\eta^\top \mathbf{h}_t + b_\eta,$$

$$[\mathbf{h}_1, \mathbf{h}_2 \cdots, \mathbf{h}_T, \mathbf{h}_*] = F([\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_T, \mathbf{v}_*])$$

$$\mathbf{v}_t = \mathbf{e}_t + \mathbf{r}_t$$

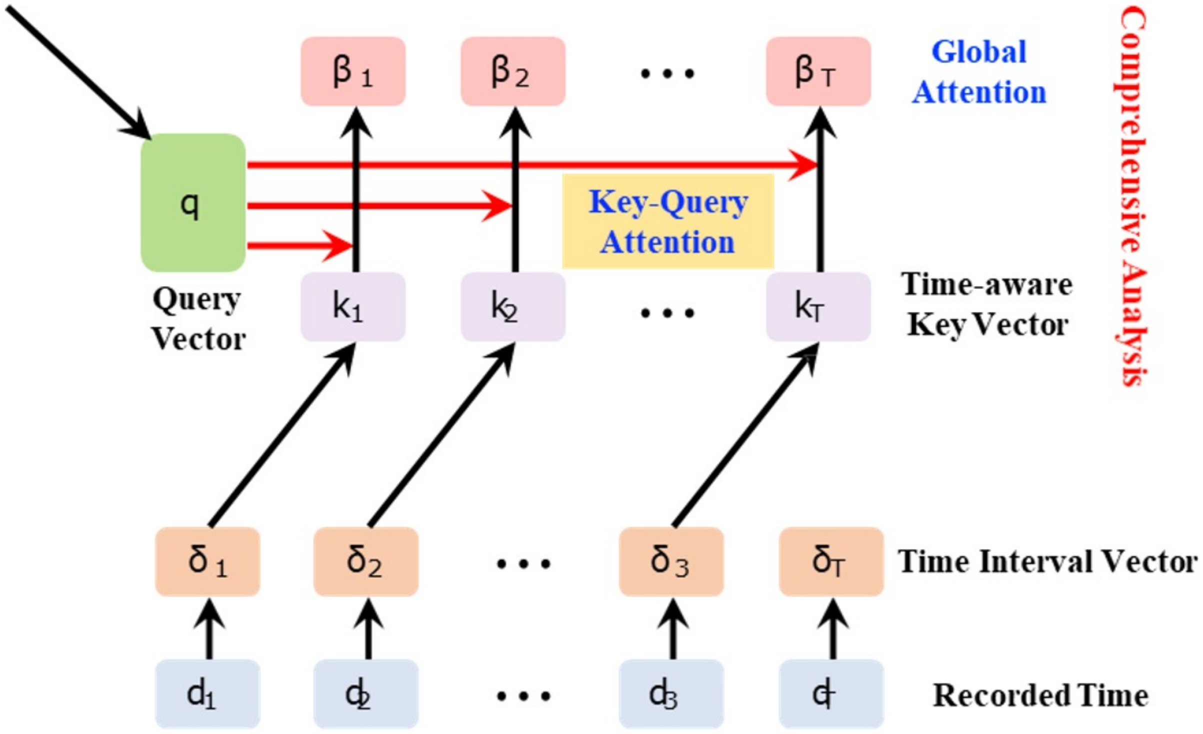
$$\mathbf{f}_t = 1 - \tanh((\mathbf{W}_f \frac{\delta_t}{180} + \mathbf{b}_f)^2),$$

$$\mathbf{r}_t = \mathbf{W}_r \mathbf{f}_t + \mathbf{b}_r,$$

Represent different preference of temporal distances

❖ Luo et al. *HiTANet: Hierarchical time-aware attention networks for risk prediction on electronic health records*. KDD 2020.

Comprehensive Analysis



$$\boldsymbol{\beta} = \mathrm{Softmax}(\boldsymbol{\phi}) = [\beta_1, \beta_2, \cdots, \beta_T]$$

$$\phi_t = \frac{\mathbf{q}^\top \mathbf{k}_t}{\sqrt{s}}$$

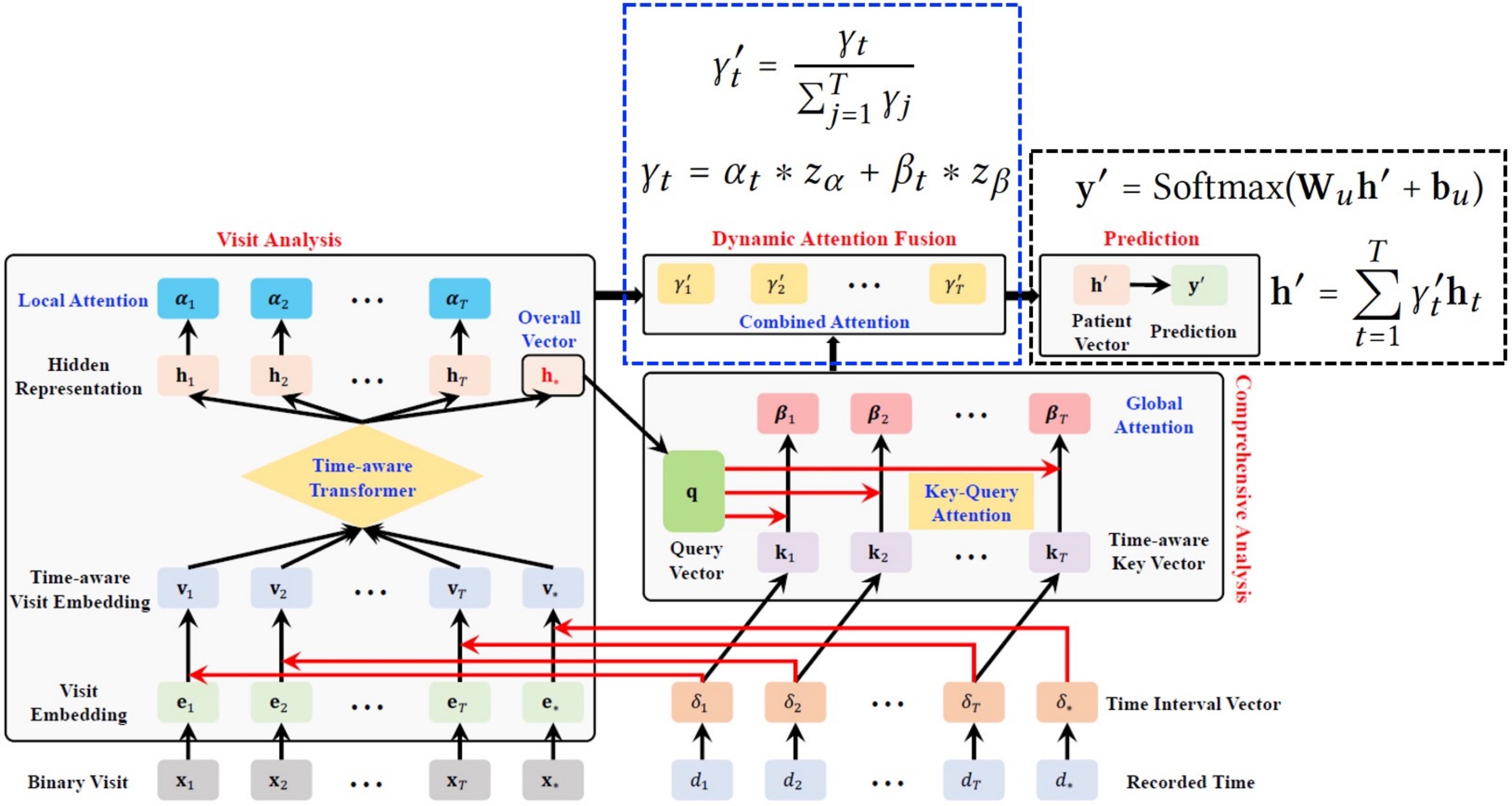
$$\mathbf{q} = \mathrm{ReLU}(\mathbf{W}_q \mathbf{h}_* + \mathbf{b}_q)$$

$$\mathbf{o}_t = \mathbf{1} - \tanh((\mathbf{W}_o \frac{\delta_t}{180} + \mathbf{b}_o)^2),$$

$$\mathbf{k}_t = \tanh(\mathbf{W}_k \mathbf{o}_t + \mathbf{b}_k),$$

❖ Luo et al. *HiTANet: Hierarchical time-aware attention networks for risk prediction on electronic health records*. KDD 2020.

Attention Fusion & Prediction



Experiments

Table 2: Average Performance on Three Disease Prediction Tasks

Method		COPD					Heart Failure					Kidney Diseases				
		Acc	Pre	Recall	F1	Auc	Acc	Pre	Recall	F1	Auc	Acc	Pre	Recall	F1	Auc
Classical Methods	SVM	0.804	0.713	0.319	0.441	0.639	0.784	**0.757**	0.327	0.457	0.644	0.840	**0.777**	0.545	0.641	0.745
	LR	0.678	0.328	0.319	0.324	0.556	0.716	0.489	0.466	0.477	0.639	0.772	0.558	0.636	0.594	0.728
	RF	0.798	0.664	0.334	0.444	0.640	0.779	0.746	0.310	0.438	0.635	0.819	0.758	0.452	0.567	0.701
Plain RNNs	LSTM	0.807	0.680	0.461	0.548	0.693	0.812	0.640	0.510	0.561	0.708	0.823	0.680	0.572	0.616	0.739
	GRU	0.820	0.694	0.462	0.553	0.698	0.794	0.679	0.490	0.567	0.700	0.818	0.678	0.591	0.629	0.745
Attention-based Models	Dipole−	0.818	0.699	0.440	0.538	0.690	0.795	0.689	0.481	0.565	0.698	0.826	0.679	0.635	0.656	0.764
	Dipole	0.821	0.687	0.477	0.562	0.704	0.794	0.713	0.445	0.542	0.687	0.843	0.771	0.571	0.656	0.755
	Retain	0.821	0.696	0.463	0.555	0.699	0.784	0.655	0.474	0.549	0.689	0.821	0.706	0.544	0.614	0.732
	SAnD	0.810	0.653	0.462	0.539	0.692	0.785	0.661	0.466	0.544	0.686	0.823	0.690	0.592	0.636	0.748
Time-based Models	RetainEx	0.829	**0.728**	0.470	0.570	0.707	0.799	0.730	0.438	0.546	0.688	0.827	0.745	0.520	0.612	0.728
	T-LSTM	0.818	0.687	0.525	0.595	0.722	**0.831**	0.695	0.527	0.598	0.727	0.832	0.728	0.524	0.608	0.729
	TimeLine	0.812	0.654	0.478	0.550	0.698	0.792	0.661	0.510	0.574	0.705	0.827	0.697	0.607	0.648	0.756
Ours	HiTANet	**0.840**	0.707	**0.583**	**0.637**	**0.752**	0.823	0.724	**0.587**	**0.647**	**0.750**	**0.851**	0.743	**0.668**	**0.702**	**0.792**

Attention Analysis

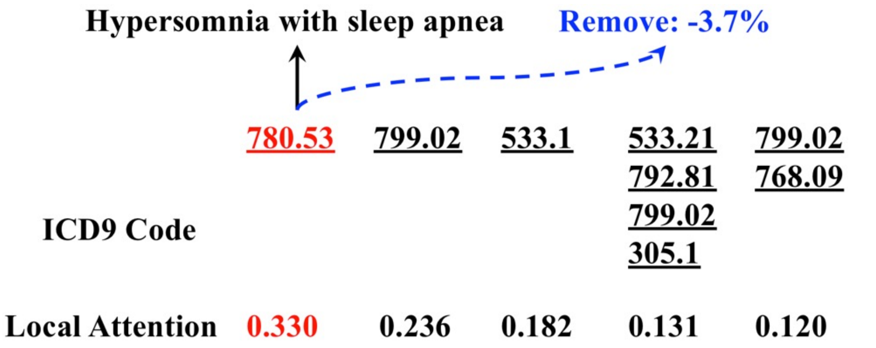


Figure 3: A positive example from the Heart Failure testing set. HiTANet assigns a higher attention to the first visit, which contains Hypersomnia, a common signal of Heart Failure problems. If we remove this record, then the probability of predicting as a positive case will drop 3.7%.

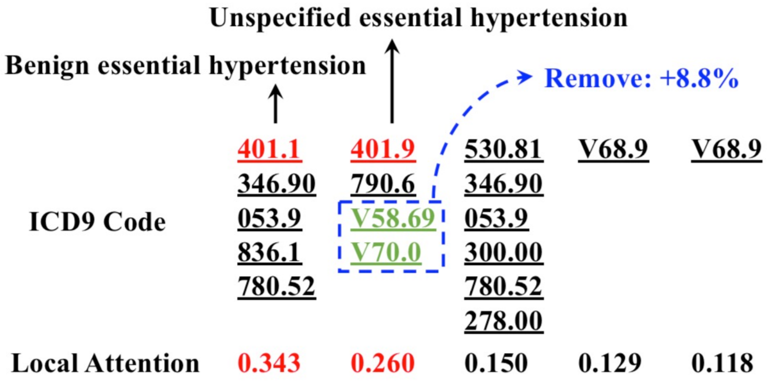


Figure 4: A negative example from the Heart Failure testing set. HiTANet assigns high attention weights to the first two visits. They both contain hypertension related diagnosis codes marked in red, which are the risk factors for Heart Failure. Codes marked in green means the adopted treatments. If we remove the treatment codes, the probability of being positive will increase 8.8%.

❖ Luo et al. *HiTANet: Hierarchical time-aware attention networks for risk prediction on electronic health records*. KDD 2020.

MedPath

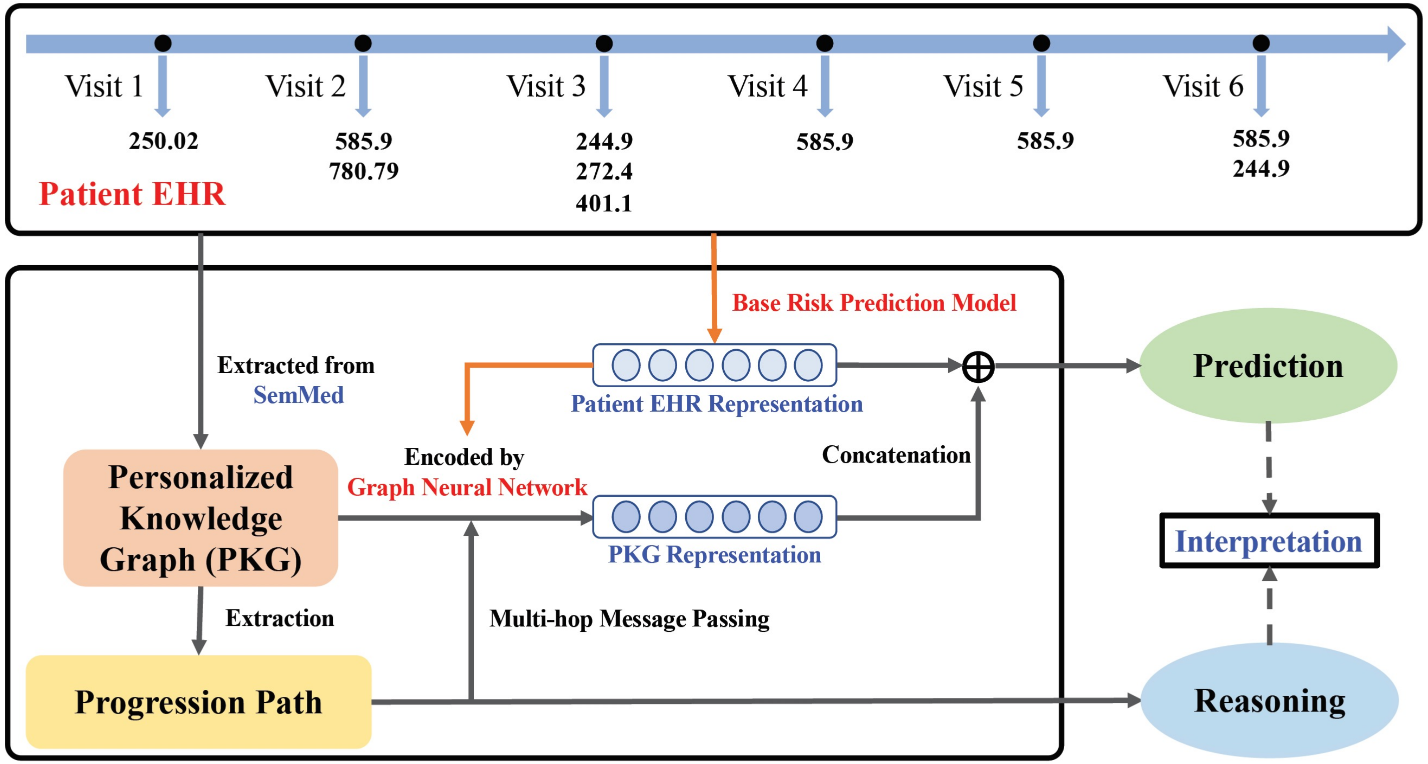
- Necessity of Incorporating Personalized Knowledge Graph
  - The number of overlapping medical codes between individual patients' EHR data and the entire KG is very small.
  - The leading causes of a specific target disease for different patients vary a lot.

- Explicit Reasoning over Disease Progression Paths
  - Enhance the representation learning of medical codes.
  - Implicit reasoning with attention weights.
  - Multi-hop explicit disease progression paths in KG.

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

MedPath

- Augmenting Health Risk Prediction via Medical Knowledge Paths

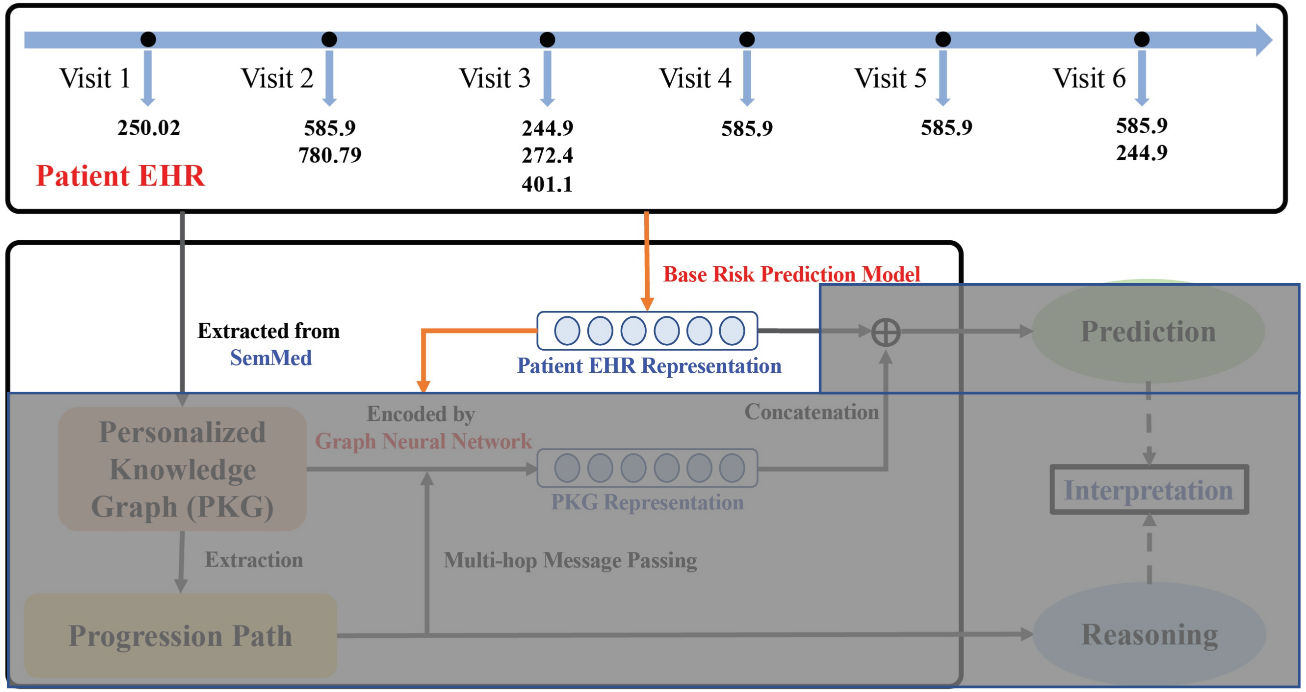


❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

EHR Encoder

- Any of existing risk prediction model
  - Retain [NeurIPS 2016]
  - Dipole [KDD 2017]
  - GRAM [KDD 2017]
  - HiTANet [KDD 2020]
  - ...

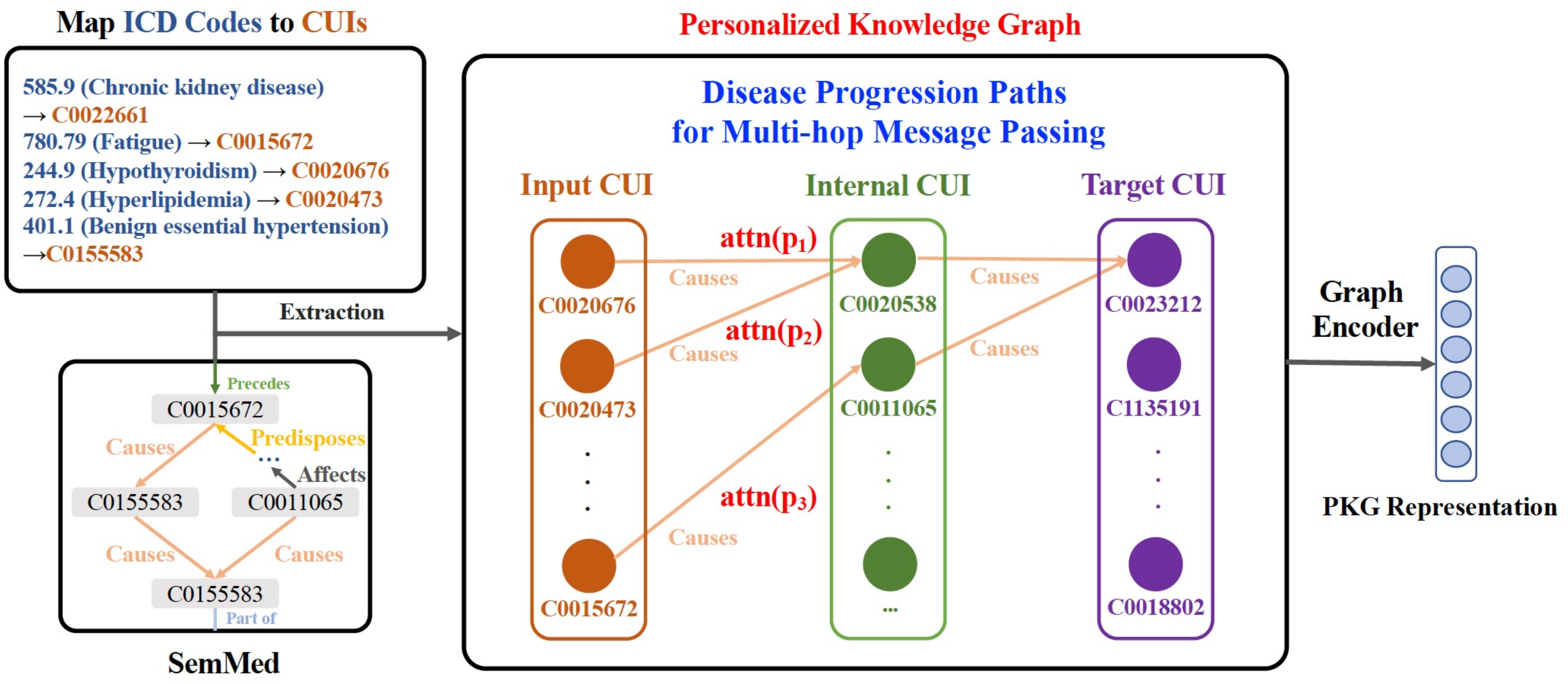
$$\mathbf{s} = F_e(X)$$



Personalized Graph Extraction

- Medical Knowledge Graph
  - SemMed: Semantic MEDLINE (https://skr3.nlm.nih.gov/SemMed/)
- Unification of ICD Codes and SemMed Entities
  - SemMed: Concept Unique Identifiers (CUIs)
  - EHR data: ICD codes
  - Mapping: SNOMED CT
- Path Extraction
  - Source: CUIs from input EHR data
  - Target: CUIs of our target disease/condition

Graph Encoder



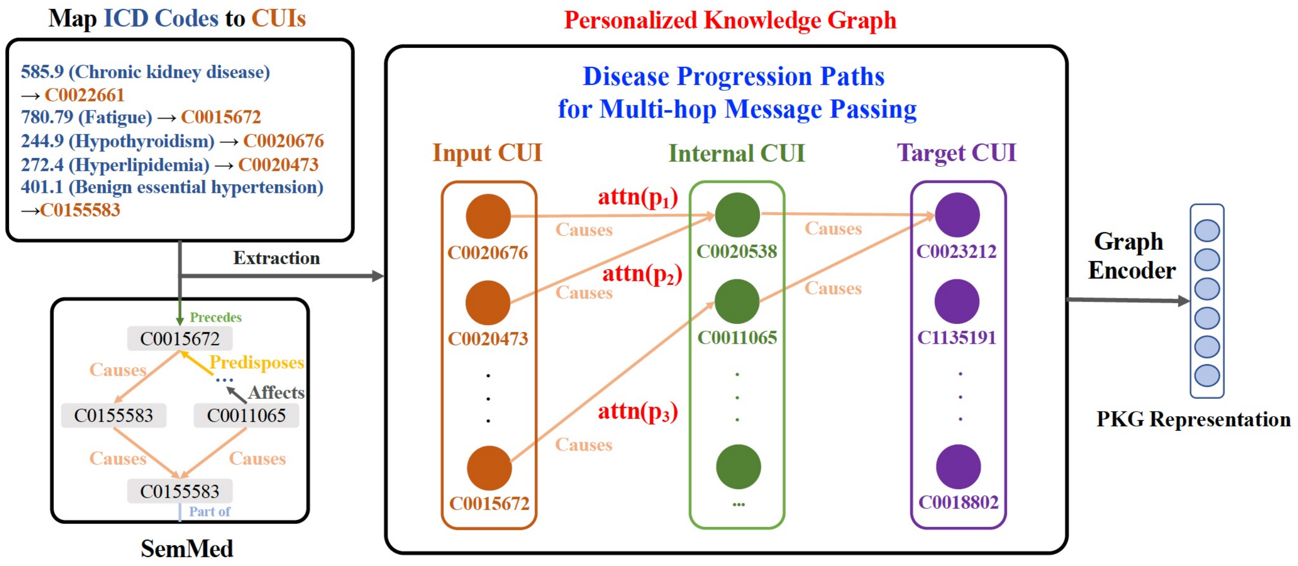
❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

Type-Specific Transformation

- Input CUIs
- Target CUIs
- Internal CUIs

$$\mathbf{v}_j = \mathbf{U}_t \mathbf{h}_j + \mathbf{b}_t$$

$\mathbf{h}_j$ *is the pretrained node embedding with TransE.*

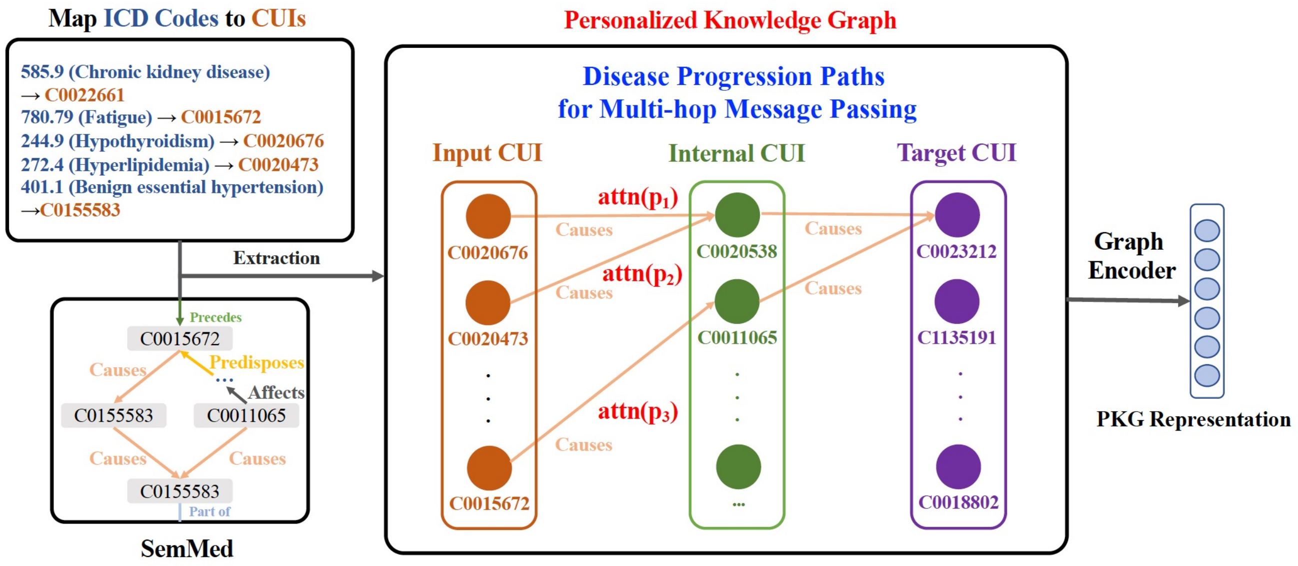


Multi-hop Message Passing

- K-hop paths

$$P_k = \{(e_s, r_1, \cdots, r_k, e_d) | (e_s, r_1, e_1), \cdots, (e_{k-1}, r_k, e_d) \in \mathcal{G}\}, (1 \leq k \leq K)$$

$$e_s \in \{Input\ CUIs\}$$
$$e_d \in \{Target\ CUIs\}$$
r_j *is the jth relation in the path.*



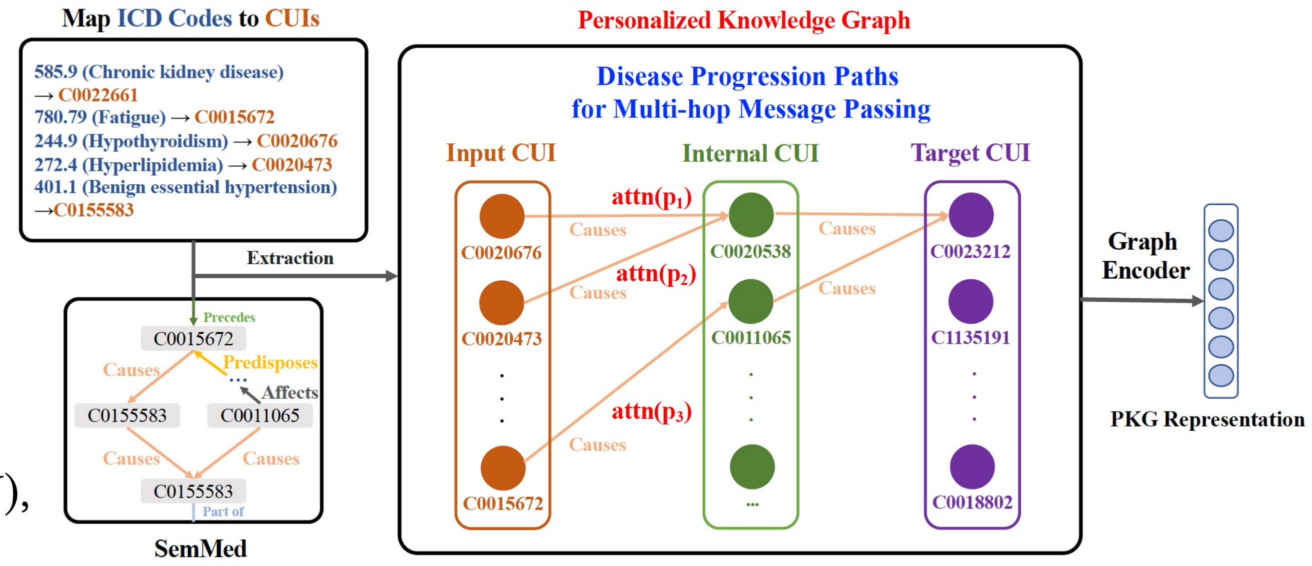
Multi-hop Message Passing

- Graph node embedding

$$\mathbf{h}'_j = \sigma(\mathbf{T} \cdot \mathbf{h}_j + \mathbf{T}' \cdot \mathbf{z}_j),$$

$$\mathbf{z}_j = \sum_{k=1}^{K} \text{Softmax}(\text{bilinear}(\mathbf{s}, \mathbf{z}_j^k)) \cdot \mathbf{z}_j^k,$$

$$\mathbf{z}_d^k = \sum_{p \in P_k} \text{attn}(p) \cdot W_0^K \cdots W_0^{k+1} W_{r_k}^k \cdots W_{r_1}^1 \mathbf{v}_s, (1 \le k \le K),$$



Transformation matrix $W_{r_l}^t$:
how this relation passes the information from source node e_s to e_d

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

Structured Relational Attention

- Transition Matrix-based Attention

$$\text{attn}(p) = \text{probability}(p|\mathbf{s}),$$

  - A probabilistic graphical model
  - Conditional random field

$$\begin{aligned}&\text{probability}(p|\mathbf{s})\\ &\propto \exp\Big(\mu(\phi(e_s),\mathbf{s}) + \sum_{t=1}^{k}\delta(r_t,\mathbf{s}) + \sum_{t=1}^{k-1}\tau(r_t,r_{t+1}) + \nu(\phi(e_d),\mathbf{s})\Big)\\ &\triangleq \underbrace{\beta(r_1,\cdots,r_k,e_d)}_{\text{Relation Type Attention}} \cdot \underbrace{\gamma(\phi(e_s),\phi(e_d),\mathbf{s})}_{\text{Node Type Attention}},\end{aligned}$$

where function $\phi(\cdot)$ outputs the node type of the input node. In implementation, functions $\mu(\cdot)$, $\nu(\cdot)$ and $\delta(\cdot)$ are learned by two-layer multilayer perceptrons (MLPs) and $\tau(\cdot)$ by a transition matrix $\in \mathbb{R}^{m\times m}$, where m is the number of relations.

