Part II: Learning from EHR data

Stockholms universitet

Panagiotis Papapetrou, PhD
Professor, Stockholm University
Adjunct Professor, Aalto University

Who are we?

DSV @ Stockholm University

- DSV: Data- och SystemVetenskap (Computer and Systems Sciences)
- # of students: approx. 5400
- # of staff members: 176 (60 profs. /associate profs. / lecturers)

Stockholms universitet

Our AI research arena @ DSV

RESEARCH AREAS @DSV

AI ACROSS FIELDS

	AI across fields	BPM and Enterprise Modeling	Cyber Security	Data Science	Digital Games	Digital Innovation	E-government and E-democracy	Health Informatics	Immersive Networking	Interaction Design	IT Management and Governance	Language Technology	Risk and Decision Analysis	Technology Enhanced Learning
Methods	Knowledge representation	■				■						■		
Methods	Decision theory and reasoning		■				■				■		■	
Methods	Machine learning			■				■				■		■
Methods	Natural language processing	■		■				■				■		
Methods	Simulation				■									
Methods	Human-computer interaction									■				
Methods	Agent systems				■								■	
Methods	Distributed systems		■						■					
Applications	Healthcare	■	■	■		■		■	■			■		
Applications	Automotive			■		■			■					
Applications	Governance	■	■				■				■		■	
Applications	Education				■									■
Applications	Security		■										■	
Applications	Games				■									
Applications	Art								■	■				
Applications	Telecommunications			■										

Research at DSV – data science group

Main research areas:

- Sequential and temporal data mining
- Interpretability and explainability of machine learning methods
- Ethics and bias in machine learning
- Machine learning for healthcare applications
- Clinical text mining and natural language processing

Current projects:

- **EXTREMUM (2020-2024):** explainable and ethical ML for healthcare
- **Covid-Sim (2020-2021):** reinforcement learning for simulation of pandemics
- **TEMPOMiner (2017-2020):** temporal data mining for detecting ADEs in healthcare

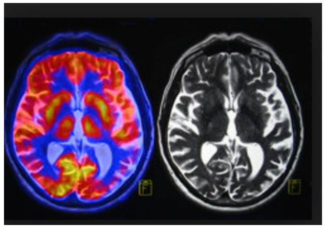
Part II - Outline

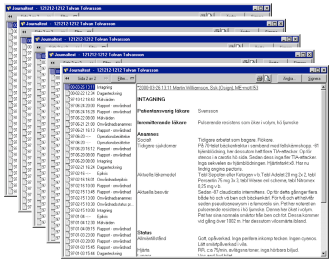
- **Temporal abstractions** for EHR data
- **Actionable models** and **counterfactual explanations** for EHR data
- **Attention-based** deep learning for healthcare event prediction
- **Interpretable ranking** and **classification** of **radiography exams**

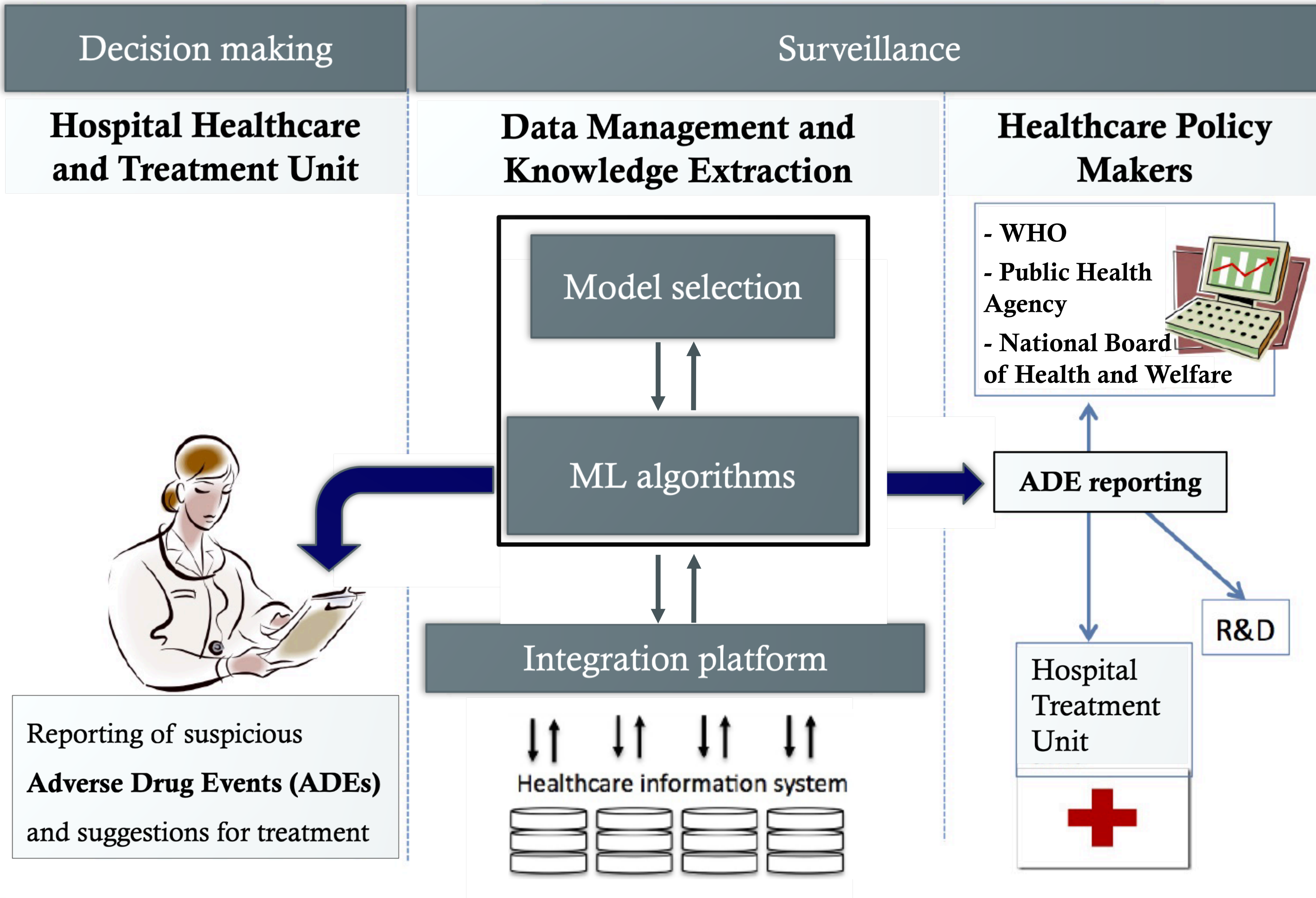
Electronic Health Records: content

Longitudinal collection of electronic health information about individual patients and populations

- **Diagnoses** I25.110
- **Drug prescriptions** A01AD05
- **Clinical tests**
- **More complex structures**
  - clinical notes
  - medical images
  - MRIs
  - ECGs
  - …







Decision making
Surveillance
Hospital Healthcare and Treatment Unit
Data Management and Knowledge Extraction
Healthcare Policy Makers
Model selection
ML algorithms
Integration platform
Healthcare information system
- WHO
- Public Health Agency
- National Board of Health and Welfare
ADE reporting
R&D
Hospital Treatment Unit
Reporting of suspicious
Adverse Drug Events (ADEs)
and suggestions for treatment

ICD10* codes

- 10th revision of the International Classification of Diseases and Related Health Problems
- a classification system that is used to record medical activity
- the system enables classification and quantification of diseases and other health-related issues

3 -7 characters long

* http://www.ahima.org/icd10

ICD10 codes: examples (total of 22 chapters)

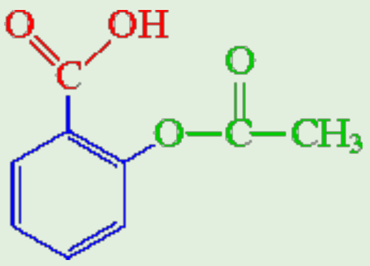
Chapter	Code Range	Estimated # of Codes	Description
1	A00-B99	1,056	Certain infectious and parasitic diseases
2	C00-D49	1,620	Neoplasms
3	D50-D89	238	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4	E00-E89	675	Endocrine, nutritional and metabolic diseases
5	F01-F99	724	Mental, Behavioral and Neurodevelopmental disorders
6	G00-G99	591	Diseases of the nervous system
7	H00-H59	2,452	Diseases of the eye and adnexa
8	H60-H95	642	Diseases of the ear and mastoid process
9	I00-I99	1,254	Diseases of the circulatory system
10	J00-J99	336	Diseases of the respiratory system
11	K00-K95	706	Diseases of the digestive system

ICD10 codes: examples

Code	Description
Combination Codes	
I25.110	Atherosclerotic heart disease of native coronary artery with unstable angina pectoris
Increased Specificity	
S72.044G	Non-displaced fracture of base of neck of right femur, subsequent encounter for closed fracture with delayed healing
Laterality	
C50.511	Malignant neoplasm of lower-outer quadrant of right female breast
C50.512	Malignant neoplasm of lower-outer quadrant of left female breast
"X" Placeholder	
H40.11X2	Primary open-angle glaucoma, moderate stage

ATC* codes

- Anatomical Therapeutic Chemical codes, first published in 1976
- Used for classification of active ingredients of drugs
- Based on the organ/system on which they act
  - therapeutic
  - pharmacological and chemical properties



acetylsalicylic acid (aspirin)

- Controlled by the World Health Organization Collaborating Centre (WHOCC) for drug statistics methodology

* http://www.whocc.no/atc_ddd_index/

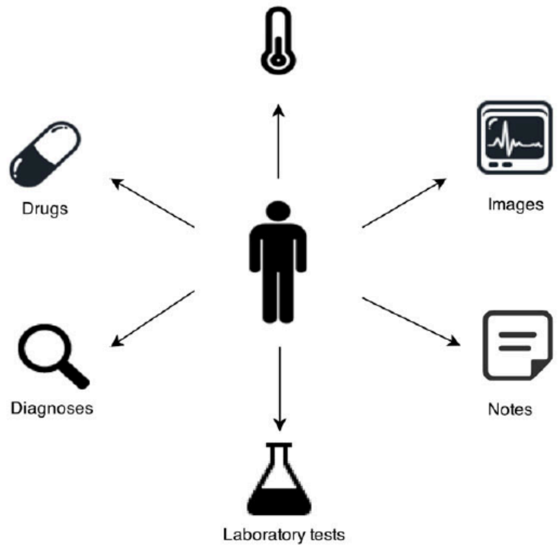
ATC codes – example: A10BA02

ATC codes classify drugs into five different levels

Level	Content	Type	Example
I	anatomical main group	1 letter	A: alimentary tract and metabolism
II	therapeutic subgroup	2 digits	A10: diabetes drugs
III	pharmacological subgroup	1 letter	A10B: blood glucose lowering drugs, excl. insulins
IV	chemical subgroup	1 letter	A10BA: biguanides
V	chemical substance	2 digits	A10BA02: metformin

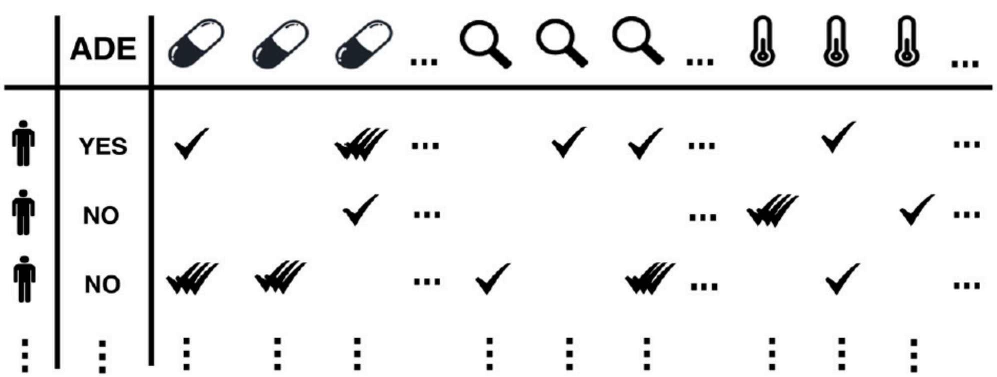
Extracting features from EHRs

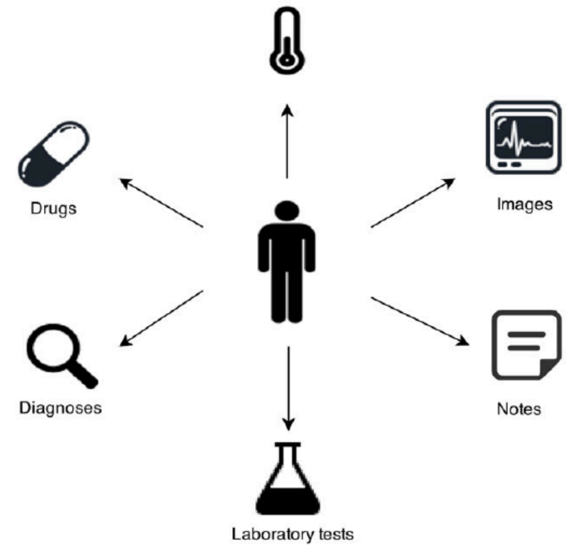
- **Mainly two lines of approaches:**
  - static features
  - temporal features



