RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records

Bum Chul Kwon, Min-Je Choi, Joanne Taery Kim, Edward Choi,
Young Bin Kim, Soonwook Kwon, Jimeng Sun, and Jaegul Choo

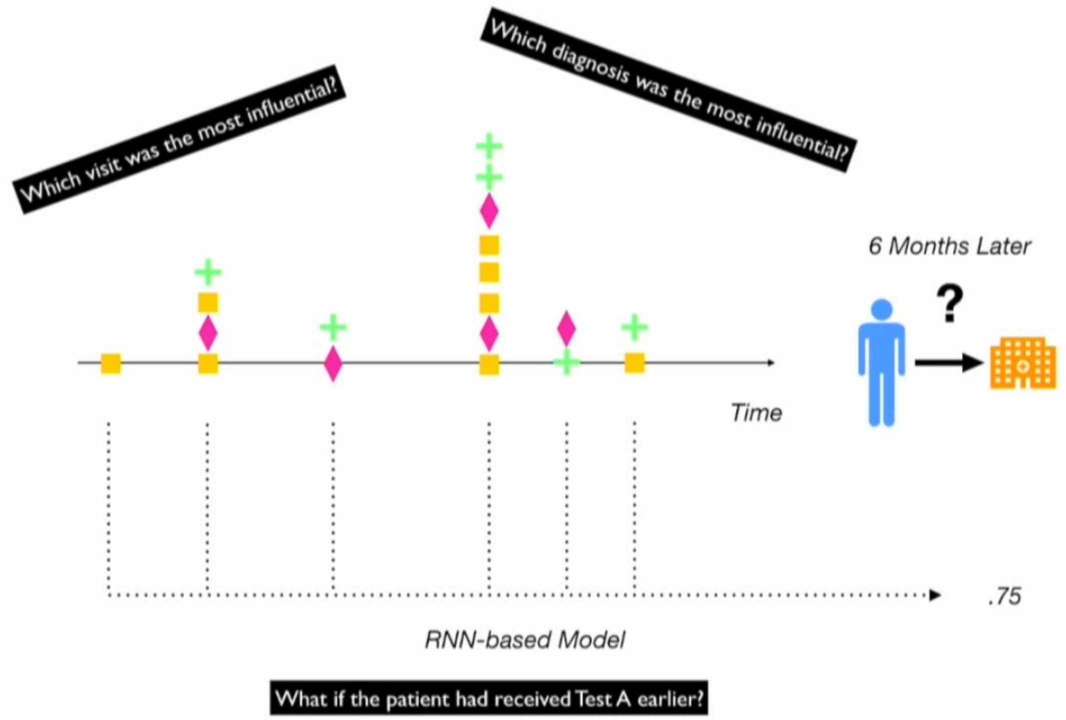
- *Bum Chul Kwon is with IBM T.J. Watson Research Center. E-mail: bumchul.kwon@us.ibm.com.*
- *Min-Je Choi, Joanne Taery Kim, and Jaegul Choo, who is the corresponding author, are with Korea University. E-mail: devnote5676, tengyee, jchoo@korea.ac.kr.*
- *Edward Choi and Jimeng Sun are with Georgia Institute of Technology. E-mail: mp2893, jsun@gatech.edu.*
- *Young Bin Kim is with Chung-Ang University. E-mail: ybkim85@cau.ac.kr.*
- *Soonwook Kwon is with Catholic University of Daegu. E-mail: anatomy3@cu.ac.kr.*

1) Design of a RNN-based model, ***RetainEX***, improved for Interpretability and Interactivity
2) Design of ***RetainVis*** that couples multiple views with model
3) Quantitative and Qualitative Evaluation of Model and Case Study

RetainVis, VAST 2018 https://vimeo.com/303201095
RetainVis: Demonstration https://vimeo.com/272587219
1st Author’s website https://www.bckwon.com/

http://solarisailab.com/archives/1515

How can we leverage users' domain knowledge and prior knowledge as inputs for steering the RNN-based model?



Design Study with

Artificial Intelligence, *Medical*, and Visual Analytics Experts

https://vimeo.com/303201095

Dataset

- The Health Insurance Review and Assessment Service (HIRA)
- The national patient sampling (HIRA-NPS) contains 1.4 million patients through age- and gender-stratified random sampling
- The Korean Standard Classification of Disease, Ninth Revision (KCD-9)

User Questions - Individual

- What’s the expected outcome of Patient A?
  - What’s the diagnostic risk of Patient A as of ith visit?
- Why? Which visits or procedure contribute the most to prediction?
- What if the patient had procedure Y instead of X?
  - What if the patient had it more frequently?
  - What if the patient had it more recently?

User Questions - Population

- What's the expected outcomes of **selected group of patients**?

- Why? Which visits or procedure contribute the most to prediction of **the group of patients**?
  - What's the most contributing procedure, respectively?
  - Is the result in line with medical knowledge?

Fig. 1. A screenshot of RetainVis consisting of five areas: (A) *Overview* shows an overview of all patients (left) and an attribute summary view (right) of patients. (B) *Patient Summary* shows the summary of the patient cohort built from (A). (C) *Patient List* shows individual patients in a row of rectangles. In Patient List, users can select a patient of interest to view details in (E) *Patient Details*. Users can open (D) *Patient Editor* to conduct a what-if analysis, and (E) *Patient Details* shows the updated results.

https://vimeo.com/303201095

Main User Tasks

- T1: View Patient Demographics and Medical Records Summary.
- T2: Select Interesting Patient Cohorts.
- T3: View Summary of Selected Patients.
- T4: Investigate Single Patient History.
- T5: View Contributions of Medical Records for Prediction.
- T6: Conduct What-If Case Analyses.
- T7: Evaluate and Steer Predictive Model.

Model Improvement

We investigated how we can improve the RNN-based model to support user tasks.