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

Structured Relational Attention

- Relational Self-Attention
  - For modeling the differences among different patients, we need to use a dynamic score matrix for each relation type at each hop conditioned on the source node s, instead of using a fixed relation transition matrix $\tau(\cdot)$.

hop-specific transformation: $\mathbf{a}_j = \mathbf{M}_j \mathbf{s}, \quad \mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_k]$

$$\mathrm{SelfAttention}(\mathbf{A}) = \mathrm{Softmax}\left(\frac{\mathbf{A}_q \mathbf{A}_k^\top}{\sqrt{d}}\right)\mathbf{A}_v,$$

$$\beta(r_1, \cdots, r_k, \mathbf{s}) = \mathrm{Softmax}(\mathbf{M}_l \cdot \mathrm{SelfAttention}(\mathbf{A})).$$

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

Prediction

- Attentive pooling over all the target CUI entity features to obtain graph embeddings $\boldsymbol{g}$
- Concatenate $\boldsymbol{g}$ and $\mathbf{s}$ to compute the final output by FC($\mathbf{s} \oplus \boldsymbol{g}$)
- Cross-entropy loss

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

Results

- MedPath-TA: Transition Matrix-based Attention
- MedPath-SA: Relational Self-Attention

Table 2: Performance Comparison (with the p-values of significance test) in terms of AUC. The average AUC scores of our MedPath variants MedPath-TA and MedPath-SA for each dataset are followed by the percentage improvement (↑) over Vanilla models.

Dataset	Heart Failure			COPD			Kidney Disease		
Method	Vanilla	MedPath-TA	MedPath-SA	Vanilla	MedPath-TA	MedPath-SA	Vanilla	MedPath-TA	MedPath-SA
LSTM	0.708	0.716 (1e-10)	**0.739** (6e-10)	0.693	0.703 (4e-9)	**0.707** (7e-9)	0.739	0.762 (4e-10)	**0.774** (1e-10)
Dipole	0.687	0.744 (2e-8)	**0.751** (2e-8)	0.704	0.714 (2e-10)	**0.728** (1e-10)	0.755	0.765 (3e-7)	**0.768** (2e-7)
Retain	0.689	0.733 (2e-8)	**0.735** (5e-8)	0.699	0.723 (6e-10)	**0.730** (6e-10)	0.732	**0.766** (1e-7)	0.764 (3e-7)
SAnD	0.686	0.733 (1e-7)	**0.745** (1e-7)	0.692	0.736 (7e-10)	**0.737** (9e-11)	0.748	0.769 (2e-7)	**0.790** (5e-8)
RetainEx	0.688	0.738 (6e-9)	**0.751** (2e-9)	0.707	**0.746** (2e-9)	0.743 (2e-9)	0.728	0.772 (2e-8)	**0.786** (3e-9)
Timeline	0.705	**0.735** (3e-9)	0.729 (2e-8)	0.698	**0.713** (4e-9)	0.704 (1e-9)	0.756	0.761 (6e-9)	**0.769** (7e-9)
LSAN	0.738	0.729 (9e-8)	**0.745** (1e-7)	0.723	**0.728** (4e-6)	0.720 (2e-6)	0.766	0.765 (9e-7)	**0.782** (5e-8)
HiTANet	0.750	**0.785** (4e-8)	**0.785** (3e-8)	0.752	0.787 (7e-11)	**0.799** (1e-10)	0.792	0.800 (8e-8)	**0.810** (4e-7)
Average	0.706	0.739 (↑4.7%)	**0.748 (↑5.9%)**	0.709	0.731 (↑3.1%)	**0.734 (↑3.5%)**	0.752	0.770 (↑2.4%)	**0.780 (↑3.7%)**

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

Case Study

Table 4: Case study results of heart failure for showing the explicit interpretability that MedPath has.

EHR Data		**Visit 1**: 250.02 (Diabetes mellitus); **Visit 2**: 585.9 (Chronic kidney disease) and 780.79 (Fatigue); **Visit 3**: 244.9 (Hypothyroidism), 272.4 (Hyperlipidemia), and 401.1 (Benign essential hypertension); **Visit 4**: 585.9 (Chronic kidney disease); **Visit 5**: 585.9 (Chronic kidney disease); **Visit 6**: 585.9 (Chronic kidney disease) and 244.9 (Hypothyroidism)
1st Highest Attention Weighted Path	Weight: 0.0189	Hypothyroidism $\xrightarrow[E1]{CAUSES}$ Hypertensive disease $\xrightarrow[E2]{CAUSES}$ Left heart failure
	Evidence E1	*Animal studies suggest that hypertension leads to cardiac tissue hypothyroidism a condition that can by itself lead to heart failure.*
	Evidence E2	*Left ventricular failure in some SA/OHS patients may be the result of hypertensive cardiac disease.*
2nd Highest Attention Weighted Path	Weight: 0.0178	Hyperlipidemia $\xrightarrow[E3]{CAUSES}$ Hypertensive disease $\xrightarrow[E4]{CAUSES}$ Left heart failure
	Evidence E3	*A literature search indicates that Anglo-Saxon countries report alarming hyperplastic changes particularly in the liver blood clots hyperlipidemia leading to high blood pressure porphyria atypical leiomyomas and cervical hyperplasia.*
	Evidence E4	*Left ventricular failure in some SA/OHS patients may be the result of hypertensive cardiac disease.*
3rd Highest Attention Weighted Path	Weight: 0.0150	Fatigue $\xrightarrow[E5]{CAUSES}$ Cessation of life $\xrightarrow[E6]{CAUSES}$ Left heart failure
	Evidence E5	*In light of the magnitude of this sleep debt it is not surprising that fatigue is a factor in 57% of accidents leading to the death of a truck driver and in 10% of fatal car accidents and results in costs of up to 56 billion dollars per year.*
	Evidence E6	*Though rare death due to myocardial stunning and LV power failure can occur during ICD insertion.*
One of the Lowest Attention Weighted Path	Weight: 0.0000	Heart failure $\xrightarrow[E7]{CAUSES}$ Hypertensive disease $\xrightarrow[E8]{CAUSES}$ Left heart failure
	Evidence E7	*These findings suggest that the ATF3 activator tBHQ may have therapeutic potential for the treatment of pressure-overload heart failure induced by chronic hypertension or other pressure overload mechanisms.*
	Evidence E8	*Left ventricular failure in some SA/OHS patients may be the result of hypertensive cardiac disease.*

KDD

❖ Ye et al. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. WWW 2021.

MedRetriever

- ICD to CUI mapping
  - 70% ICD codes have 1 to 1 maps
- Explanation
  - Attention
    - Hard to be understood by humans
  - Path
    - No evidence

❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

Rethinking of risk prediction task

- ICD-9 401.1: Benign essential hypertension

Patient Care & Health Information › Diseases & Conditions

High blood pressure (hypertension)

Symptoms & causes | Diagnosis & treatment | Doctors & departments

Print

We're welcoming patients at Mayo Clinic
See our safety precautions in response to COVID-19.
Request an appointment.

Overview

High blood pressure (hypertension) is a common condition in which the long-term force of the blood against your artery walls is high enough that it may eventually cause health problems, such as heart disease.

Blood pressure is determined both by the amount of blood your heart pumps and the amount of resistance to blood flow in your arteries. The more blood your heart pumps and the narrower your arteries, the higher your blood pressure. A blood pressure reading is given in millimeters of mercury (mm Hg). It has two numbers.

- **Top number (systolic pressure).** The first, or upper, number measures the pressure in your arteries when your heart beats.
- **Bottom number (diastolic pressure).** The second, or lower, number measures the pressure in your arteries between beats.

You can have high blood pressure for years without any symptoms. Uncontrolled high blood pressure increases your risk of serious health problems, including heart attack and stroke. Fortunately, high blood pressure can be easily detected. And once you know you have high blood pressure, you can work with your doctor to control it.

Complications

The excessive pressure on your artery walls caused by high blood pressure can damage your blood vessels as well as your organs. The higher your blood pressure and the longer it goes uncontrolled, the greater the damage.

Uncontrolled high blood pressure can lead to complications including:

- **Heart attack or stroke.** High blood pressure can cause hardening and thickening of the arteries (atherosclerosis), which can lead to a heart attack, stroke or other complications.
- **Aneurysm.** Increased blood pressure can cause your blood vessels to weaken and bulge, forming an aneurysm. If an aneurysm ruptures, it can be life-threatening.
- **Heart failure.** To pump blood against the higher pressure in your vessels, the heart has to work harder. This causes the walls of the heart's pumping chamber to thicken (left ventricular hypertrophy). Eventually, the thickened muscle may have a hard time pumping enough blood to meet your body's needs, which can lead to heart failure.
- **Weakened and narrowed blood vessels in your kidneys.** This can prevent these organs from functioning normally.
- **Thickened, narrowed or torn blood vessels in the eyes.** This can result in vision loss.
- **Metabolic syndrome.** This syndrome is a group of disorders of your body's metabolism, including increased waist size, high triglycerides, decreased high-density lipoprotein (HDL) cholesterol (the "good" cholesterol), high blood pressure and high insulin levels. These conditions make you more likely to develop diabetes, heart disease and stroke.
- **Trouble with memory or understanding.** Uncontrolled high blood pressure may also affect your ability to think, remember and learn. Trouble with memory or understanding concepts is more common in people with high blood pressure.
- **Dementia.** Narrowed or blocked arteries can limit blood flow to the brain, leading to a certain type of dementia (vascular dementia). A stroke that interrupts blood flow to the brain also can cause vascular dementia.

Target: Heart Failure

Risk factors

A single risk factor may be enough to cause heart failure, but a combination of factors also increases your risk.

Risk factors include:

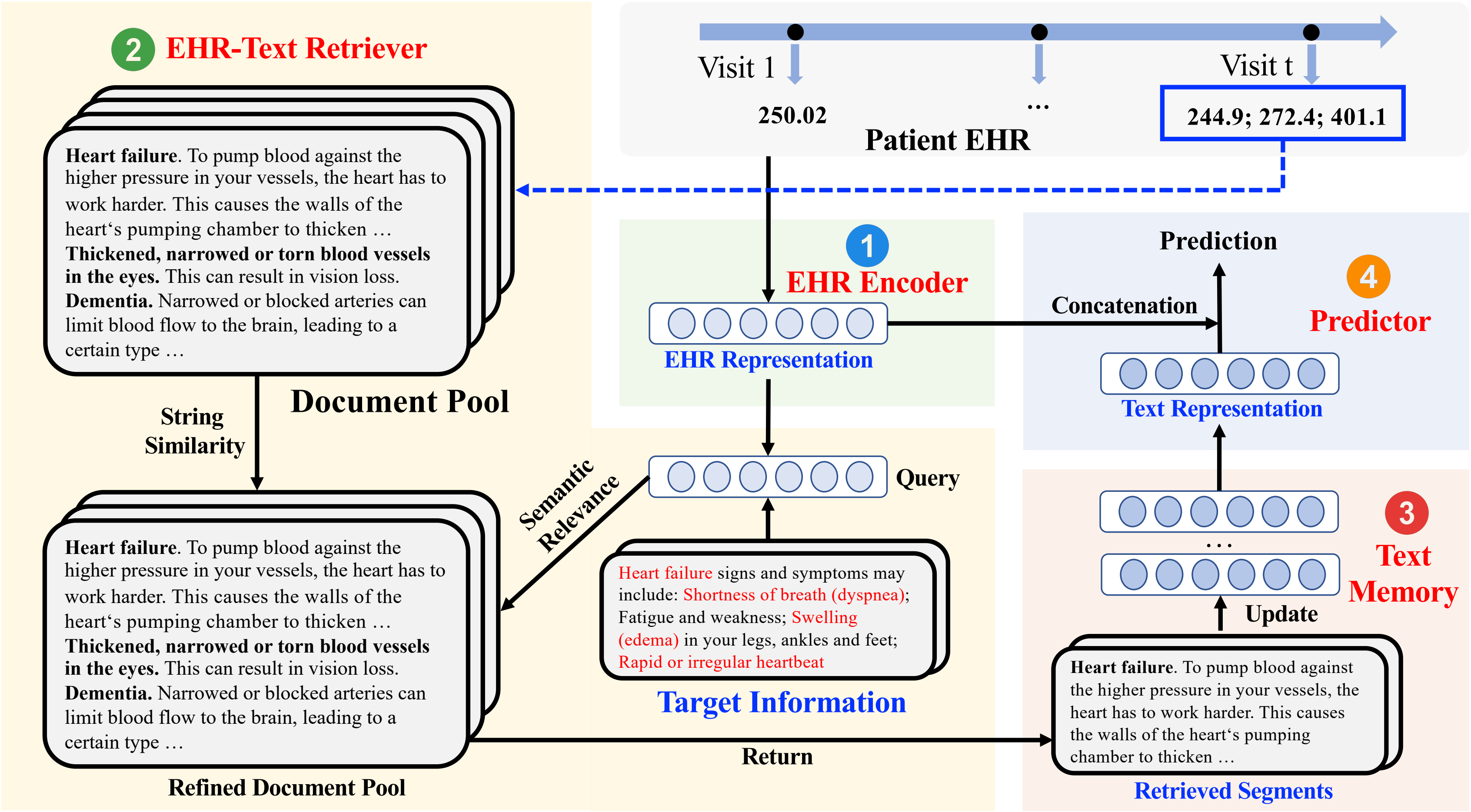
- **High blood pressure.** Your heart works harder than it has to if your blood pressure is high.
- **Coronary artery disease.** Narrowed arteries may limit your heart's supply of oxygen-rich blood, resulting in weakened heart muscle.
- **Heart attack.** A heart attack is a form of coronary disease that occurs suddenly. Damage to your heart muscle from a heart attack may mean your heart can no longer pump as well as it should.
- **Diabetes.** Having diabetes increases your risk of high blood pressure and coronary artery disease.

https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410
https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142

❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

KDD

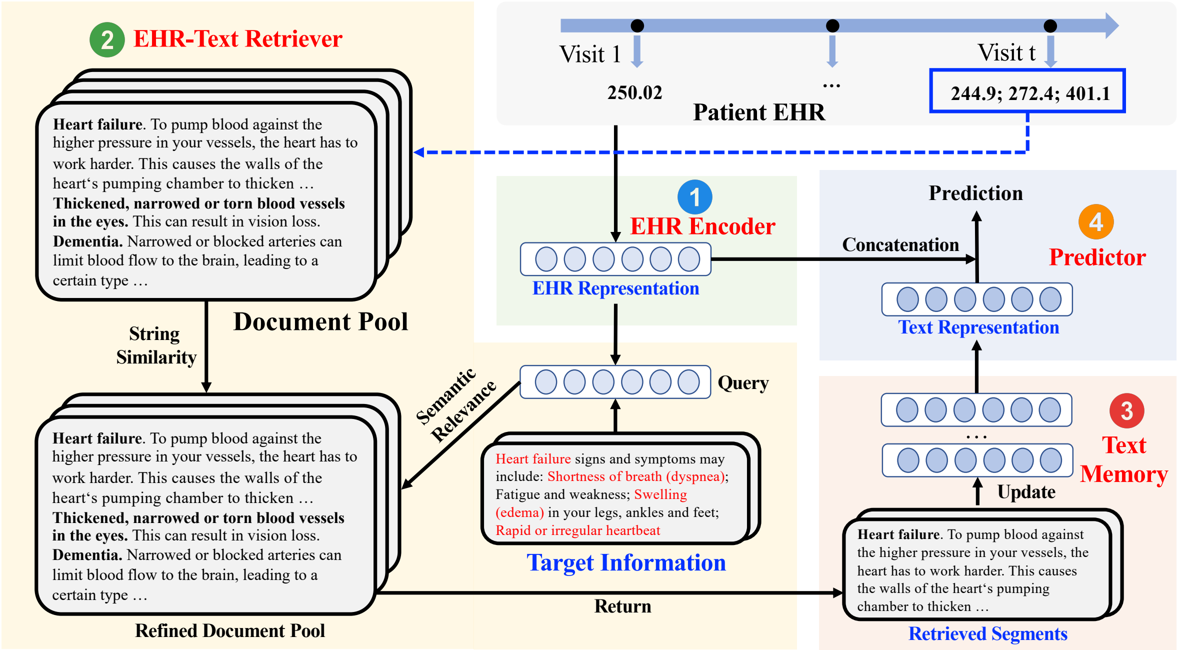
MedRetriever



❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

1. EHR Encoder

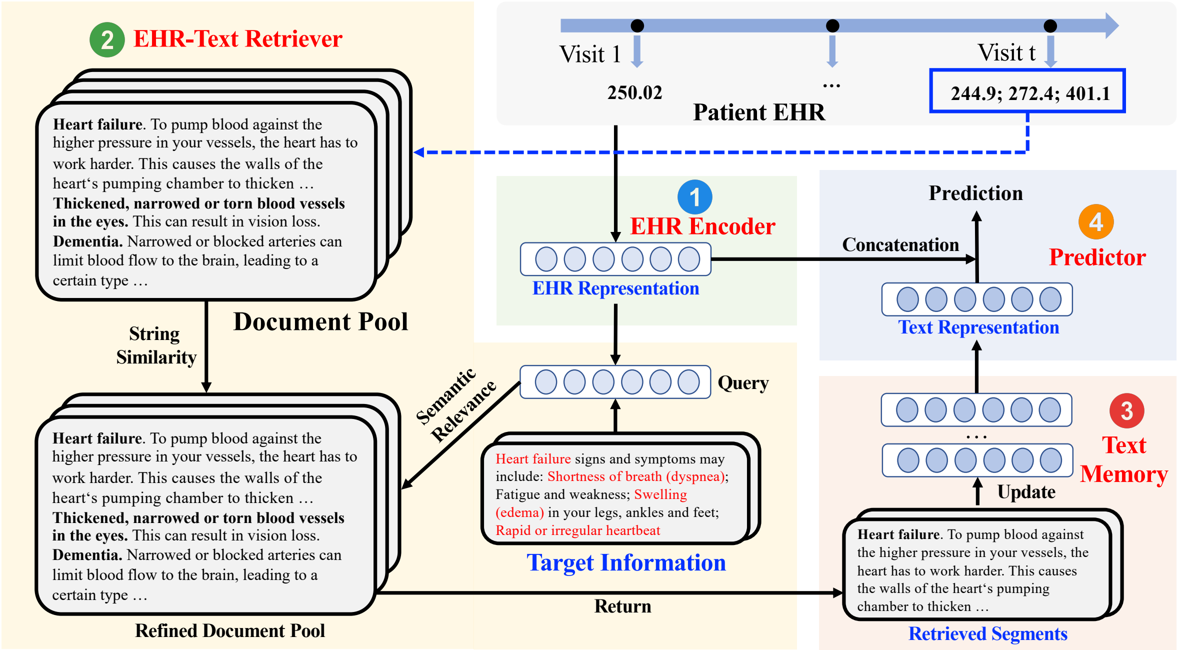
- RNN-based models
  - LSTM
  - Dipole
  - Retain
  - RetainEx
  - Timeline
- Transformer-based models
  - SAnD
  - LSAN
  - HiTANet
- ICD ontology-based model
  - GRAM



❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

2. EHR-Text Retriever

- Medical Text
  - Mayo Clinic
  - WebMD
- Preliminary Retrieval by String Similarity
  - Levenshtein distance
- Refined Retrieval by Semantic Relevance

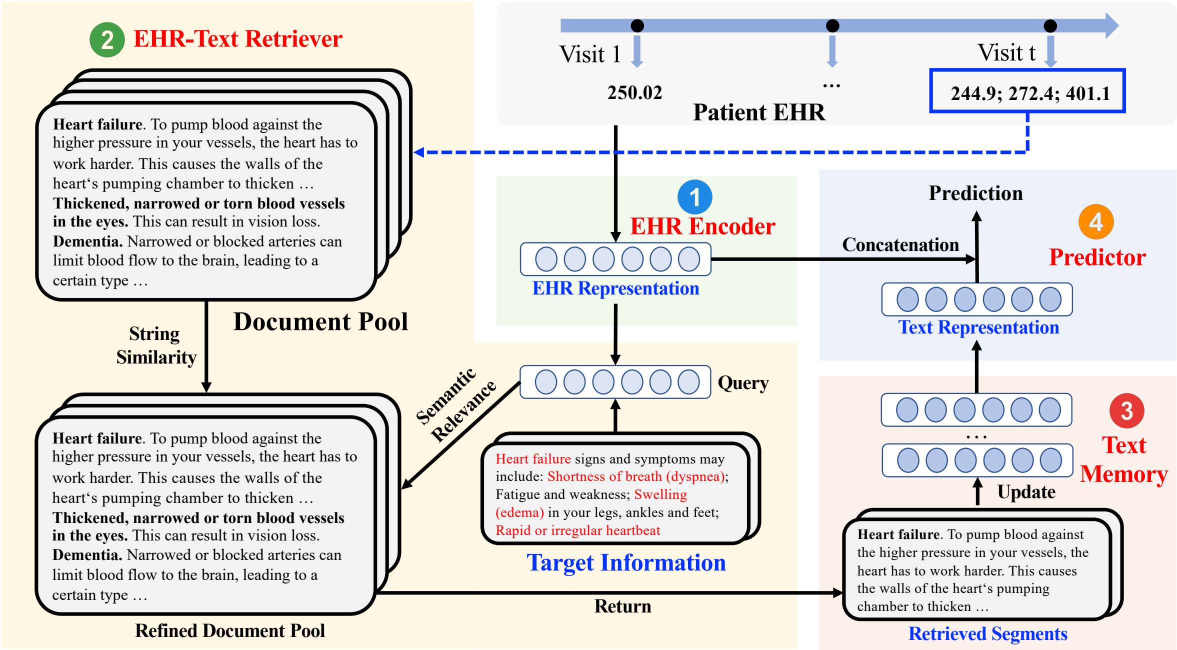


❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

KDD

3. Text Memory

- Dynamic updates
- Fixed size

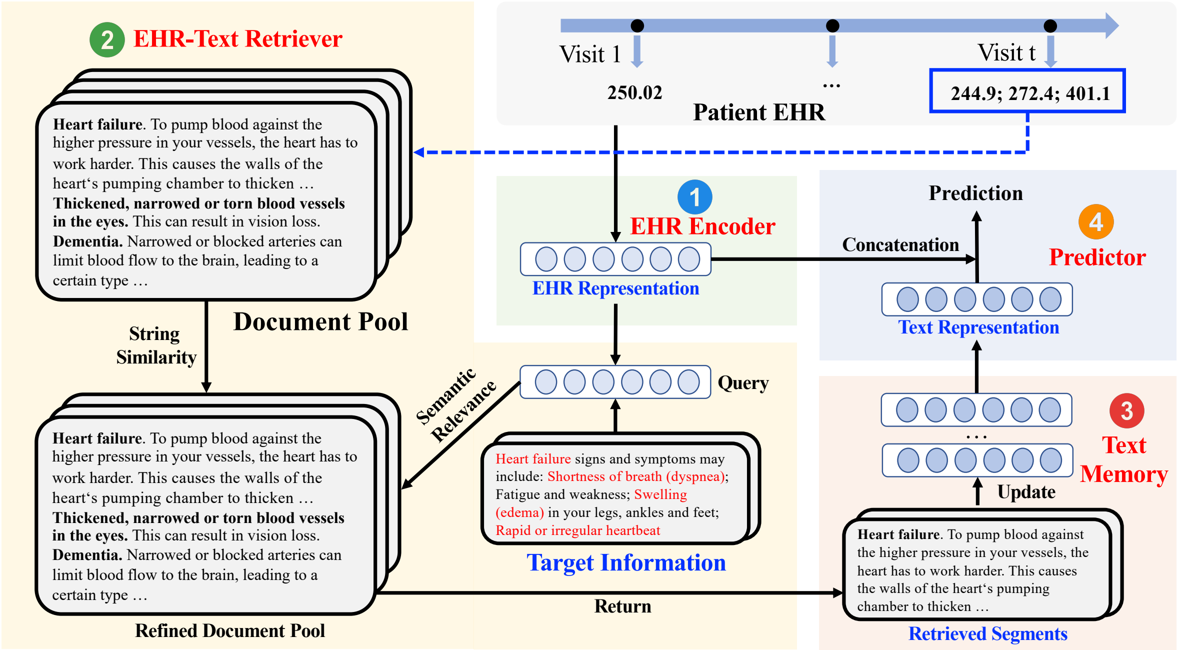


❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

4. Predictor

- Max pooling over segments stored in the memory to learn the text representation.

- EHR representation and text representation are used to make a prediction.

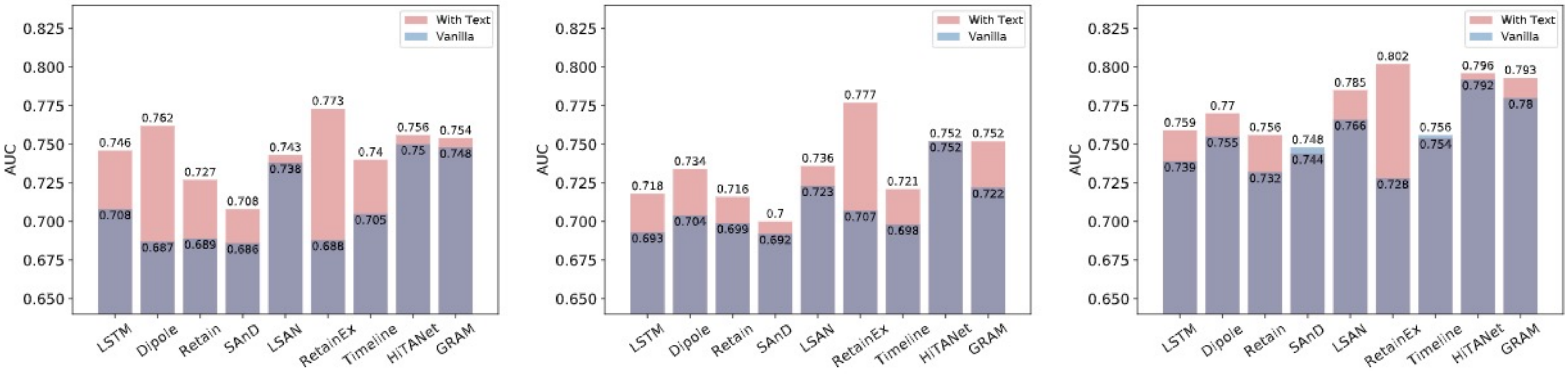


Experimental Results

Dataset	Heart Failure				COPD				Kidney Disease			
Metrics	AUC	Precision	Recall	F1	AUC	Precision	Recall	F1	AUC	Precision	Recall	F1
LSTM	0.708	0.640	0.510	0.561	0.693	0.680	0.461	0.548	0.739	0.680	0.572	0.616
Dipole	0.687	0.713	0.445	0.542	0.704	0.687	0.477	0.562	0.755	**0.771**	0.571	0.656
Retain	0.689	0.655	0.474	0.549	0.699	0.696	0.463	0.555	0.732	0.706	0.544	0.614
SAnD	0.686	0.661	0.466	0.544	0.692	0.653	0.462	0.539	0.748	0.690	0.592	0.636
LSAN	0.738	0.621	0.626	0.623	0.723	0.661	0.500	0.574	0.766	0.651	0.672	0.661
RetainEx	0.688	0.730	0.438	0.546	0.707	**0.728**	0.470	0.570	0.728	0.745	0.520	0.612
Timeline	0.705	0.661	0.510	0.574	0.698	0.654	0.478	0.550	0.756	0.697	0.607	0.648
HiTANet	0.750	**0.724**	0.587	0.647	0.752	0.707	0.583	0.637	0.792	0.743	0.668	**0.702**
GRAM	0.748	0.570	0.698	0.628	0.722	0.603	0.562	0.582	0.780	0.681	0.672	0.677
MedRetriever (std)	**0.773** (7e-3)	0.595 (4e-2)	**0.746** (3e-2)	**0.660** (1e-2)	**0.777** (6e-3)	0.576 (2e-2)	**0.725** (3e-2)	**0.645** (2e-3)	**0.802** (7e-3)	0.636 (5e-2)	**0.763** (4e-2)	0.688 (1e-2)

❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

Experimental Results



(a) Heart Failure (b) COPD (c) Kidney Disease

Figure 2: Comparison of AUC values with different baselines as the backbone for EHR embedding.

❖ Ye et al. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. CIKM 2021.