Static features

- **Mainly two lines of approaches:**
  - static features
  - temporal features



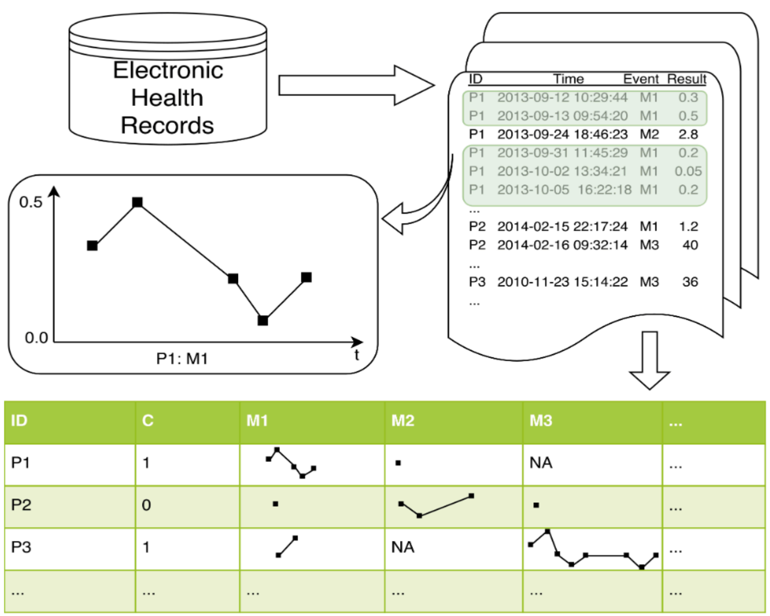


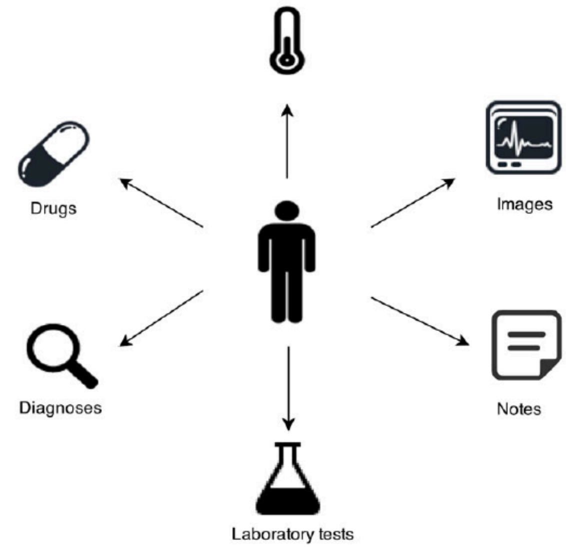
The class labels assigned depending on task at hand, e.g., ADE detection

- **Existing out-of-the-box classifiers are used**
  - **Decision trees, random forests, SVMs, deep learning architectures** [Chazard2011, Zhao2013, Karlsson2013, Shickel2018, Bampa2019]

Temporal features

- **Mainly two lines of approaches:**
  - static features
  - temporal features

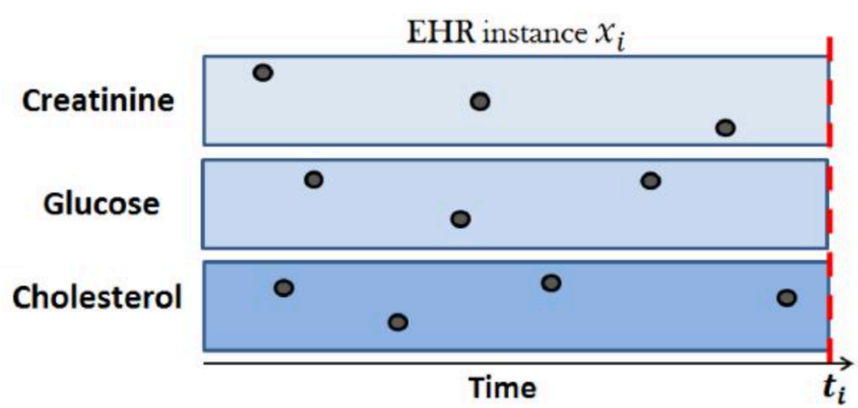




Clinical measurements:
- different units
- times of measurement
- sparsity

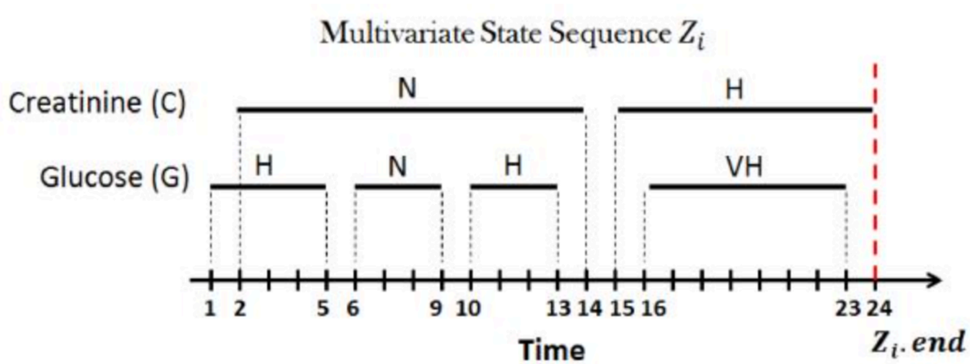
Temporal abstractions of EHRs

- Database: **EHRs of Patients**
- **Each EHR:**
  - Multiple temporal variables registered and evolving concurrently
  - Each variable with multiple readings until a critical time point t_i, e.g., glucose, creatinine, cholesterol
  - **Class label:** Disease/symptom detected at time t_i (event of interest)



Temporal abstractions of EHRs

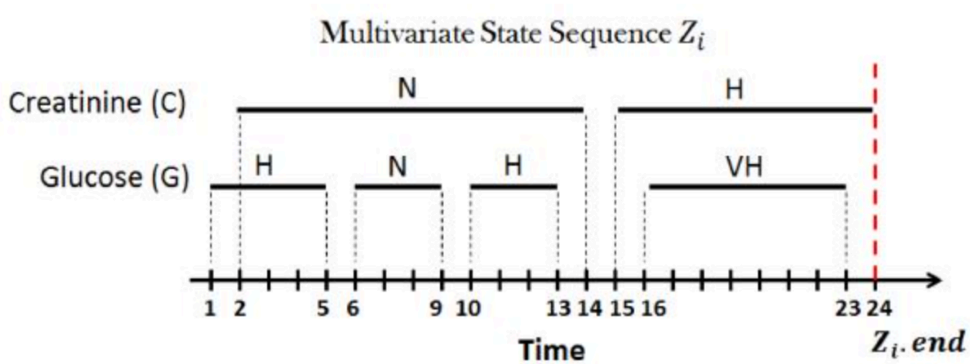
- Database: **EHRs of Patients**
- **Each EHR:**
  - Multiple temporal variables registered and evolving concurrently
  - Each variable with multiple readings until a critical time point t_i, e.g., glucose, creatinine, cholesterol
  - **Class label:** Disease/symptom detected at time t_i (event of interest)



Two types of temporal abstractions

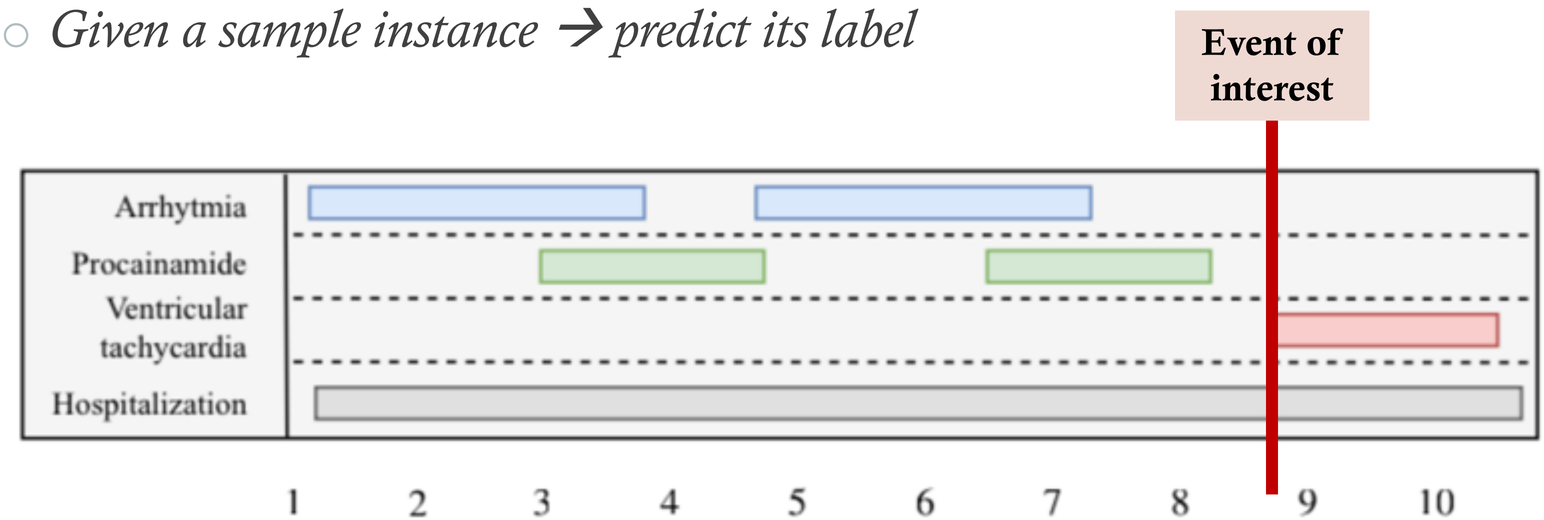
- **Trend abstraction:**
  - e.g., *decreasing*, *steady*, *increasing*
  - time series segmentation + identify slopes

- **Value abstraction**
  - e.g., *very low*, *low*, *normal*, *high*, *very high*
  - use *10^{th}*, *25^{th}*, *75^{th}*, and *90^{th}* percentiles on the lab values to define [Batal2012]



Temporal abstractions of EHRs

- **Supervised temporal prediction [Batal2012, Rebane2019]**
  - Given a labeled dataset of temporal instances up to time t_i
  - Find frequently occurring "temporal patterns" for each label
  - *Given a sample instance → predict its label*

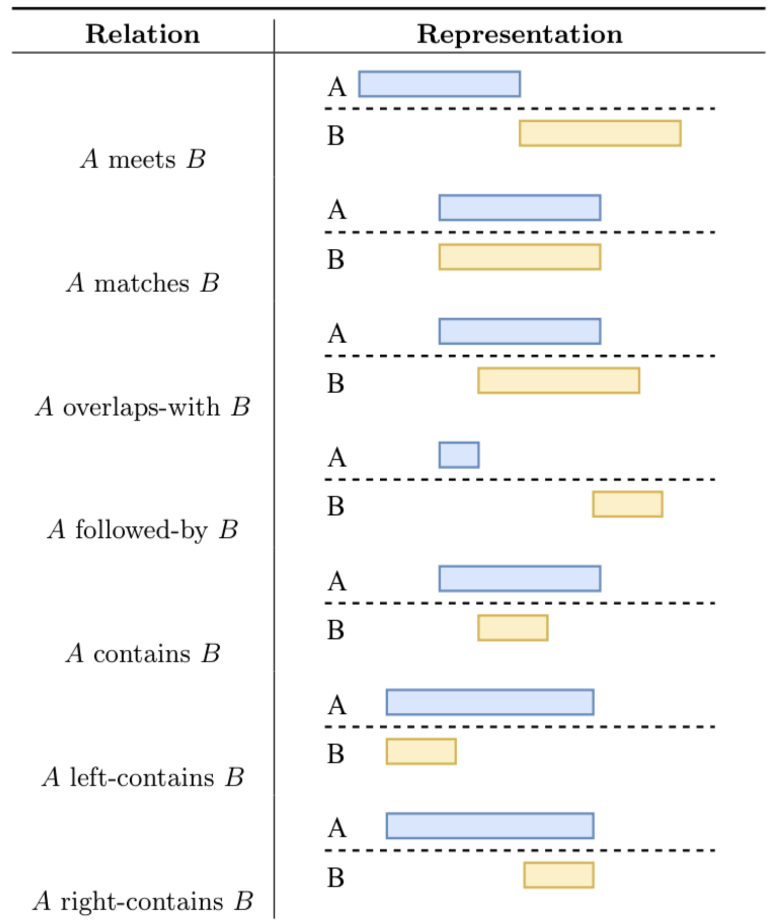


Temporal abstraction patterns

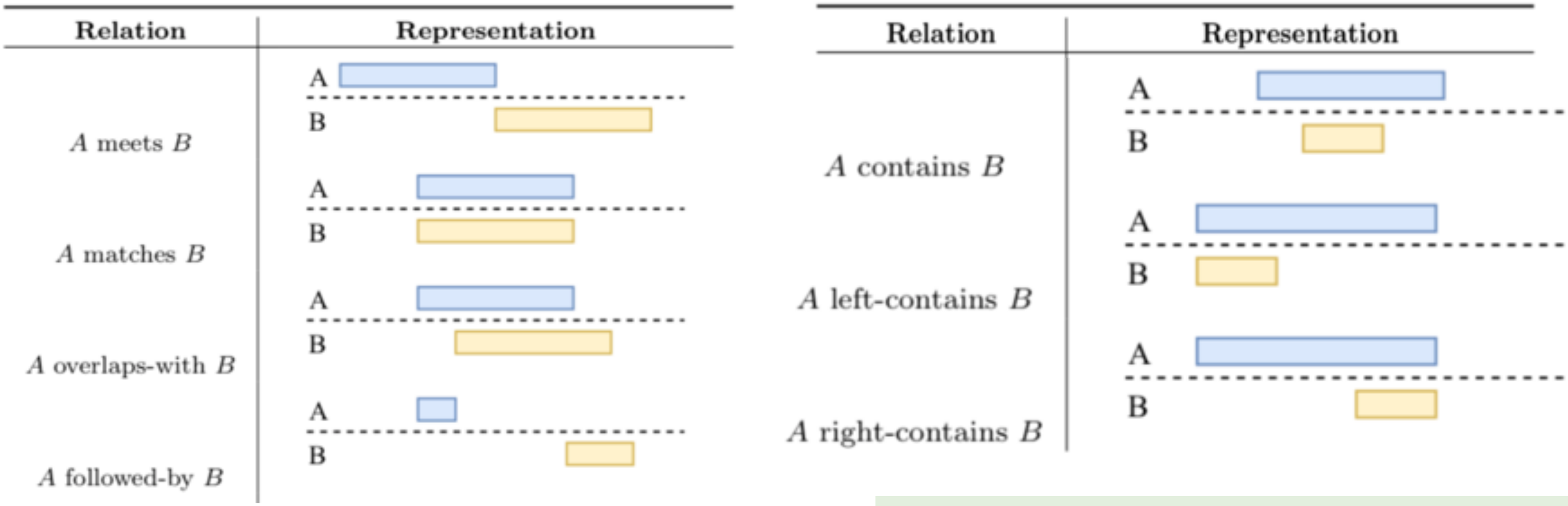
- **What is a temporal pattern?**
  - a sequence of ***"temporal relations"*** between state intervals A and B