Interpretability:
1) Linear relationship between input and output

Interactivity:
1) Testing what-if scenarios by making changes to input
2) Retraining model based on user feedback on results

Performance:
1) Maintain the performance level of other RNNs

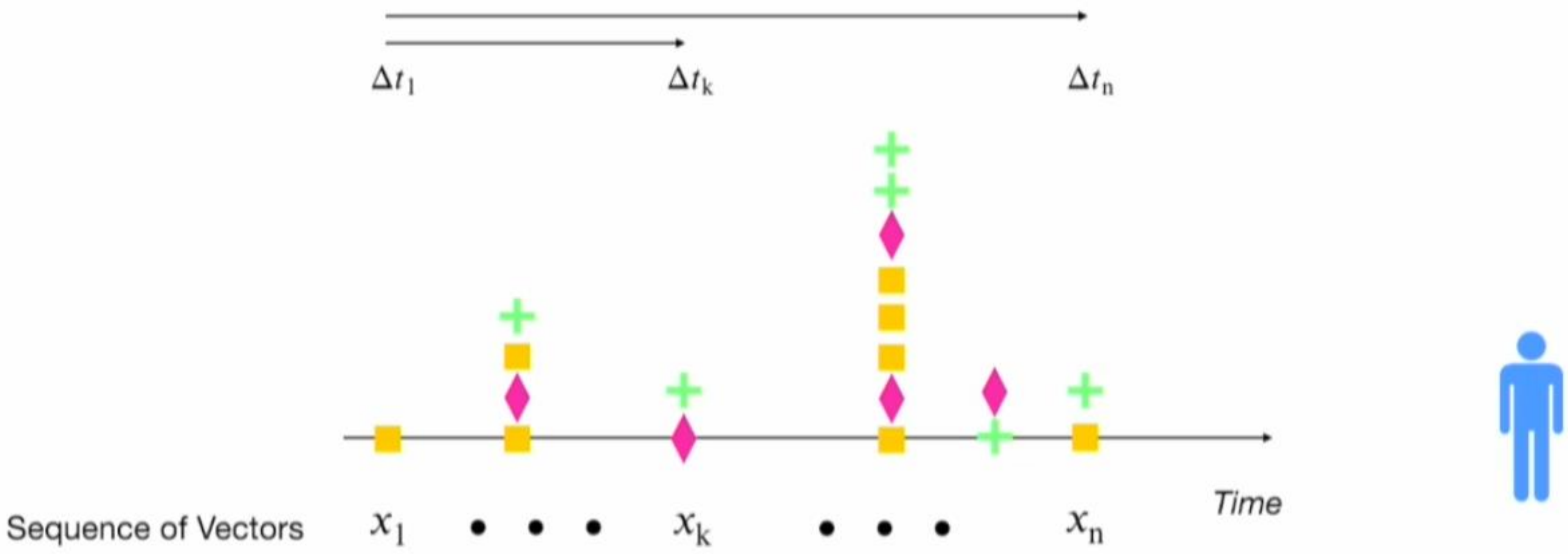
Base Model: **Retain**

RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism

Edward Choi, Mohammad Taha Bahadori, Joshua A. Kulas, Andy Schuetz, Walter F. Stewart, Jimeng Sun

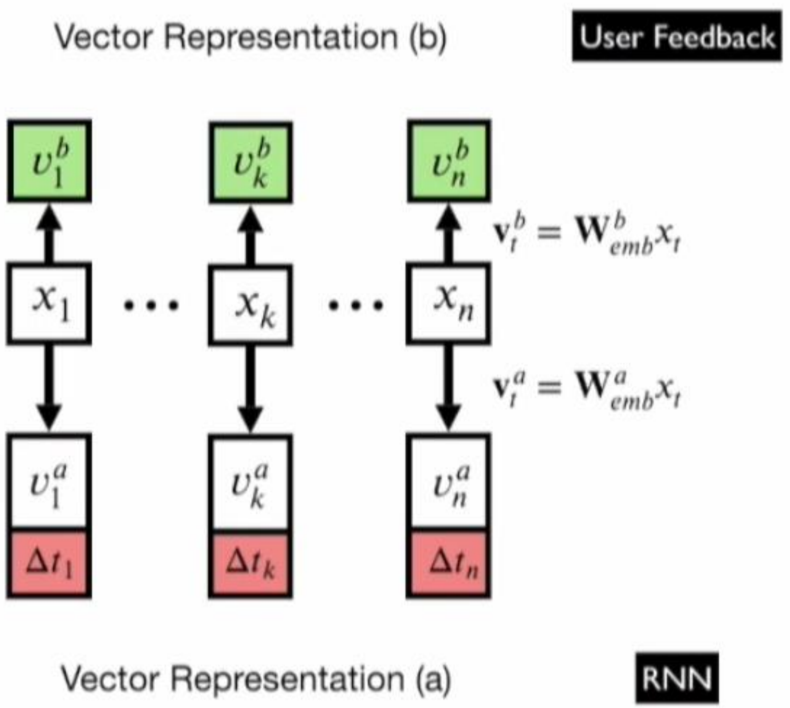
(Submitted on 19 Aug 2016 (v1), last revised 26 Feb 2017 (this version, v4))