Case Study

EHR		**Visit 1:** Diabetes mellitus (250.00), Atrial fibrillation (427.31), Vaginitis and vulvovaginitis (616.10), Benign essential hypertension (401.1) **Visit 2:** Senile osteoporosis (733.01) **Visit 3:** Benign essential hypertension (401.1), Diabetes mellitus (250.00), Atrial fibrillation (427.31) **Visit 4:** Atrial fibrillation (427.31) **Visit 5:** Coronary atherosclerosis of native coronary artery (414.01), Atrial flutter (427.32), Diseases of tricuspid valve (397.0)
Visit 1	Target Disease Text	1. Other diseases. Chronic diseases — such as diabetes, HIV, hyperthyroidism, hypothyroidism, or a buildup of iron (hemochromatosis) or protein (amyloidosis) — also may contribute to heart failure. *(Weight: 0.02384)* 2. Coronary artery disease. Narrowed arteries may limit your heart's supply of oxygen-rich blood, resulting in weakened heart muscle. *(Weight: 0.02381)* 3. Diabetes. Having diabetes increases your risk of high blood pressure and coronary artery disease. *(Weight: 0.02379)*
	Text Memory	1. Age. The older you are, the greater your risk of developing atrial fibrillation. *(Weight: 0.0701)* 2. Inactivity. The less active you are, the greater your risk. Physical activity helps you control your weight, uses up glucose as energy and makes your cells more sensitive to insulin. *(Weight: 0.0698)* 3. Weight. Being overweight before pregnancy increases your risk of diabetes. *(Weight: 0.0686)*
Visit 2	Target Disease Text	1. Diabetes. Having diabetes increases your risk of high blood pressure and coronary artery disease. *(Weight: 0.02383)* 2. But heart failure can occur even with a normal ejection fraction. This happens if the heart muscle becomes stiff from conditions such as high blood pressure. *(Weight: 0.02380)* 3. Congenital heart defects. Some people who develop heart failure were born with structural heart defects. *(Weight: 0.02380)*
	Text Memory	1. Race. You're at greatest risk of osteoporosis if you're white or of Asian descent. *(Weight: 0.0537)* 2. Age. The older you get, the greater your risk of osteoporosis. *(Weight: 0.0521)* 3. Inactivity. The less active you are, the greater your risk. Physical activity helps you control your weight, uses up glucose as energy and makes your cells more sensitive to insulin. *(Weight: 0.0519)*
Visit 3	Target Disease Text	1. High blood pressure. Your heart works harder than it has to if your blood pressure is high. *(Weight: 0.02389)* 2. Valvular heart disease. People with valvular heart disease have a higher risk of heart failure. *(Weight: 0.02384)* 3. Heart rhythm problems. Heart rhythm problems (arrhythmias) can be a potential complication of heart failure. *(Weight: 0.02381)*
	Text Memory	1. Cardiovascular disease. Diabetes dramatically increases the risk of various cardiovascular problems, including coronary artery disease with chest pain (angina), heart attack, stroke and narrowing of arteries (atherosclerosis). If you have diabetes, you're more likely to have heart disease or stroke. *(Weight: 0.0526)* 2. Age. The older you get, the greater your risk of osteoporosis. *(Weight: 0.0525)* 3. Other chronic conditions. People with certain chronic conditions such as thyroid problems, sleep apnea, metabolic syndrome, diabetes, chronic kidney disease or lung disease have an increased risk of atrial fibrillation. *(Weight: 0.0514)*
Visit 4	Target Disease Text	1. Irregular heartbeats. These abnormal rhythms, especially if they are very frequent and fast, can weaken the heart muscle and cause heart failure. *(Weight: 0.02385)* 2. Diabetes. Having diabetes increases your risk of high blood pressure and coronary artery disease. *(Weight: 0.02384)* 3. Valvular heart disease. People with valvular heart disease have a higher risk of heart failure. *(Weight: 0.02383)*
	Text Memory	1. Cardiovascular disease. Diabetes dramatically increases the risk of various cardiovascular problems, including coronary artery disease with chest pain (angina), heart attack, stroke and narrowing of arteries (atherosclerosis). If you have diabetes, you're more likely to have heart disease or stroke. **(appear twice)** *(Weight: 0.0526)* 3. Heart failure. Atrial fibrillation, especially if not controlled, may weaken the heart and lead to heart failure — a condition in which your heart can't circulate enough blood to meet your body's needs. *(Weight: 0.0524)*
Visit 5	Target Disease Text	1. Irregular heartbeats. These abnormal rhythms, especially if they are very frequent and fast, can weaken the heart muscle and cause heart failure. *(Weight: 0.02382)* 2. Diabetes. Having diabetes increases your risk of high blood pressure and coronary artery disease. *(Weight: 0.02380)* 3. Coronary artery disease. Narrowed arteries may limit your heart's supply of oxygen-rich blood, resulting in weakened heart muscle. *(Weight: 0.02380)*
	Text Memory	1. Heart failure. Atrial fibrillation, especially if not controlled, may weaken the heart and lead to heart failure — a condition in which your heart can't circulate enough blood to meet your body's needs. **(appear twice)** *(Weight: 0.0519)* 2. Heart disease. Anyone with heart disease — such as heart valve problems, congenital heart disease, congestive heart failure, coronary artery disease, or a history of heart attack or heart surgery — has an increased risk of atrial fibrillation. *(Weight: 0.0516)*

KDD

Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding/Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

Multimorbidity

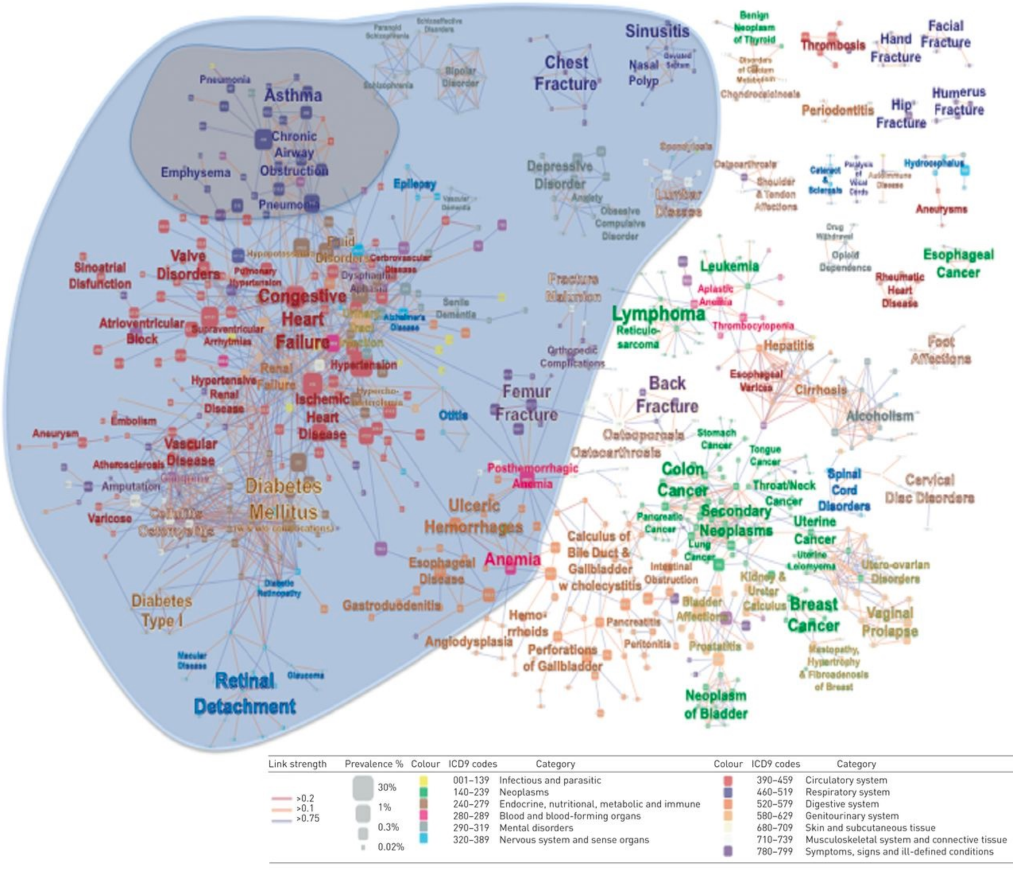
- Co-occurrence of multiple medical conditions
- Traditional way of prescribing is based on doctors’ intuition.
- Clinical decisions can be sub-optimal due to knowledge gaps.

Challenges of Managing Multimorbidity

- Adverse drug reactions:
  - 6.7% of patients in US suffer from serious drug reactions
  - 0.32 of such are fatal
  - Leading to a yearly cost of over $136 billion
- Solution:
  - Computer-assisted treatment recommendation?

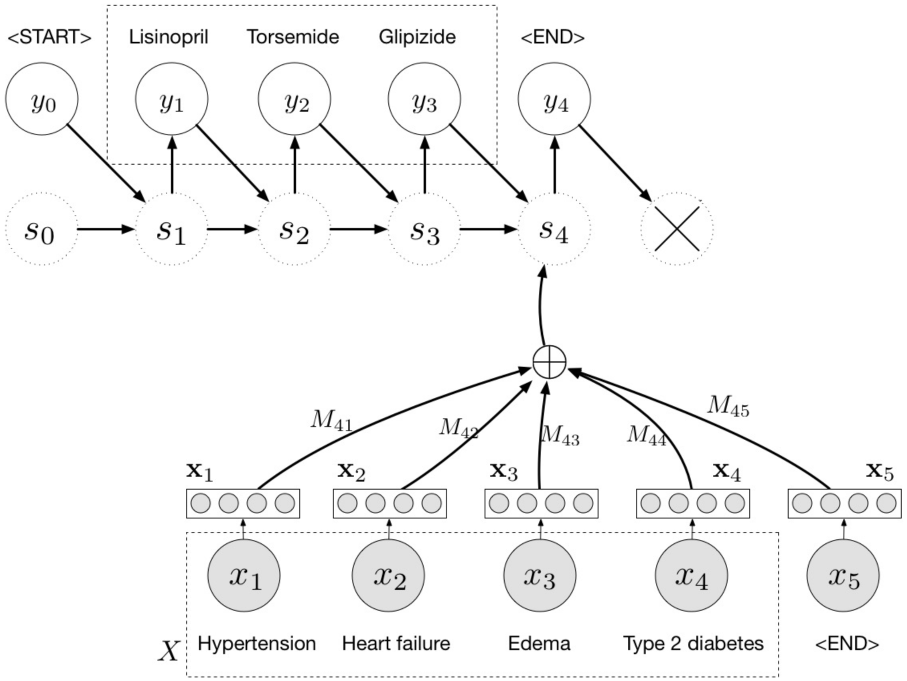
Hidden Knowledge from Electronic Health Records

- EHRs capture comprehensive medical histories of patients:
  - Diagnosis
  - Medications
  - Treatment plans
  - Lab test results…
- Discover hidden knowledge from existing EHR data



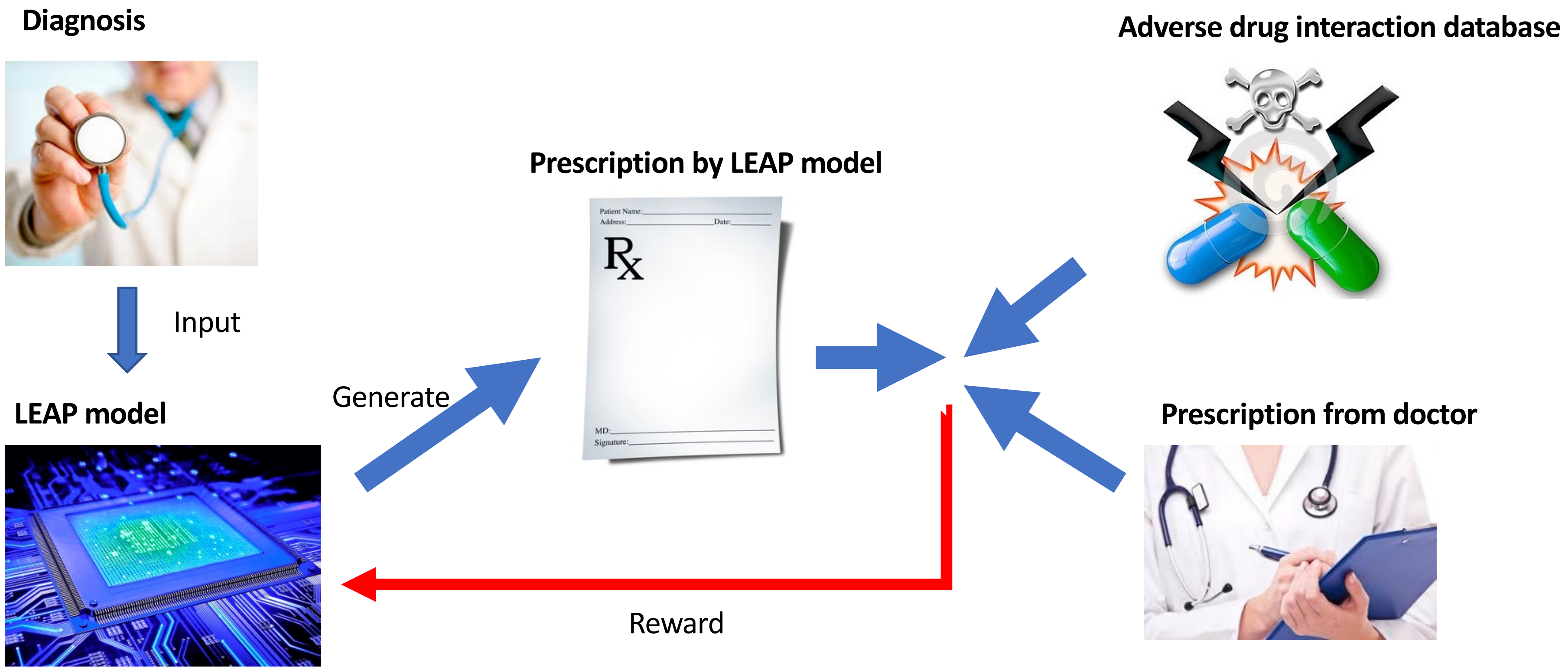
LEAP

- Decompose treatment recommendation into sequential decision making.
- Learning prescribing practice from EHR data
- Use distributed representation to encode diagnoses and medications.
- Use Recurrent Neural Network (RNN) to model the generation probability of the next medication in the treatment plan.

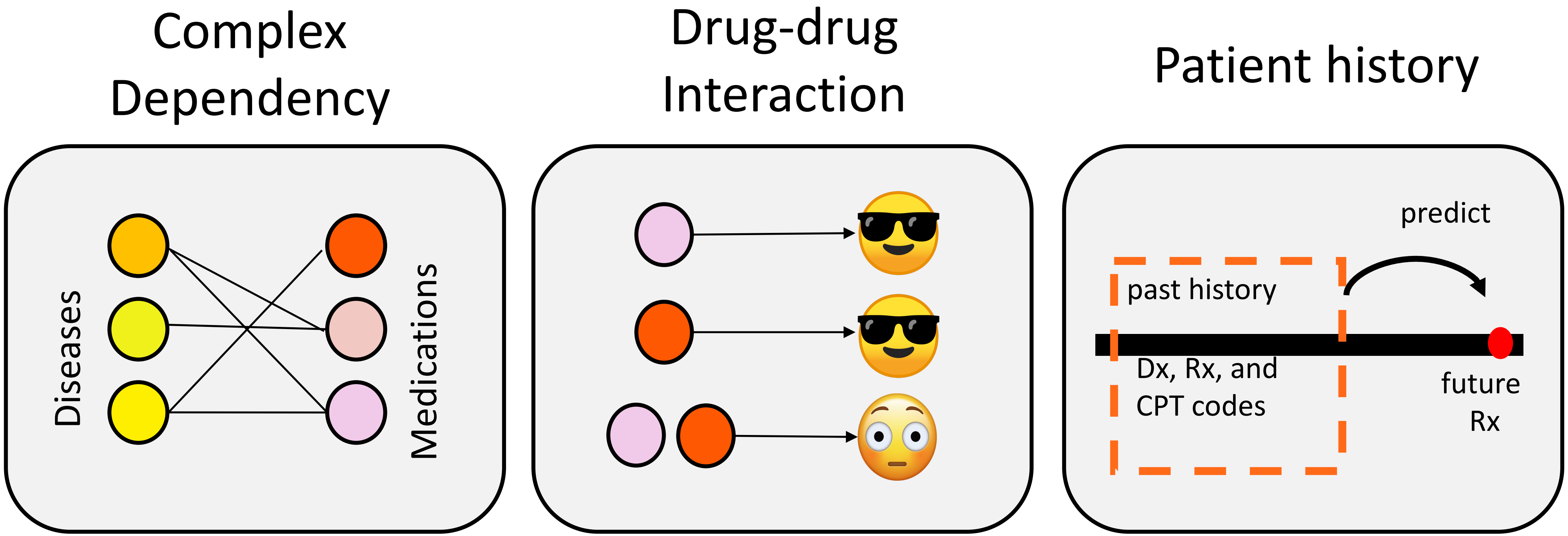


❖ Zhang et al. *LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity*. KDD 2017.

Reinforcement Fine-Tuning



Challenges for Medication Recommendation



❖ Shang et al. *GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination*. AAAI 2019.

GAMENet: Graph Augmented Memory Networks

Patient Representation

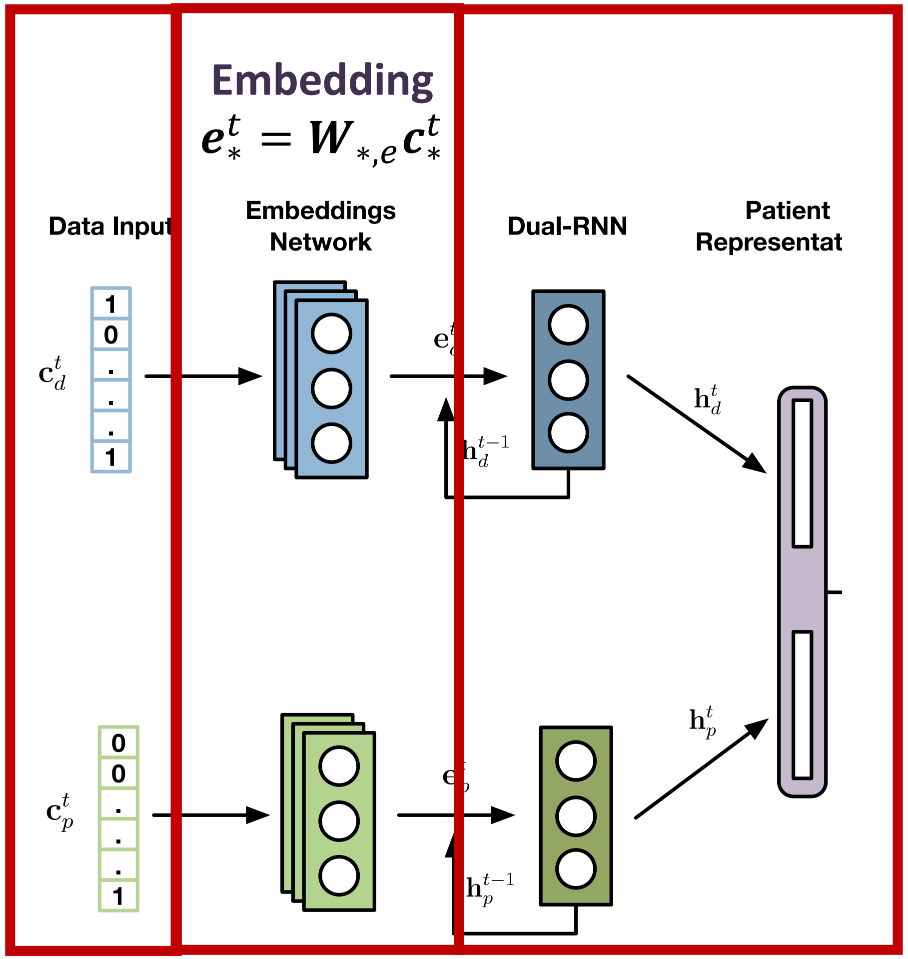
Graph Augmented Memory Network

❖ Shang et al. *GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination*. AAAI 2019.

Patient Representation Module

INPUT

Visit codes $\boldsymbol{c}_*^t$



OUTPUT

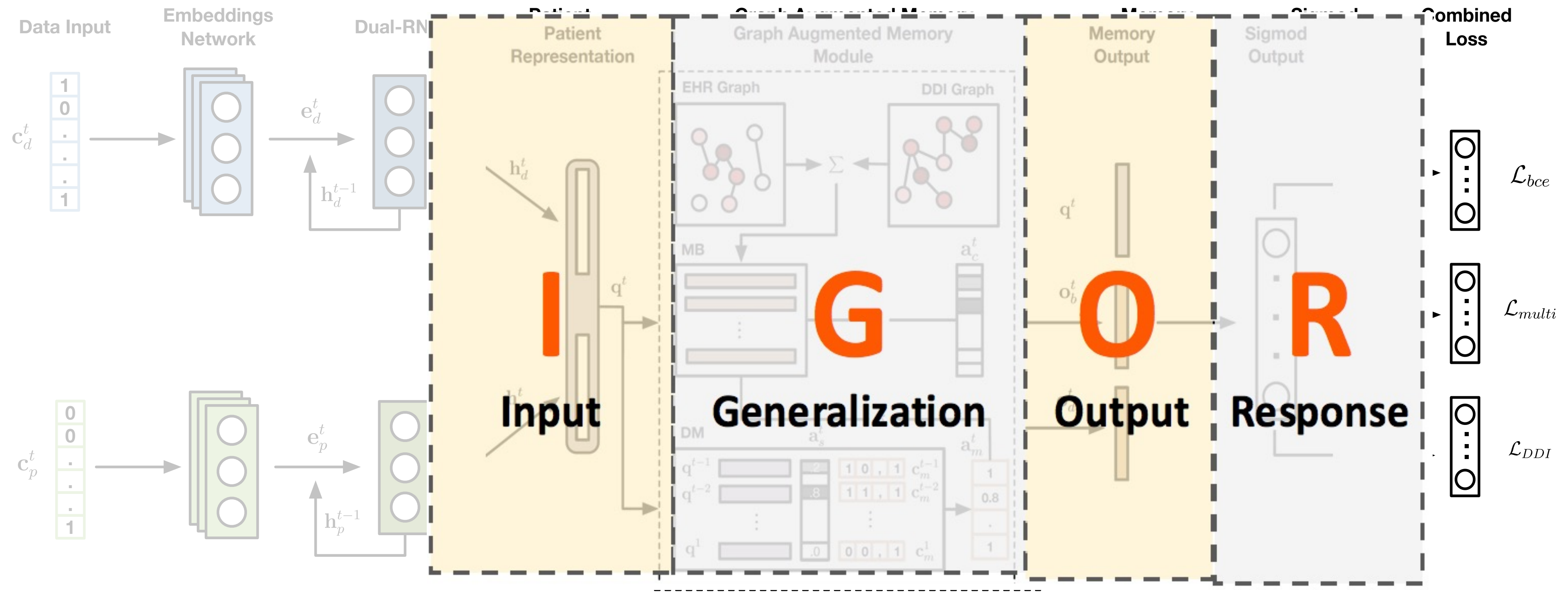
Patient Representation

$[\boldsymbol{h}_d^t, \boldsymbol{h}_p^t]$

$\boldsymbol{h}_d^t = RNN_d(\boldsymbol{e}_d^1, \cdots, \boldsymbol{e}_d^t)$ (diagnosis)

$\boldsymbol{h}_p^t = RNN_p(\boldsymbol{e}_p^1, \cdots, \boldsymbol{e}_p^t)$ (procedure)

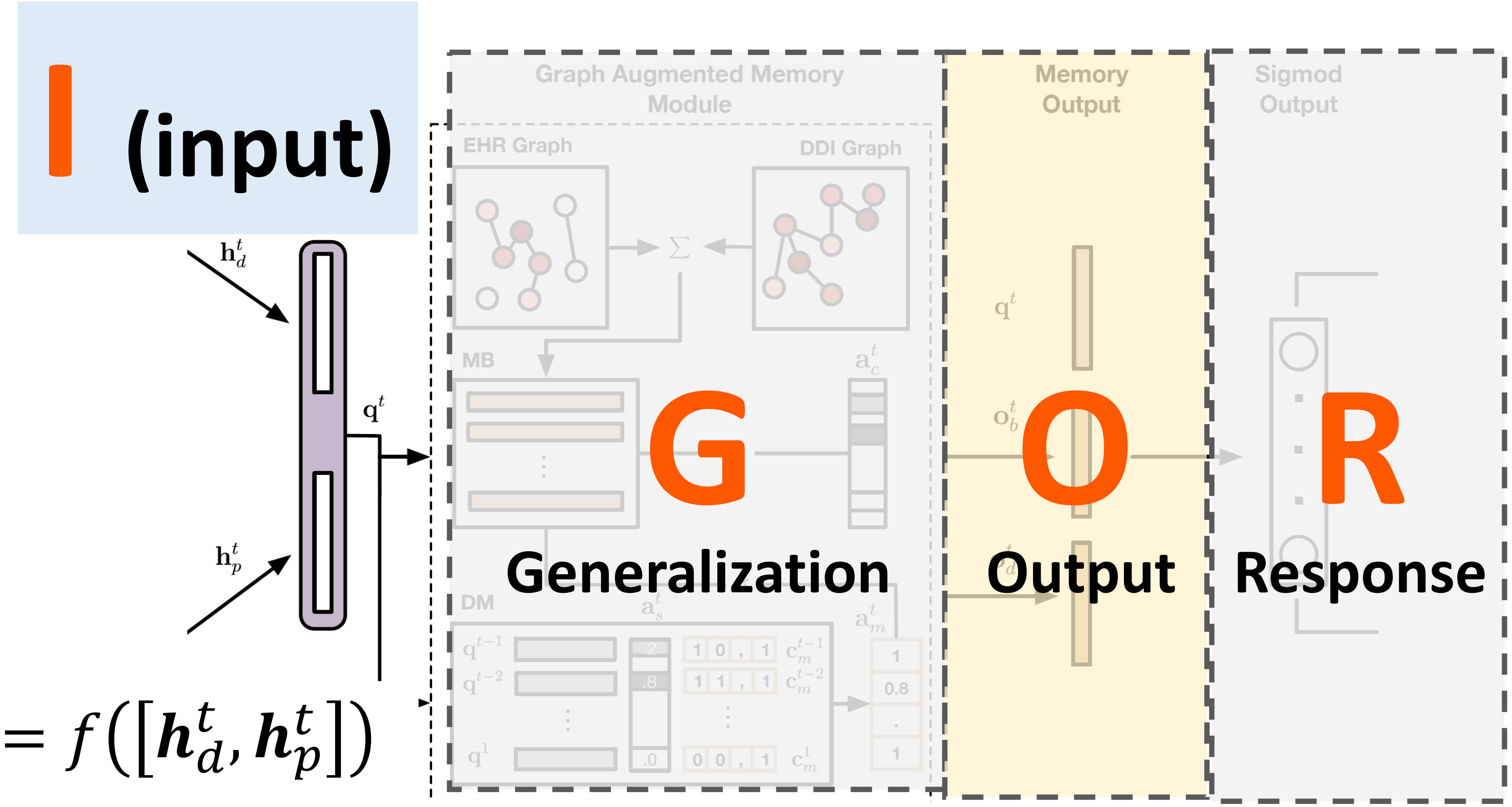
Graph Augmented Memory Module (I, G, O, R)



Graph augmented memory network that comprises of memory components **I, G, O, R.**

❖ Shang et al. *GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination*. AAAI 2019.

Graph Augmented Memory Module (I, G, O, R)

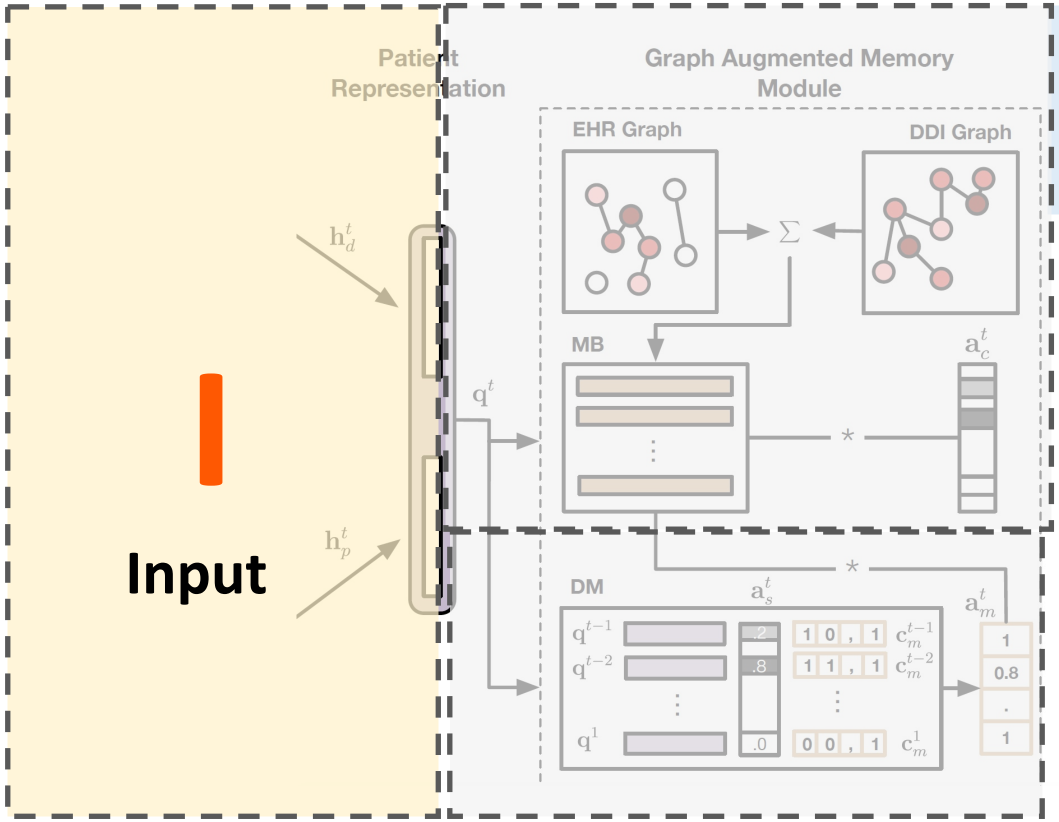


$$q^t = f([h_d^t, h_p^t])$$

Medical embedding h_d^t, h_p^t generates patient query q^t.

❖ Shang et al. *GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination*. AAAI 2019.