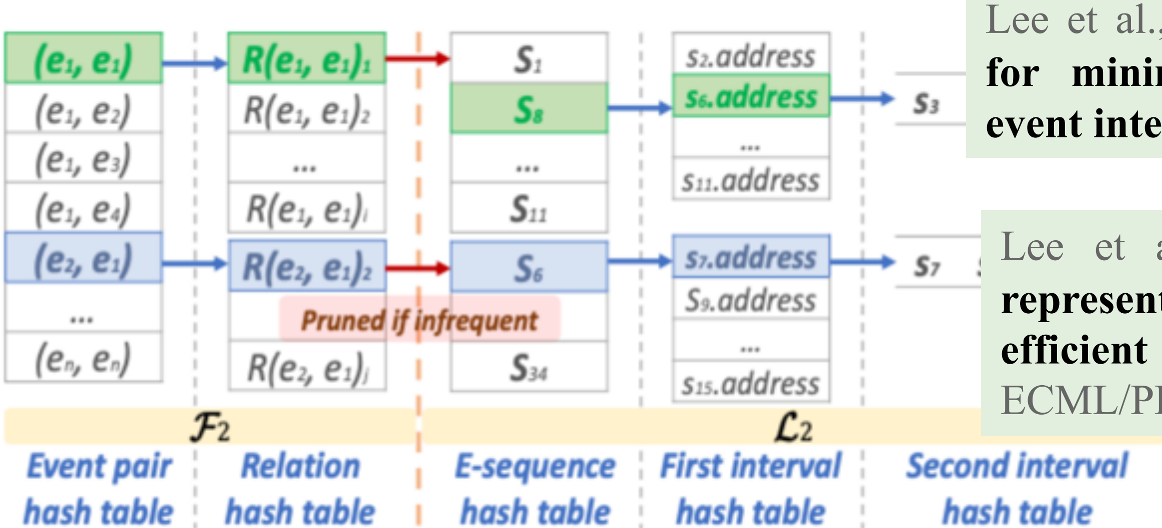
- **What kind of "temporal relations"?**

Based on Allen's temporal logic



Mining temporal abstraction patterns

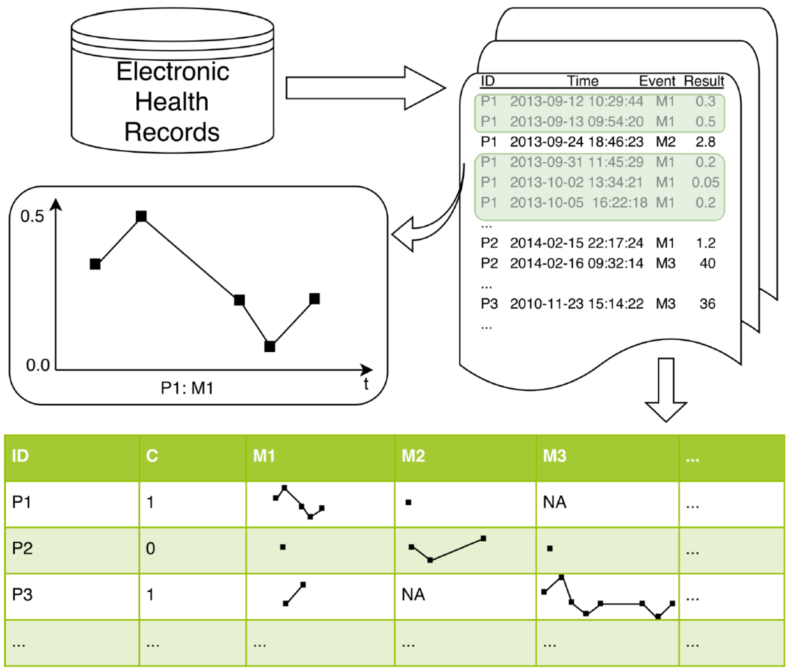




Lee et al., **Z-Miner: an efficient method for mining frequent arrangements of event intervals**, KDD 2020

Lee et al., **Z-Embedding: A spectral representation of event intervals for efficient clustering and classification**, ECML/PKDD 2020

Temporal abstractions of sparse EHRs



Hielscher et al. **Mining Longitudinal Epidemiological**

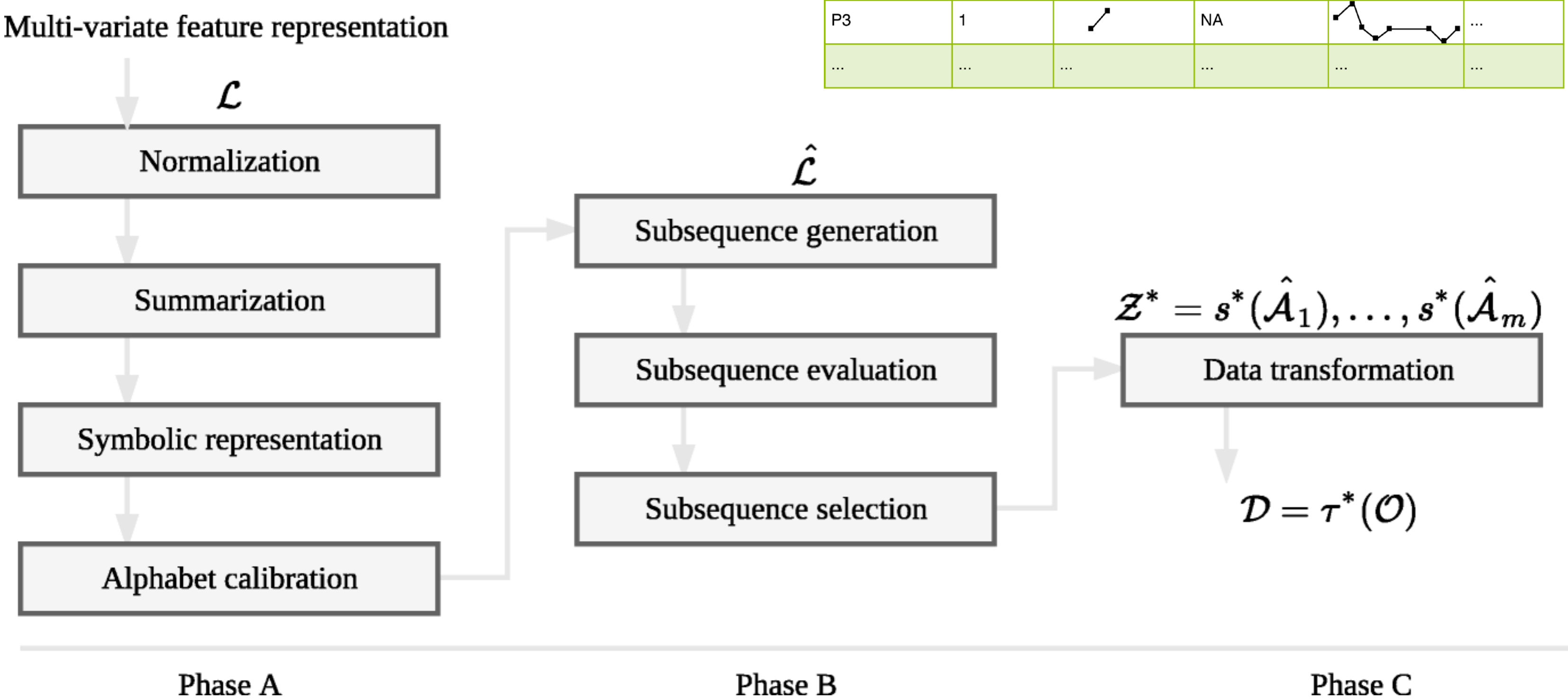
Zhao et al. **Learning from Heterogeneous Temporal**

Bagattini et al. **A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records**, BMC Medical Informatics and Decision Making, 2019

Rebane et al. **SMILE: a feature-based temporal abstraction framework for event-interval sequence classification**, Data Mining and Knowledge Discovery, accepted [pre-print online]

Framework overview

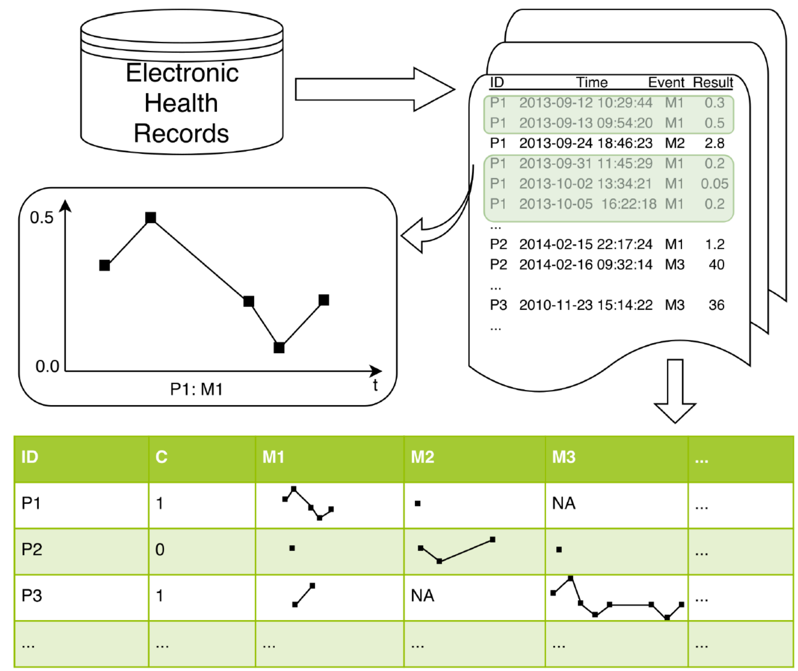
ID	C	M1	M2	M3	...
P1	1			NA	...
P2	0				...
P3	1		NA		...
...	...	...	...	...	...



Phase A: normalization

- Z-normalization
  - Each multi-variate feature S is z-normalized:

$$S := \frac{\sum_{i=1}^{|S|}\{s_i - \mu(S)\}}{\sigma(S)}$$



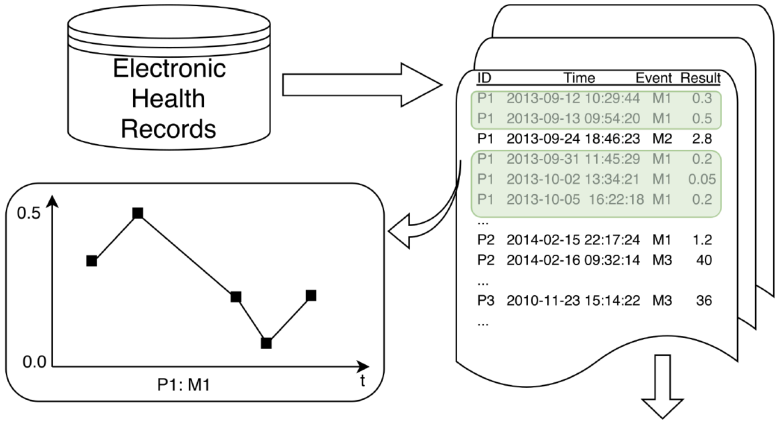
Phase A: summarization

- Z-normalization
  - Each multi-variate feature S is z-normalized:

$$S := \frac{\sum_{i=1}^{|S|}\{s_i - \mu(S)\}}{\sigma(S)}$$

- Summarization
  - Piecewise Aggregate Approximation (PAA)
  - Dimensionality reduction from d to w

$$\overline{S} = \{\overline{s}_1, \ldots, \overline{s}_w\}$$

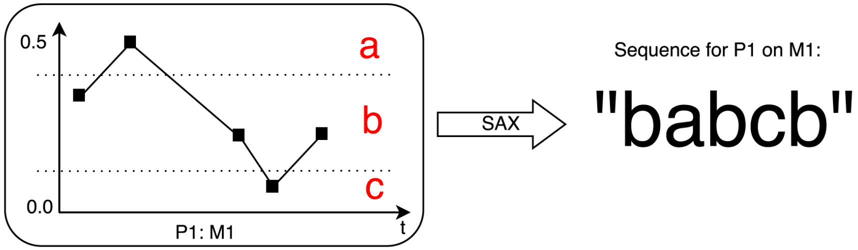


ID	C	M1	M2	M3	...
P1	1			NA	...
P2	0				...
P3	1		NA		...
...	...	...	...	...	...

$$\overline{s}_i = \frac{w}{d} \sum_{j=\frac{d}{w}(j-1)+1}^{\frac{d}{w}i} s_j$$

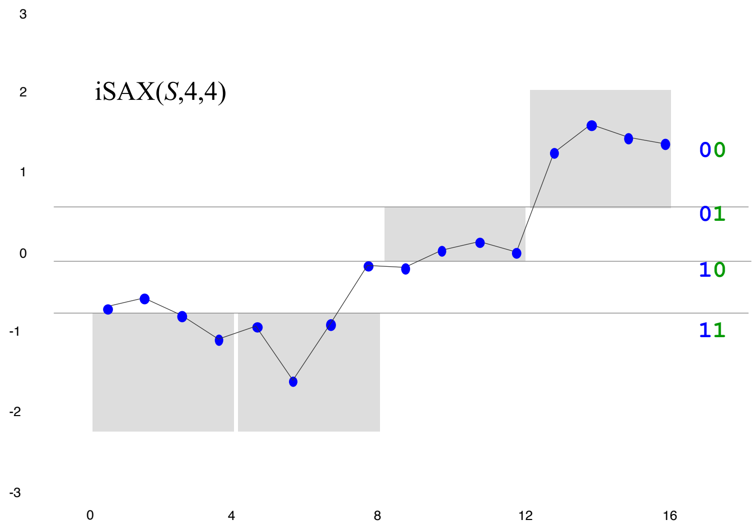
Phase A: symbolic representation

- SAX mapping
  - each record is mapped to a string using SAX
  - **length:** number of measurements
  - **alphabet:** 2 – 5, or set using domain knowledge