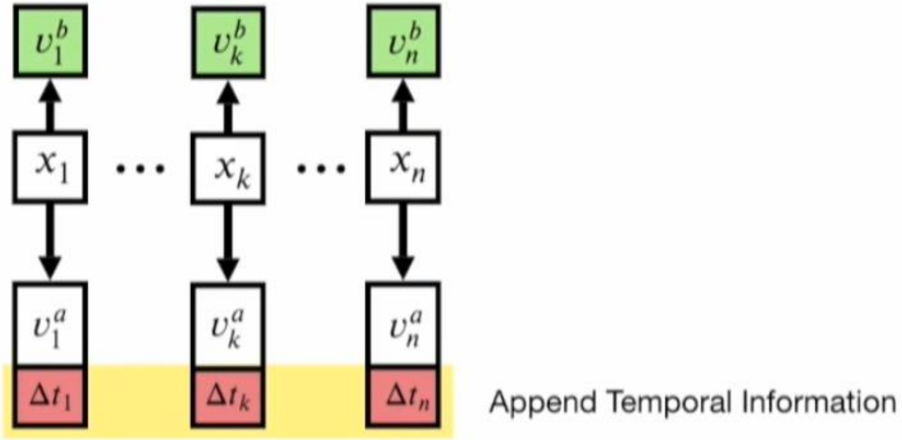
Accuracy and interpretability are two dominant features of successful predictive models. Typically, a choice must be made in favor of complex black box models such as recurrent neural networks (RNN) for accuracy versus less accurate but more interpretable traditional models such as logistic regression. This tradeoff poses challenges in medicine where both accuracy and interpretability are important. We addressed this challenge by developing the REverse Time AttentioN model (RETAIN) for application to Electronic Health Records (EHR) data. RETAIN achieves high accuracy while remaining clinically interpretable and is based on a two-level neural attention model that detects influential past visits and significant clinical variables within those visits (e.g. key diagnoses). RETAIN mimics physician practice by attending the EHR data in a reverse time order so that recent clinical visits are likely to receive higher attention. RETAIN was tested on a large health system EHR dataset with 14 million visits completed by 263K patients over an 8 year period and demonstrated predictive accuracy and computational scalability comparable to state-of-the-art methods such as RNN, and ease of interpretability comparable to traditional models.



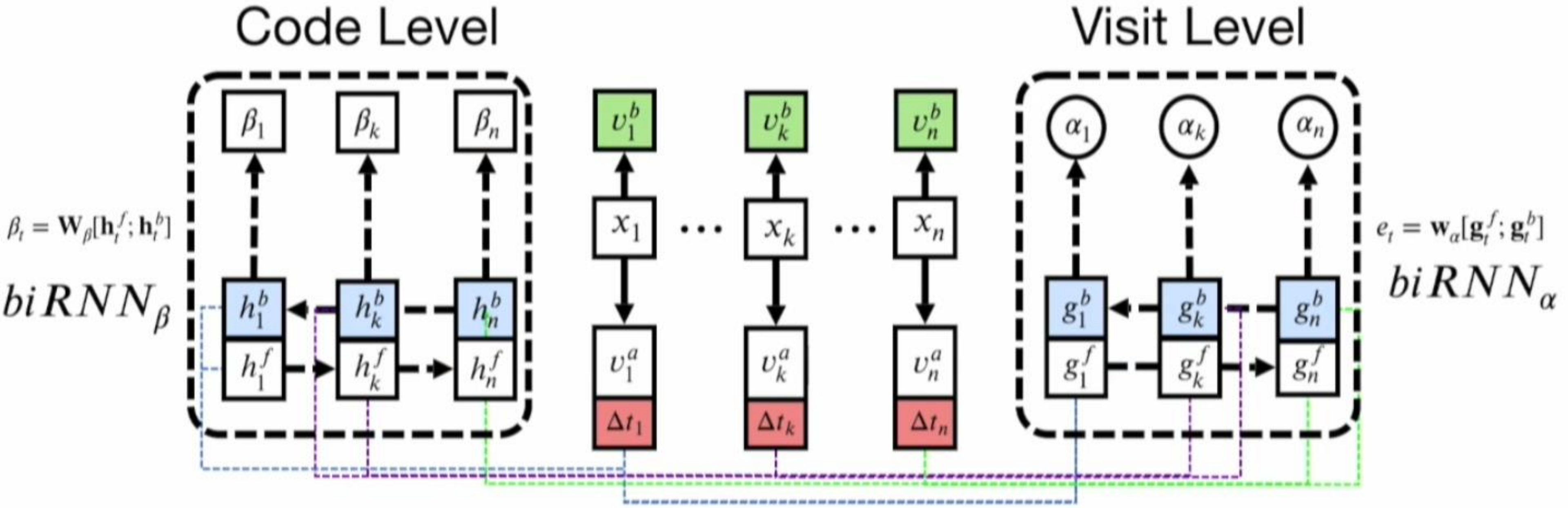
1400-dimensional binary vector

(KCD-9 code)