Graph Augmented Memory Module (I, G, O, R)

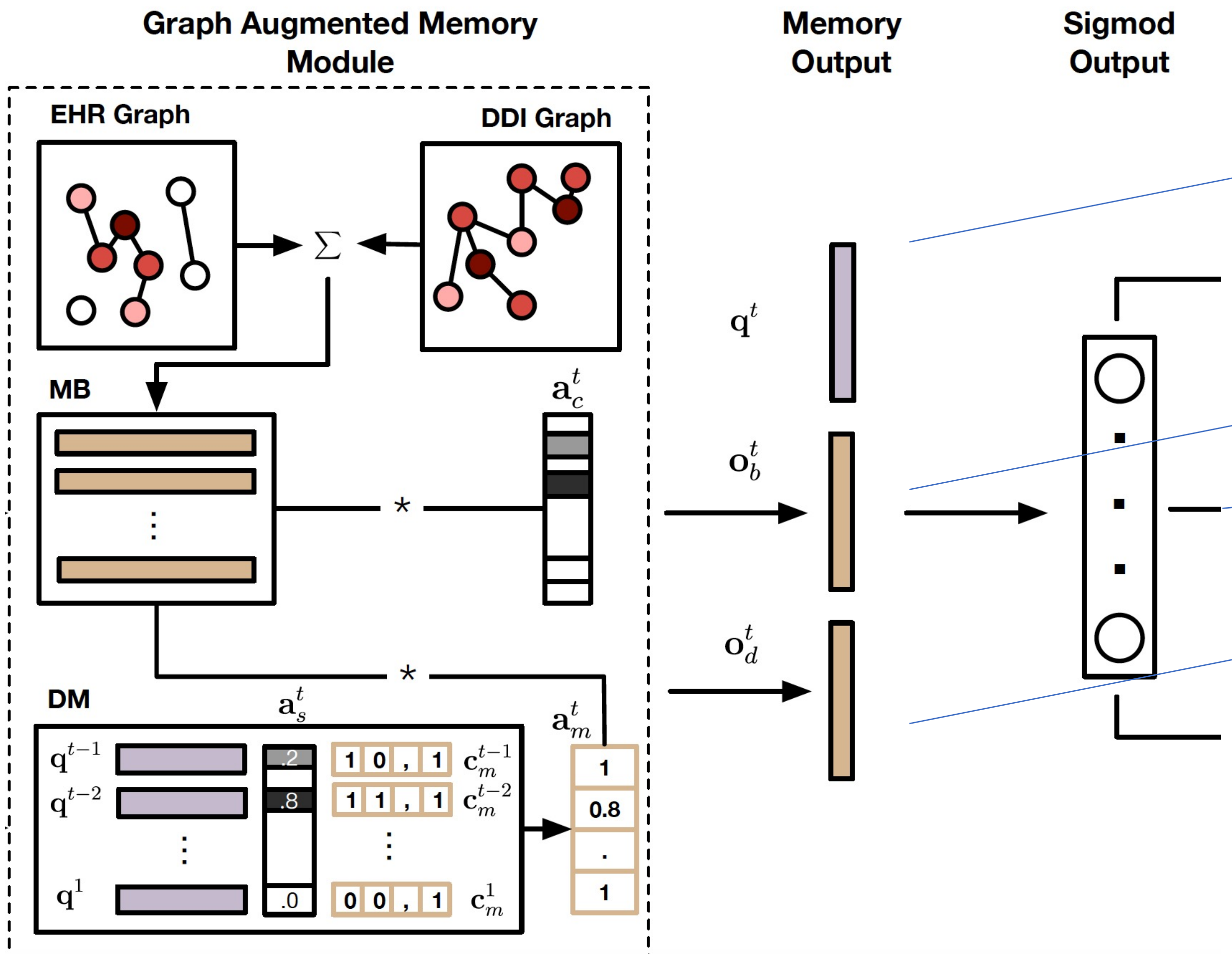


G (generalization)

Memory Bank (MB)

Dynamic Memory (DM)

Output and Response Module (I, G, O, R)



Patient query

Output from static memory

Output from dynamic memory

Medication prediction
$\hat{y}_t = \sigma([\boldsymbol{q}^t, \boldsymbol{o}_b^t, \boldsymbol{o}_d^t])$

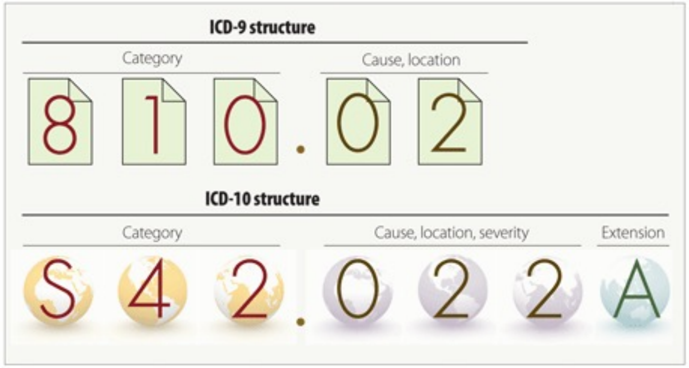
Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding/Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

ICD Coding

- International Classification of Diseases (ICD)
- The World Health Organization (WHO) currently develops and maintains the list for use by Member States.

Gross anatomy of ICD-9 and ICD-10 codes



Source: American Health Information Management Association

ICD-9	ICD-10
3-5 characters in length	3-7 characters in length
Approximately 13,000 codes	Approximately 68,000 available codes
First digit may be alpha (E or V) or numeric; digits 2-5 are numeric	Digit 1 is alpha; digits 2 and 3 are numeric; digits 4-7 are alpha or numeric (alpha digits are not case sensitive)
Limited space for adding new codes	Flexible for adding new codes
Lacks detail	Very specific
Lacks laterality	Has laterality (i.e., codes identifying right vs. left side of the body)
Use same code for every visit	Has possibility of identifying initial encounter, subsequent encounter; or sequela
Only 4 codes were reported on a claim form	Up to 12 codes can be reported on a claim form

Diagnosis	ICD-9	ICD-10
Cervical Sprain, initial encounter	847.0	**S13.4xxA**
Thoracic Sprain, initial encounter	847.1	**S23.3xxA**
Lumbar Sprain, initial encounter	847.2	**S33.5xxA**
Cervical Degenerative Disc Disease	722.4	**M50**
Thoracic Degenerative Disc Disease	722.51	**M51**
Lumbar Degenerative Disc Disease	722.52	**M51.2**

Clinical Notes

- A key component to communicate the current status of a patient.
- Support transitions of care, care planning, quality reporting, and billing.

- Include:
  - Discharge summary
  - Attending and/or Resident
  - Nurse
  - Specialist
    - Radiology, Pathology, ECG, Nutrition, Respiratory, Social work, …
  - Consultant
  - Referring physician
  - Emergency Department

Admission Date :
⟨ deidentified ⟩
Discharge Date :
⟨ deidentified ⟩
Date of Birth :
⟨ deidentified ⟩ Sex :
F
Service :
SURGERY
Allergies :
Patient recorded as having No Known Allergies to Drugs
Attending :
⟨ deidentified ⟩
Chief Complaint :
Dyspnea
Major Surgical or Invasive Procedure :
Mitral Valve Repair
History of Present Illness :
Ms. ⟨ deidentified ⟩ is a 53 year old female who presents after a large bleed rhythmically lag to 2 dose but the patient was brought to the Emergency Department where he underwent craniotomy with stenting of right foot under the LUL COPD and transferred to the OSH on ⟨ deidentified ⟩ .
The patient will need a pigtail catheter to keep the sitter daily .

Automated ICD Coding Task

1. Multiple Codes
2. Noisy text inputs
3. synonym

- Multilabel Classification Task

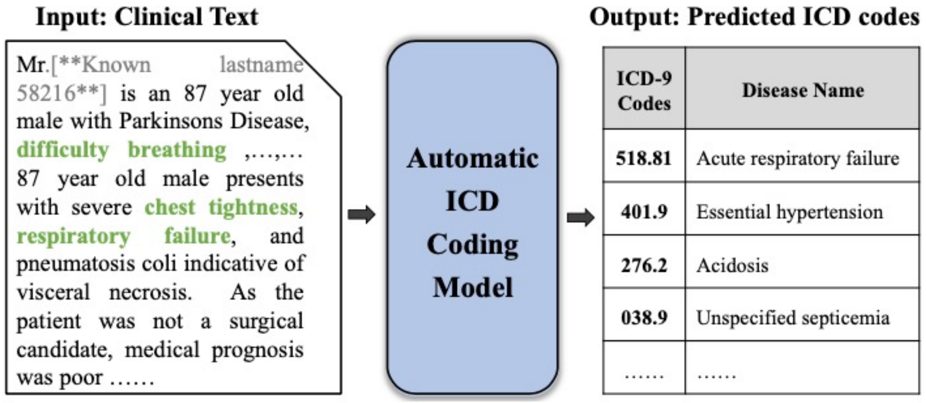


Figure 1: An example of automatic ICD coding task. The input and output of the automatic ICD coding model are clinical text and predicted ICD codes, respectively. For better understanding, we add the corresponding disease name for each code.

Source: Cao et al., HyperCore, ACL 2020

ICD-9	ICD-10
3-5 characters in length	3-7 characters in length
Approximately 13,000 codes	Approximately 68,000 available codes
First digit may be alpha (E or V) or numeric; digits 2-5 are numeric	Digit 1 is alpha; digits 2 and 3 are numeric; digits 4-7 are alpha or numeric (alpha digits are not case sensitive)
Limited space for adding new codes	Flexible for adding new codes
Lacks detail	Very specific
Lacks laterality	Has laterality (i.e., codes identifying right vs. left side of the body)
Use same code for every visit	Has possibility of identifying initial encounter, subsequent encounter; or sequela
Only 4 codes were reported on a claim form	Up to 12 codes can be reported on a claim form

Models

- C-MemNN [Prakash et al., AAAI'17]
- CAML [Mullenbach et al., NAACL'18]
- MultiResCNN [Li et al., AAAI'20]
- MSATT-KG [Xie et al., CIKM'19]
- HyperCore [Cao et al., ACL'20]
- Fusion [Luo et al., Findings of ACL'21]

Condensed Memory Networks for Clinical Diagnostic Inferencing

- Input

Medical Note (partially shown)
Date of Birth: [**2606–2–28**] Sex: M Service: Medicine **Chief Complaint:** Admitted from rehabilitation for hypotension (systolic blood pressure to the 70s) and decreased urine output. **History of present illness:** The patient is a 76-year-old male who had been hospitalized at the [**Hospital1 3007**] from [**8–29**] through [**9–6**] of 2002 after undergoing a left femoral-AT bypass graft and was subsequently discharged to a rehabilitation facility. On [**2682–9–7**], he presented again to the [**Hospital1 3087**] after being found to have a systolic blood pressure in the 70s and no urine output for 17 hours.

- Output

Diagnosis	
Cardiorespiratory arrest.	(427.5)
Non-Q-wave myocardial infarction.	(410.7)
Acute renal failure.	(584)

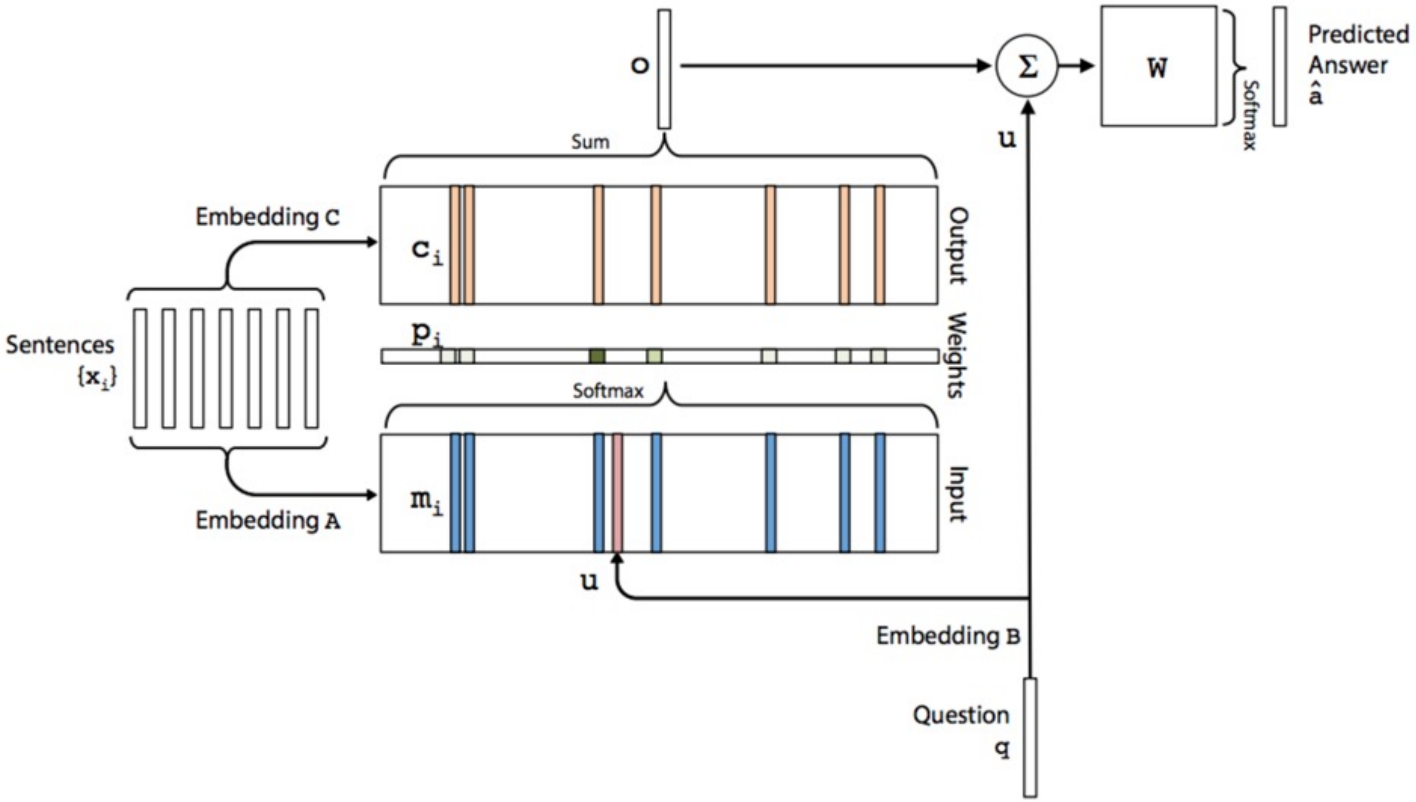
+

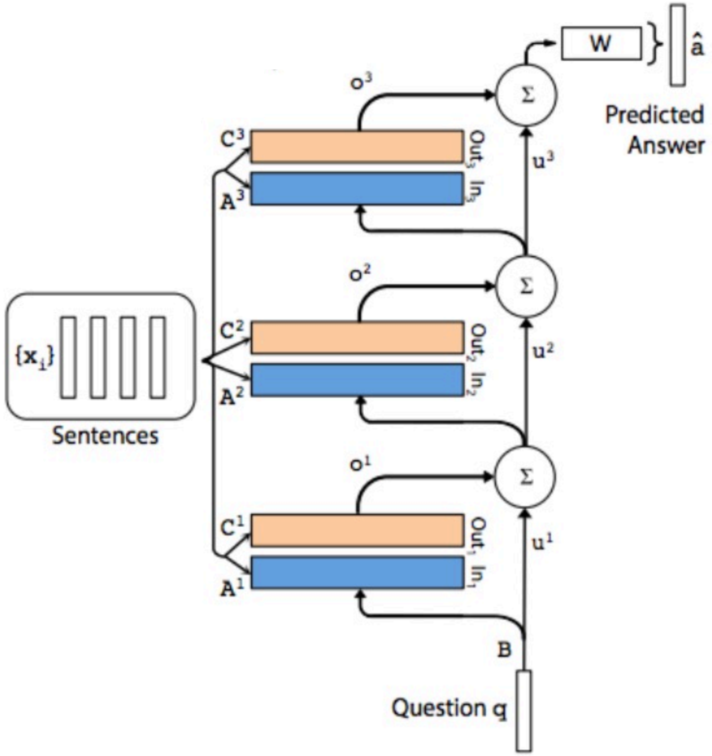
Cardiac arrest
Cardiac arrest is a sudden stop in effective blood circulation due to the failure of the heart to contract effectively or at all[1]. A cardiac arrest is different from (but may be caused by) a myocardial infarction (also known as a heart attack), where blood flow to the muscle of the heart is impaired such that part or all of the heart tissue dies. . . **Signs and symptoms** Cardiac arrest is sometimes preceded by certain symptoms such as fainting, fatigue, blackouts, dizziness, chest pain, shortness of breath, weakness, and vomiting. The arrest may also occur with no warning . . .

Partially shown example of a relevant Wikipedia page

End-to-End Memory Networks

```
Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.
Q: Where is the apple?
A. Bedroom
```





❖ Sukhbaatar et al. *End-To-End Memory Networks*. In NeurIPS 2015.

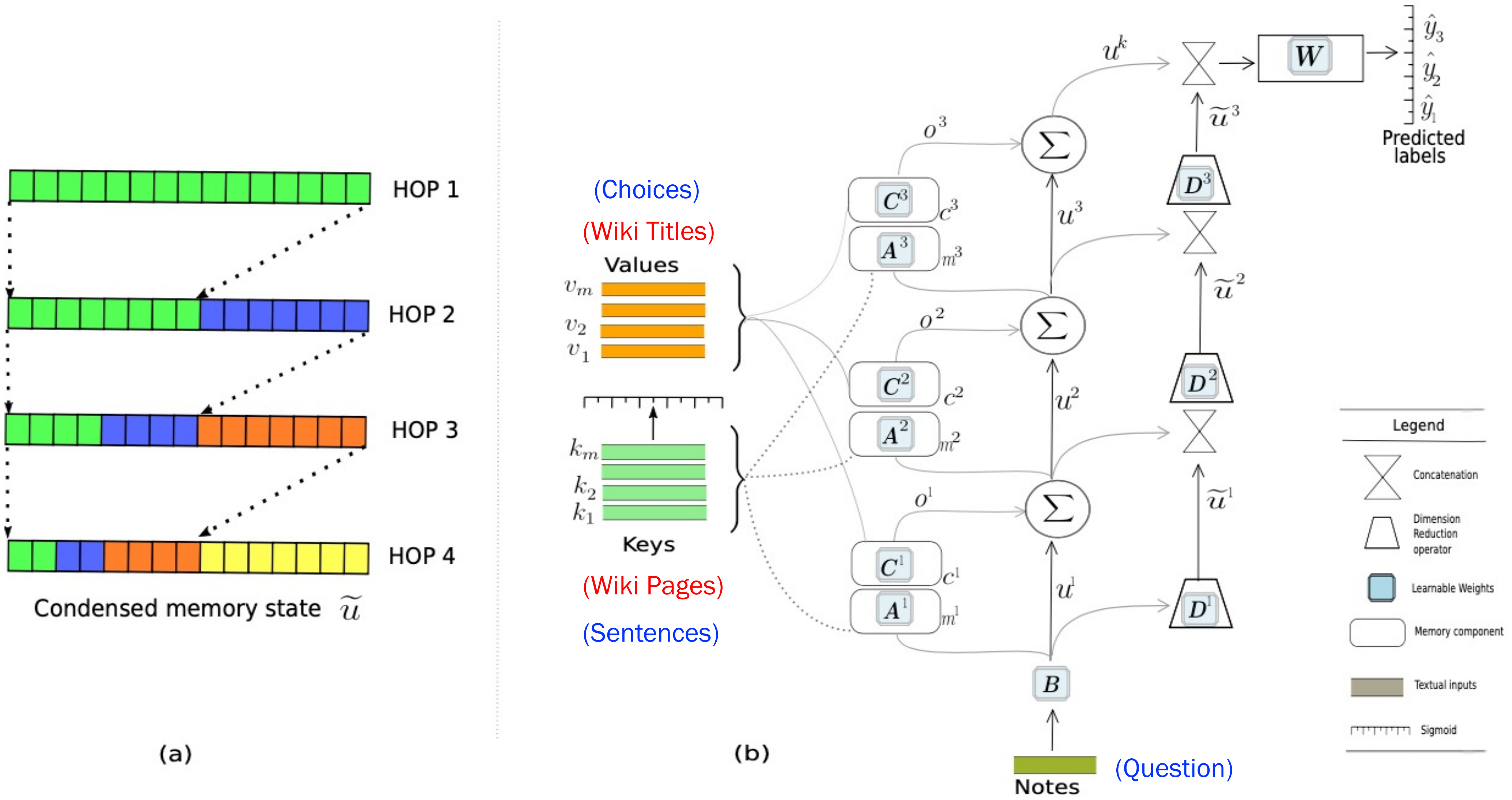


Figure 2: (a) Abstract view of transformation of memory representation over multiple hops. (b) Structural overview of end-to-end model for condensed memory networks.

Condensed Memory Networks for Clinical Diagnostic Inferencing

# Hops	Model	# classes = 50 AUC (macro) ↑	Average Precision @5 ↑	Hamming Loss ↓	# classes = 100 AUC (macro) ↑	Average Precision @5 ↑	Hamming Loss ↓
3	End-to-End	0.759	0.32	0.06	0.664	0.23	0.15
	KV MemNN	0.761	0.36	**0.05**	0.679	0.24	0.14
	A-MemNN	0.762	0.36	0.06	0.675	0.23	0.14
	C-MemNN	**0.785**	**0.39**	**0.05**	**0.697**	**0.27**	**0.12**
4	End-to-End	0.760	0.33	0.04	0.672	0.24	0.15
	KV MemNN	0.776	0.35	0.04	0.683	0.24	0.13
	A-MemNN	0.775	0.37	0.03	0.689	0.23	0.11
	C-MemNN	**0.795**	**0.42**	**0.02**	**0.705**	**0.27**	**0.09**
5	End-to-End	0.761	0.34	0.04	0.683	0.25	0.14
	KV MemNN	0.775	0.36	0.03	0.697	0.25	0.11
	A-MemNN	0.804	0.40	0.02	0.720	0.29	0.11
	C-MemNN	**0.833**	**0.42**	**0.01**	**0.767**	**0.32**	**0.05**

Table 3: Evaluation results of various memory networks on `MIMIC-III` dataset.

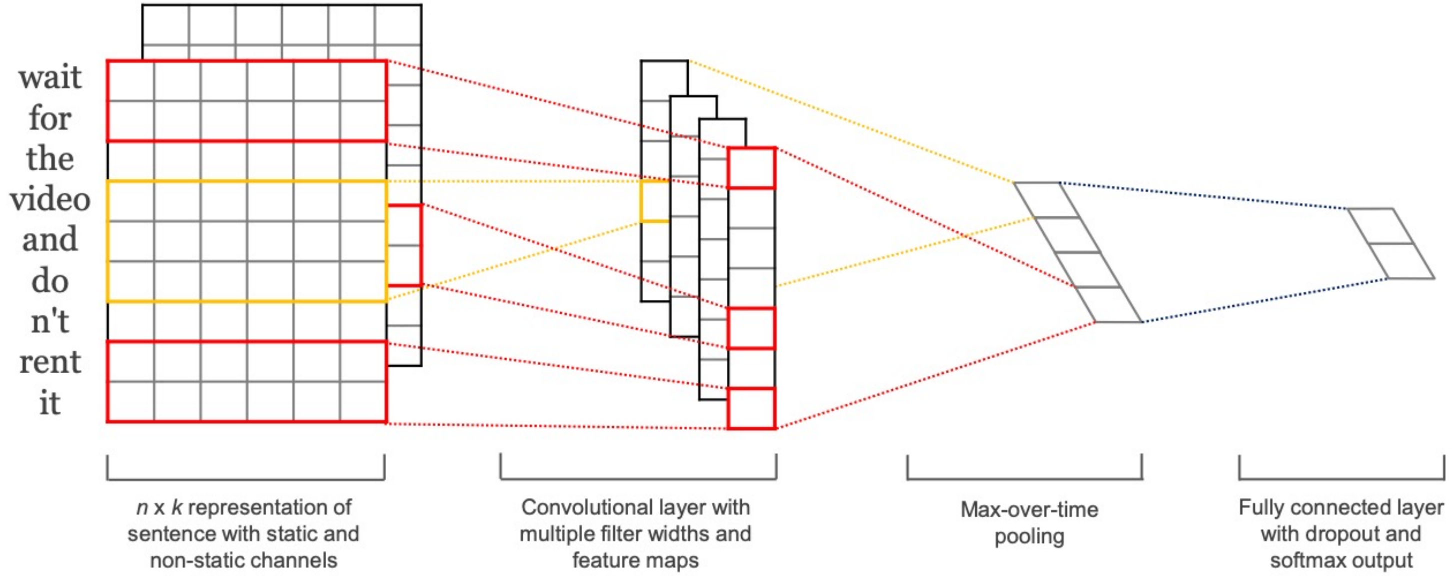
❖ Prakash et al. *Condensed Memory Networks for Clinical Diagnostic Inferencing*. AAAI'17.

Models

- C-MemNN [Prakash et al., AAAI'17]
- CAML [Mullenbach et al., NAACL'18]
- MultiResCNN [Li et al., AAAI'20]
- MSATT-KG [Xie et al., CIKM'19]
- HyperCore [Cao et al., ACL'20]
- Fusion [Luo et al., Findings of ACL'21]

Explainable Prediction of Medical Codes from Clinical Text

- Motivation:
  - Important information for code assignment usually contained in short snippets of text.
  - Convolutional Neural Networks (CNN)



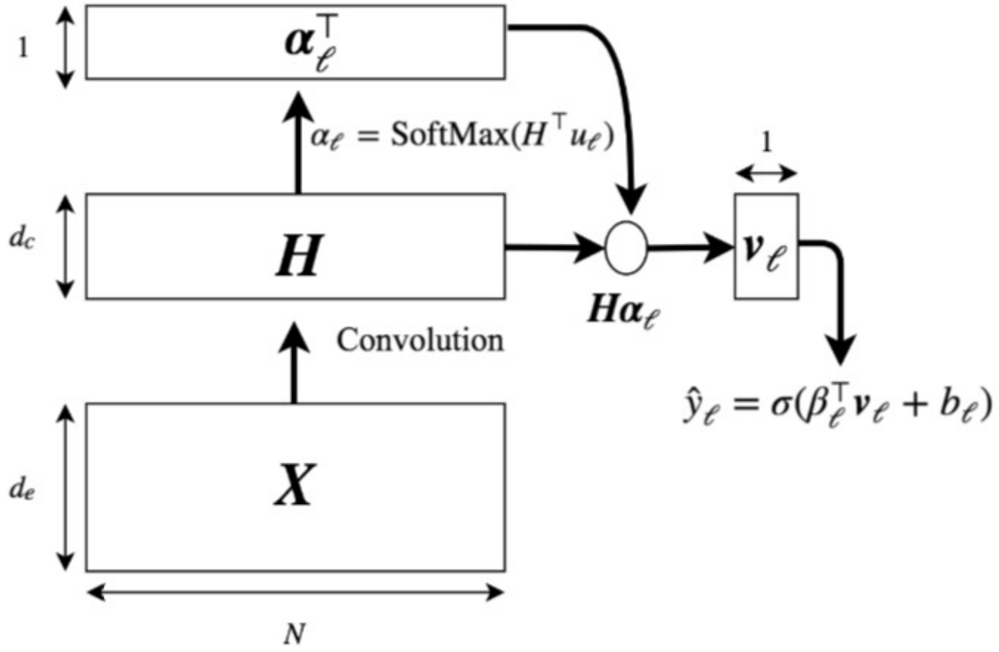
❖ Mullenbach et al. *Explainable Prediction of Medical Codes from Clinical Text*, NAACL'18.
❖ Kim, Yoon. *Convolutional Neural Networks for Sentence Classification*, EMNLP'17.

Explainable Prediction of Medical Codes from Clinical Text

- Challenge 1:
  - Large label space

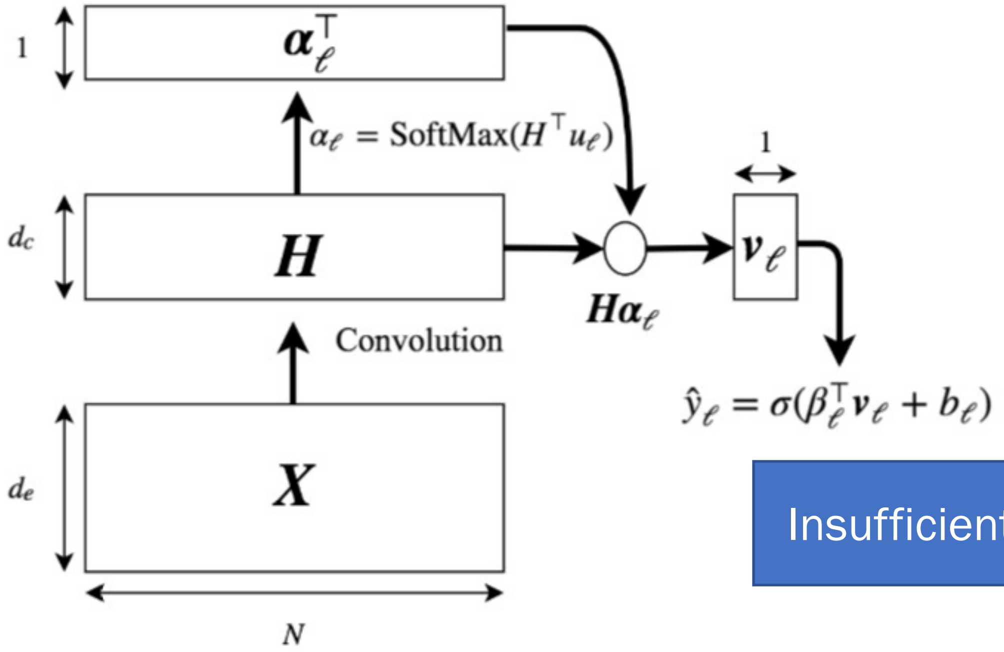
ICD-9	ICD-10
3-5 characters in length	3-7 characters in length
Approximately 13,000 codes	Approximately 68,000 available codes
First digit may be alpha (E or V) or numeric; digits 2-5 are numeric	Digit 1 is alpha; digits 2 and 3 are numeric; digits 4-7 are alpha or numeric (alpha digits are not case sensitive)
Limited space for adding new codes	Flexible for adding new codes
Lacks detail	Very specific
Lacks laterality	Has laterality (i.e., codes identifying right vs. left side of the body)
Use same code for every visit	Has possibility of identifying initial encounter, subsequent encounter; or sequela
Only 4 codes were reported on a claim form	Up to 12 codes can be reported on a claim form

- Code-wise Attention or Per-label attention



Explainable Prediction of Medical Codes from Clinical Text

- Challenge 2:
  - Small training set problem: some labels only have few training samples.



$$L_{\text{BCE}}(X, y) = -\sum_{\ell=1}^{\mathcal{L}} y_\ell \log(\hat{y}_\ell) + (1 - y_\ell)\log(1 - \hat{y}_\ell)$$

Insufficient training on β_l

Explainable Prediction of Medical Codes from Clinical Text

- Solution:
  - ICD code description
  - Add a regularizer
    - If code ℓ is rarely observed in the training data, this regularizer will encourage its parameters to be similar to those of other codes with similar descriptions.

Code	Description
Diagnosis codes	
996.41	Mechanical loosening of prosthetic joint
996.42	Dislocation of prosthetic joint
996.43	Prosthetic joint implant failure/breakage
996.44	Periprosthetic fracture around prosthetic joint
996.45	Periprosthetic osteolysis
996.46	Articular bearing surface wear of a prosthetic joint
996.47	Other mechanical complication of prosthetic joint implant
996.49	Other mechanical complication of other internal orthopedic device, implant, or graft

$$L(\boldsymbol{X}, \boldsymbol{y}) = L_{\mathrm{BCE}} + \lambda \frac{1}{n_y} \sum_{\ell : y_\ell = 1}^{\mathcal{L}} \|\boldsymbol{z}_\ell - \boldsymbol{\beta}_\ell\|_2$$

Obtained by a max-pooling CNN

Models

- C-MemNN [Prakash et al., AAAI'17]
- CAML [Mullenbach et al., NAACL'18]
- MultiResCNN [Li et al., AAAI'20]
- MSATT-KG [Xie et al., CIKM'19]
- HyperCore [Cao et al., ACL'20]
- Fusion [Luo et al., Findings of ACL'21]

MultiResCNN Model

- Motivation:
  - Lengths of text and grammar vary a lot in the MIMIC-III dataset.
  - It may not be sufficient to learn decent document representations from a flat and fixed-length convolutional architecture.

Table 1: Examples of clinical text fragments and their corresponding ICD codes.

998.32: Disruption of external operation wound ... wound infection, and **wound breakdown** ...
428.0: Congestive heart failure ... DIAGNOSES: 1. **Acute congestive heart failure** 2. Diabetes mellitus 3. Pulmonary edema ...
202.8: Other malignant lymphomas ... a 55 year-old female with **non Hodgkin's lymphoma** and acquired C1 esterase inhibitor deficiency ...
770.6: Transitory tachypnea of newborn ... Chest x-ray was consistent with **transient tachypnea of the newborn** ...
424.1: Aortic valve disorders ... mild **aortic stenosis with an aortic valve area** of 1.9 cm squared and 2+ **aortic insufficiency** ...