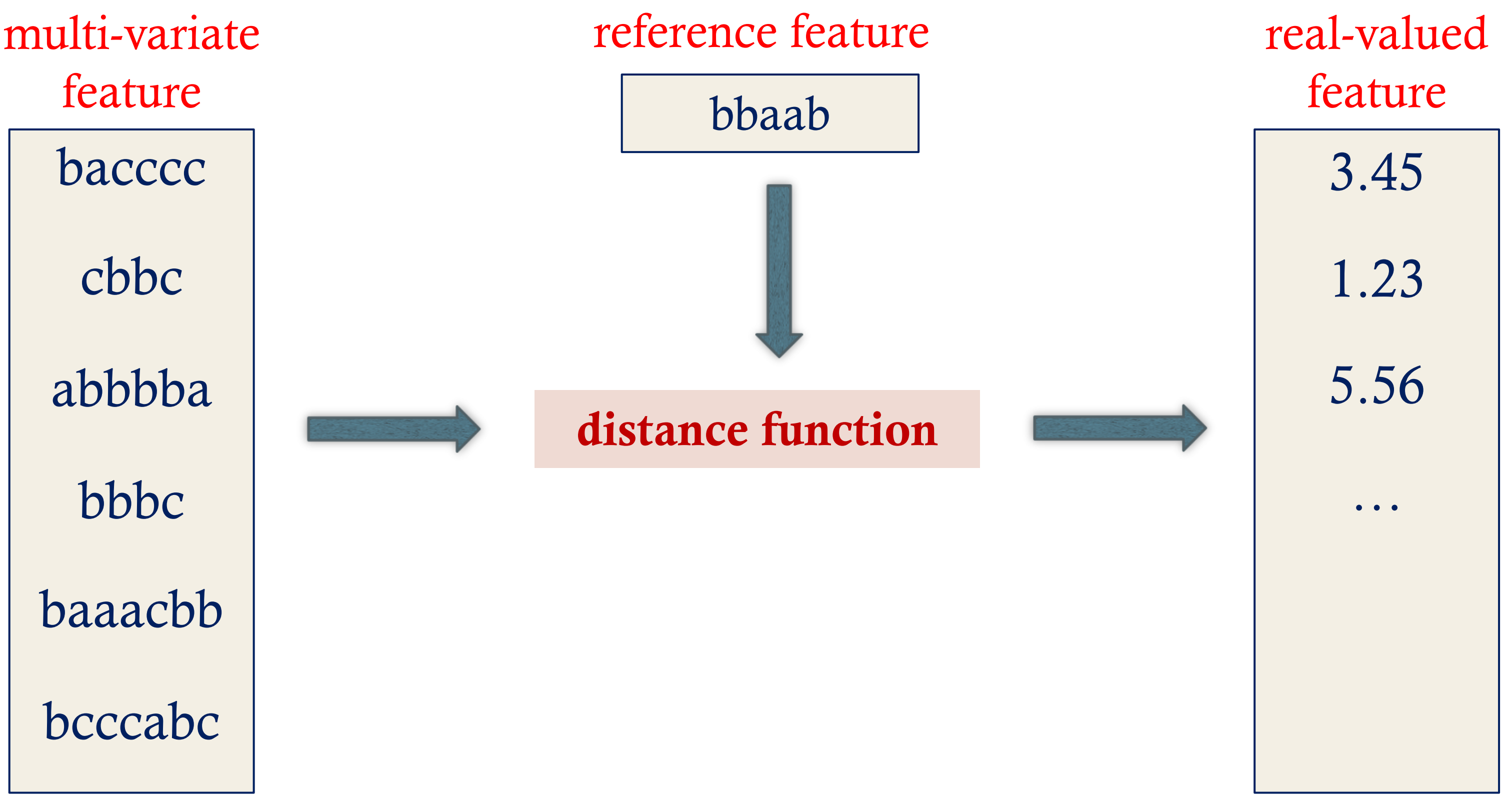


Discretize into a vector of symbols

- breakpoints map to small alphabet ***a*** of symbols



Phase B: mapping to real features



Phase B: subsequence enumeration

- *s-shapelet* generation:
  - random subsequences ***s*** of length t ∈ [1, l_{max}]

l_{max}: max length of a feature sequence

- *s-shapelet* evaluation:
  - each *s* is converted to a real value based on its distance to each multi-variate feature sequence

$$Dist\left(s, \widehat{S}\right) := \min_{s' \subseteq \widehat{S}, |s'| = |s|} \left\{D\left(s, s'\right)\right\}$$

Phase B: subsequence selection

- For each mutli-variate feature:
  - select the s-shapelet s^* with the max utility:

$$s^* := \arg\max_{s \in \widehat{\mathcal{L}}} Gain\left(s, \delta_{osp}(s), \widehat{\mathcal{L}}\right) \qquad s^*_\alpha := \arg\max_{s \in S_\alpha} Gain\left(s, \delta_{osp}(s), \widehat{\mathcal{L}}\right)$$

  - select the alphabet size with the max utility

$$\alpha^* := \arg\max_{\alpha \in \mathcal{I}} Gain\left(s^*_\alpha, \delta_{osp}(s^*_\alpha), \widehat{\mathcal{L}}\right)$$

- Final set of s-shapelets:

$$\mathcal{Z}^* = \left\{s^*\left(\hat{\mathcal{A}}_1\right), \ldots, s^*\left(\hat{\mathcal{A}}_m\right)\right\}$$

Phase C: transformation

- A function τ* is learned:
  - transform any data object of the original multi-variate space to a set of real-valued features

$$\tau^* : \mathcal{A} \rightarrow \mathbb{R}^m$$

- Each data example is transformed using τ*:

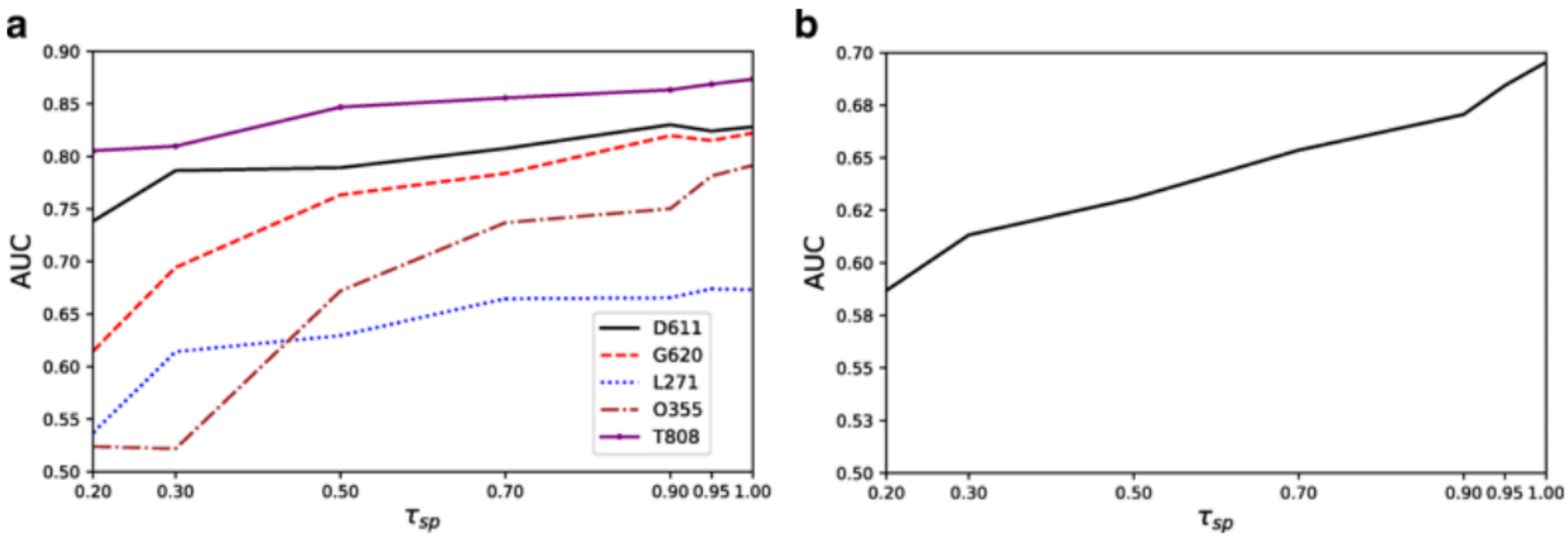
$$\tilde{O} = \tau^*(O)$$

Dataset	Class label description	Pos.	Neg.	Feat.
D611	Drug-induced aplastic anaemia	593	105	285
D642	Drug-induced secondary sideroblastic anaemia	217	9673	513
D695	Secondary thrombocytopenia	1246	2148	450
E273	Drug-induced adrenocortical insufficiency	70	259	229
G620	Drug-induced polyneuropathy	96	783	258
I952	Drug-induced hypotension	115	1287	324
L270	Drug-induced generalized skin eruption	182	468	314
L271	Drug-induced localized skin eruption	151	498	311
M804	Drug-induced osteoporosis with pathological fracture	52	1170	282
M814	Drug-induced osteoporosis	57	5097	434
O355	Maternal care for (suspected) damage to fetus by drugs	146	260	148
R502	Drug-induced fever	80	6434	498
T782	Adverse effects: anaphylactic shock	131	856	293
T783	Adverse effects: angioneurotic oedema	283	720	293
T784	Adverse effects: allergy	574	415	294
T801	Vascular complications following infusion, transfusion and therapeutic injection	66	609	229
T808	Other complications following infusion, transfusion and therapeutic injection	538	138	229
T886	Drug-induced anaphylactic shock	89	1506	363
T887	Unspecified adverse effect of drug or medicament	1047	550	363

Dataset	Class label description	Pos.	Neg.	Feat.
D611	Drug-induced aplastic anaemia	593	105	285
D642	Drug-induced secondary sideroblastic anaemia	217	9673	513
D695	Secondary thrombocytopenia	1246	2148	450
E273	Drug-induced adrenocortical insufficiency	70	259	229
G620	Drug-induced polyneuropathy	96	783	258
I952	Drug-induced hypotension	115	1287	324
L270	Drug-induced generalized skin eruption	182	468	314
T808	Other complications following infusion, transfusion and therapeutic injection	538	138	229
T886	Drug-induced anaphylactic shock	89	1506	363
T887	Unspecified adverse effect of drug or medicament	1047	550	363

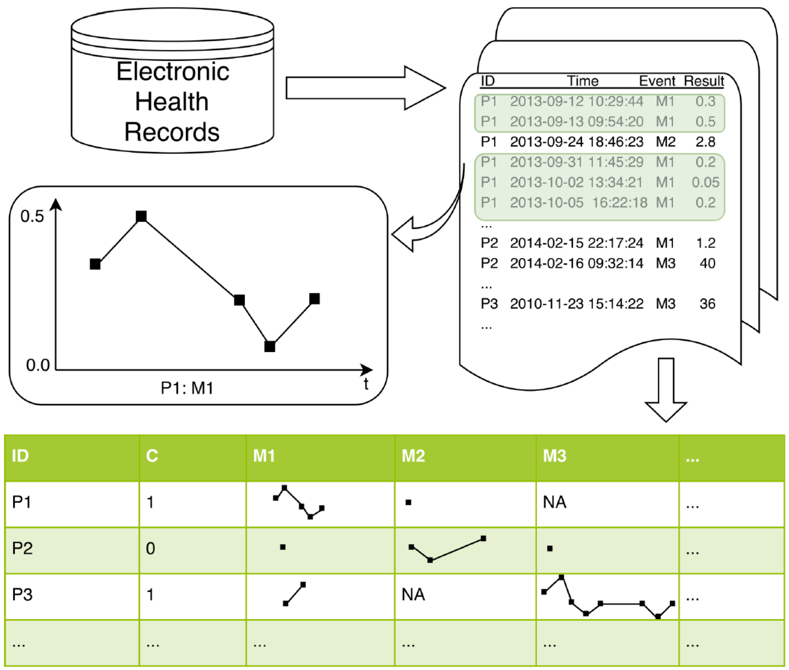
- Adverse Drug Events (ADEs) are injuries that occur from the use of a drug, such as overdoses or dose reductions, or drug interactions
- They account for 3.7% of hospital admissions around the world
- ADEs have been estimated to come at a cost of $3.5 billion/year in the U.S alone, despite ADEs being preventable

lr: AUC vs feature sparsity



- As the % of **feature sparsity** increases, **AUC** also increases!
- Shorter s-shapelets (i.e., 2-8) are preferable to longer ones (> 20)