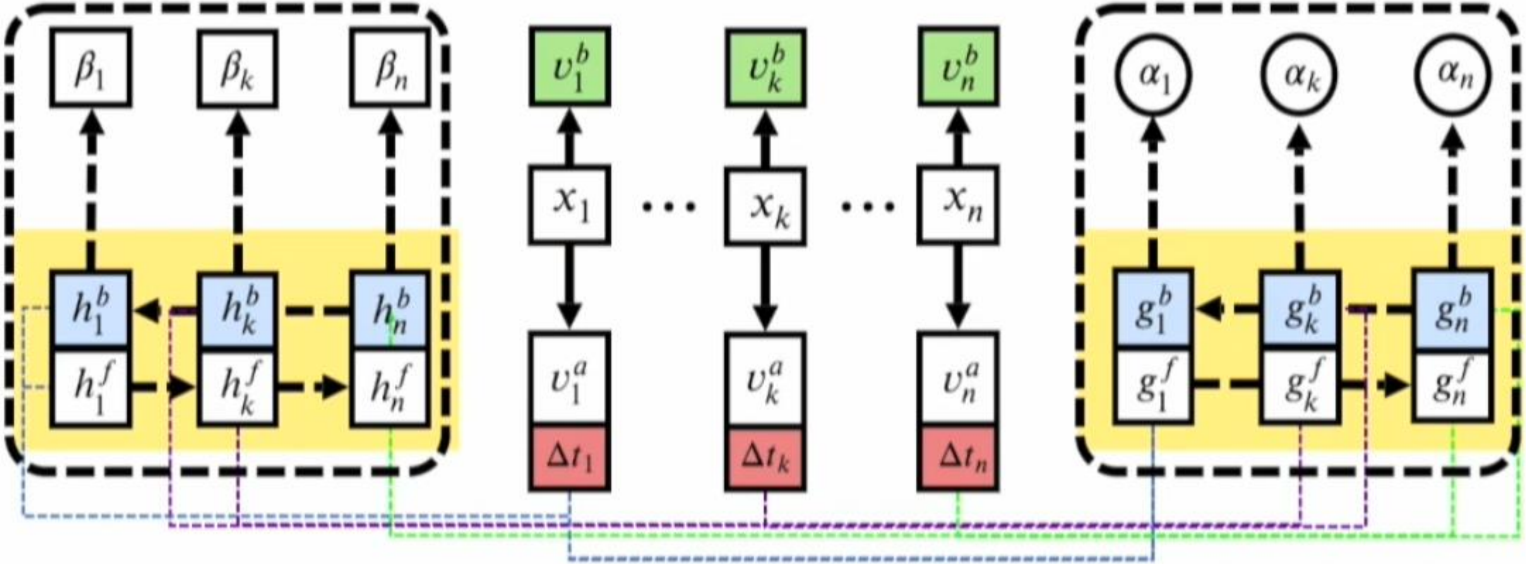


Retain model에는 없었음



Code Level
Visit Level
$\beta_t = \mathbf{W}_\beta[\mathbf{h}_t^f; \mathbf{h}_t^b]$
$biRNN_\beta$
β_1 β_k β_n
h_1^b h_k^b h_n^b
h_1^f h_k^f h_n^f
v_1^b v_k^b v_n^b
x_1 … x_k … x_n
v_1^a v_k^a v_n^a
Δt_1 Δt_k Δt_n
α_1 α_k α_n
g_1^b g_k^b g_n^b
g_1^f g_k^f g_n^f
$e_t = \mathbf{w}_\alpha[\mathbf{g}_t^f; \mathbf{g}_t^b]$
$biRNN_\alpha$

Physicians observe patient's history in a **chronological** order to see progression and trace **backward** to identify possible cues



Hidden States of Bidirectional RNN

Hidden States of Bidirectional RNN

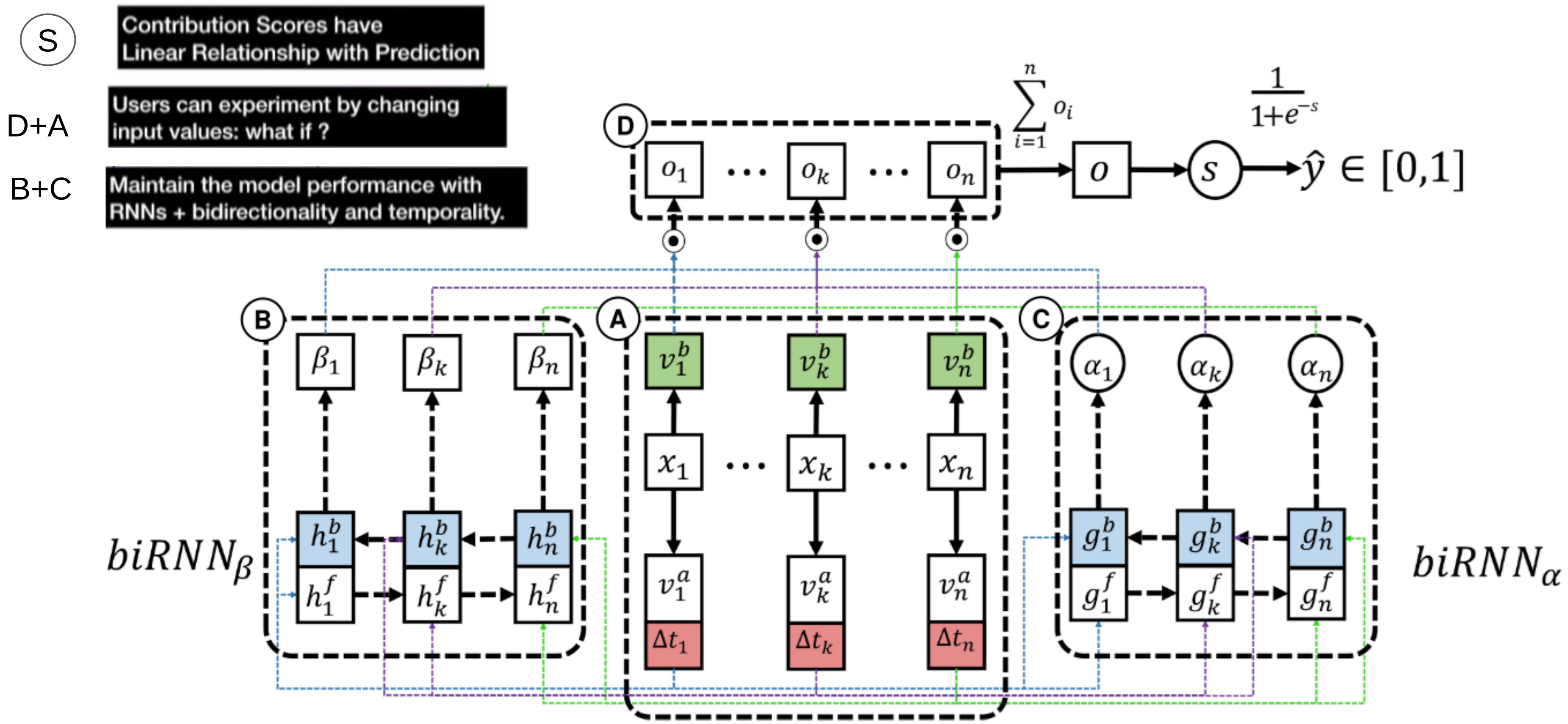
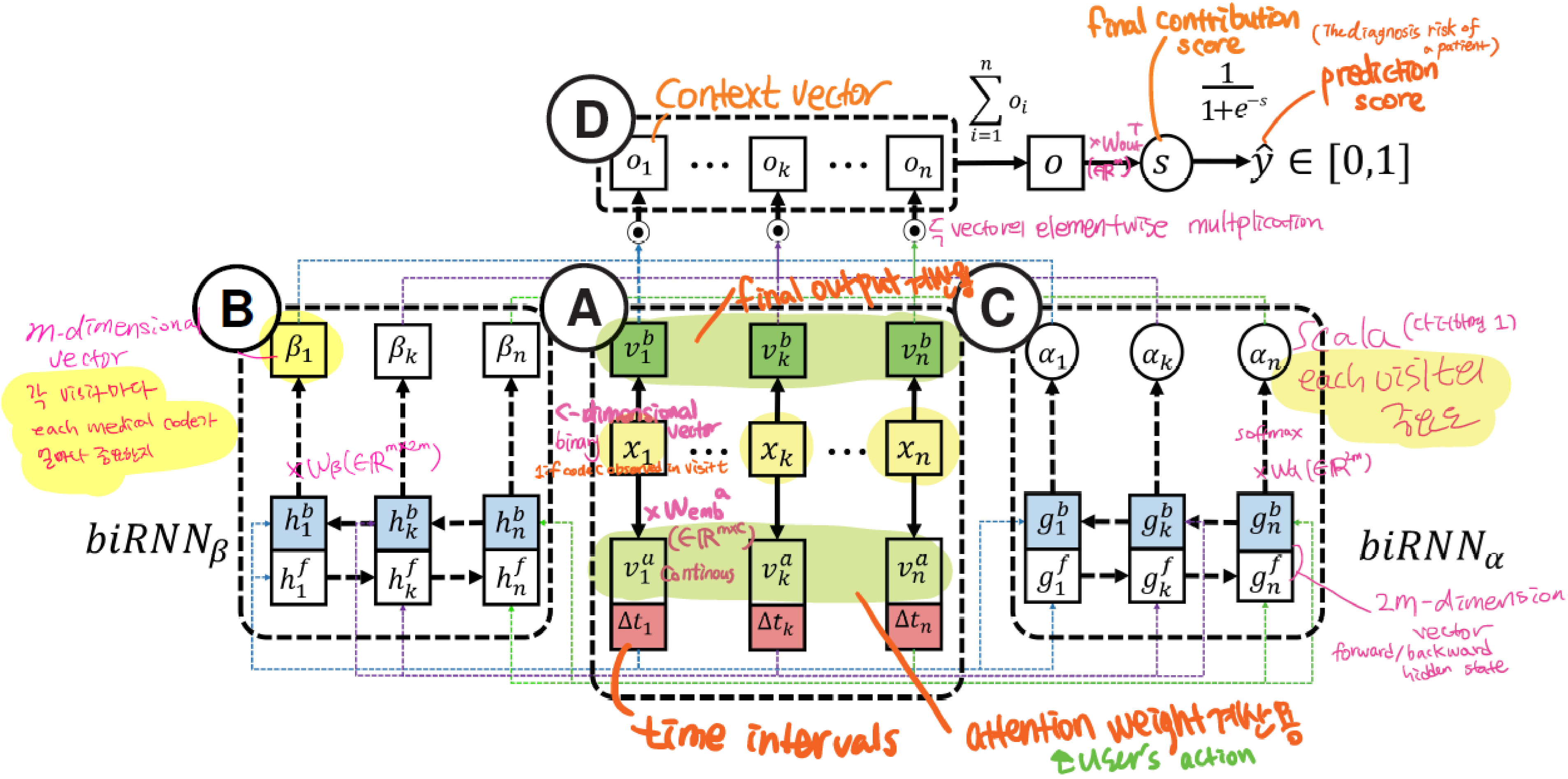