❖ Li et al., *ICD Coding from Clinical Text Using Multi-Filter Residual Convolutional Neural Network* , AAAI'20.

MultiResCNN Model

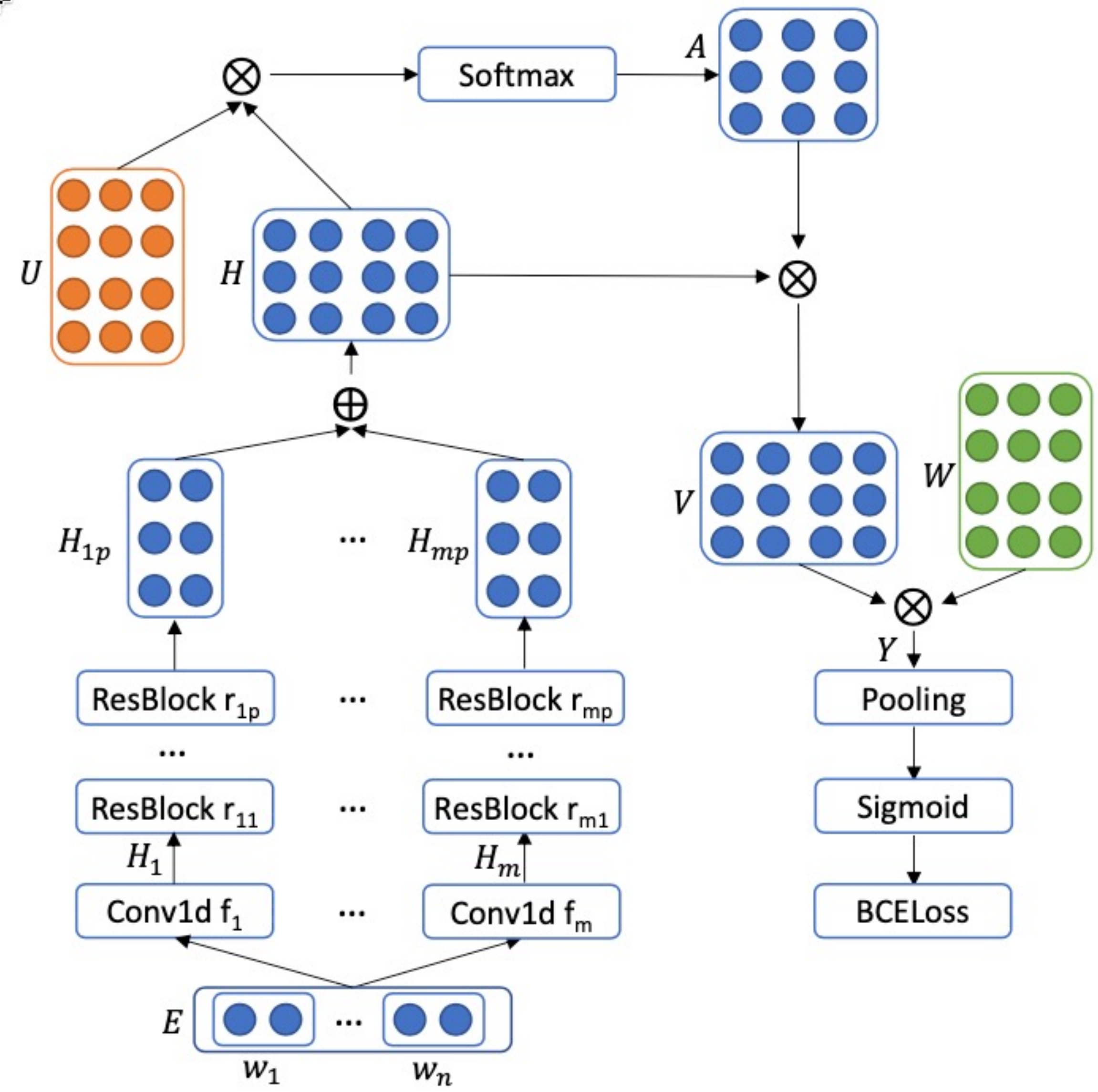
- Motivation:
  - Lengths of text and grammar vary a lot in the MIMIC-III dataset
- Solution:
  - Multi-Filter Residual Convolutional Neural network (Multi-ResCNN)
    - Multi-filter convolutional layers are used to capture the change of scaling.
    - A residual convolutional layer is used to enlarge receptive field (i.e., increasing the dimension of features or making feature more abstract).

KDD

❖ Li et al., *ICD Coding from Clinical Text Using Multi-Filter Residual Convolutional Neural Network*, AAAI'20.

MultiResCNN Model

Figure 1: The architecture of our MultiResCNN model. "Conv1d" represents the 1-dimensional convolution, "Res-Block" represents the residual block, "⊕" represents the concatenation operation and "⊗" represents the matrix multiplication. Here we use orange and green for U and W to denote they are learnable parameters, and to distinguish with other matrices (e.g., H) which are not parameters.



❖ Li et al., *ICD Coding from Clinical Text Using Multi-Filter Residual Convolutional Neural Network*, AAAI'20.

Multi-Filter Convolutional Layer

$$H_1 = f_1(E) = \bigwedge_{j=1}^{n} tanh(W_1^T E^{j:j+k_1-1}),$$

$$\ldots$$

$$H_m = f_m(E) = \bigwedge_{j=1}^{n} tanh(W_m^T E^{j:j+k_m-1}),$$

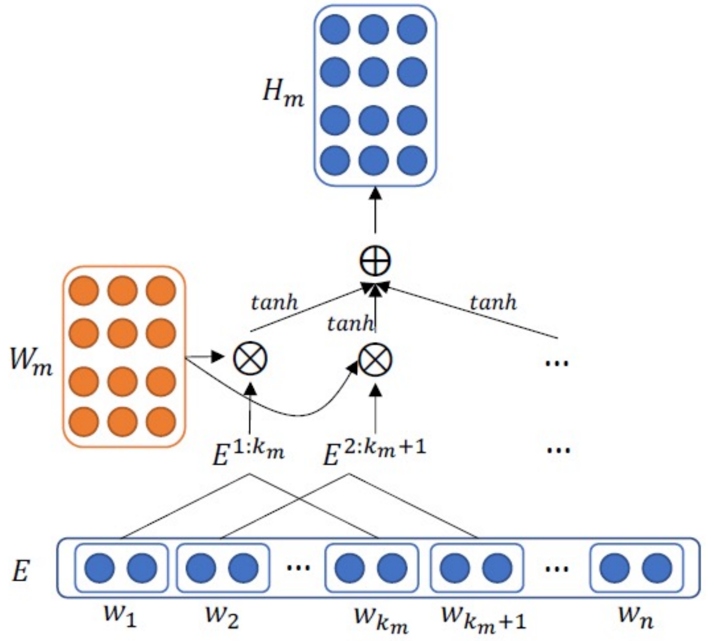


Figure 2: The architecture of a 1-dimensional convolution filter f_m. "$\oplus$" represents the concatenation operation and "$\otimes$" represents the matrix multiplication.

❖ Li et al., *ICD Coding from Clinical Text Using Multi-Filter Residual Convolutional Neural Network* , AAAI'20.

Residual Convolutional Layer

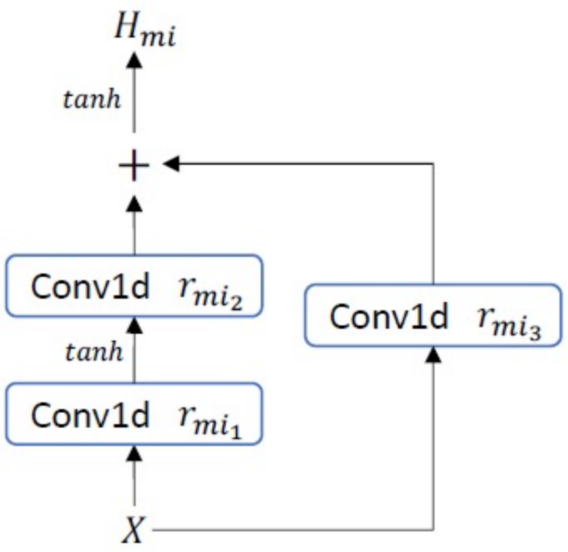


Figure 3: The architecture of a residual block r_{mi}. "+" represents the element-wise addition.

❖ Li et al., *ICD Coding from Clinical Text Using Multi-Filter Residual Convolutional Neural Network* , AAAI'20.

Models

- C-MemNN [Prakash et al., AAAI’17]
- CAML [Mullenbach et al., NAACL’18]
- MultiResCNN [Li et al., AAAI’20]
- MSATT-KG [Xie et al., CIKM’19]
- HyperCore [Cao et al., ACL’20]
- Fusion [Luo et al., Findings of ACL’21]

MSATT-KG

- Motivation:
  - Clinical note is composed of multiple long and heterogeneous textual narratives.
  - The code label space is large and the label distribution is extremely unbalanced.
- Solution:
  - Multi-scale Feature Attention and Structured Knowledge Graph Propagation
    - A densely connected convolutional neural network is used to produce variable n-gram features layer by layer.
    - Multi-scale feature attention is used to adaptively select most informative n-gram features.
    - Graph convolutional neural network to capture the hierarchical relationships among medical codes and the semantics of each code.

MSATT-KG

- The method is mainly composed of three parts:
- (1) clinical document multi-scale featu extraction;
- (2) two-level attention mechanism for better document representation learning;
- (3) structured knowledge graph propagation.

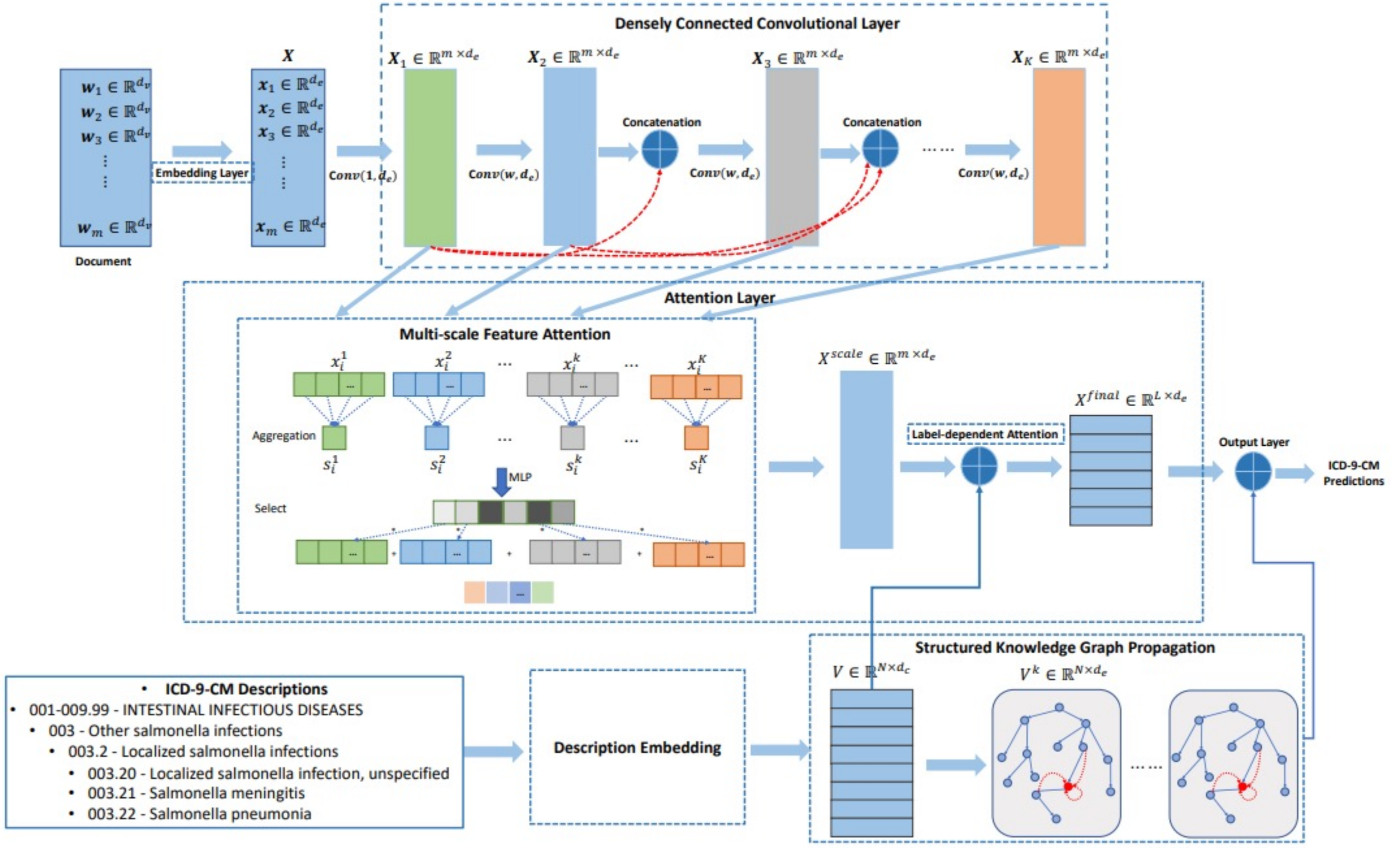


Figure 3: An overall pipeline of our proposed model.

❖ Xie et al., *EHR Coding with Multi-scale Feature Attention and Structured Knowledge Graph Propagation*, CIKM’19.

Models

- C-MemNN [Prakash et al., AAAI'17]
- CAML [Mullenbach et al., NAACL'18]
- MultiResCNN [Li et al., AAAI'20]
- MSATT-KG [Xie et al., CIKM'19]
- HyperCore [Cao et al., ACL'20]
- Fusion [Luo et al., Findings of ACL'21]

HyperCore

- Motivation:
  - Most of existing methods independently predict each code, ignoring two important characteristics: Code Hierarchy and Code Co-occurrence.
- Solution:
  - Hyperbolic and Co-graph Representation
    - Code Hierarchy: ICD codes are organized under a tree-like hierarchical structure.
    - Code Co-occurrence: To capture the correlations of codes.
    - A hyperbolic representation learning method to learn the Code Hierarchy Relation.

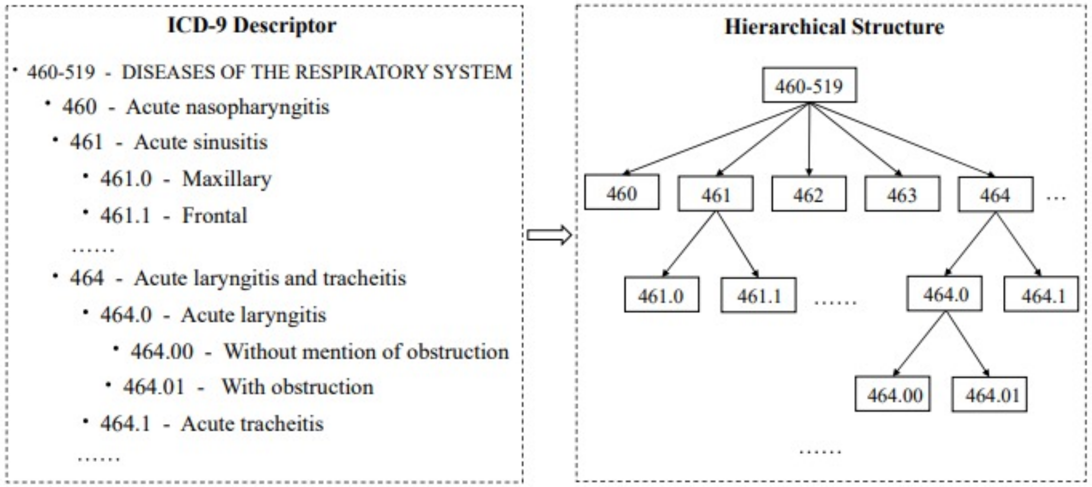
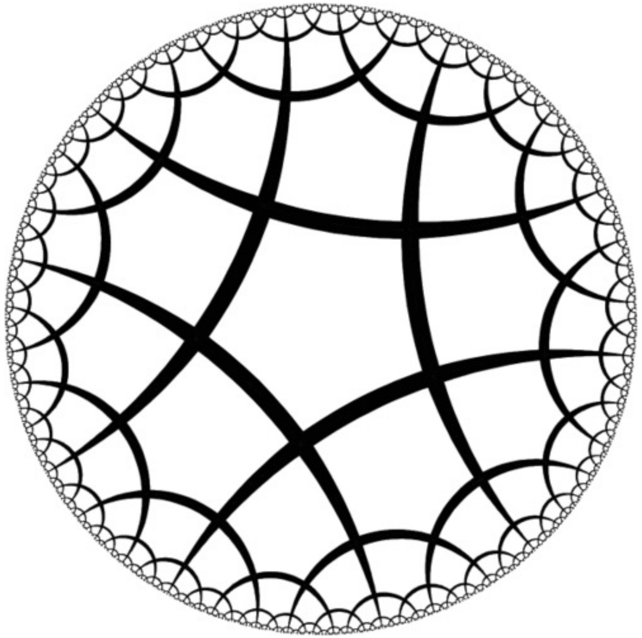


Figure 2: An example of ICD-9 descriptors and the derived hierarchical structure.

❖ Cao et al., *HyperCore: Hyperbolic and Co-graph Representation for Automatic ICD Coding*, ACL'20.

HyperCore

- Hyperbolic Space:
  - The density is less at the edge of the space.



$$g_{\boldsymbol{x}} = \left(\frac{2}{1-||\boldsymbol{x}||^2}\right)^2 g^E \quad (5)$$

where $\boldsymbol{x} \in \mathcal{B}^n$. g^E denotes the Euclidean metric tensor. Furthermore, the distance between two points $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{B}^n$ is given as:

$$d(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}(1 + 2\frac{||\boldsymbol{u}-\boldsymbol{v}||^2}{(1-||\boldsymbol{u}||^2)(1-||\boldsymbol{v}||^2)}) \quad (6)$$

❖ Cao et al., *HyperCore: Hyperbolic and Co-graph Representation for Automatic ICD Coding*, ACL'20.

HyperCore

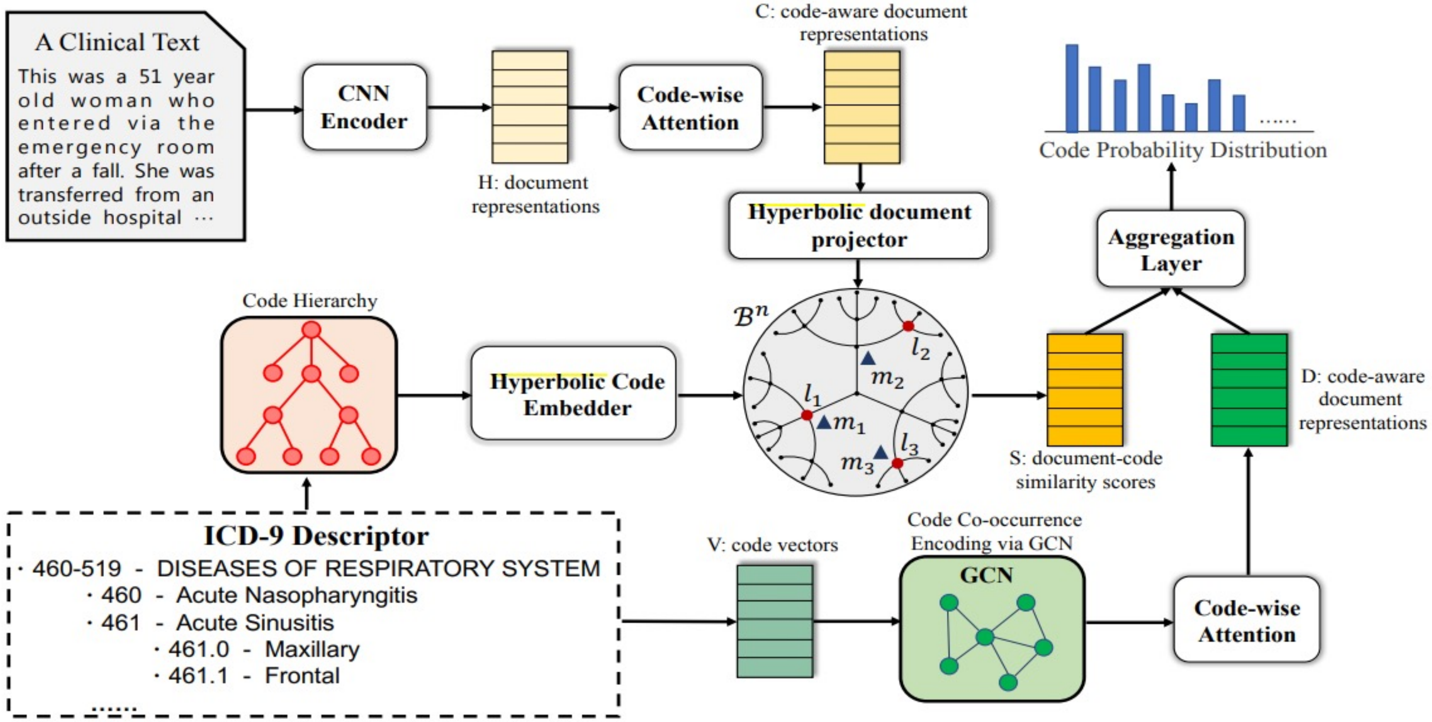


Figure 3: The architecture of **Hyper**bolic and **Co**-graph **Re**presentation method (HyperCore). In the Poincaré ball $\mathcal{B}^n$, we show the embeded code hierarchy (i.e., tree-like hierarchical structure). The dots l_i $(i = 1, 2, 3)$ on the tree-like hierarchical structure and triangles m_i $(i = 1, 2, 3)$ in the Poincaré ball denote hyperbolic code embeddings and hyperbolic document representations, respectively.

❖ Cao et al., *HyperCore: Hyperbolic and Co-graph Representation for Automatic ICD Coding*, ACL'20.

Models

- C-MemNN [Prakash et al., AAAI'17]
- CAML [Mullenbach et al., NAACL'18]
- MultiResCNN [Li et al., AAAI'20]
- MSATT-KG [Xie et al., CIKM'19]
- HyperCore [Cao et al., ACL'20]
- Fusion [Luo et al., Findings of ACL'21]

Fusion

- Motivation:
  - The clinical notes are noisy and complex, where only some key phrases are highly related to the coding.
  - Most existing only use the local features for coding obtained using different filters. The inner-relations between different local features are not considered.
- Solution:
  - A feature compressed ICD coding model: Fusion
    - Attention-based Soft-pooling is used to remove redundant information and keep the key information.
    - A Feature Aggregation Layer is used to model the inner-reactions between different local features.

Table 1: Examples of clinical text fragments and their corresponding ICD codes.

998.32: Disruption of external operation wound ... wound infection, and **wound breakdown** ...
428.0: Congestive heart failure ... DIAGNOSES: 1. **Acute congestive heart failure** 2. Diabetes mellitus 3. Pulmonary edema ...
202.8: Other malignant lymphomas ... a 55 year-old female with **non Hodgkin's lymphoma** and acquired C1 esterase inhibitor deficiency ...
770.6: Transitory tachypnea of newborn ... Chest x-ray was consistent with **transient tachypnea of the newborn** ...
424.1: Aortic valve disorders ... mild **aortic stenosis with an aortic valve area** of 1.9 cm squared and 2+ **aortic insufficiency** ...

❖ Luo et al., *Fusion: Towards Automated ICD Coding via Feature Compression* , Findings of ACL'21.

Fusion

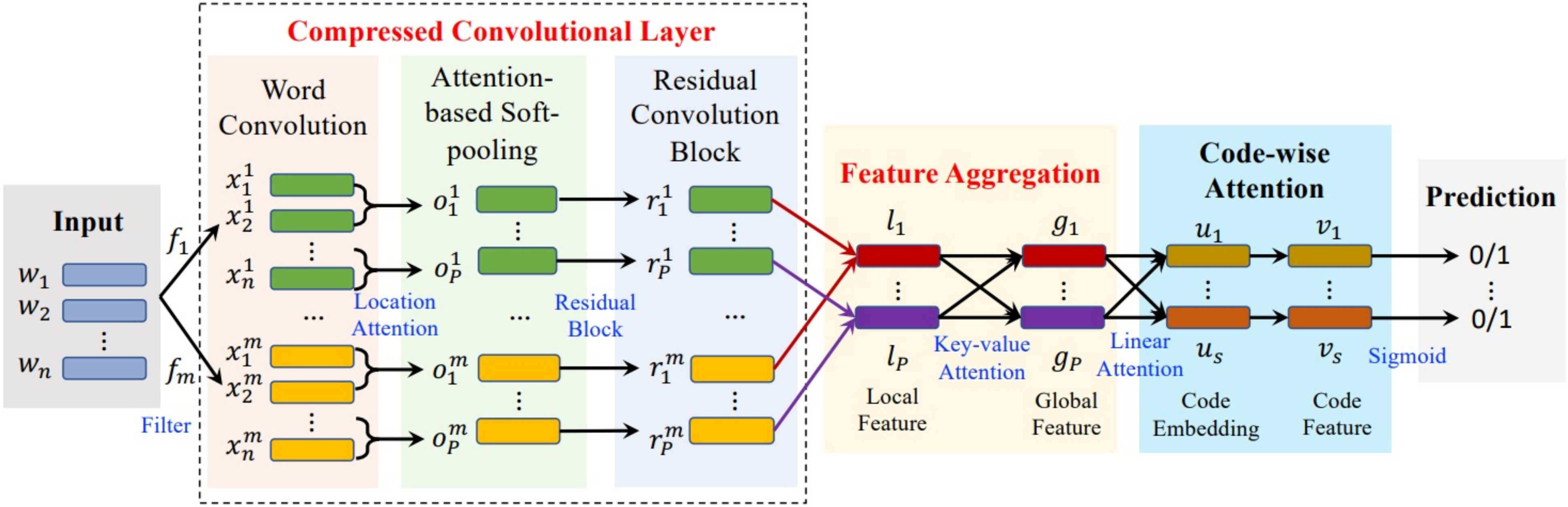


Figure 1: Overview of the proposed Fusion.

- This model consists of five modules: the input layer, the compressed convolutional layer, the feature aggregation layer, the code-wise attention layer, and the prediction layer.

Experimental Results

Dataset		MIMIC-III 50					MIMIC-III Full				
Setting	Model	AUC		F1		P@N	AUC		F1		P@N
		Macro	Micro	Macro	Micro	5	Macro	Micro	Macro	Micro	8
Note Only	Fusion	**0.931**	**0.950**	**0.683**	**0.725**	**0.679**	0.915	0.987	0.083	**0.554**	**0.736**
	C-MemNN	0.833	–	–	–	0.420	–	–	–	–	–
	C-LSTM-ATT	–	0.900	–	0.532	–	–	–	–	–	–
	CAML	0.875	0.909	0.532	0.614	0.609	0.895	0.986	0.088	0.539	0.709
	DR-CAML	0.884	0.916	0.576	0.633	0.618	0.897	0.985	0.086	0.529	0.690
	MultiResCNN	0.899	0.928	0.606	0.670	0.641	0.910	0.986	0.085	0.552	0.734
Note + Ontology	HyperCore	0.895	0.929	0.609	0.663	0.632	**0.930**	0.989	**0.090**	0.551	0.722
	MSATT-KG	0.914	0.936	0.638	0.684	0.644	0.910	**0.992**	**0.090**	0.553	0.728

Table 1: Experiment results on MIMIC-III 50 and MIMIC-III Full datasets.

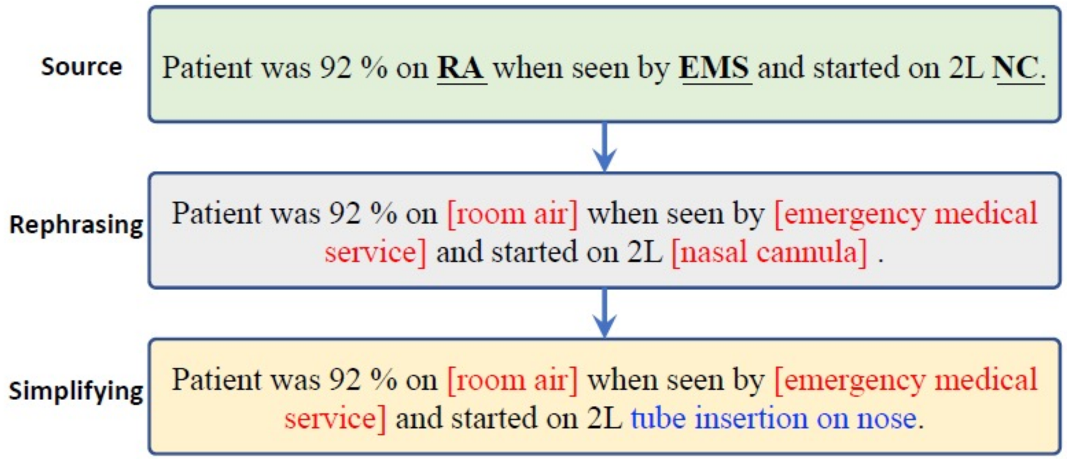
❖ Luo et al., *Fusion: Towards Automated ICD Coding via Feature Compression* , Findings of ACL'21.

Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

Task

- Background:
  - Medical notes are hard to understand for the ordinary users due to the medical jargons and abbreviations.
- Target:
  - Automatically translate the professional medical notes into layman style.



Unsupervised Clinical Language Translation

- Motivation:
  - Professional, clinical jargon makes it hard for patients to access their medical records.
  - Existing methods are limited by expert curation, like the dictionary.
- Solution:
  - The two-step unsupervised translation method
    - A word translation system that translates professional words into consumer-understandable words.
    - Language models and back-translation to consider the contextual lexical and syntactic information for better quality of translation.

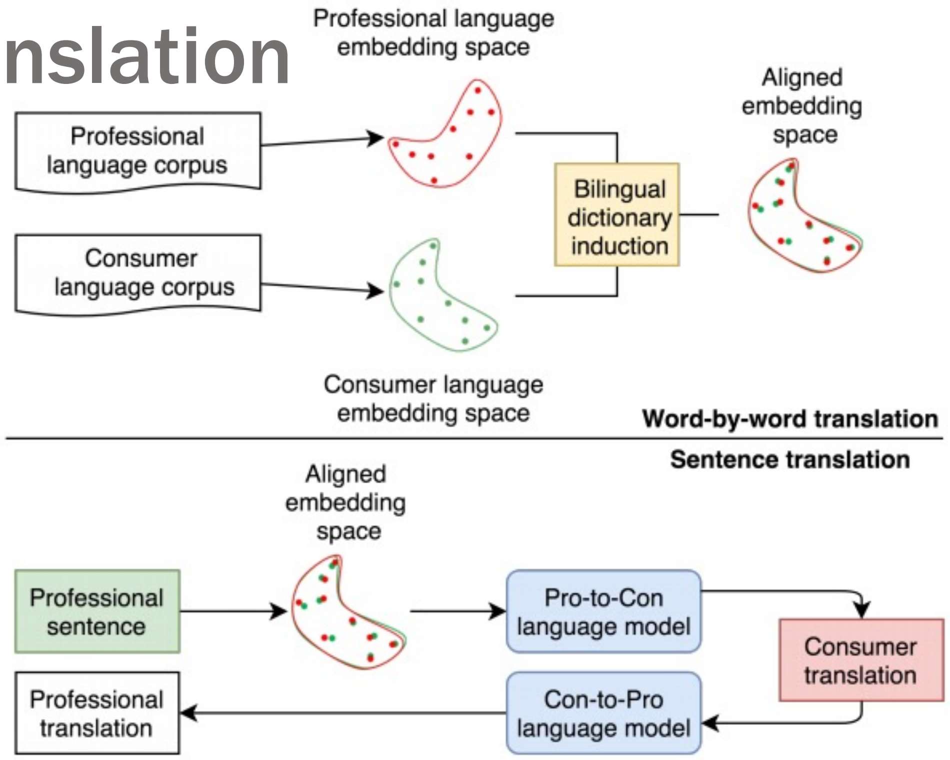


Figure 1: Overview of our framework. The framework is composed of two steps: (1) word translation through unsupervised word representation learning and bilingual dictionary induction (BDI), and (2) sentence translation, which is initialized by the BDI-aligned word embedding spaces and refined by a statistical language model and back-translation.