Temporal abstractions of sparse EHRs



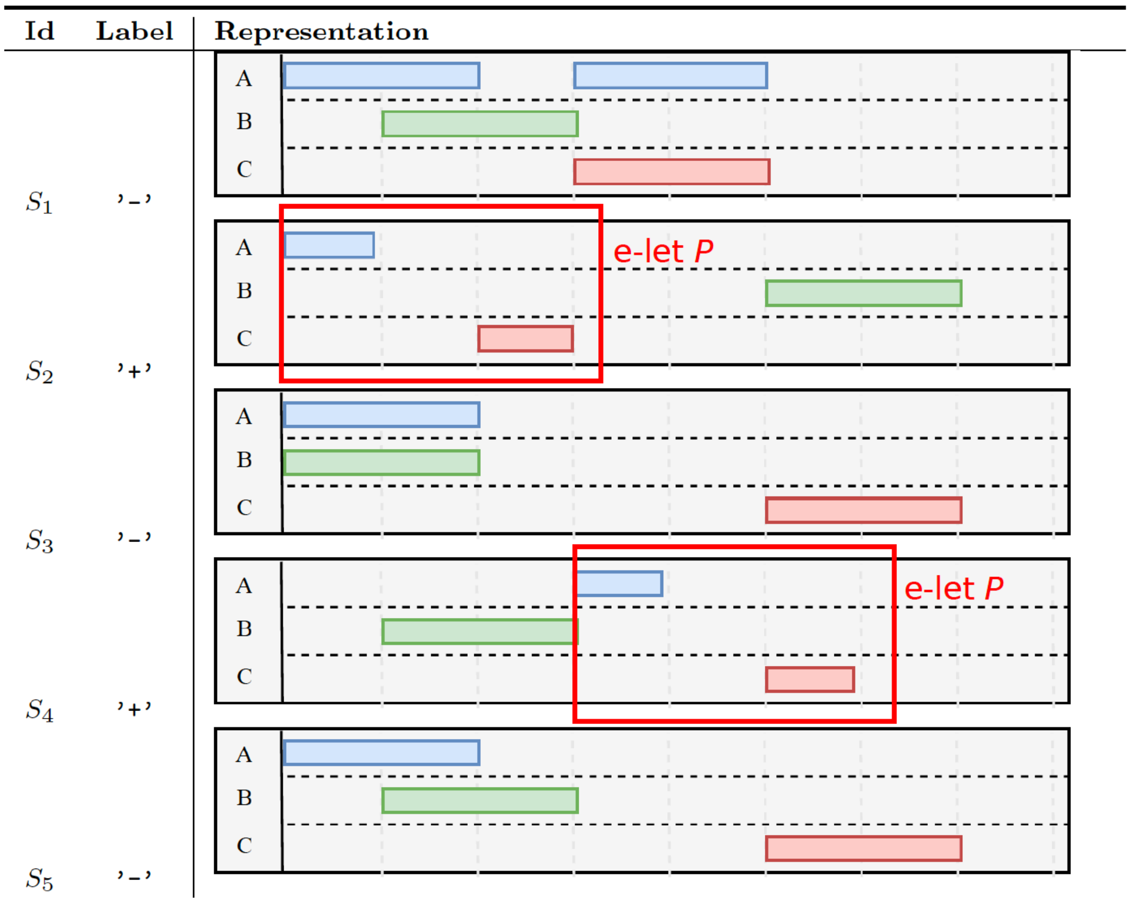
Hielscher et al. **Mining Longitudinal Epidemiological**

Zhao et al. **Learning from Heterogeneous Temporal**

Bagattini et al. **A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records**, BMC Medical Informatics and Decision Making, 2019

Rebane et al. **SMILE: a feature-based temporal abstraction framework for event-interval sequence classification**, Data Mining and Knowledge Discovery, accepted [pre-print online]

The notion of e-lets



Class-distinctive Temporal abstraction subsequences

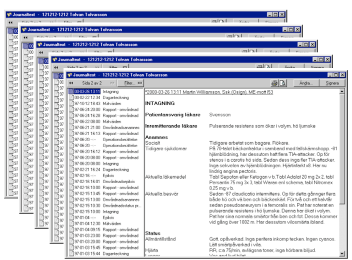
SMILE: a formulation that uses e-lets as classification features!

Dataset	**Accuracy**				**Area under ROC**			
	RF	LR	DT	SVM	RF	LR	DT	SVM
AUSLAN2	**0.485**	0.335	0.465	0.305	0.686	0.668	0.674	**0.696**
BLOCKS	**1.000**	0.995	0.995	0.986	**1.000**	0.994	0.975	**1.000**
CONTEXT	0.988	0.979	0.963	**0.996**	**1.000**	0.999	0.969	0.997
HEPATITIS	**0.831**	0.735	0.769	0.823	**0.890**	0.789	0.788	0.813
PIONEER	**0.988**	0.981	0.950	0.969	**1.000**	**1.000**	0.894	0.969
SKATING	**0.977**	0.970	0.972	0.847	**0.999**	**0.999**	0.973	0.966
D611	**0.945**	0.874	0.921	0.929	**0.878**	0.744	0.628	0.523
D642	**0.988**	0.976	0.981	0.976	**0.991**	0.943	0.897	0.793
D695	**0.745**	0.726	0.688	0.729	**0.813**	0.793	0.659	0.726
E273	0.586	**0.619**	0.581	0.607	**0.690**	0.628	0.580	0.517
G620	**0.854**	0.750	**0.854**	0.816	**0.902**	0.661	0.767	0.500
I952	**0.876**	0.751	0.826	**0.876**	**0.670**	0.543	0.568	0.500
L270	**0.672**	0.611	0.605	0.561	**0.734**	0.637	0.611	0.509
L271	**0.746**	0.647	0.709	0.692	**0.800**	0.611	0.624	0.521
O355	**0.875**	0.713	0.839	0.826	**0.944**	0.773	0.826	0.853
R502	**0.977**	0.964	0.975	**0.977**	**0.816**	0.591	0.507	0.500
T801	**0.917**	0.892	0.896	0.913	**0.819**	0.623	0.716	0.501
T782	0.750	**0.763**	0.727	0.723	0.508	**0.628**	0.575	0.500
T783	**0.857**	0.686	0.732	0.796	**0.750**	0.568	0.642	0.558
T886	0.771	0.788	**0.838**	0.786	**0.834**	0.759	0.696	0.502
Avg.	**0.841**	0.788	0.814	0.807	**0.836**	0.748	0.728	0.672

Interpretable and actionable models

- Trade-offs between interpretability + accur
- Ability to understand the irable outcomes without compromising p

I can tell you **what changes need to be done** to the patient record, so that I can **change my prediction** ☺



black box classifier

The patient will die from HF in **2 days**!

Now what? Please tell me **why**?

ADE prediction (Yes/No)

- **Main task:**

  predict the presence or absence of an ADE in a patient's next visit given EHR data entries from all previous visits!

Adverse drug event ICD-10 codes	
D611	Drug-induced aplastic anaemia
D642	Drug-induced secondary sideroblastic anaemia
D695	Secondary thrombocytopenia
E273	Drug-induced adrenocortical insufficiency
G620	Drug-induced polyneuropathy
I952	Drug-induced hypotension
L270	Drug-induced generalized skin eruption
L271	Drug-induced localized skin eruption
M804	Drug-induced osteoporosis with pathological fracture
M814	Drug-induced osteoporosis
O355	Maternal care for (suspected) damage to fetus by drugs
R502	Drug-induced fever
T782	Adverse effects: anaphylactic shock
T783	Adverse effects: angioneurotic edema
T784	Adverse effects: allergy
T801	Vascular complications following infusion, transfusion and therapeutic injection
T886	Drug-induced anaphylactic shock

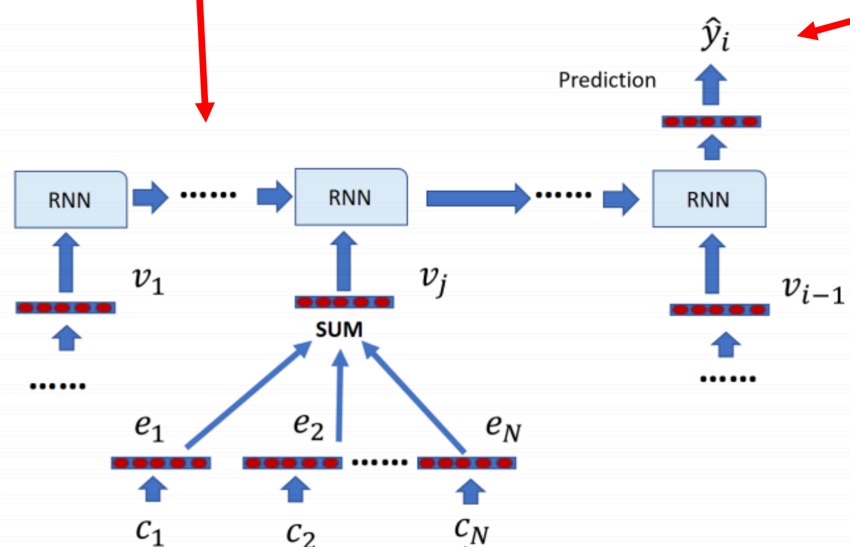
Main goals

- Empirically evaluate which code-level interpretable deep learning architecture provides the best performance for ADE prediction
- Examine which data sources (diagnoses, medications, lab tests) best aid in ADE predictive performance and medical interpretability
- Determine the extent in which code-level **attention mechanisms** contribute to interpretability for ADE predictions

Methods (Vanilla RNN)

Pass info from one visit to the next within network

Predict ADE yes/no in the future



Input: medical codes *c* for each medical visit *v* to train the network across patients

Limitations of Vanilla RNN

- Standard seq2seq models are normally composed of an encoder-decoder architecture

- **Encoder:** processes the input sequence and summarizes the information into a context vector of fixed length

- This representation is expected to be a good summary of the entire input sequence

- **Decoder:** initialized with the context vector and uses it to generate the transformed output

Limitations of Vanilla RNN

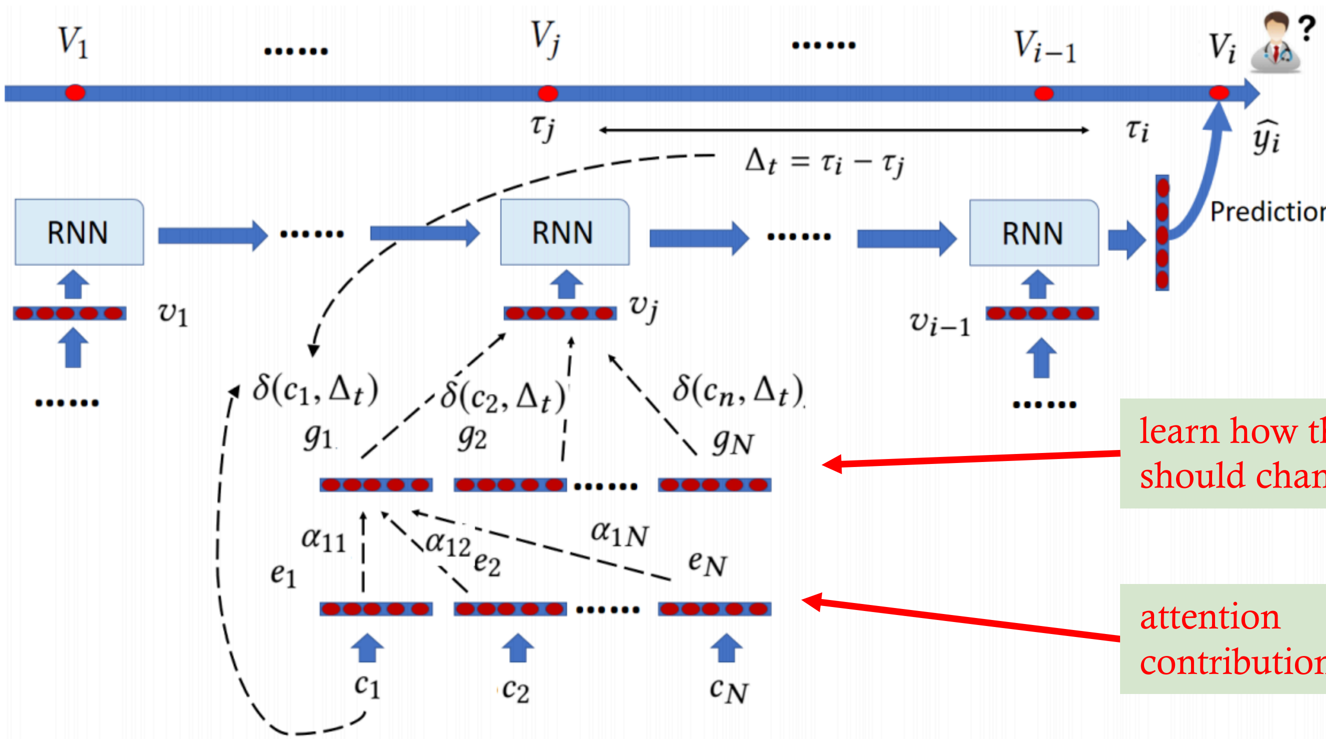
- Structrural limitation:
  - fixed-length context vector
- Why?
  - inability of remembering longer sequences
  - earlier parts of the sequence are forgotten once the entire sequence is processed
- The attention mechanism concept was born to resolve this problem
- **Attention mechanism:** keep the intermediate encoder states and utilize all of them in order to construct the context vectors required by the decoder to generate the output sequence

Medical ”attention”

- We may want the decoder to focus more on, e.g., visits 1 and 3, while paying less attention to the remaining visits of the patient

- Solution:
  - Train a feed forward neural network
  - learn to identify relevant encoder states
  - generate a high score for the visits for which attention is to be paid while low score for the visits which are to be ignored

Methods (Timeline, Bai 2018)



learn how the code contributions should change over time

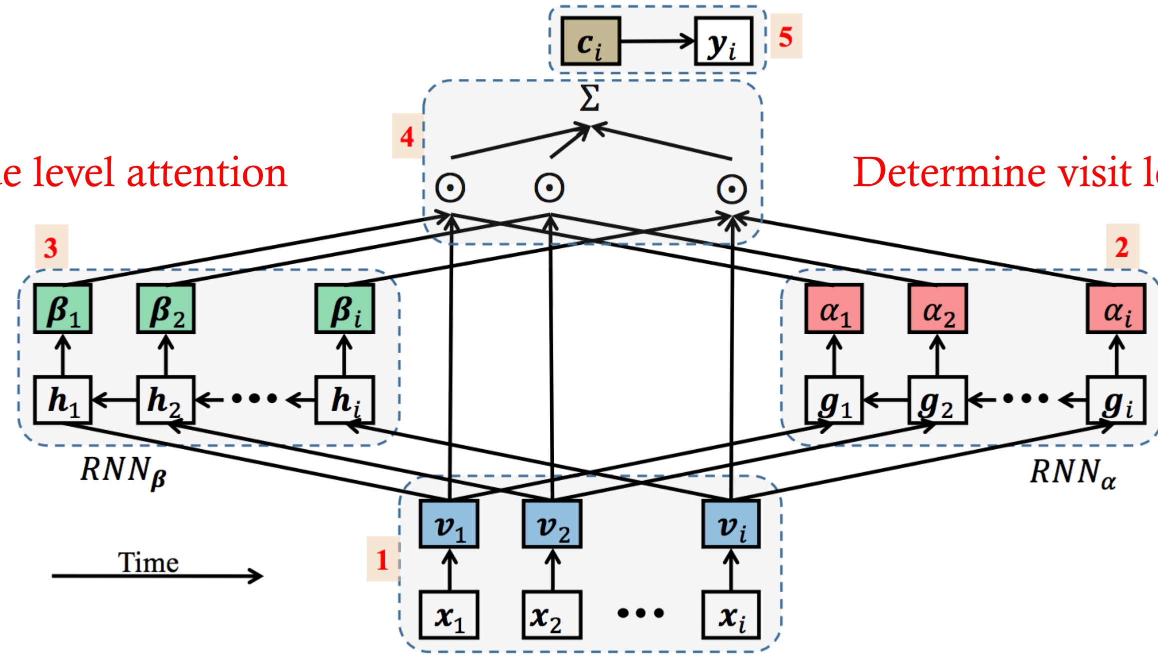
attention values for code contribution

$$\phi(c_x) = \sum_{j=1}^{N} \delta(c_j, \Delta_t)\alpha_{j,x}$$

Methods (RETAIN, Choi 2016)