Fig. 6. Overview of RetainEX. (A) Using separate embedding matrices, the binary vectors $\mathbf{x}_1, \ldots, \mathbf{x}_T$ are transformed into embedding vectors $\mathbf{v}^a{}_1, \ldots, \mathbf{v}^a{}_T$ and $\mathbf{v}^b{}_1, \ldots, \mathbf{v}^b{}_T$, with time interval information appended to the former. (B) $\mathbf{v}^a{}_1, \ldots, \mathbf{v}^a{}_T$ are fed into a bidirectional RNN to produce scalar weights α. (C) $\mathbf{v}^a{}_1, \ldots, \mathbf{v}^a{}_T$ are fed into another biRNN, this time to generate vector weights β. (D) α, β and $\mathbf{v}^b$ are multiplied over all timesteps, then are summed to form a single vector $\mathbf{o}$, which goes through linear and nonlinear transformation to produce a probability score $\hat{y}$.



D
o_1 ⋯ o_k ⋯ o_n
Context vector
$\sum_{i=1}^{n} o_i$
o
$\times W_{out}^T$ ($\in \mathbb{R}^m$)
s
final contribution score
$\frac{1}{1+e^{-s}}$
$\hat{y} \in [0,1]$
prediction score
(The diagnosis risk of a patient)
← vector의 elementwise multiplication
A
final output
v_1^b v_k^b v_n^b
c-dimensional vector
binary
1 if code c observed in visit t
x_1 ⋯ x_k ⋯ x_n
$\times W_{emb}^a$ ($\in \mathbb{R}^{m \times c}$)
continous
v_1^a v_k^a v_n^a
Δt_1 Δt_k Δt_n
time intervals
attention weight 계산됨
← user's action
B
m-dimensional vector
각 visit마다 each medical code가 얼마나 중요한지
β_1 β_k β_n
$\times W_\beta$ ($\in \mathbb{R}^{m \times 2m}$)
h_1^b h_k^b h_n^b
h_1^f h_k^f h_n^f
$biRNN_\beta$
C
α_1 α_k α_n
scala (다 더하면 1)
each visited 중요도
softmax
$\times W_\alpha$ ($\in \mathbb{R}^{2m}$)
g_1^b g_k^b g_n^b
g_1^f g_k^f g_n^f
$biRNN_\alpha$
2m-dimension vector
forward/backward hidden state