MedLane

- Motivation:
  - The simplification of the medical text is popular area but lacks of proper benchmark and data.
- Solution:
  - A new dataset named MedLane to support the development and evaluation of automated clinical language understanding approaches.
  - A new model called Declare that follows the human annotation procedure as the new SOTA baseline.
  - New evaluation metric named AScore.

❖Luo et al., *Benchmarking Automated Clinical Language Understanding*, EMNLP'21 (under review).

MedLane

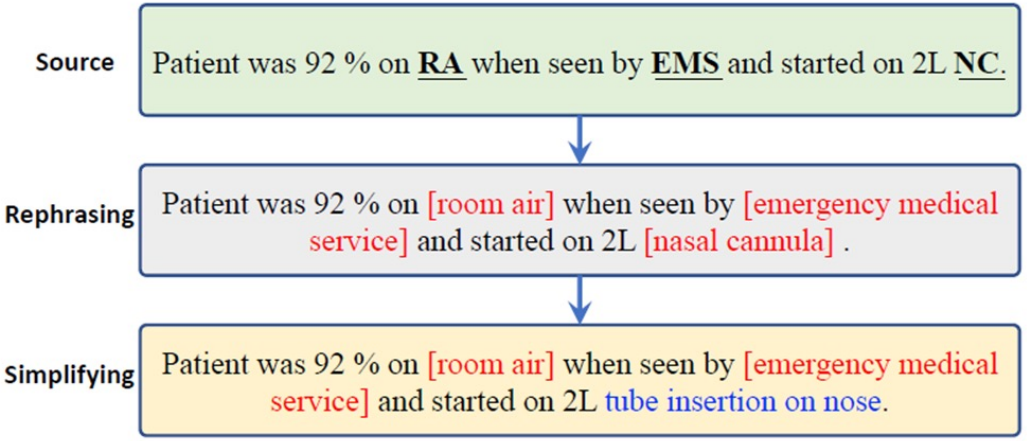


Figure 1: An example of annotating a source sentence by a work using two steps, i.e., rephrasing and simplifying. In the rephrasing step, three abbreviations are replaced by full forms. In the simplifying step, the full form "nasal cannula" is replaced by "tube insertion on nose".

# of tokens in the source sentences	14,780
# of tokens in the target sentences	14,278
# of overlapped tokens between source & target	12,501
Avg. length of the source sentences	20.6
Avg. length of the target sentences	24.0
Avg. # of abbreviations in validation & testing sets	1.2

Table 1: `MedLane` data statistics.

❖Luo et al., *Benchmarking Automated Clinical Language Understanding*, EMNLP'21 (under review).

Declare

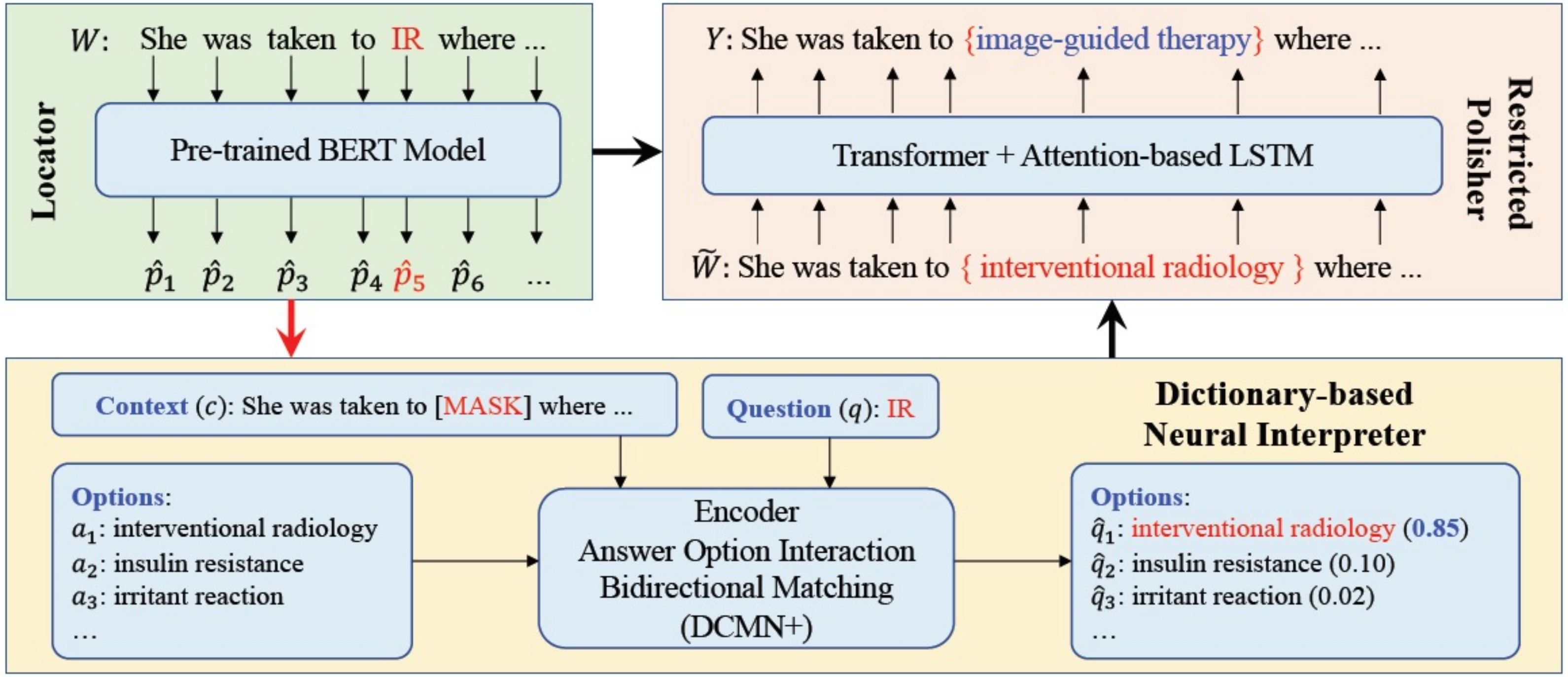


Figure 2: Overview of the proposed Declare model.

Given a tokenized professional medical sentence W = [w_1,w_2, ... , w_n], where n denotes the number of tokens, the locator aims to dig out possible phrases that need to be simplified or translated. In the neural interpreter, the chosen phrases will be replaced with full-term expressions selected from the medical dictionary. Finally, the replaced sentence will pass the polisher to generate the final output Y. These three parts tightly work together and enhance each other.

Experiment

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	BLEU	METEOR	ROUGE-L	CIDEr	HIT	CWR	AScore
Dictionary	0.7158	0.6364	0.5684	0.5076	0.6070	0.3933	0.7308	4.2037	0.5572	0.6407	0.5948
Moses	0.7880	0.7130	0.6530	0.6016	0.6889	0.4237	0.8188	5.1046	0.6823	0.7543	0.6859
Seq2seq	0.7136	0.6322	0.5969	0.5160	0.6147	0.3533	0.7609	4.1299	0.7388	0.7980	0.6648
Seq2seq-	0.5066	0.3315	0.2373	0.1787	0.3135	0.1859	0.4948	1.2670	0.6427	**0.8367**	0.4070
Seq2seq-S	0.7180	0.6386	0.5778	0.5267	0.6153	0.3604	0.7683	4.2635	0.7331	0.7953	0.6630
PointerNet	0.6870	0.5904	0.5158	0.4541	0.5618	0.3338	0.7285	3.9458	0.6414	0.7555	0.5949
BERT-MT	0.8003	0.7428	0.6952	0.6531	0.7228	0.4566	0.8218	5.3293	0.7808	0.7358	0.7417
Declare	**0.8624**	**0.8291**	**0.8004**	**0.7737**	**0.8165**	**0.5290**	**0.8894**	**6.7212**	**0.7986**	0.7328	**0.7983**
↑	+7.8%	+11.6%	+15.7%	+18.5%	+12.9%	+15.9%	+8.2%	+26.1%	+2.2%	-12.4%	+7.6%

Table 2: Performance evaluation of all the baselines with different metrics. ↑ denotes the percentage of performance gain compared with the best baselines.

❖Luo et al., *Benchmarking Automated Clinical Language Understanding*, EMNLP'21 (under review).

Experiment

Source:	**NSTEMI/CAD** - history of **3V-CABG** with only **RCA** graft still patent .
Reference 1:	[non-ST-elevation myocardial infarction]/[coronary artery disease] - history of [coronary artery bypass graft] with only [right coronary artery] graft still patent .
Reference 2:	heart attack/heart disease - history of heart bypass surgery with only right heart artery graft still patent .
Declare	heart attack attack/heart disease-history of coronary artery bypass graft with only right heart artery graft still patent .
BERT-MT	heart attack/heart disease - history of 3v - heart bypass surgery with only right right heart artery graft still patent .
Seq2Seq	- history of with only right heart artery graft .
Seq2Seq–	- - history of with history only - when are only .
PointerNet	- history of right heart disease graft with two-vessel coronary artery still patent .
Moses	nstemi/cad - history of 3v-cabg with only still patent artery graft .

Table 3: An example that Declare outperforms other baselines.

Source:	# **cirrhosis** : patient with history of alcoholic vs **nash cirrhosis** complicated by esophagel , gastric , and rectal varices
Reference 1:	# [chronic disease of the liver] : patient with history of alcoholic vs [non-alcoholic steatohepatitis] [chronic disease of the liver] complicated by esophagel , gastric , and rectal varices .
Reference 2:	# [chronic disease of the liver] : patient with history of alcoholic vs liver inflammation and damage complicated by esophagel , gastric , and rectal varices .
Declare	cirrhosis : patient with history of alcoholic vs [non-alcoholic] cirrhosis cirrhosis cirrhosis complicated by esophagel , gastric , and rectal varices .
BERT-MT	cirrhosis cirrhosis : patient with history of alcoholic vs late disease disease complicated by food , , , , , , , esesesesesesesesesesesesesesesses .
Seq2Seq	# cirrhosis : patient with history of alcoholic vs cirrhosis cirrhosis , gastric , gastric , and , , , varices . 22
Seq2Seq–	# surgical history of patient with history by , history , patient , by surgical by and by surgical tract .
PointerNet	# cirrhosis : patient with history of painful cell function cirrhosis complicated by , , , , , , and rectal rectal in rectal varices .
Moses	# cirrhosis : patient with history of alcoholic cirrhosis , complicated by nash esophagel , acid , and rectal and .

Table 4: A hard example that all the approaches cannot translate accurately.

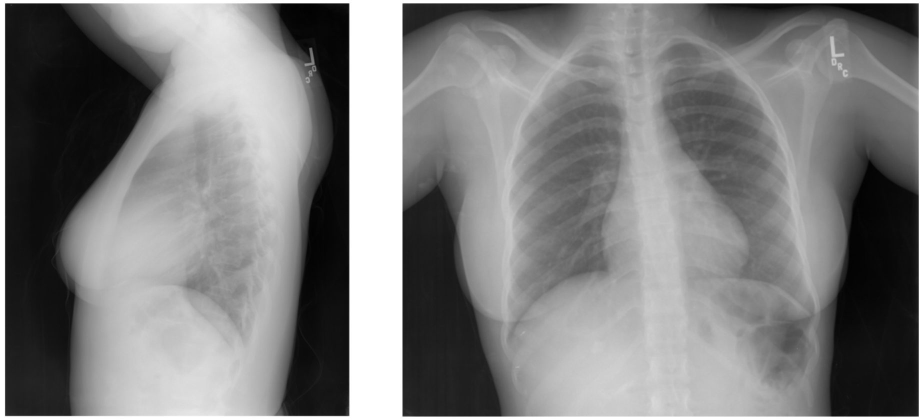
❖Luo et al., *Benchmarking Automated Clinical Language Understanding*, EMNLP'21 (under review).

Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

Task Description

- **Medical Report Generation**: Computer generates medical description that contains the clinical findings and treatment suggestions given medical images.



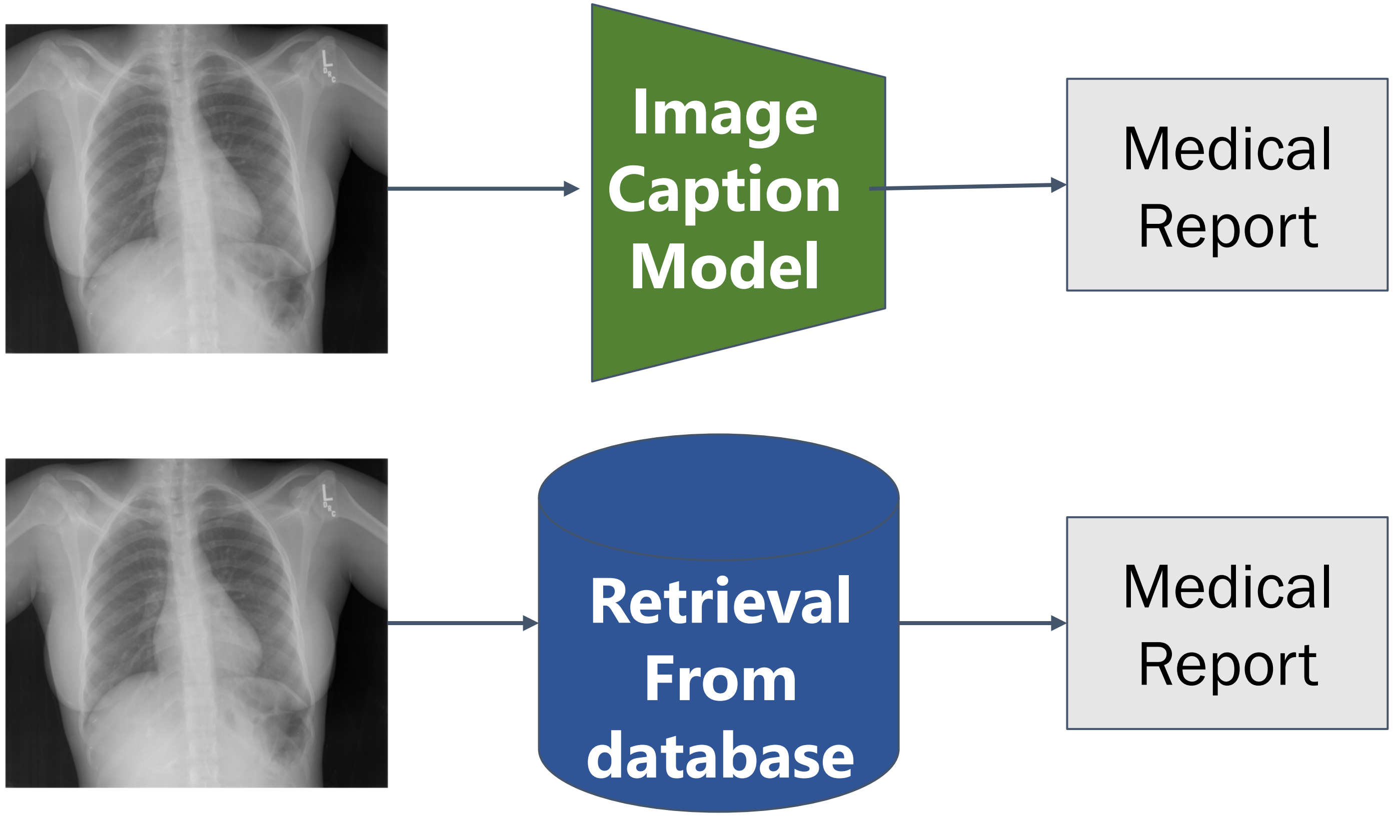
- Highly standardized and structured text
- Reflecting clinical findings (Importance)

FINDINGS: The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of a pleural effusion. There is no evidence of pneumothorax.
IMPRESSION: Normal chest x-XXXX.

Generation and Retrieval

Generation-Based

Retrieval-Based



Models

- Generation:
  - TieNet [Wang et al., CVPR'18]
  - CoAtt [Jing et al., ACL'18]
  - MvH [Yuan et al., MICCAI'19]
  - SentSAT + KG [Zhang et al., AAAI'20]
- Retrieval
  - HRGR-Agent [Li et al., NeurIPS'18]
  - KERP [Li et al., AAAI'19]
  - MedWriter [Yang et al., ACL'21]

TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays

- Main Contributions:
  - TieNet, A CNN-RNN **text-image embedding network**
  - Boost the disease classification with generated text
  - Design **multi-level attention** for embedding extraction (Image spatial attention and text attention)

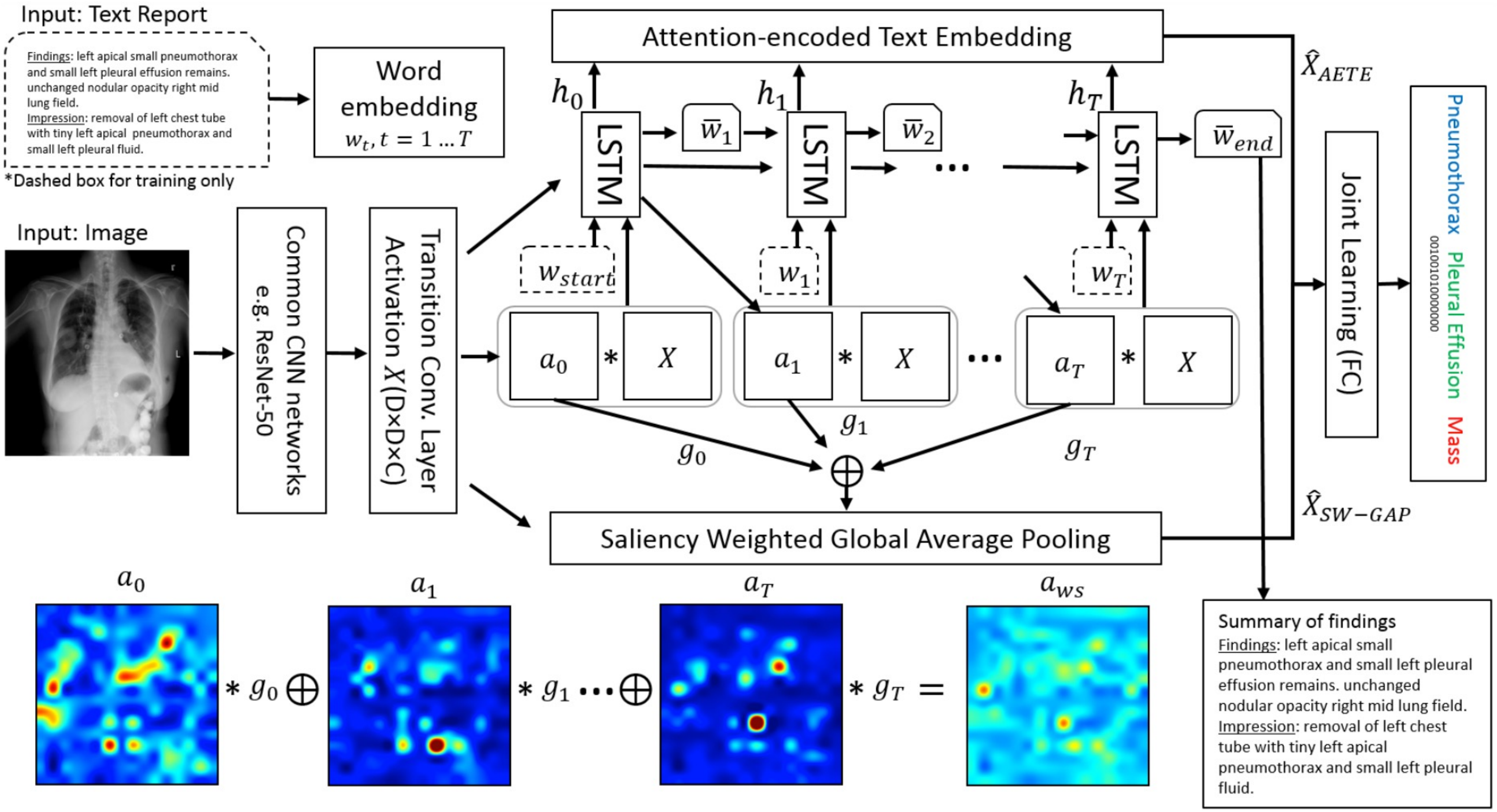


Figure 2. Framework of the proposed chest X-ray auto-annotation and reporting framework. Multi-level attentions are introduced to produce saliency-encoded text and image embeddings.

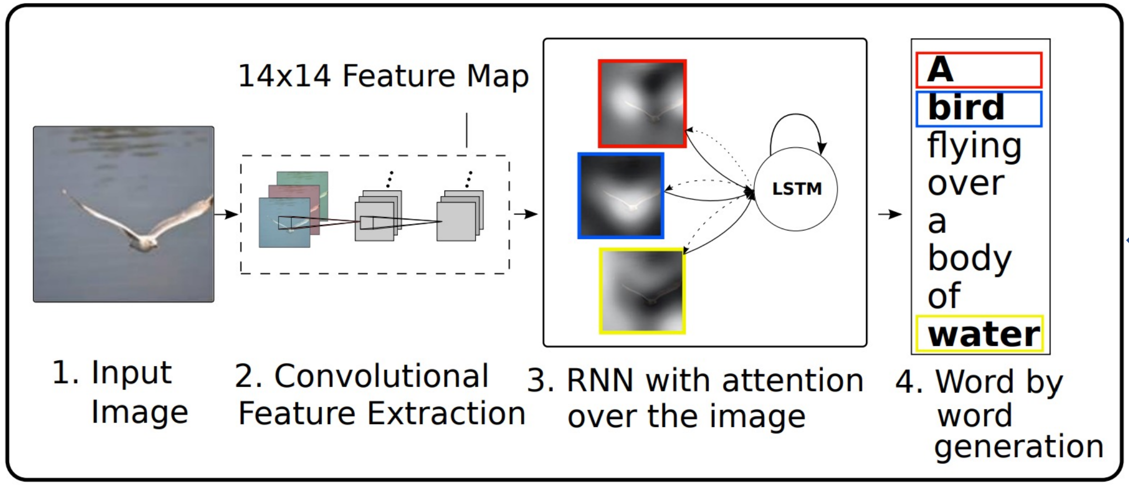
❖Wang et al., *TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays*, CVPR 2018.

TieNet

Attention: $\mathbf{G} = softmax(\mathbf{W}_{s2}\, tanh(\mathbf{W}_{s1}\, \mathbf{H}))$

$\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_T) \in \mathrm{R}^{d_h \times T}$

$\mathbf{M} = \mathbf{GH}$



Xu et al., *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention*, ICML 2015.

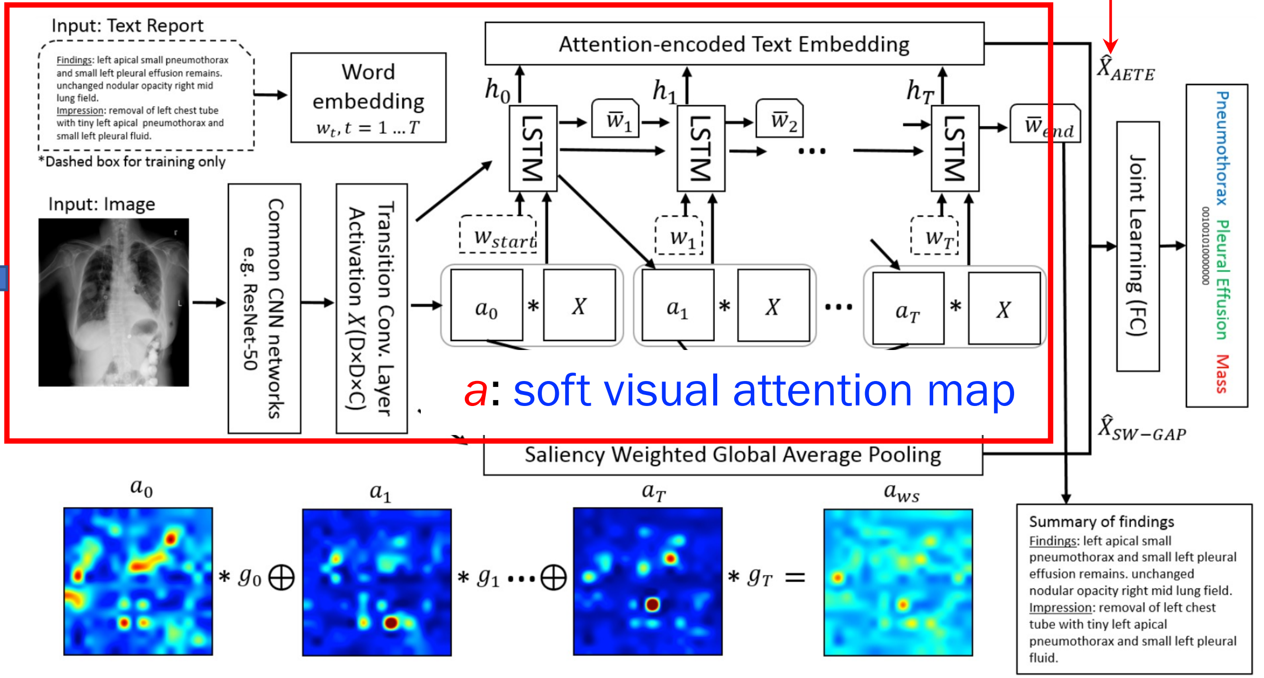


Figure 2. Framework of the proposed chest X-ray auto-annotation and reporting framework. Multi-level attentions are introduced to produce saliency-encoded text and image embeddings.

TieNet

Attention: $\mathbf{G} = softmax(\mathbf{W}_{s2}\, tanh(\mathbf{W}_{s1}\, \mathbf{H}))$

$\mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_T) \in \mathrm{R}^{d_h \times T}$

$\mathbf{M} = \mathbf{GH}$

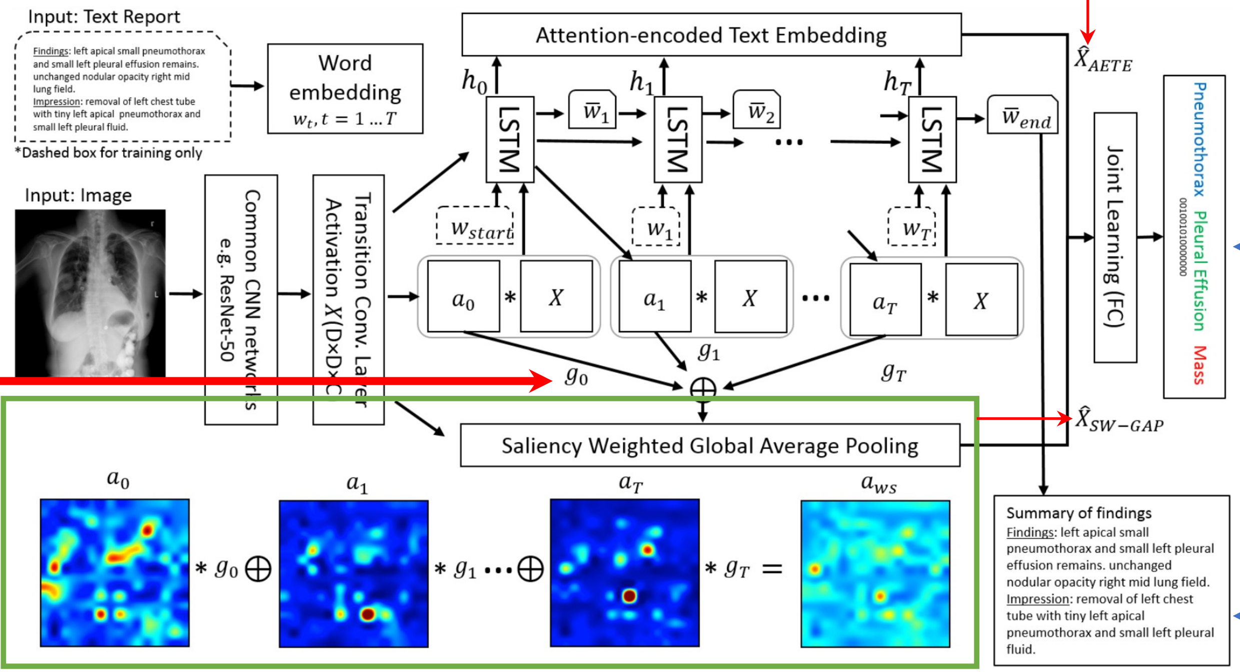


Figure 2. Framework of the proposed chest X-ray auto-annotation and reporting framework. Multi-level attentions are introduced to produce saliency-encoded text and image embeddings.

$L_{overall} = \alpha L_C + (1 - \alpha) L_R$

Limitation:
Medical reports contain several sentences, and one LST may not work well.

❖Wang et al., *TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays*, CVPR 2018.

On the Automatic Generation of Medical Imaging Reports

- Main Contributions:
  - A multi-task learning framework which can simultaneously predict the tags and text descriptions.
  - A co-attention mechanism for localizing sub-regions related to different diseases.
  - We build a hierarchical LSTM to generate long paragraphs.

On the Automatic Generation of Medical Imaging Reports

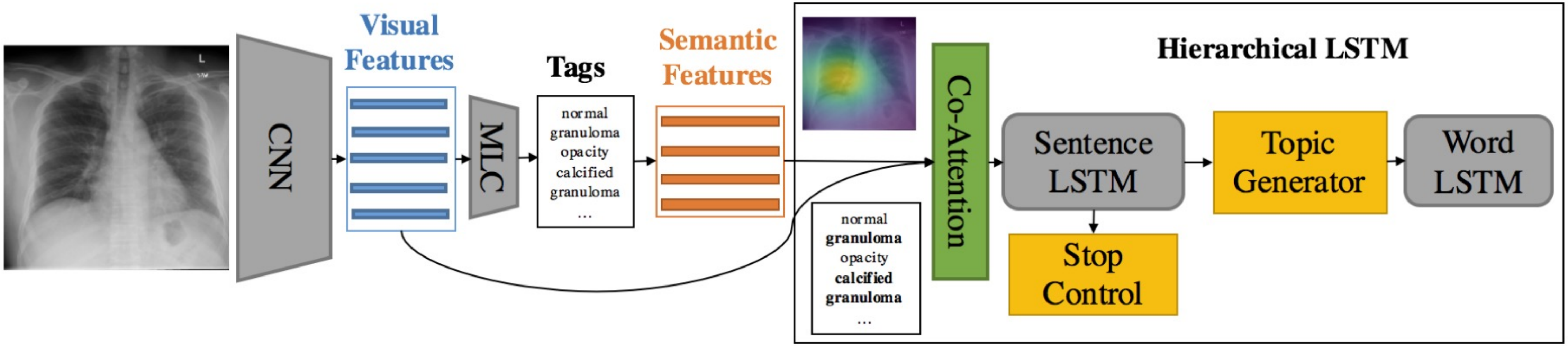


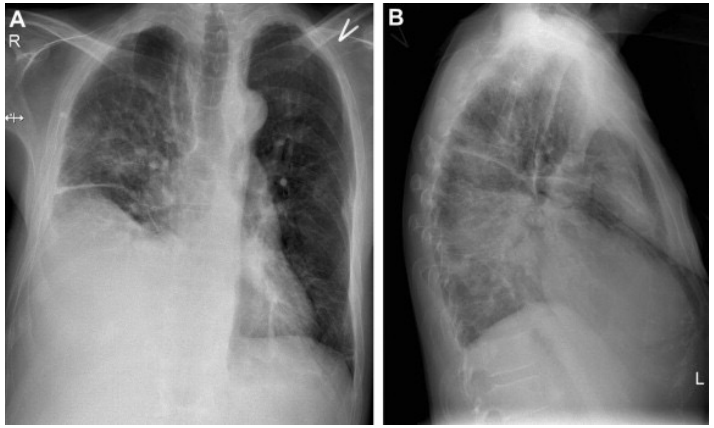
Figure 2: Illustration of the proposed model. MLC denotes a *multi-label classification* network. Semantic features are the word embeddings of the predicted tags. The boldfaced tags "calcified granuloma" and "granuloma" are attended by the co-attention network.