Determine code level attention

Determine visit level attention



$$\omega(y_i, x_{j,k}) = \alpha_j W(\beta_j \odot W_{emb}[:, k])x_{j,k}$$

Experimental Setup

- RETAIN and Timeline:
  - proven to be competitive state-of-the-art architectures which permit thorough interpretability down to the code-level
  - trained for ADE prediction using an original data source consisting of information for 1,4 million patients obtained from HealthBank at Stockholm University

Experimental Setup

- Non-ADE ICD-10 and ATC codes were reduced to higher level hierarchical categories by selecting only the first three characters
- Such categories correspond to main categories of ICD-10 codes and to therapeutic subgroups in the case of ATC codes
- # of ICD-10 categories: 1692
- # of ATC subgroups: 109
- Visits defined on a monthly basis
- Patients also needed at least three such visits to be included
- Two data sets: including or excluding medication data

Results: AUC / F1

Dataset	Area under ROC		Micro F1-Score	
	RETAIN	Timeline	RETAIN	Timeline
Without medication	**0.765**	0.668	**0.789**	0.699
With medication	0.759	0.693	0.775	0.754

RETAIN was determined to be the best performing architecture under the conditions of using diagnoses data

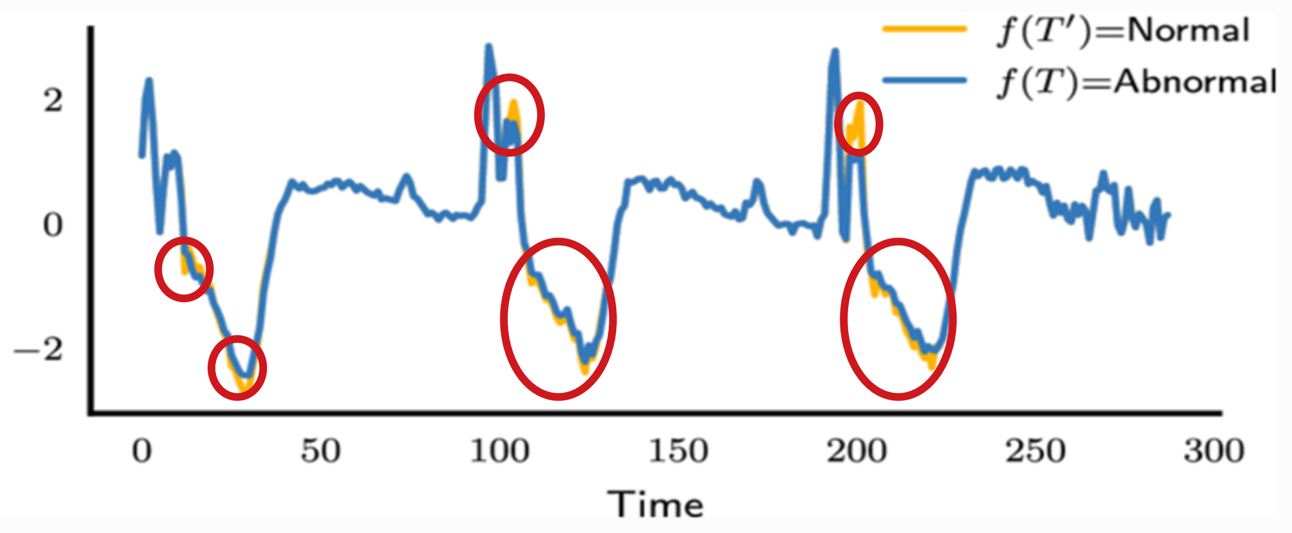
Interpretability of RETAIN

Code	Description	Score
	Visit 1	
L50	Urticaria	0.214
	Visit 2	
R42	Dizziness and giddiness	0.034
A02†	Drugs for acid related disorders	0.049
	Visit 3	
L50	Urticaria	0.239
R06†	Antihistamines for systemic use	0.344
H02†	Corticosteroids for systemic use	0.321
	Visit 4	
L50	Urticaria	0.225
R06†	Antihistamines for systemic use	0.322
C01†	Cardiac therapy	0.205
H02†	Corticosteroids for systemic use	0.230
	Prediction	
T784	Adverse effects: allergy	0.891

Rebane et al. **Exploiting Complex Medical Data with Interpretable Deep Learning for Adverse Drug Event Prediction**. Journal of AI in Medicine

Very high risk from given history

Time series counterfactuals

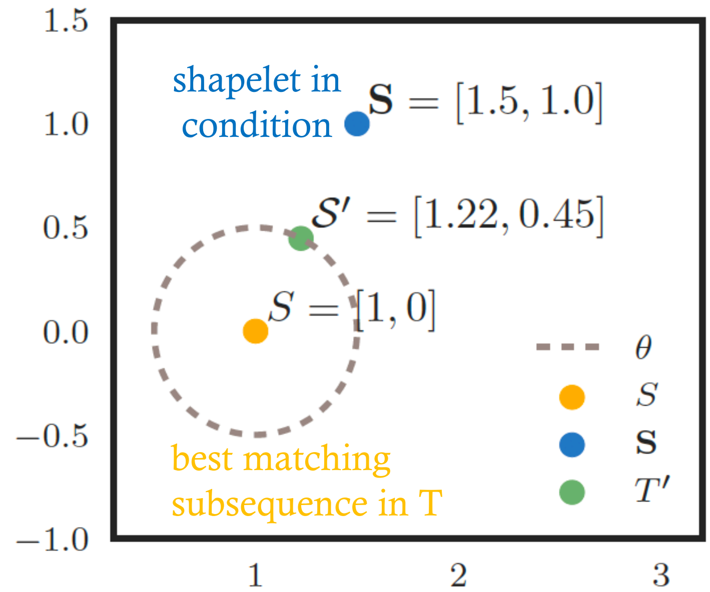


Problem: What is the min # of changes to the abnormal time series to convert it to normal ?

Karlsson et al. **Locally and globally explainable time series tweaking**. In Knowledge and Information Systems 2020

Solution outline

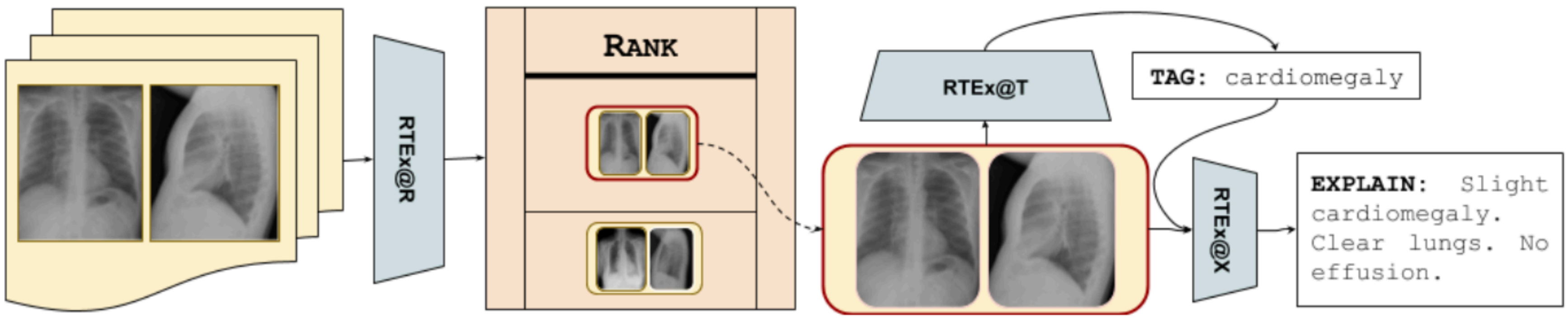
- Consider S as an m-dimensional point
- Define an m-sphere with S as its center and radius θ

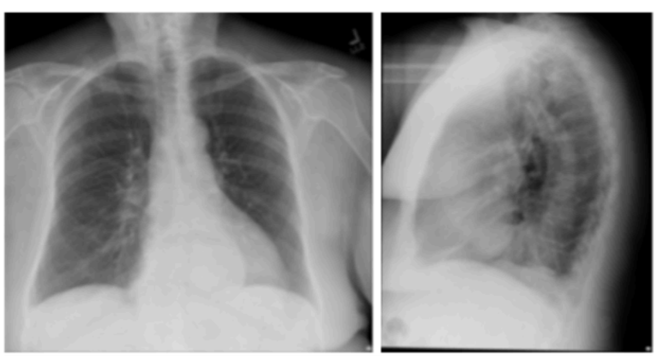


- The transformed time series counterpart of **S** is given by the following equation:

$$\tau_{\mathcal{S}}(\mathbf{S}, p_{ik}^j, \epsilon) = \mathcal{S}_k^j + \frac{\mathcal{S}_k^j - \mathbf{S}}{\|\mathcal{S}_k^j - \mathbf{S}\|_2}(\theta_k^j + (\epsilon\delta_{ik}^j))$$

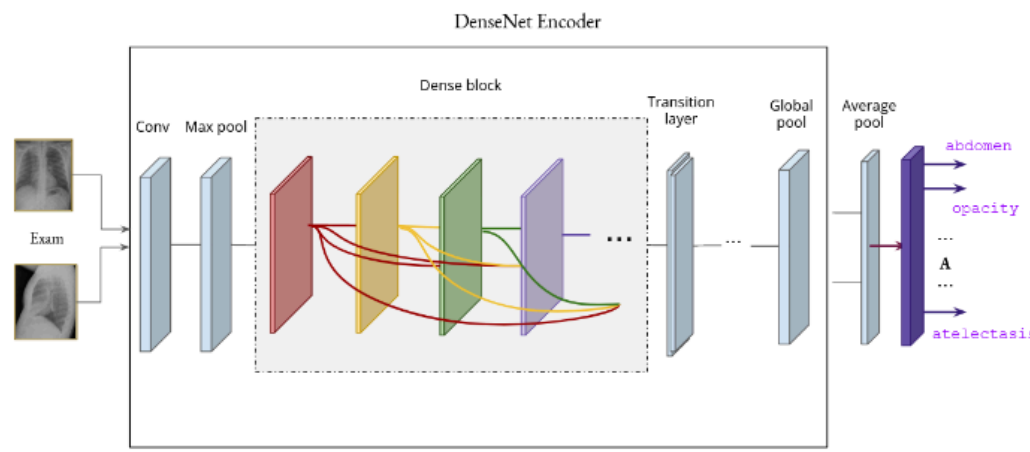
Automated ranking/tagging/captioning of radiography exams



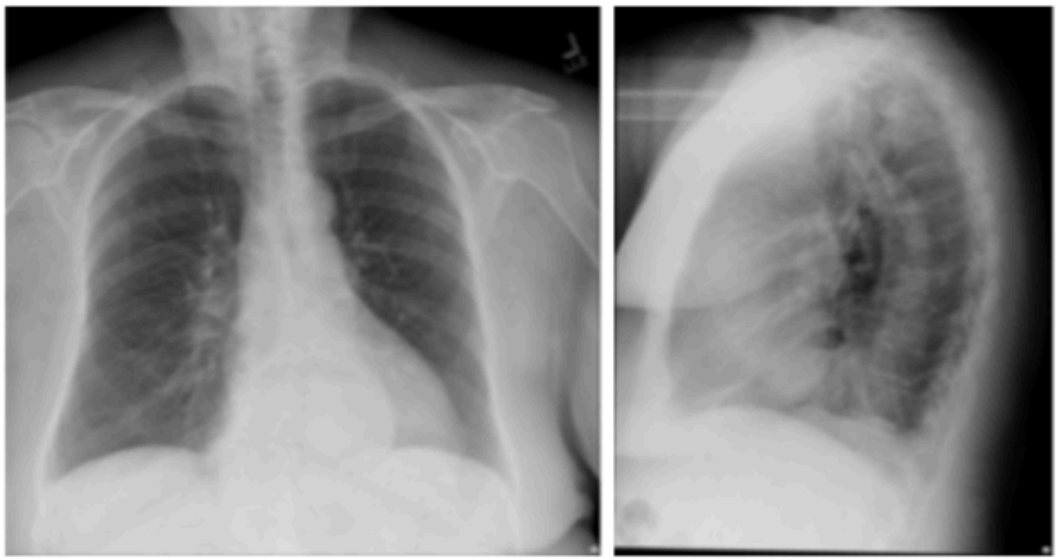


Diagnostic text: The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

Abnormality tags: Atelectases, Cervical Arthritis, Atelectasis, Spondylarthritis, Thoracic Spondylosis.