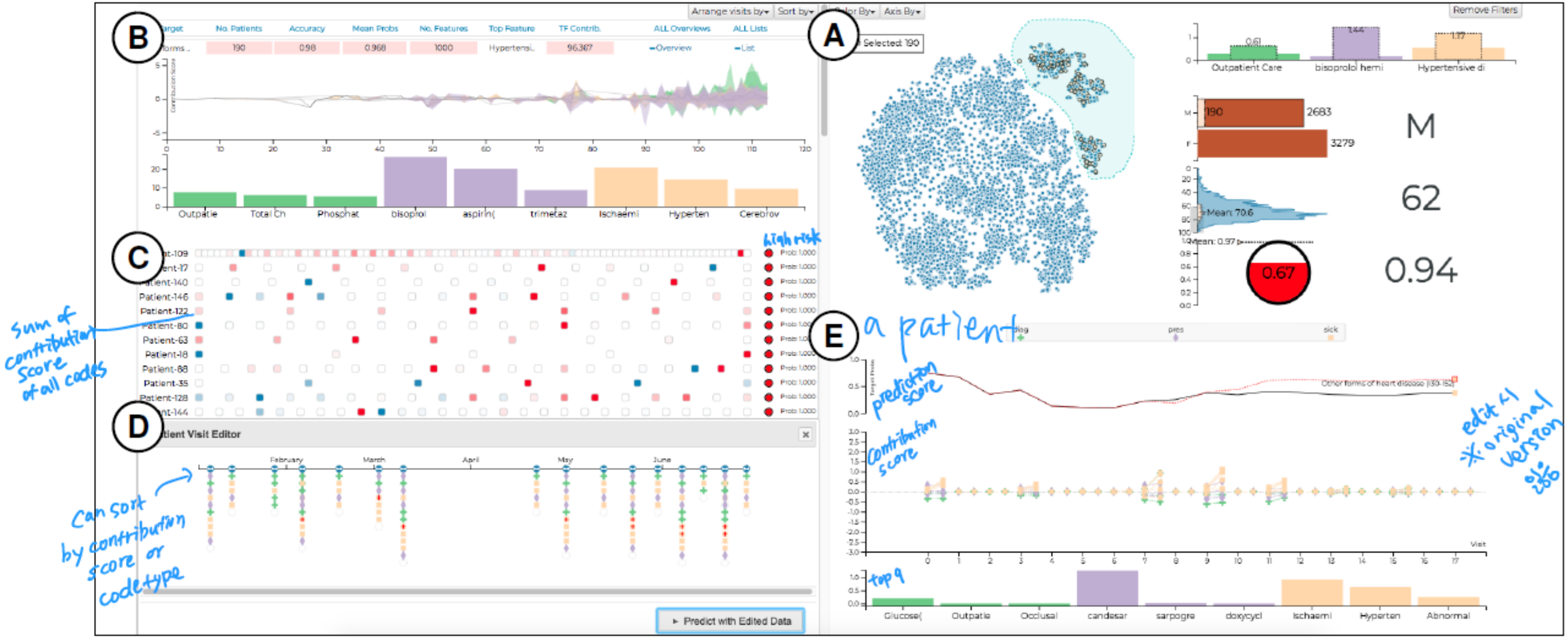
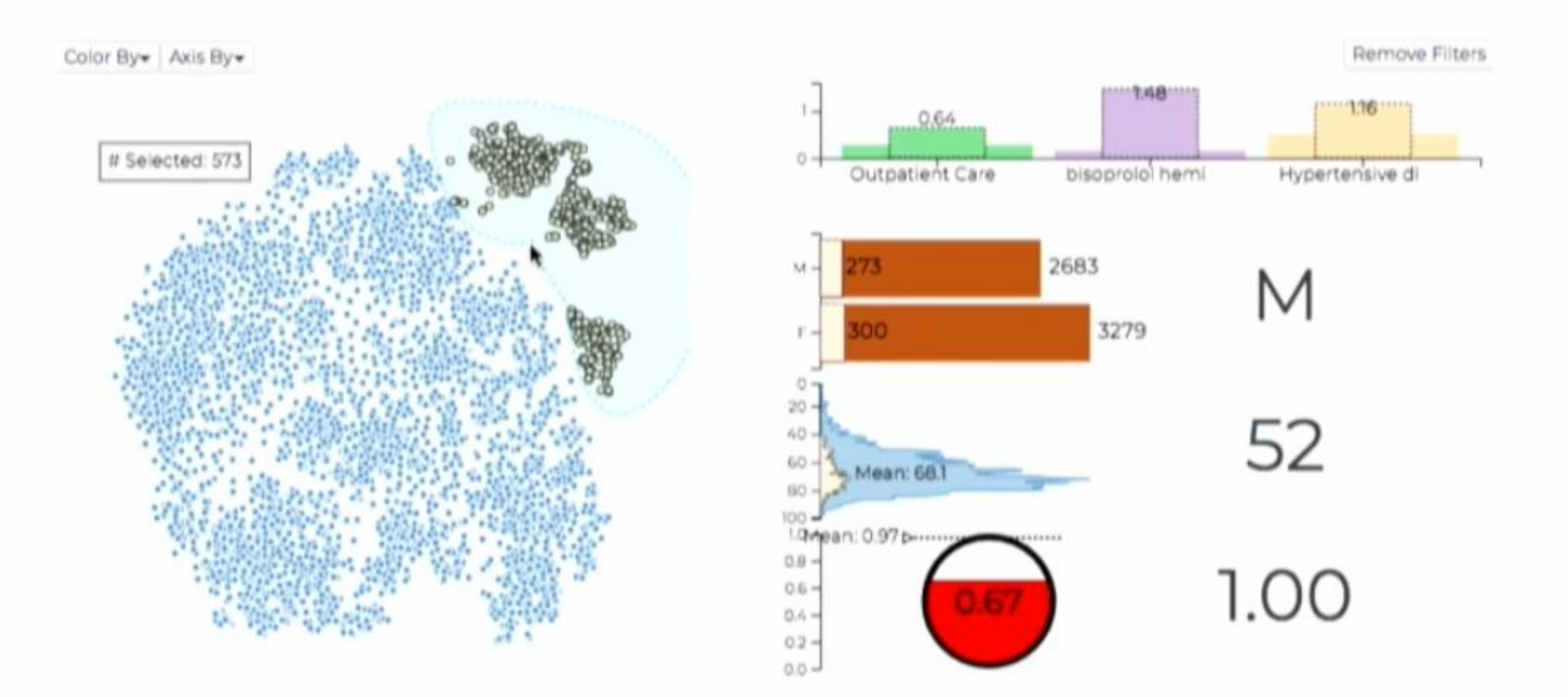