❖Jing et al., *On the Automatic Generation of Medical Imaging Reports*, ACL 2018.

Automatic Radiology Report Generation based on Multi-view Image Fusion and Medical Concept Enrichment

- **Motivation:**

frontal



lateral

- **Main Contributions:**
  - Large scale **CNN encoder pretraining** with chest x-ray images
  - **Multi-view visual feature consistency** with sentence-level attentions
  - Apply **medical concepts** to the decoder with word-level attentions

❖Yuan et al., *Automatic Radiology Report Generation based on Multi-view Image Fusion and Medical Concept Enrichment*, MICCAI 2019.

Automatic Radiology Report Generation based on Multi-view Image Fusion and Medical Concept Enrichment

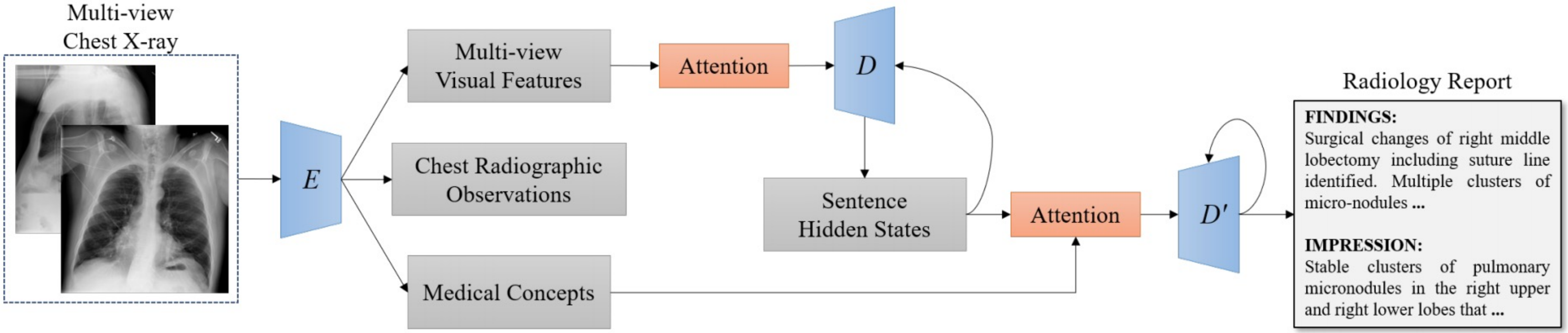
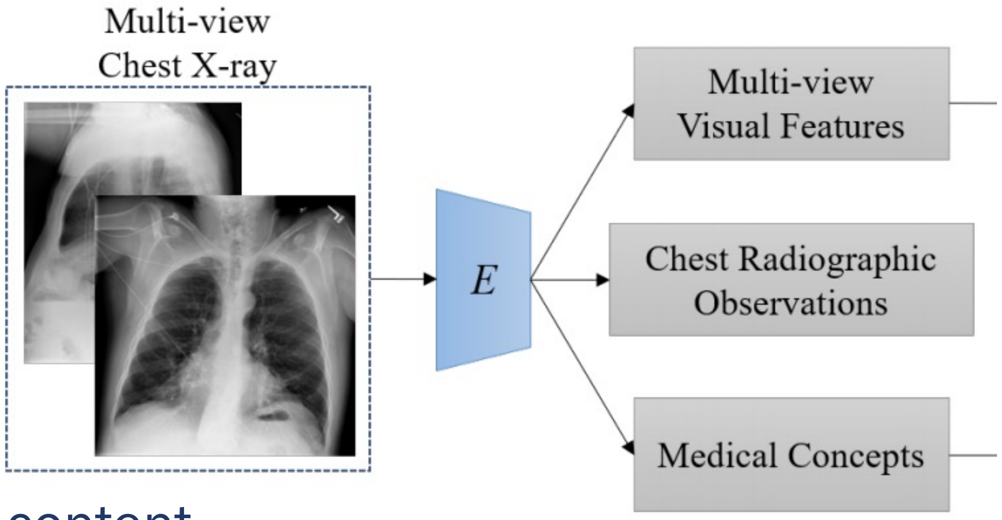


Fig. 1. Overall framework of the proposed encoder and decoder with attentions. E, D, and D' denote the encoder, sentence decoder, and word decoder, respectively.

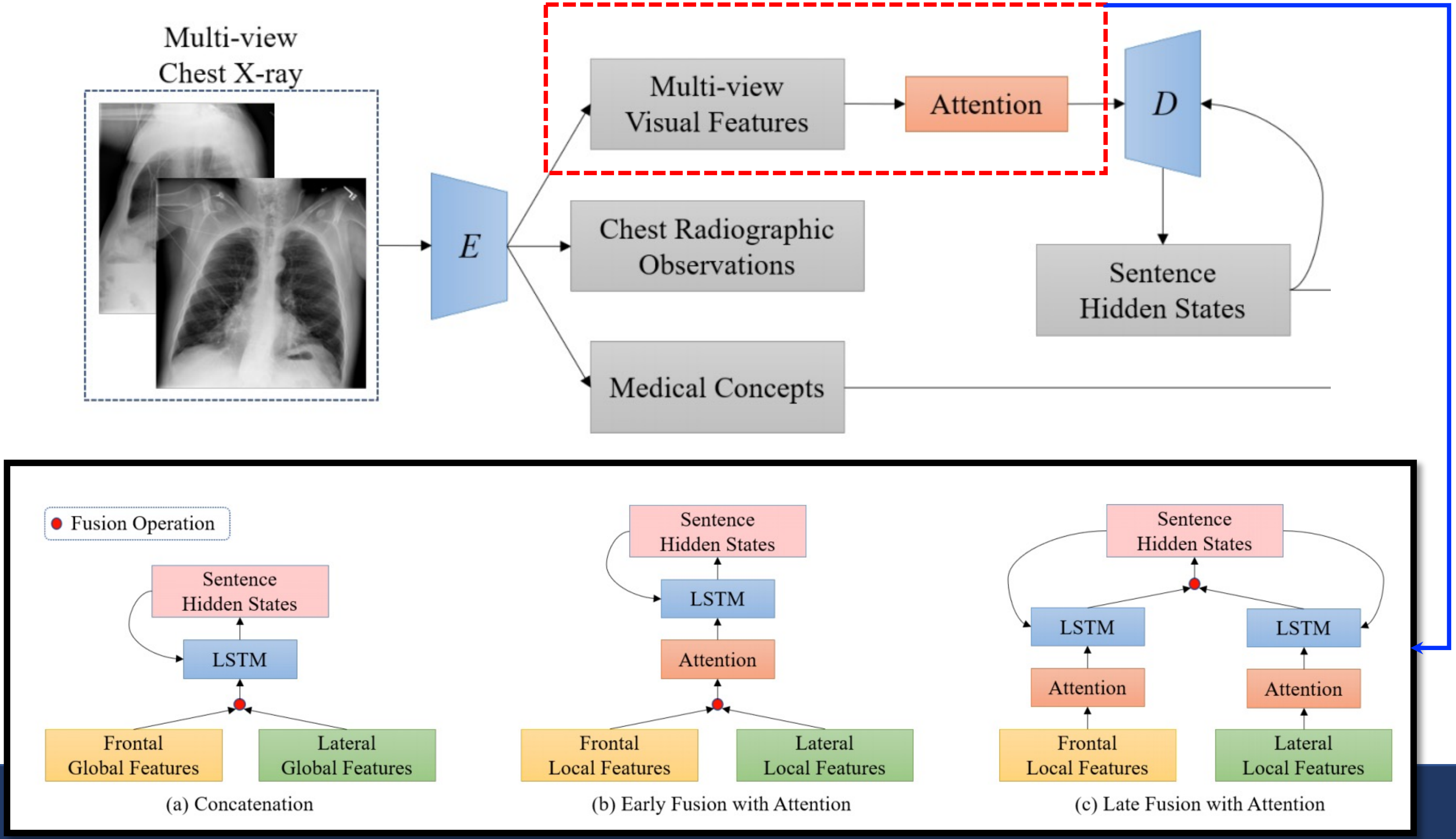
❖Yuan et al., *Automatic Radiology Report Generation based on Multi-view Image Fusion and Medical Concept Enrichment*, MICCAI 2019.

Automatic Radiology Report Generation based on Multi-view Image Fusion and Medical Concept Enrichment

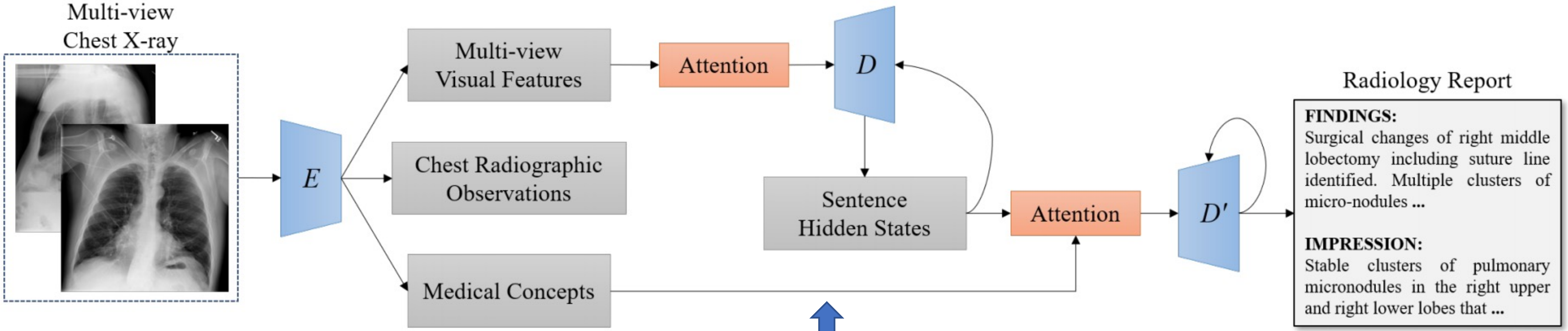
- Image Encoder
  - Resnet-152
- Chest Radiographic Observations
  - Multi-label classification
- Medical Concepts
  - Descriptive information related to the visual content
  - Medical text indexer (MTI) in Open-I
  - Multi-label classification



Sentence Decoder with Attentions



Word Decoder with Attentions



$$\mathbf{a}_i^c = \mathbf{W}_{a^c}\left[\tanh\left(\hat{y_i^c}\mathbf{W}_c\mathbf{c} + \mathbf{W}_w\hat{\mathbf{h}}_{t_w-1}\right)\right], \quad \alpha_i^c = softmax(\mathbf{a}_i^c)$$

medical concept embeddings

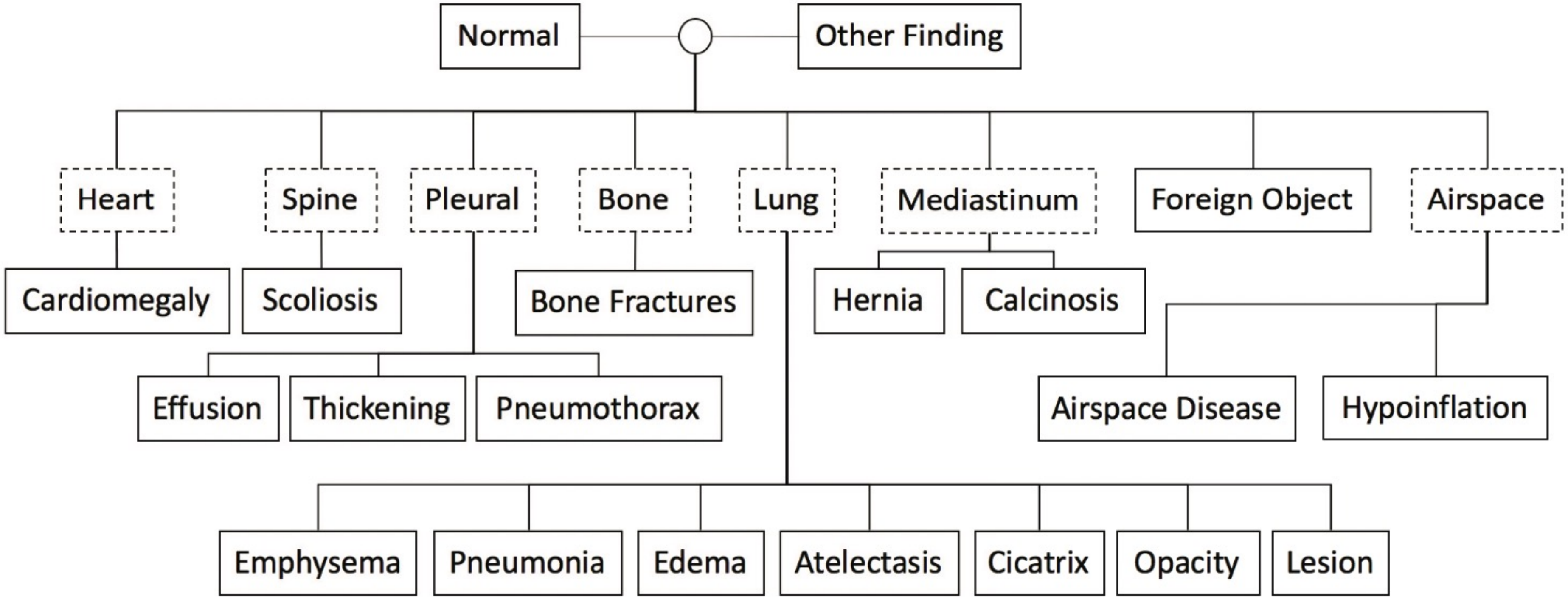
word hidden state

When Radiology Report Generation Meets Knowledge Graph

Main Contributions:

- Utilize a **pre-constructed graph neural network** on multiple disease findings to assist the generation of reports
- **New evaluation metric** for radiology image reporting with the assistance of the same composed graph

Graph Construction with Prior Knowledge



The solid boxes are classes which have corresponding nodes in graph. The dotted boxes are organs or tissues and are not part of target classes. Classes linked to the same organ or tissue are connected to each other in the graph.

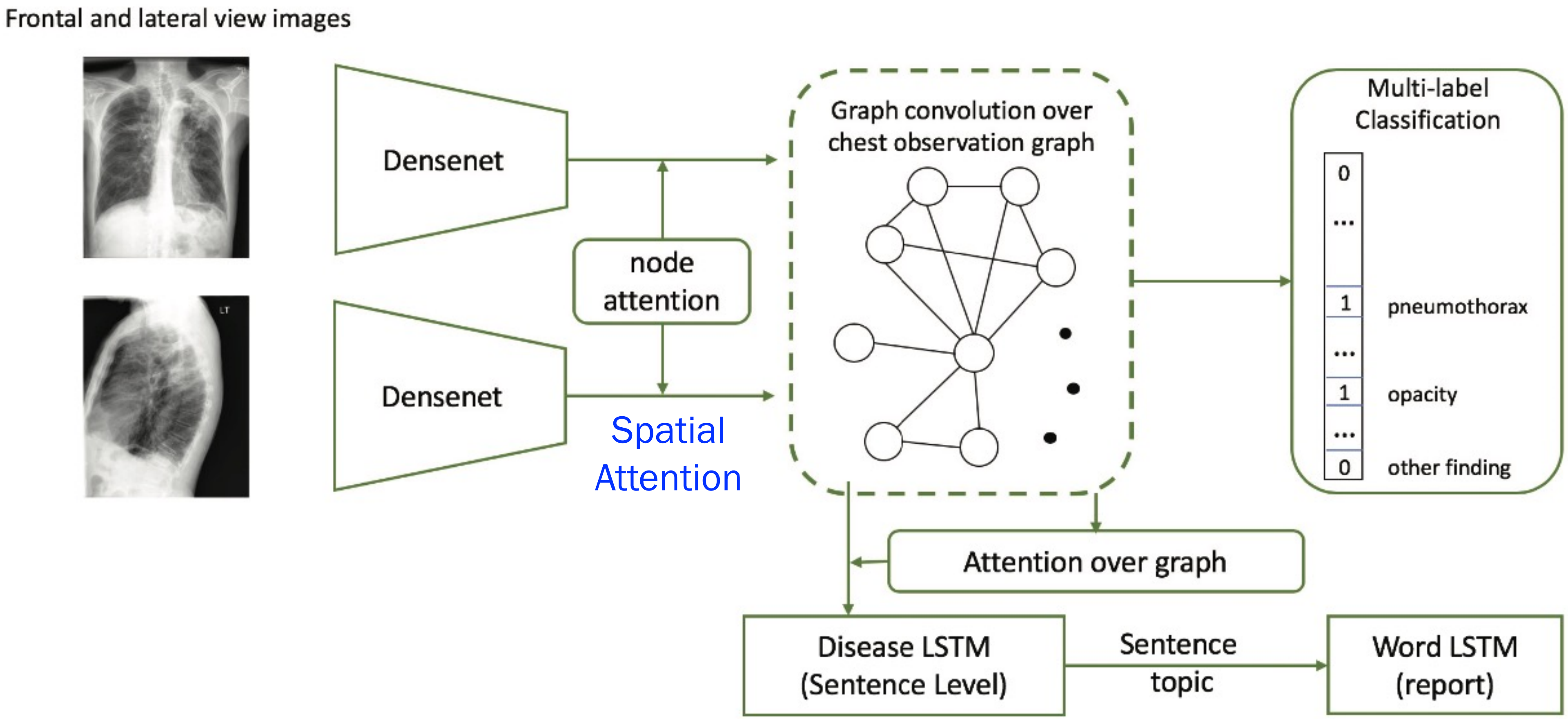


Figure 2: Overview of the proposed framework. Graph node features are extracted from CNN features, followed by graph convolution layers. There are two branches after graph convolution: one for classification and one for report generation.

❖Zhang et al., *When Radiology Report Generation Meets Knowledge Graph*, AAAI 2020.

Models

- Generation:
  - TieNet [Wang et al., CVPR’18]
  - CoAtt [Jing et al., ACL’18]
  - MvH [Yuan et al., MICCAI’19]
  - SentSAT + KG [Zhang et al., AAAI’20]
- Retrieval:
  - HRGR-Agent [Li et al., NeurIPS’18]
  - KERP [Li et al., AAAI’19]
  - MedWriter [Yang et al., ACL’21]

Hybrid Retrieval-Generation Reinforced Agent for Medical Image Report Generation

Main Contributions:

- HRGR-Agent employs a **retrieval policy module**, which chooses to either retrieve a template sentence or generate a new sentence.
- HRGR-Agent is **updated via reinforcement learning**, guided by sentence-level and word-level rewards.

❖Li et al., *Hybrid Retrieval-Generation Reinforced Agent for Medical Image Report Generation*, NeurIPS 2018.

Hybrid Retrieval-Generation Reinforced Agent for Medical Image Report Generation

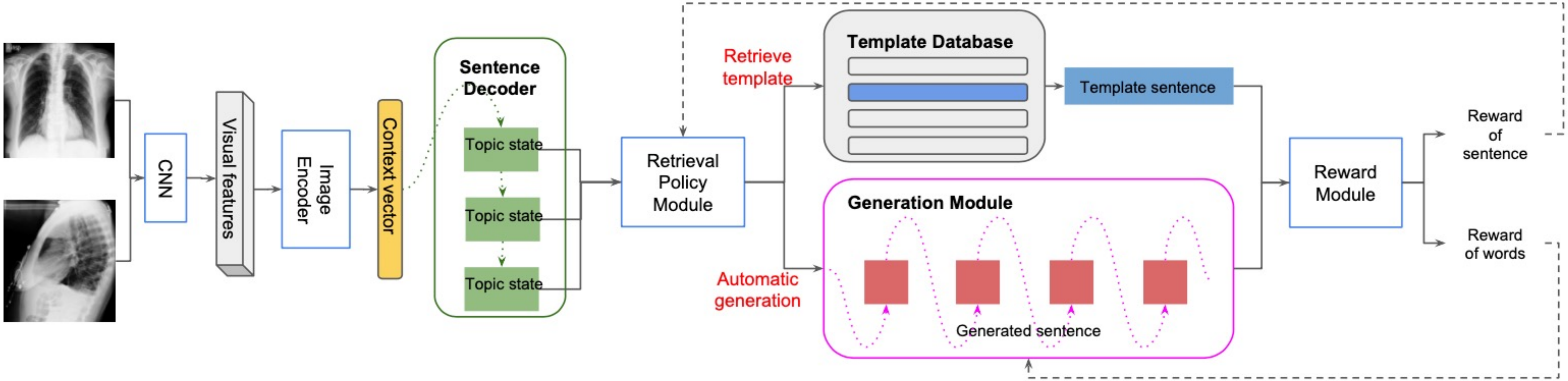


Figure 2: Hybrid Retrieval-Generation Reinforced Agent. Visual features are encoded by a CNN and image encoder, and fed to a sentence decoder to recurrently generate hidden topic states. A retrieval policy module decides for each topic state to either automatic generate a sentence, or retrieve a specific template from a template database. Dashed black lines indicate hierarchical policy learning.

❖Li et al., *Hybrid Retrieval-Generation Reinforced Agent for Medical Image Report Generation*, NeurIPS 2018.

Knowledge-driven encode, retrieve, paraphrase for medical image report generation

Main Contributions:

- KERP = **abnormality graph construction + graph-to-report paraphrase**
- KERP first employs an **Encode module** that transforms visual features into a structured abnormality graph using retrieved text templates.
- KEPR uses a **Paraphrase module** that rewrites the templates according to the extracted graph.

❖Li et al., *Knowledge-driven encode, retrieve, paraphrase for medical image report generation*, AAAI 2019.

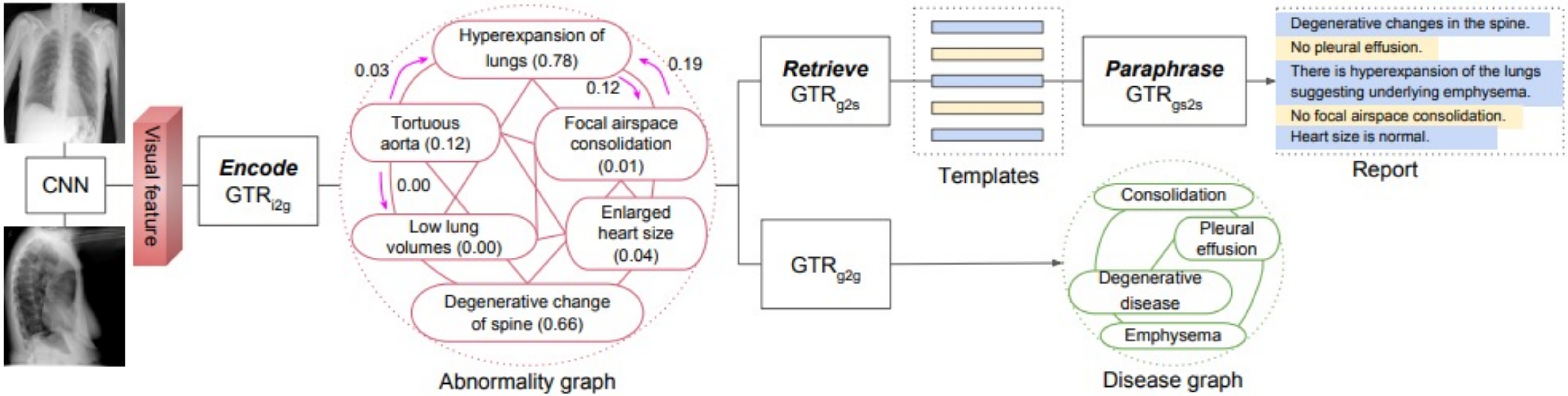
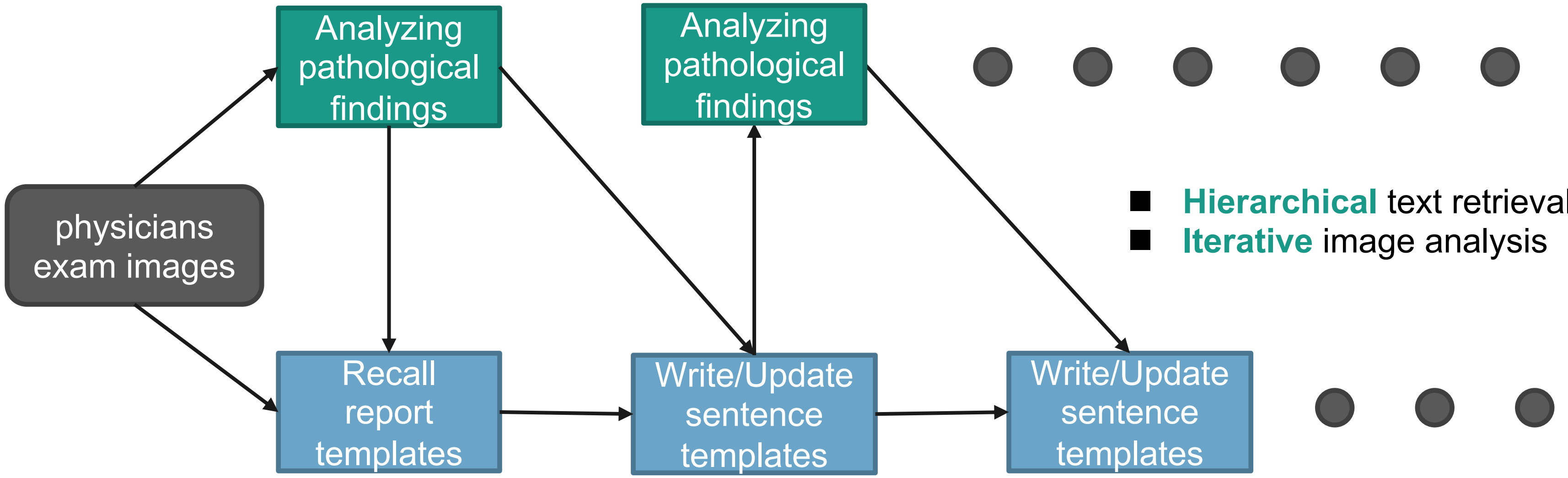


Figure 3: Architecture of KERP. Image features are first extracted from a CNN, and further encoded as an abnormality graph via *Encode* GTR_{i2g}. *Retrieve* GTR_{g2s} decodes the abnormality graph as a template sequence, the words of which are then retrieved and paraphrased by *Paraphrase* GTR_{gs2s} as the generated report. Simultaneously, a GTR_{g2g} decodes the abnormality graph as a disease graph, and predicts disease categories via extra classification layers. In the abnormality graph, values inside parentheses are probabilities of the corresponding nodes predicted by extra classification layers taking latent semantic features of nodes as input. Values along the directed arrows indicate attention scores of source nodes on target nodes.

❖Li et al., *Knowledge-driven encode, retrieve, paraphrase for medical image report generation*, AAAI 2019.

Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation

- How physicians write medical reports in real life?



- **Hierarchical** text retrieval
- **Iterative** image analysis

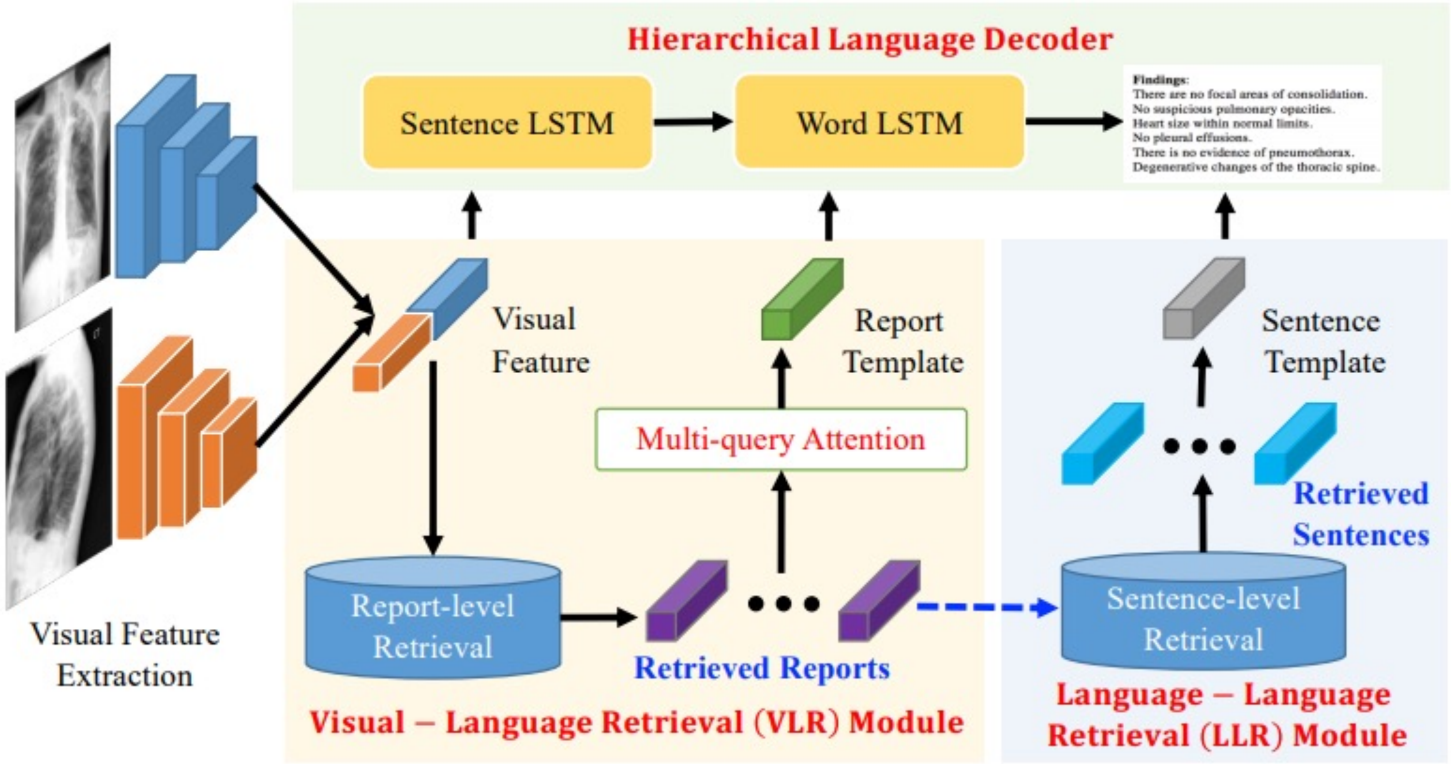
❖Yang et al., *Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation*, ACL 2021.