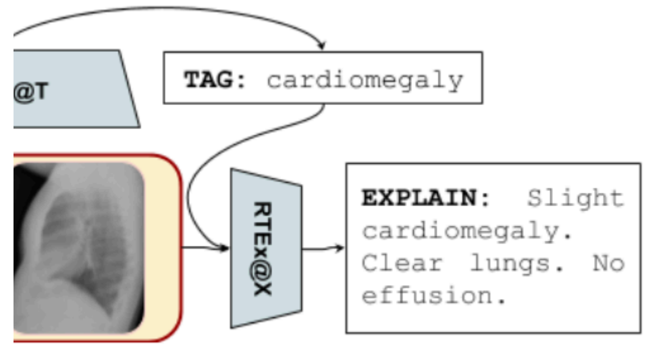


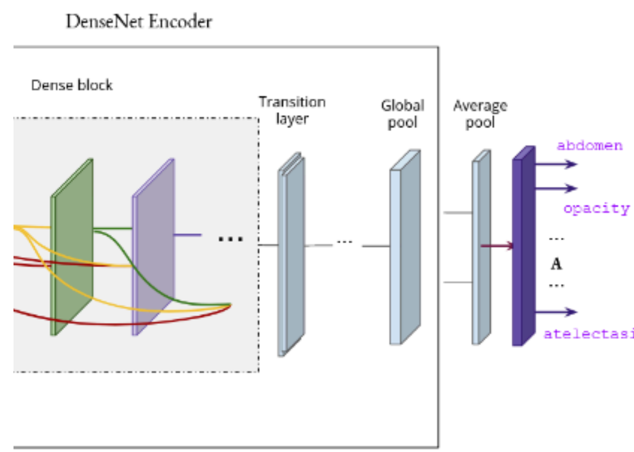
Automated ranking/tagging/captioning of radiography exams



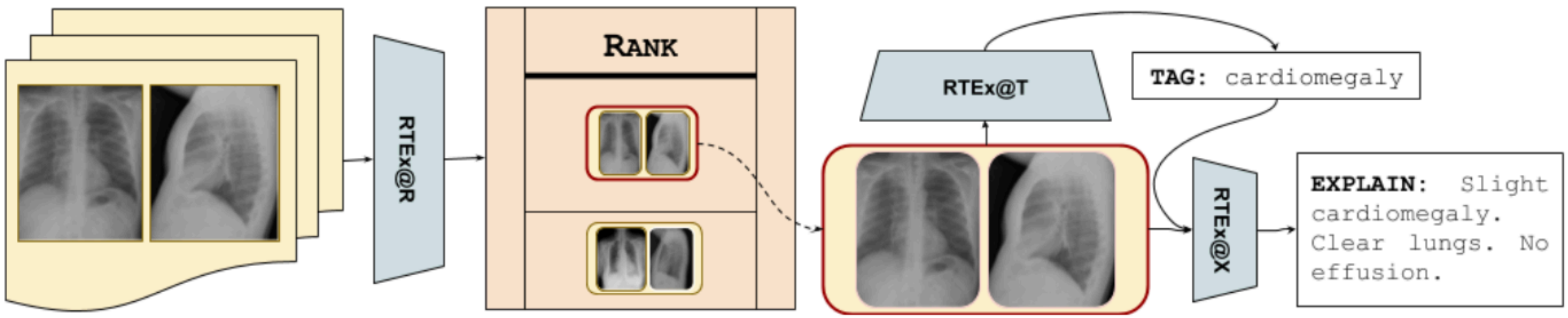
Diagnostic text: The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

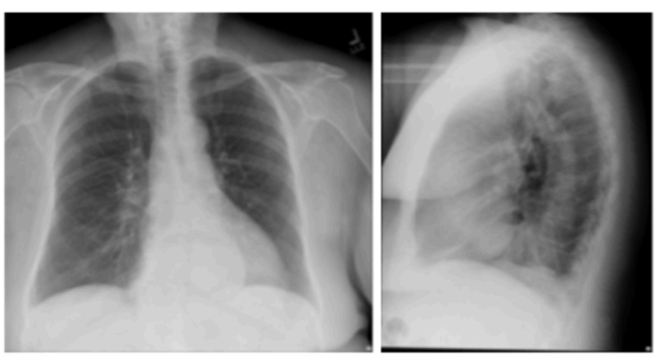
Abnormality tags: Atelectases, Cervical Arthritis, Atelectasis, Spondylarthritis, Thoracic Spondylosis.





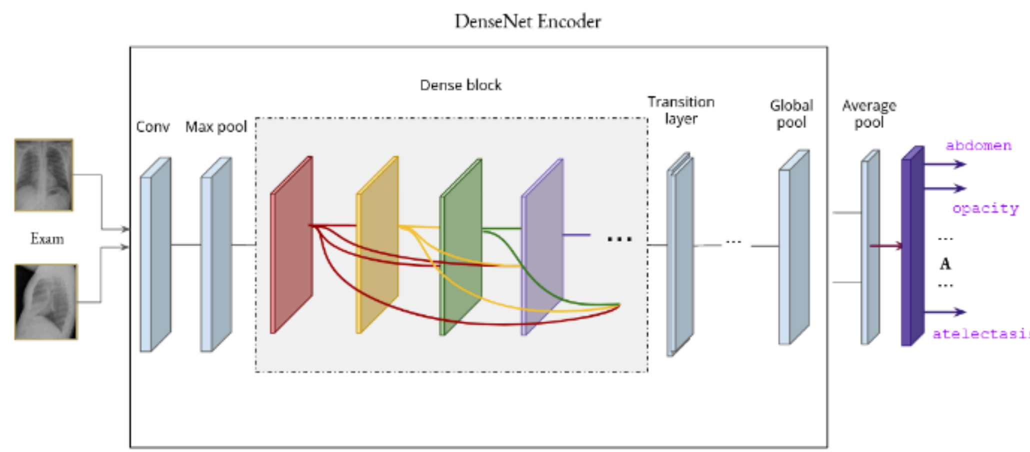
Automated ranking/tagging/captioning of radiography exams



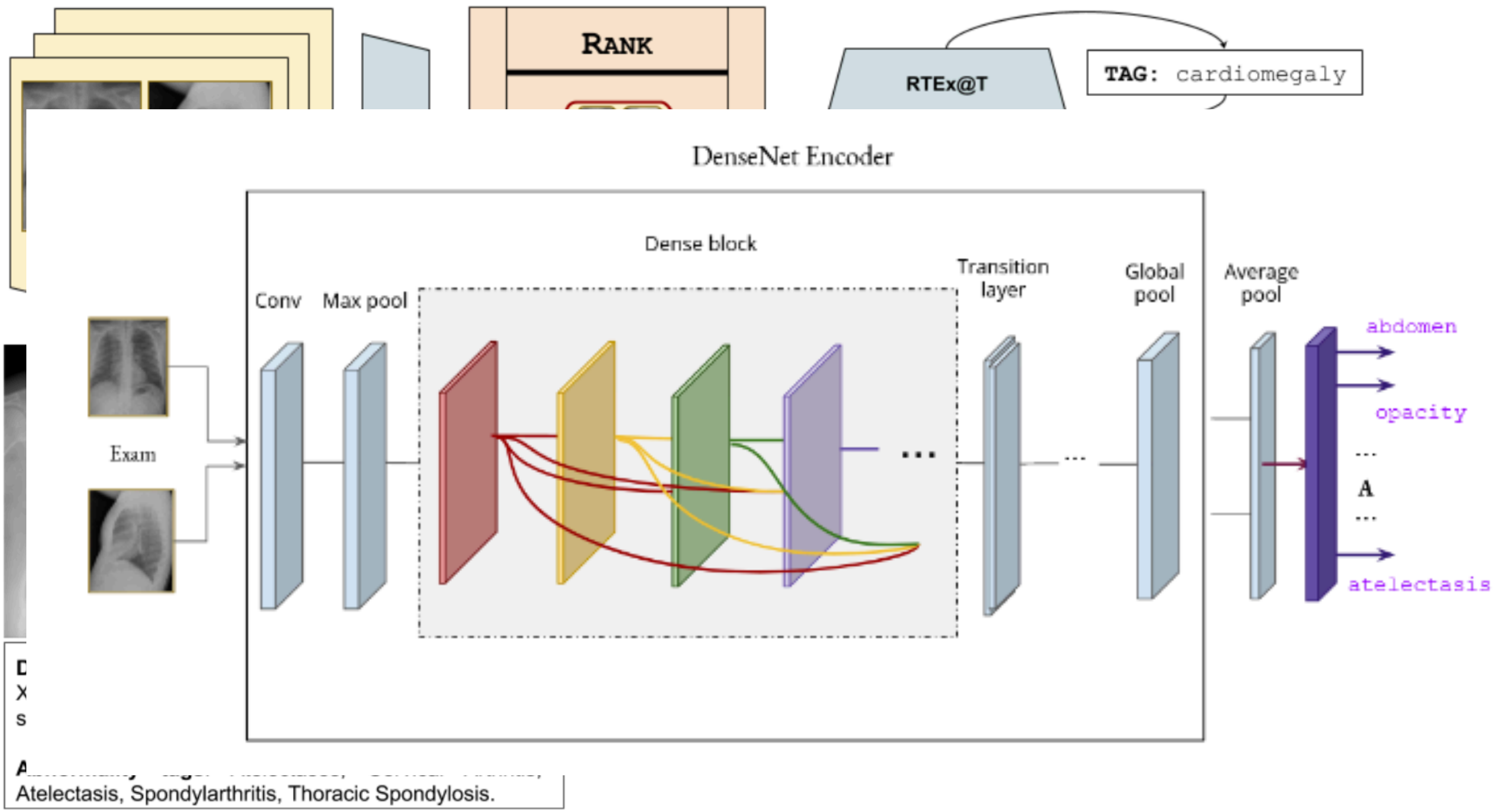


Diagnostic text: The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

Abnormality tags: Atelectases, Cervical Arthritis, Atelectasis, Spondylarthritis, Thoracic Spondylosis.



Automated ranking/tagging/captioning of radiography exams



Closing remarks

- Security and privacy issues
- Hard to convince public authorities to make more data available to researchers
- Hard to convince doctors to adopt new “black box” models
- Cloud solutions are in many cases unacceptable
- Need to **federated learning** solutions
- Many players/systems are used by practitioners
- Need for a unified cross-border database of medical records

Thanks to…

Isak Karlsson

Jon Rebane

Maria Bampa

Aristides Gionis

Hans E. Persson

Henrik Boström

Hercules Dalianis

We are hiring!

- **Postdoc position:** two years full-time
- **Deadline:** January 15 2021
- Apply here:

https://www.su.se/english/about-the-university/work-at-su/available-jobs?rmpage=job&rmjob=13760&rmlang=UK

- More information about **our group:**

https://datascience.dsv.su.se

Bibliography

- Gunter TD, Terry NP, The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions. J Med Internet Res 2005;7(1):e3

- H.Dalianis, M. Duneld and S. Velupilai, The Stockholm EPR Corpus – Characteristics and Some Initial Findings, Swedish Language Technology Conference, (SLTC), 2009

- Skeppstedt, M., Kvist, M., Dalianis, H., Nilsson, G.H. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study. *Journal of Biomedical Informatics*, 2014

- Iyad Batal, Dmitriy Fradkin, James Harrison, Fabian Moerchen, and Milos Hauskrecht. Mining recent temporal patterns for event detection in multivariate time series data. In ACM SIGKDD, 2012

- Henriksson, A., Kvist, M., Hassel, M., Dalianis, H. Exploration of Adverse Drug Reactions in Semantic Vector Space Models of Clinical Text. In Proc. of ICML Workshop on Machine Learning for Clinical Data Analysis, 2012

- Karlsson, I., Zhao, J., Asker, L. & Boström, H. *Predicting adverse drug events by analyzing electronic patient records.* Proc. of the 14th conference on Artificial Intelligence in Medicine (AIME), Murcia, Spain, 2013

Bibliography

- Karlsson, I., and Zhao, J. Dimensionality reduction with random indexing: an application on adverse drug events detection, Submitted to CBMS, 2014
- Zhao, J., Karlsson, I., Asker, L. & Boström, H. Applying methods for signal detection in spontaneous reports to electronic patient records. KDD 2013 Workshop on Data Mining for Health care (DMH), Chicago, USA, 2013
- Asker, L., Boström, H., Karlsson, I., Panagiotis, P., & Zhao, J. Mining candidates for adverse drug interactions in electronic patient records, Submitted to PETRA, 2014
- Henriksson, A., Conway, M., Duneld, M., Chapman, W. Identifying Synonymy between SNOMED Clinical Terms of Varying Length Using Distributional Analysis of Electronic Health Records. In Proc. of Annual Symposium of the American Medical Informatics Association (AMIA), 2013
- Tran T, Luo W, Phung D, Gupta S, Rana S, Kennedy RL, Larkins A, Venkatesh S, "A framework for feature extraction from hospital medical data with applications in risk prediction", BMC Bioinformatics, 2014

Bibliography

- Klimov, D., Shknevsky, A. & Shahar, Y., Exploration of patterns predicting renal damage in diabetes type II patients using a visual temporal analysis laboratory. Journal of the American Medical Informatics Association, 22(2), pp.275-289, 2015

- Moskovitch, R. & Shahar, Y. Classification of multivariate time series via temporal abstraction and time intervals mining. Knowledge and Information Systems, 45(1), pp.35-74, 2015

- Moskovitch, R. & Shahar, Y., Classification-driven temporal discretization of multivariate time series. Data Mining and Knowledge Discovery, 29(4), pp.871-913, 2015

- S. Wu, Y. Chen, Mining Non-ambiguous Temporal Patterns for Inter-val-Based Events, IEEE Transactions on Knowledge and Data Engineering, 19 (6), 2007

- Sacchi, L. et al., Data mining with Temporal Abstractions: learning rules from time series. Data Mining and Knowledge Discovery, 15(2), pp.217-247, 2007

- C. Yi-Cheng, J. Ji-Chiang, P. Wen-Chih, L. Suh-Yin, An Efficient Algorithm for Mining Time Interval-based Patterns in Large Databases, CIKM, 2010

Extra slides

Solution

- Let x be a true-negative instance
- **Goal:** minimum number of feature tweaks (changes) so that x becomes true-positive, x'

Observe:

- If the prediction of the RF is -1, then at least half of its trees predict -1
- If the prediction of a tree is -1, then the example is passed through a negative path, i.e., a path that predicts the class to be -1
- **Solution:** revert these paths and consequently the trees!