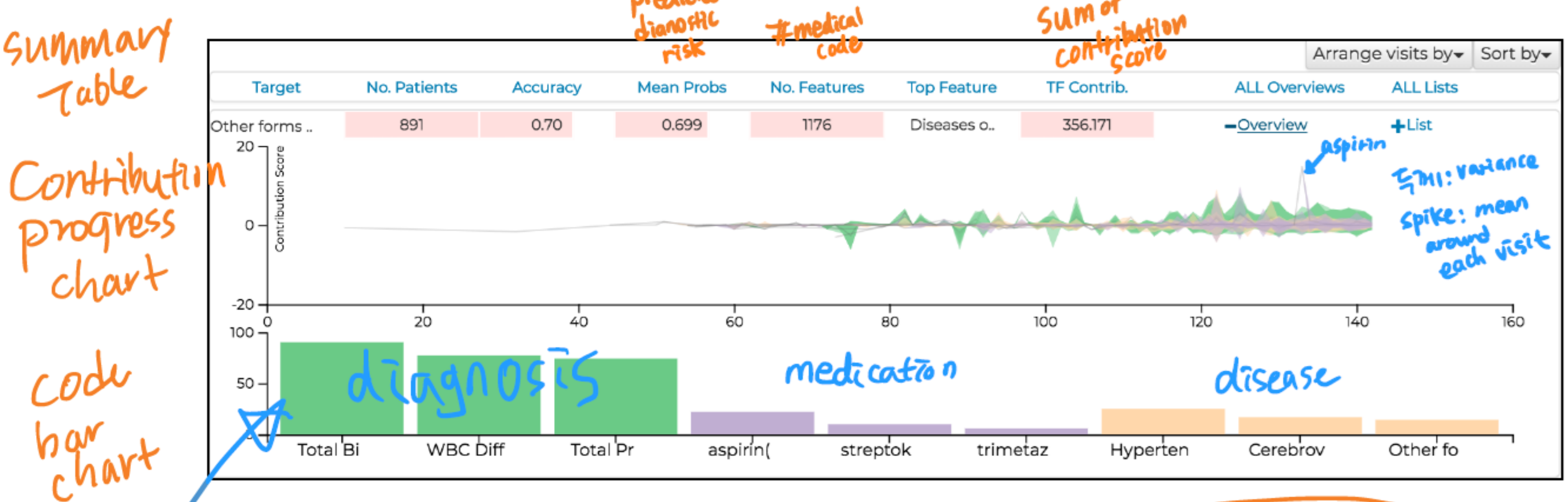
Fig. 1. A screenshot of RetainVis consisting of five areas: (A) *Overview* shows an overview of all patients (left) and an attribute summary view (right) of patients. (B) *Patient Summary* shows the summary of the patient cohort built from (A). (C) *Patient List* shows individual patients in a row of rectangles. In Patient List, users can select a patient of interest to view details in (E) *Patient Details*. Users can open (D) *Patient Editor* to conduct a what-if analysis, and (E) *Patient Details* shows the updated results.

(A) 11:40


Color By
Axis By
default: tsne
x: age, y: pred
Build Custom
Each patient can be represented with contribution vector O


Color By
Axis By
Selected: 573
Remove Filters
0.64
1.48
1.16
Outpatient Care
bisoprolol hemi
Hypertensive di
M
F
273
2683
300
3279
M
Mean: 68.1
52
Mean: 0.97
0.67
1.00

(B)

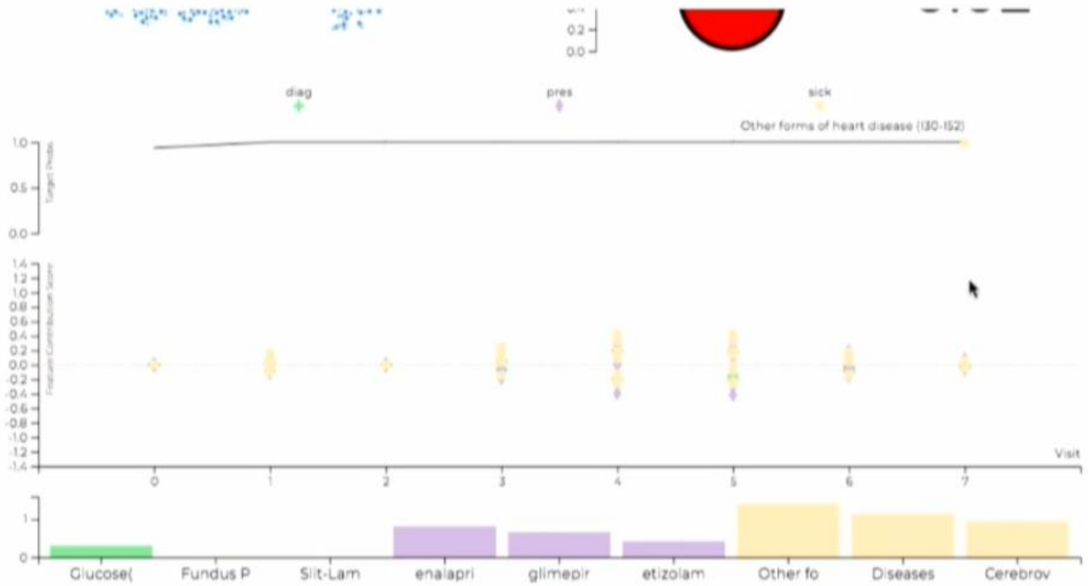


clicks

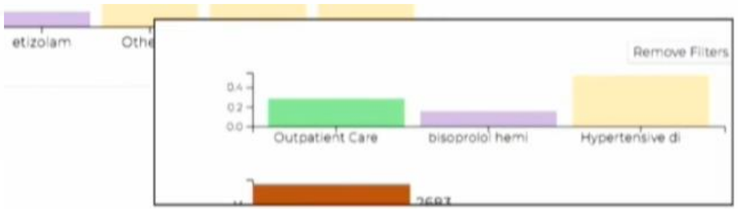
Fig. 4. *Patient Summary* shows a summary of selected patients. Table summarizes description of selected patients. In the middle, an area chart shows aggregated contribution scores of nine medical codes over time. It shows mean and standard deviation as an area. Users can also see the medical codes and their mean contribution scores in bar chart.