Contributions

- We propose MedWriter——The first to model the memory retrieval mechanism in both report and sentence levels.
- we design a new multi-query attention mechanism to fuse the retrieved information for medical report generation.
- Experiments on two large-scale medical report generation datasets, i.e., Openi and MIMIC-CXR show that MedWriter achieves better performance compared with state-of-the-art baselines.

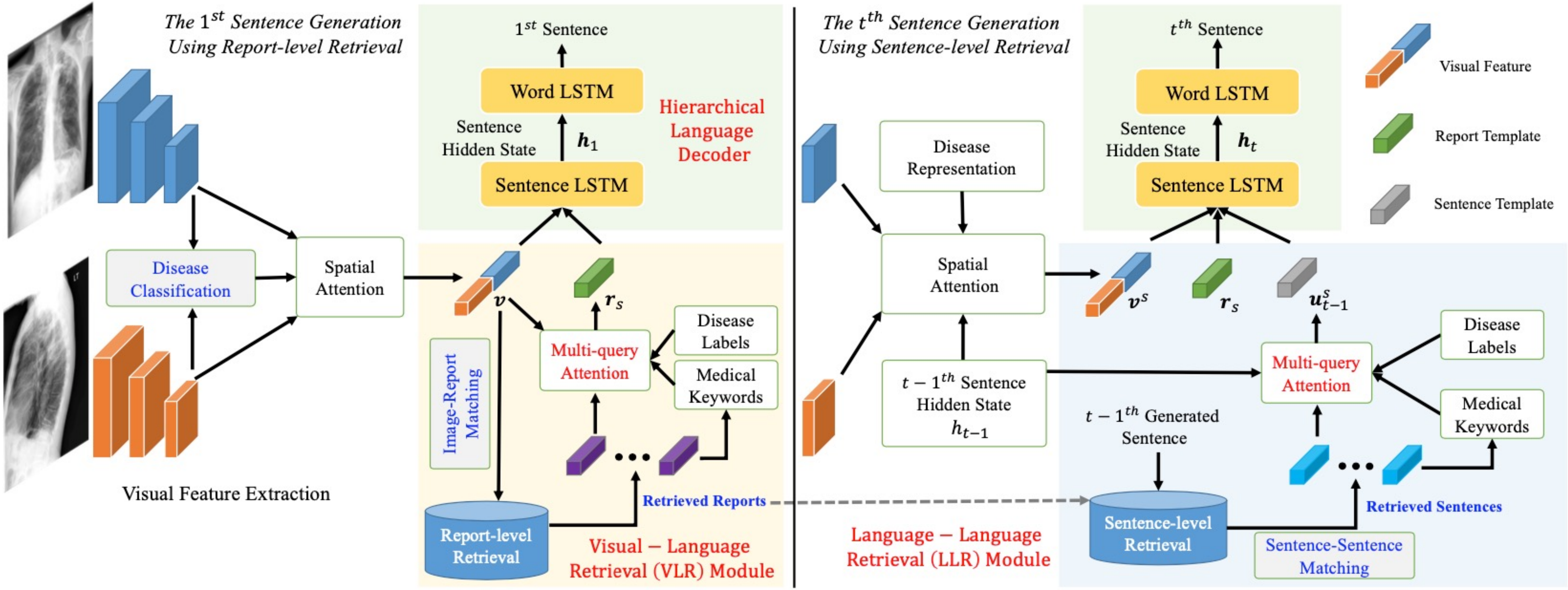
❖Yang et al., *Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation*, ACL 2021.

MedWriter: Overview



❖Yang et al., *Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation*, ACL 2021.

MedWriter



❖Yang et al., *Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation*, ACL 2021.

Experiment setup

- Datasets
  - **Open-I [1]:** 7,470 chest Xrays with 3,955 radiology reports. Sample 2,902 cases and 5,804 images.
  - **MIMIC-CXR [2]:** 377,110 chest X-rays with 227,827 radiology reports. Sample 71,386 reports and 142,772 images.

- Evaluation Metrics
  - **Language evaluation:** CIDEr, ROUGE-L, BLEU 1-4 scores
  - **Clinical evaluation**: ROC-AUC scores achieved by generated reports
  - **Human evaluation**: Two radiologists give ratings for 50 report

[1] https://openi.nlm.nih.gov/faq#collection
[2] https://physionet.org/content/mimic-cxr/2.0.0/

Results

Dataset	Type	Model	CIDEr	ROUGE-L	BLEU-1	BLEU-2	BLEU-3	BLEU-4	AUC
Open-i	Generation	CNN-RNN (Vinyals et al., 2015)	0.294	0.307	0.216	0.124	0.087	0.066	0.426
		LRCN (Donahue et al., 2015)*	0.285	0.307	0.223	0.128	0.089	0.068	–
		Tie-Net (Wang et al., 2018)*	0.279	0.226	0.286	0.160	0.104	0.074	–
		CoAtt (Jing et al., 2018)	0.277	0.369	0.455	0.288	0.205	0.154	0.707
		MvH+AttL (Yuan et al., 2019)	0.229	0.351	0.452	0.311	0.223	0.162	0.725
	Retrieval	V-L Retrieval	0.144	0.319	0.390	0.237	0.154	0.105	0.634
		HRGR-Agent (Li et al., 2018)*	0.343	0.322	0.438	0.298	0.208	0.151	–
		KERP (Li et al., 2019)*	0.280	0.339	**0.482**	0.325	0.226	0.162	–
		MedWriter	**0.345**	**0.382**	0.471	**0.336**	**0.238**	**0.166**	**0.814**
	Ground Truth		–	–	–	–	–	–	0.915
MIMIC-CXR	Generation	CNN-RNN (Vinyals et al., 2015)	0.245	0.314	0.247	0.165	0.124	0.098	0.472
		CoAtt (Jing et al., 2018)	0.234	0.274	0.410	0.267	0.189	0.144	0.745
		MvH+AttL (Yuan et al., 2019)	0.264	0.309	0.424	0.282	0.203	0.153	0.738
	Retrieval	V-L Retrieval	0.186	0.232	0.306	0.179	0.116	0.076	0.579
		MedWriter	**0.306**	0.332	**0.438**	**0.297**	**0.216**	**0.164**	**0.833**
	Ground Truth		–	–	–	–	–	–	0.923

Table 1: Automatic evaluation on the Open-i and MIMIC-CXR datasets. * indicates the results reported in (Li et al., 2019).

❖Yang et al., *Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation*, ACL 2021.

Results

- Human Evaluation
  - Randomly select 50 samples from the Open-i test set
  - Collect ground-truth reports and the generated reports from both MvH+AttL
  - Ratings: 1, 2, 3, 4, and 5 (the higher, the better)

Method	Realistic Score	Relevant Score
Ground Truth	3.85	3.82
MvH+AttL (Yuan et al., 2019)	2.50	2.57
MedWriter	3.68	3.44

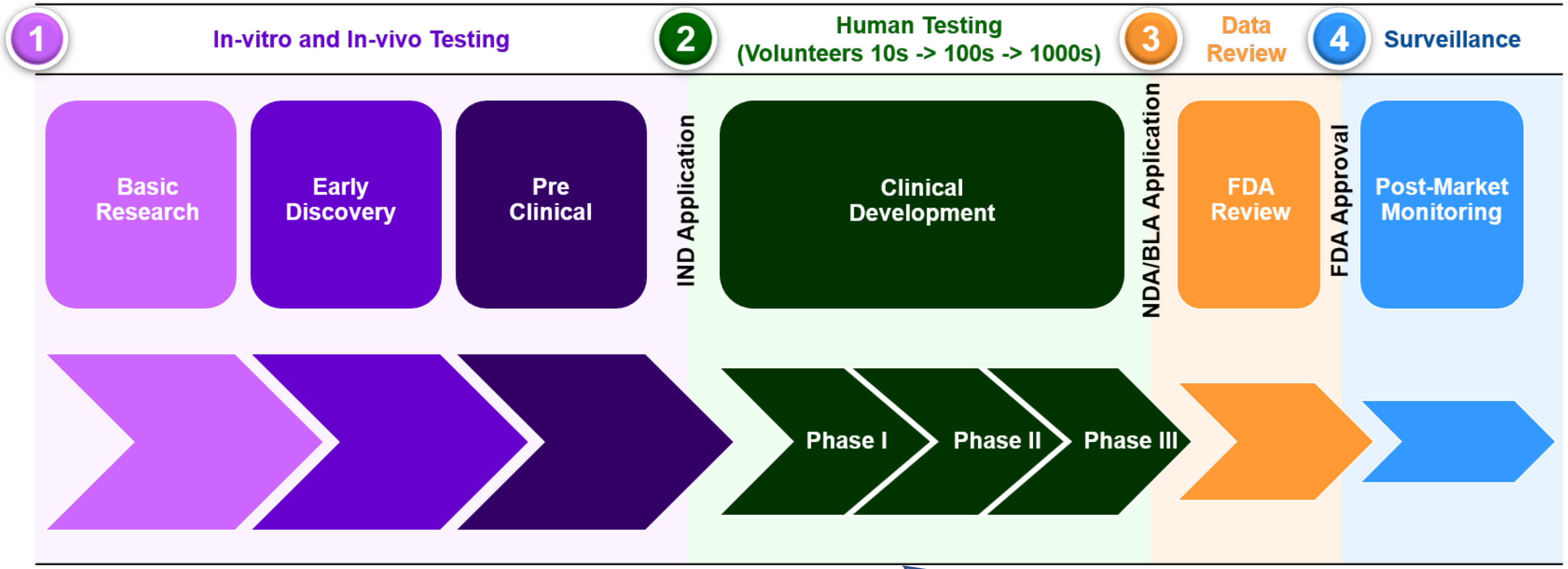
Table 2: User study conducted by two domain experts.

❖Yang et al., *Writing by Memorizing: Hierarchical Retrieval-based Medical Report Generation*, ACL 2021.

Outline

- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

Traditional Drug Discovery & Development Process

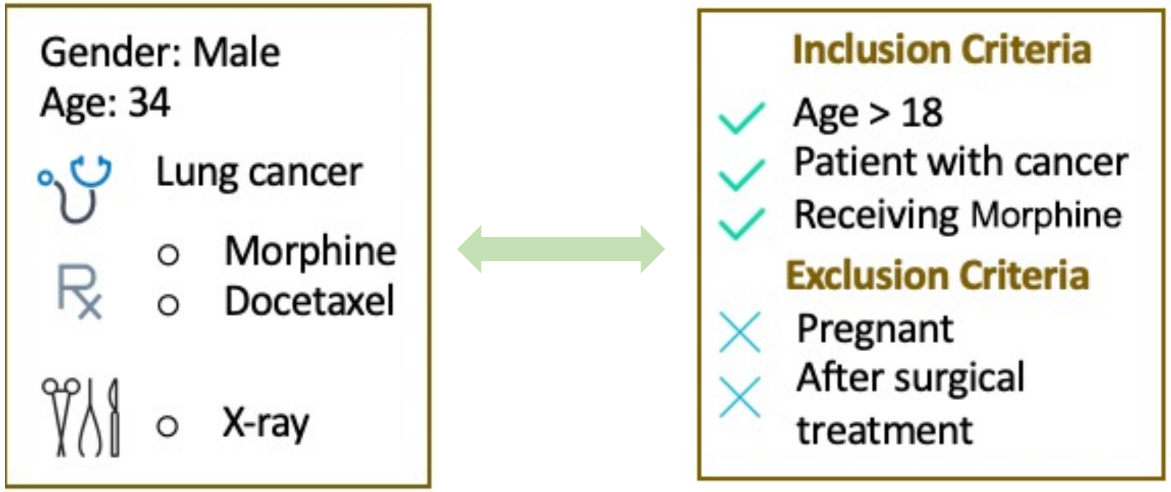


1. Statistics show 50% of trials delayed, 25% of cancer trials failed due to enrollment.
2. The recruitment cost is high, estimated around 6,000 to 7,500 USD per patient.

❖Gao et al., *COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching*, KDD 2020.

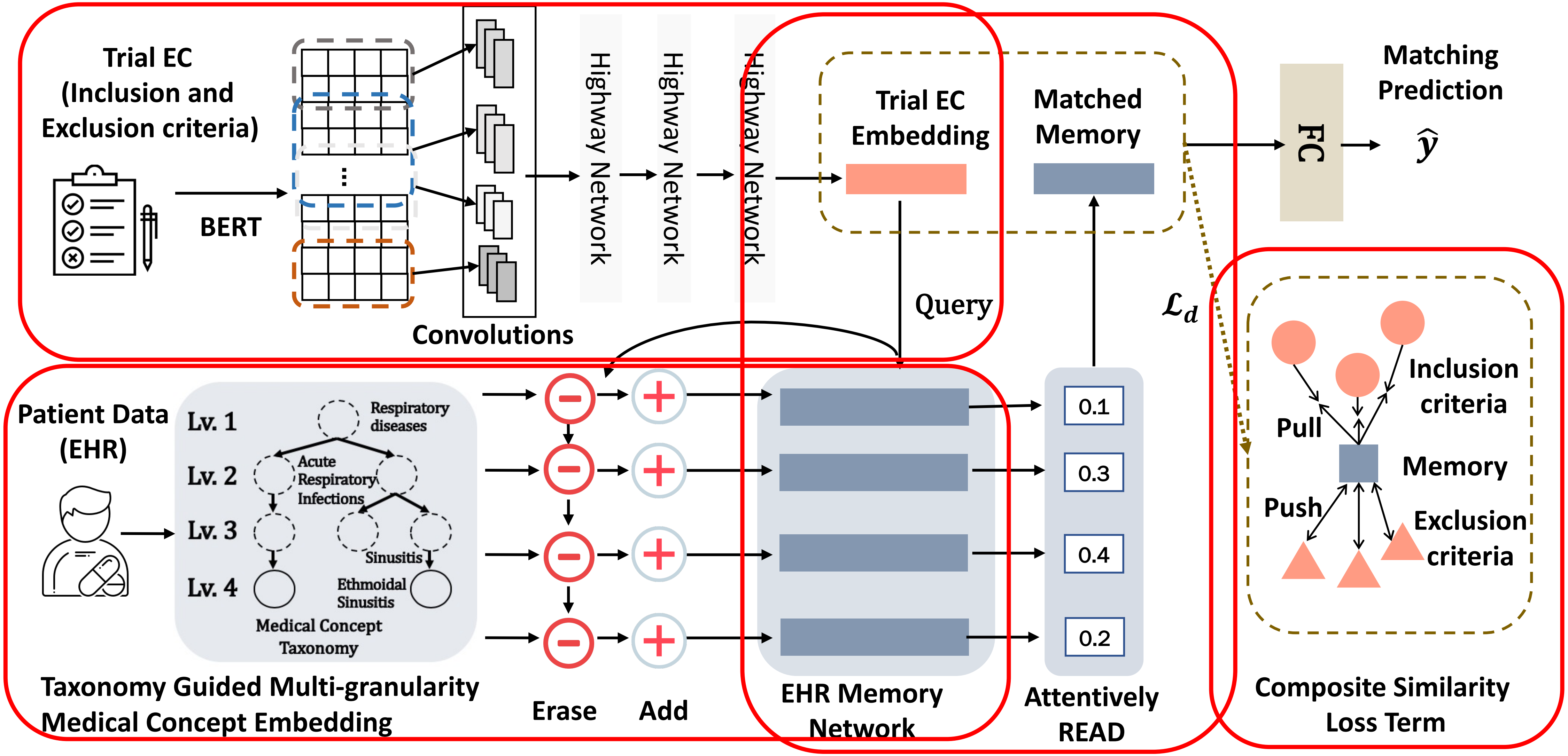
What is patient trial matching?

- Electronic Health Records (EHR): A type of high-dimensional sequence data
  - Procedures
  - Diagnosis
  - Drugs
- Clinical trials: Unstructured text data
  - Inclusion Criteria
  - Exclusion Criteria



❖Gao et al., *COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching*, KDD 2020.

COMPOSE

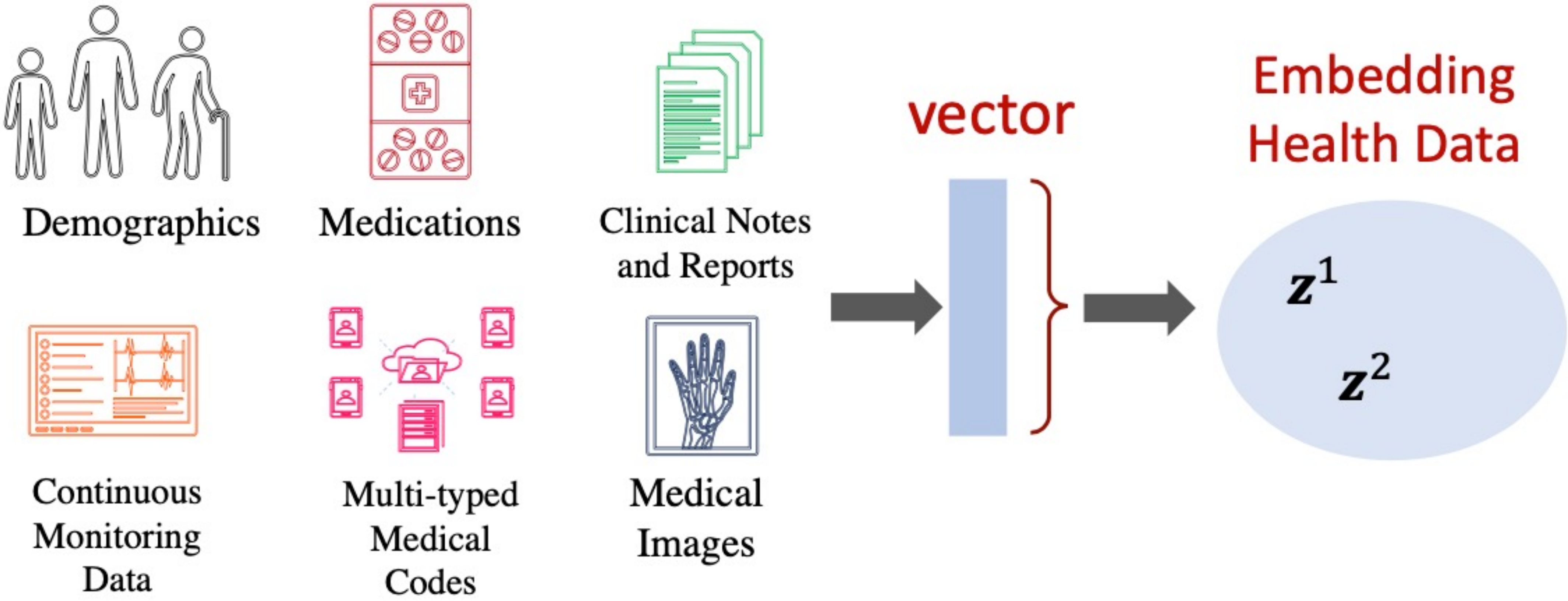


❖Gao et al., *COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching*, KDD 2020.

Outline

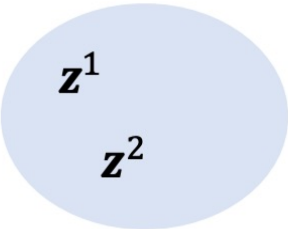
- Introduction to Electronic Healthcare Records
  - Various types of EHR data
  - Different applications
- Part I: Mining structured health data
  - Phenotyping
  - Disease detection/Risk prediction
  - Treatment recommendation
- Part II: Mining unstructured health data
  - Automated ICD coding /Disease classification
  - Understandable medical language translation
  - Medical report generation
  - Clinical trial mining
- Conclusion and Future Outlook

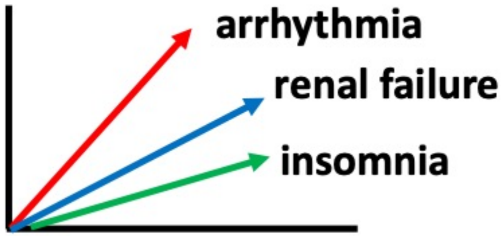
Representations Learning from Health Data



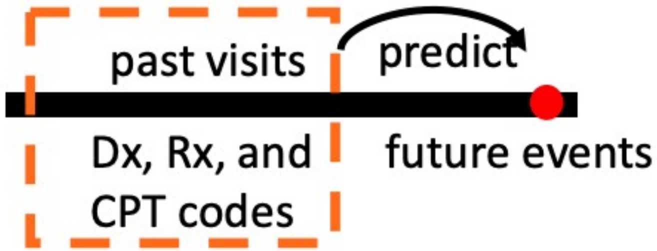
Analytics Tasks using EHR Data

Embedding
Health Data





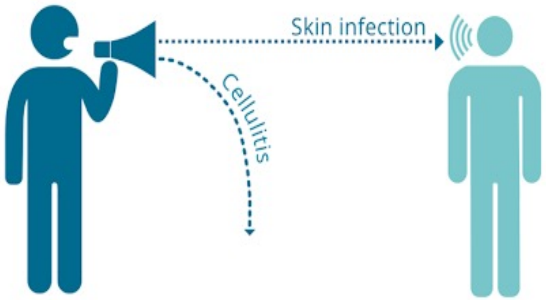
Phenotyping



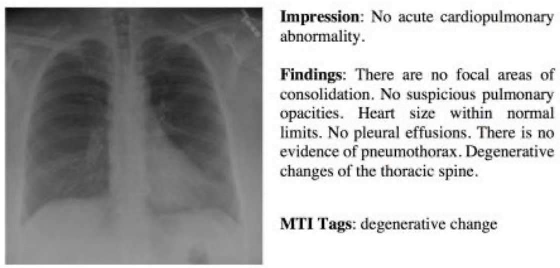
Risk Prediction

Medication
Recommendation

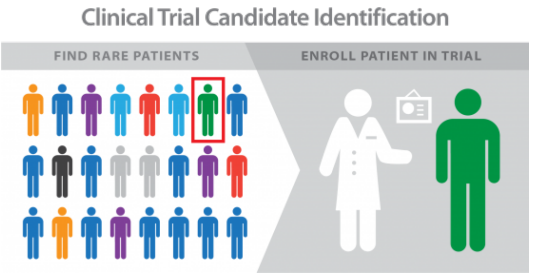
Disease
Classification



Understandable Medical
Language Translation

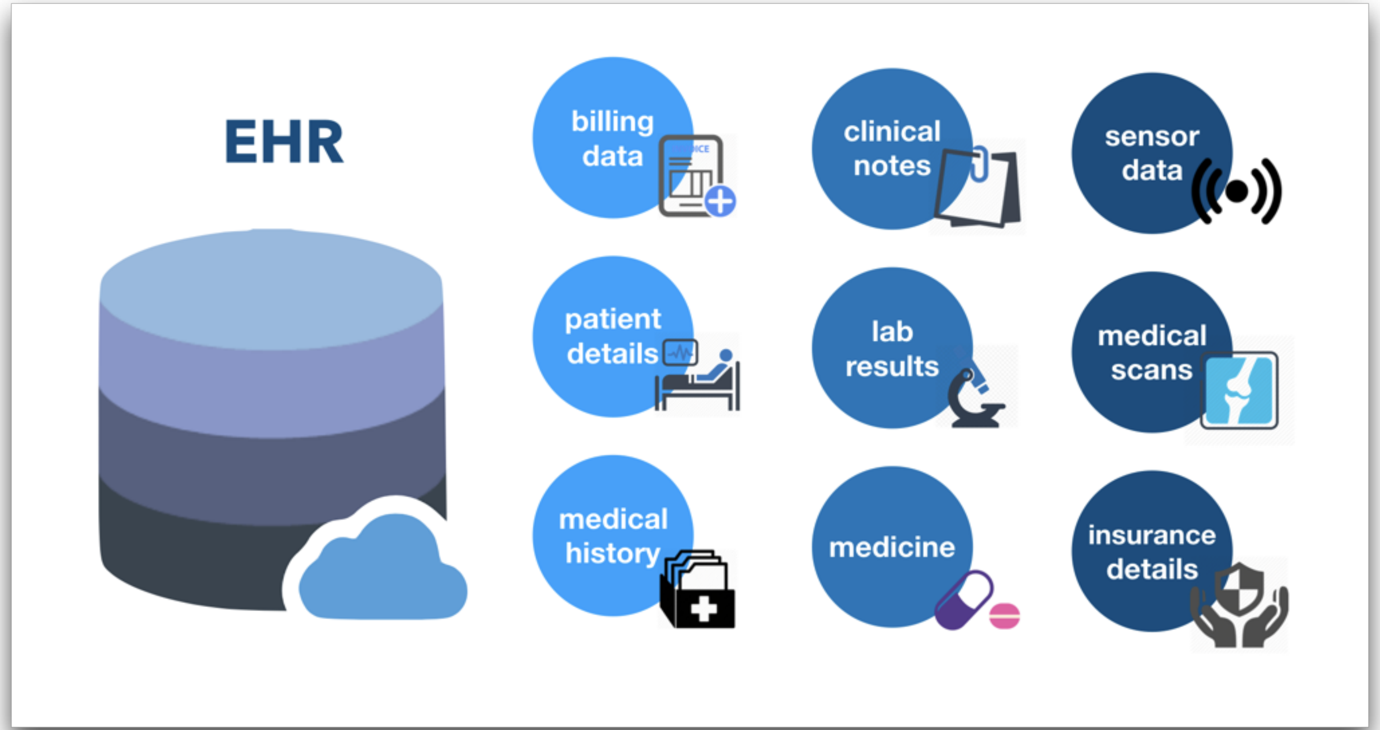


Medical Report
Generation



Clinical Trial
Mining

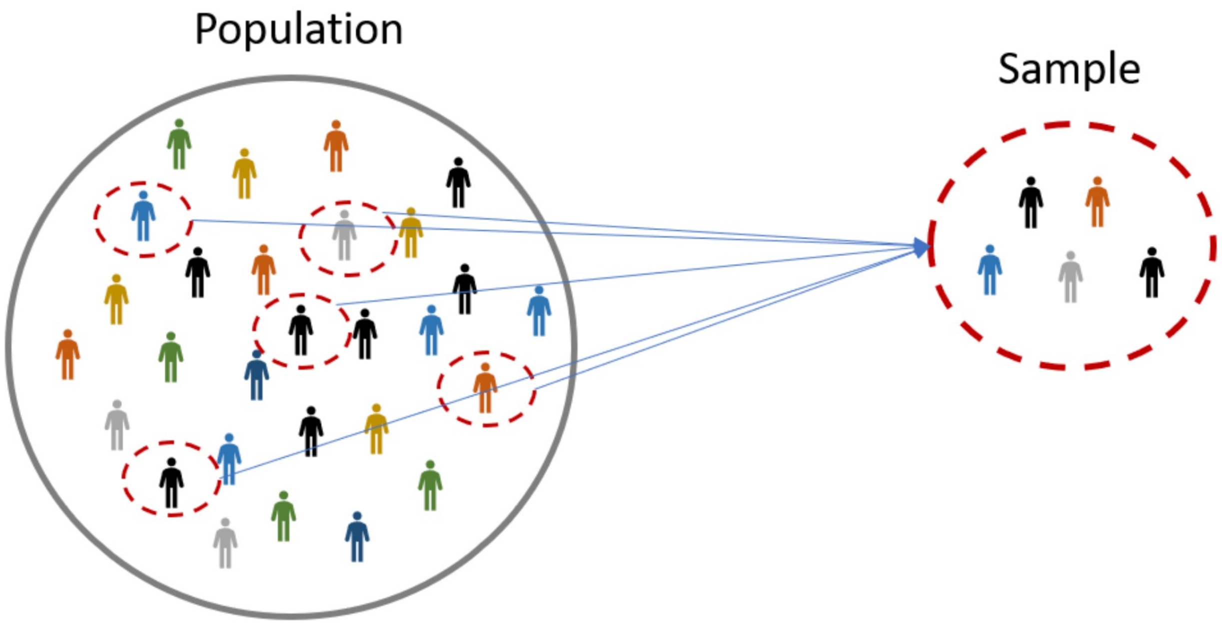
Challenges of Mining Heterogeneous Health Data



Source: https://goku.me/blog/deep-learning-with-ehr-systems

- Multi-modality

Challenges of Mining Heterogeneous Health Data



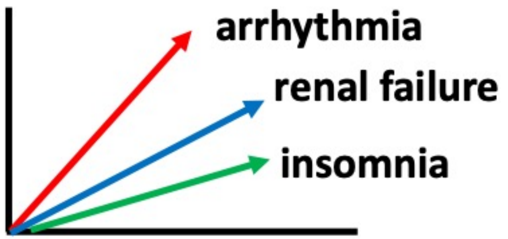
- Small Sample Size
- Lack of Label
- Fairness

Challenges of Mining Heterogeneous Health Data

- Interpretability & Robustness

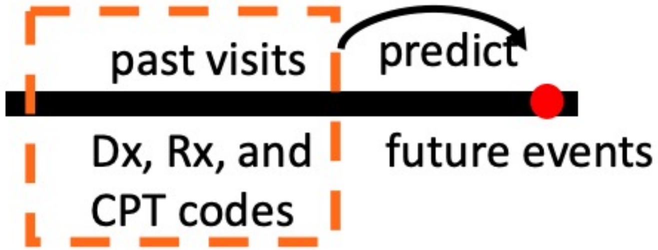
- Domain Knowledge

Open Discussions for Each Task



Phenotyping

- How to handle irregularity in EHR data?
- How to model relations between different types of medical codes?



Risk Prediction

- How to reasonably incorporate interventions?
- Do personal behaviors influence the predictions? How to model them?

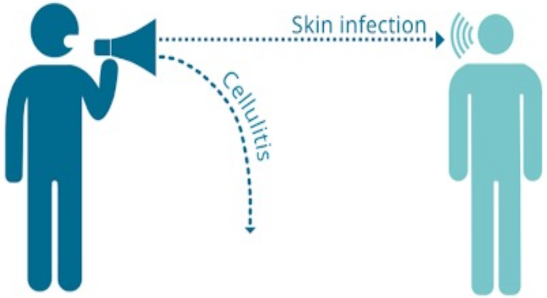
Medication Recommendation

- Will doctors preference influence results?
- Different types of insurance may cover different drugs. How to handle it?
- Socioeconomic status?

Open Discussions for Each Task

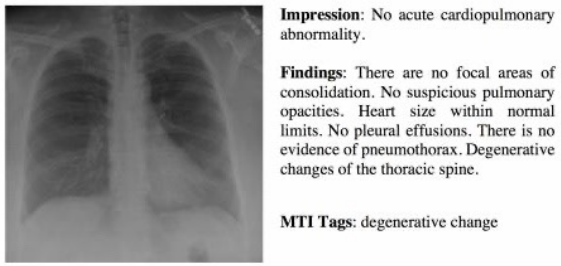
Disease Classification

- How to handle the large label size issue?
- How to use multimodal data?
- How to denoise the clinical notes?



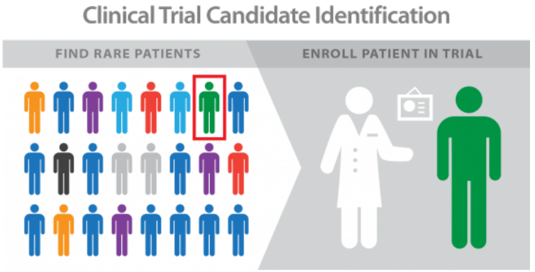
Understandable Medical Language Translation

- Personalized/user-centric translation?
- Medical Q&A?
- Medical dialogue systems?



Medical Report Generation

- How to align image and text?
- How to design fair evaluation metrics?
- How to incorporate other modalities?



Clinical Trial Mining

- Can we predict the outcome of clinical trials?
- How to find doctors?


Thank You!
Any questions?