Note: if a single transformation results in changing another tree's decision, then ignore it!

- Focus on the trees that predict -1
- For each tree: explore the positive paths, i.e., those that predict +1
- **Apply the transformations imposed by the positive path**

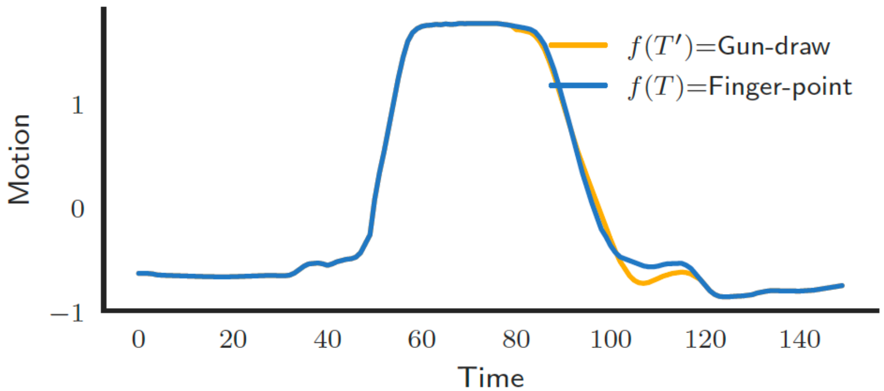
$$\mathbf{x}^+_{j(\epsilon)}[i] = \begin{cases} \theta_i - \epsilon & \text{if the } i\text{-th condition is } (x_i \leq \theta_i) \\ \theta_i + \epsilon & \text{if the } i\text{-th condition is } (x_i > \theta_i) \end{cases}$$

- Choose the transformation with the **minimum cost**

$$\mathbf{x}' = \underset{\mathbf{x}^+_{j(\epsilon)} \in \Gamma \;|\; \hat{f}(\mathbf{x}^+_{j(\epsilon)}) = +1}{\arg\min} \left\{ \delta(\mathbf{x}, \mathbf{x}^+_{j(\epsilon)}) \right\}$$

Time series tweaking

What is the minimum number of changes to apply to a time series T so that a given opaque classifier changes its prediction?



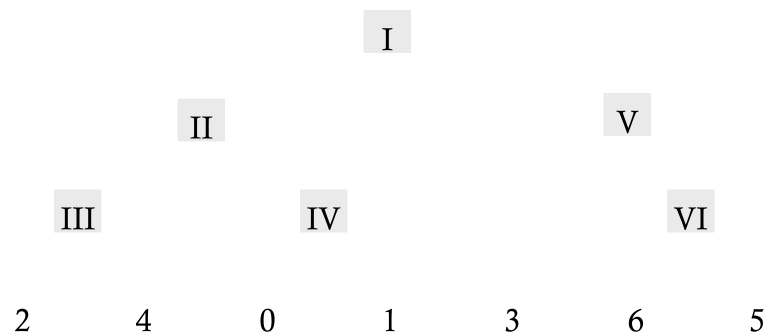
$$\mathcal{T} \rightarrow \mathcal{T}^1 \rightarrow \mathcal{T}^2 \rightarrow \ldots \rightarrow \mathcal{T}'$$

- **Reversible tweaking:** each subsequent transformation can override a previous one
- **Irreversible tweaking:** each subsequent transformation cannot override a previous one

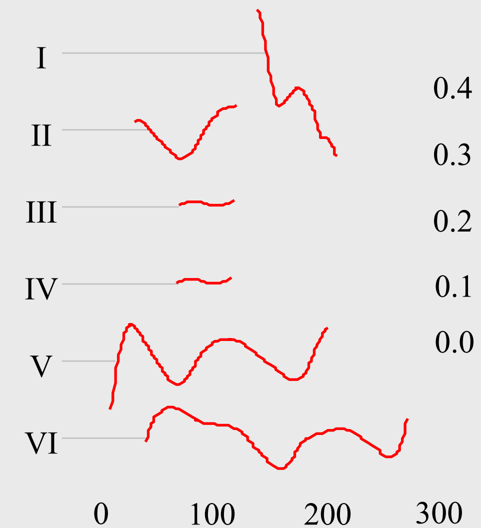
Random Shapelet Forests

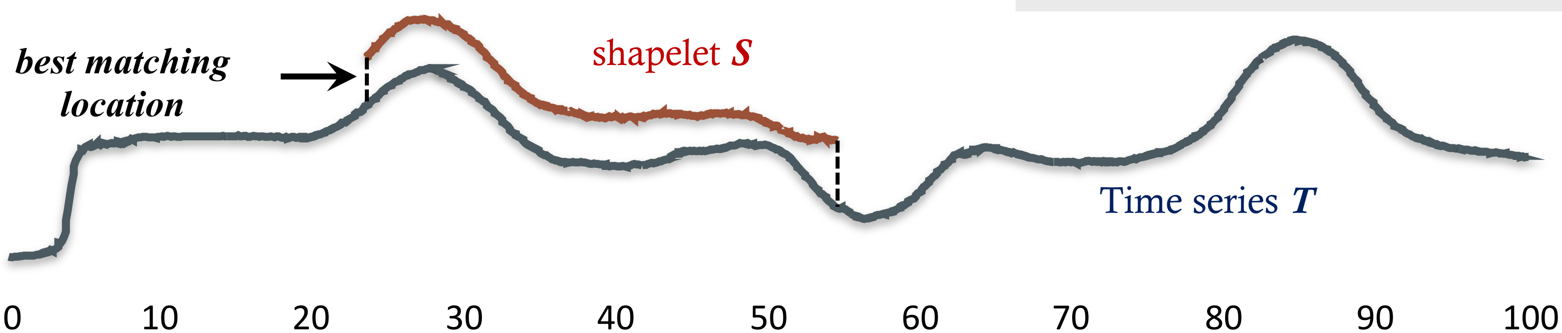
Shapelets: class-distinctive time series subsequences capturing local trends in time series

Shapelet Tree*



Shapelet Dictionary*





* *Figures taken from Eamonn Keogh*

Time series tweaking: solution

- Focus on the trees that predict -1
- For each tree, explore the positive paths, i.e., those that predict +1
- Try to force those trees to predict +1 by "tweaking" shapelet features of ***T***

Given a non-leaf node (S^j_k, θ^j_k)

- **Increase distance:**
  - If S^j_k exists in ***T***, that is $d_s(\mathcal{S}^j_k, \mathcal{T}) \leq \theta^j_k$
  - and the current node condition demands otherwise
  - ✓ Increase the distance of all matching instances of S^j_k, so that they all fall above the distance threshold θ^j_k

Time series tweaking: solution

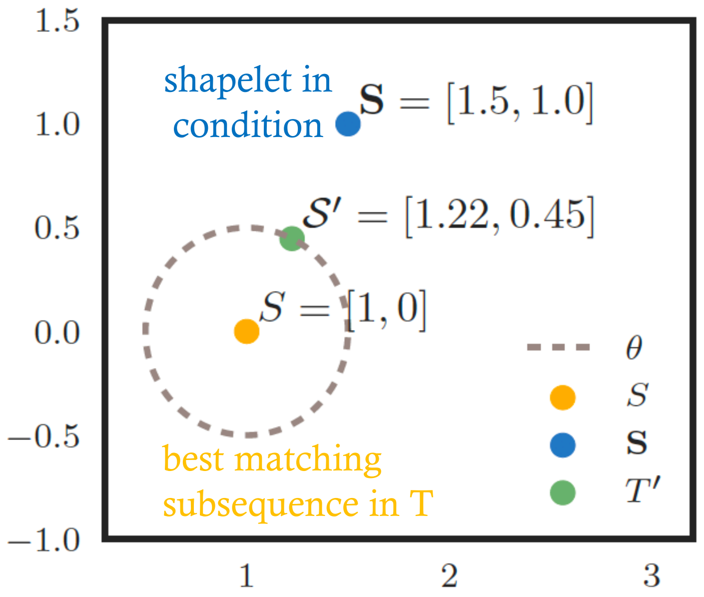
- Focus on the trees that predict -1
- For each tree, explore the positive paths, i.e., those that predict +1
- Try to force those trees to predict +1 by ”tweaking” features of $\boldsymbol{T}$

Given a non-leaf node ($\mathbf{S^j_k}$, $\boldsymbol{\theta^j_k}$)

- **Decrease distance:**
  - If S^j_k does not exist in $\boldsymbol{T}$, that is $d_s(\mathcal{S}^j_k, \mathcal{T}) > \theta^j_k$
  - and the current node condition demands otherwise
  - ✓ Decrease the distance of the best matching instance of S^j_k, so that it falls below the distance threshold θ^j_k

How to transform the time series?

- Consider S as an m-dimensional point
- Define an m-sphere with S as its center and radius θ



- The transformed time series counterpart of **S** is given by the following equation:

$$\tau_{\mathcal{S}}(\mathbf{S}, p_{ik}^{j}, \epsilon) = \mathcal{S}_{k}^{j} + \frac{\mathcal{S}_{k}^{j} - \mathbf{S}}{\|\mathcal{S}_{k}^{j} - \mathbf{S}\|_2}(\theta_{k}^{j} + (\epsilon\delta_{ik}^{j}))$$

Panagiotis Papapetrou

Experimental setup

- UCR time series repository:
  - all binary classification datasets (26 datasets)
- Competitor:
  - 1-NN under the Euclidean distance

$$\tau_{NN}(\mathcal{T}, y') = \operatorname*{arg\,min}_{\{\mathcal{T}' | (\hat{y}, \mathcal{T}') \in \mathcal{D}, \hat{y} = y'\}} d_E(\mathcal{T}, \mathcal{T}')$$

Evaluation – metrics

Average cost of successful transformation, i.e.,

how costly is the transformation?

$$c_\mu(\tau, y') = \frac{1}{n}\sum_{i=1}^{n} c(\mathcal{T}_i, \tau(\mathcal{T}_i, y'))$$

Compactness of transformation, i.e.,

how much of the time series is changed?

$$compact(\mathcal{T}, \mathcal{T}') = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} \mathit{diff}(T_i, T_i') \ ,$$

where

$$\mathit{diff}(T_i, T_i') = \begin{cases} 1, \text{ if } |T_i - T_i'| \leq e \\ 0, \text{ otherwise.} \end{cases}$$

Evaluation – result

Reversible tweaking results in the **least costly** transformations

Irreversible tweaking results in the **most compact** transformations

The baseline is too naive

Dataset	Cost			Compactness			Accuracy	
	τ_{RT}	τ_{IRT}	τ_{NN}	τ_{RT}	τ_{IRT}	τ_{NN}	RSF	NN(1)
	7.3810	**7.3810**	26.6223	**0.5737**	**0.5737**	1.0000	**0.8750**	0.7500
	4.5071	4.5098	15.6695	**0.5048**	0.5169	1.0000	**1.0000**	0.6250
	1.1447	1.1846	1.9178	0.3824	**0.1809**	1.0000	**1.0000**	**1.0000**
	2.2197	2.5132	22.4809	0.4123	**0.4044**	1.0000	**0.7000**	0.4900
	0.9314	1.1150	1.1704	0.5917	**0.4466**	0.9999	**0.7886**	0.7143
	2.2725	3.1455	30.0943	**0.7449**	0.7577	1.0000	**0.7826**	0.6630
ECG200	**1.8730**	1.9080	4.1428	0.7976	**0.7686**	1.0000	**0.8750**	0.9500
ECGFiveDays	**1.9722**	2.0158	4.2143	0.5215	**0.4913**	1.0000	**1.0000**	0.9944
GunPoint	**1.9787**	1.9942	3.6975	0.4712	**0.4460**	0.9998	**1.0000**	0.9250
H	**2.1744**	2.2187	7.8253	0.6791	**0.6621**	0.9999	**0.8605**	0.7907
	1.2492	**1.2488**	3.5817	0.4563	**0.4060**	0.9999	**0.5000**	0.3846
	1.1791	1.2645	1.3088	0.7262	**0.6397**	0.9998	**0.9726**	0.9589
	3.2741	3.9266	18.9703	0.7470	**0.7071**	1.0000	**0.6667**	0.6667
	0.6685	0.9877	0.6791	0.6182	**0.4493**	0.9999	**0.8258**	0.7753
	2.4413	2.5313	6.0249	0.5602	**0.4834**	1.0000	**0.9685**	0.9213
	0.6979	0.9568	0.7574	0.6186	**0.5116**	0.9998	**0.8421**	0.7782
ProximalPhalanxOutlineCorrect	**0.5895**	1.0056	0.5326	0.6552	**0.4121**	0.9997	**0.8315**	0.8090
SonyAIBORobotSurface1	1.7384	**1.7260**	4.7213	0.4429	**0.4394**	1.0000	0.9919	**1.0000**
SonyAIBORobotSurface2	1.8601	**1.8566**	5.6126	0.4133	**0.3584**	1.0000	0.9796	**0.9949**
	1.2082	1.3628	1.2802	0.6644	**0.5464**	0.9999	0.9695	**0.9797**
	3.1200	3.1436	14.7768	0.3871	**0.3718**	1.0000	**0.9259**	0.7407
	5.4407	5.8238	17.8733	0.6173	**0.5705**	1.0000	**0.9697**	0.7879
TwoLeadECG	**0.9112**	1.0671	1.3517	0.4966	**0.4028**	0.9999	**1.0000**	0.9957
Wafer	**3.0135**	3.1419	8.6207	0.7152	**0.6676**	0.9999	0.9958	**0.9979**
Wine	**0.5052**	0.9301	0.1708	0.7529	**0.3452**	0.9996	**1.0000**	**1.0000**
WormsTwoClass	**5.7723**	7.2023	28.7383	0.4416	**0.4219**	1.0000	**0.8269**	0.7308
Avg.	**2.3132**	**2.5329**	**8.9552**	**0.5733**	**0.4942**	**0.9999**	**0.8924**	**0.8240**