Contribution Score of Code C at Time T $s_{t,c} = \alpha_t \mathbf{w}_{out}(\mathbf{W}^b_{emb}[:, c] \bullet \beta_t)$

(E) 12:20



Contribution Score of Code C $\sum s_{i,c}$



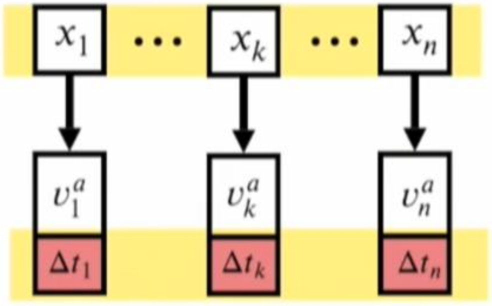
$\sum \sum s_{i,c}$ Top-k Contribution Scores of Multiple Patients

(C) 13:03

Contribution Score of Code C at Time T $s_{t,c} = \alpha_t \mathbf{w}_{out}(\mathbf{W}^b_{emb}[:, c] \bullet \beta_t)$

Contribution Score of Time T $\sum s_{t,i}$

(E) 13:30



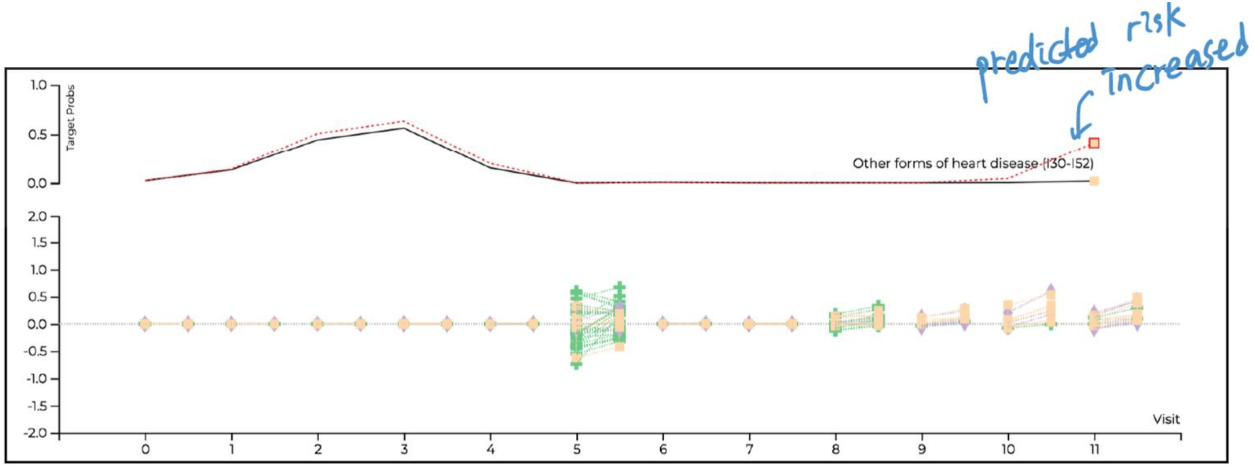


Fig. 5. The what-if analysis result shows increase in predicted diagnosis risks and contribution scores of related medical codes.

RetainEX

- Generates diagnostic risk prediction
- Generates contribution scores of each code per visit
  - Linearly correlated with prediction
- Generates patient embedding
  - Can be used for overview, similarity analysis

RetainEX Performance

- Predict diagnosis of a disease after a time period
  - Heart Failure: 63,030 patients (5,730 Case)
  - Cataract: 117,612 patients (10,692 Case)
- Split train, validation, and test sets: 0.65/0.1/0.25
- Use Ubuntu 14.04 server equipped with two Nvidia Titan X GPUs to train our models
- Compare against GRU, RETAIN, RETAINEX without Time

RetainEX Performance

	(a) heart failure			(b) cataract		
Models	AUC	AP	time (s)	AUC	AP	time (s)
GRU	0.906	0.694	997	0.953	0.834	2367
RETAIN	0.905	0.729	1114	0.959	0.835	2700
RetainEX w/o time	0.946	0.769	1143	**0.975**	0.870	2619
RetainEX	**0.954**	**0.818**	1148	**0.975**	**0.878**	2632

- Takes 3.19 hours to train using 40,964 patients in our use case for 10 epochs, with 1148 seconds per epoch.
- Takes 366 seconds to generate diagnosis risk scores as well as contribution scores.

RetainEX Performance

Top-5 contribution scores averaged over the total number of patients

Code Type	Code Name	Mean Score
Diagnosis	Obesity and other hyperalimentation	0.206
	Other infectious diseases	0.169
	Ischaemic heart diseases	0.156
	Hypertensive diseases	0.134
	Disorders of thyroid gland	0.119
Treatment	Prothrombin Time	0.299
	24hr blood pressure examination	0.278
	CA-19-9	0.253
	CK-MB	0.198
	Fibrinogen examination (functional)	0.185
Prescription	Bisoprolol hemifumarate	0.523
	Isosorbide mononitrate	0.243
	Amlodipine besylate	0.210
	Mmorphine sulfate	0.164
	Carvedilol	0.157

- Hypertensive diseases, Diseases of oesophagus, stomach and duodenum, Ischaemic heart diseases, Metabolic disorders, Cerebrovascular diseases

- Top-5 Prescriptions are highly relevant to preventive / treatment medicines for Heart Diseases

RetainVis

- helps users explore EMRs
- helps users perform predictive analytic tasks using RetainEX
- helps users conduct what-if analysis
- helps users update RetainEx with feedback

Lessons Learned

- Improving Interpretability of RNNs through Design Study
- Improving Interactivity for experiments: what if?
- Improving Interactivity for updating model
- VIS+AI for Health:
  - Minimize False Prediction
  - Why Fail? Include Health Professionals in the Loop
  - Communicate Uncertainties and Bias