By Case or Normality? Diagnosing Medical Data in Anomaly Detection Perspective

Speaker: Shenghua Liu

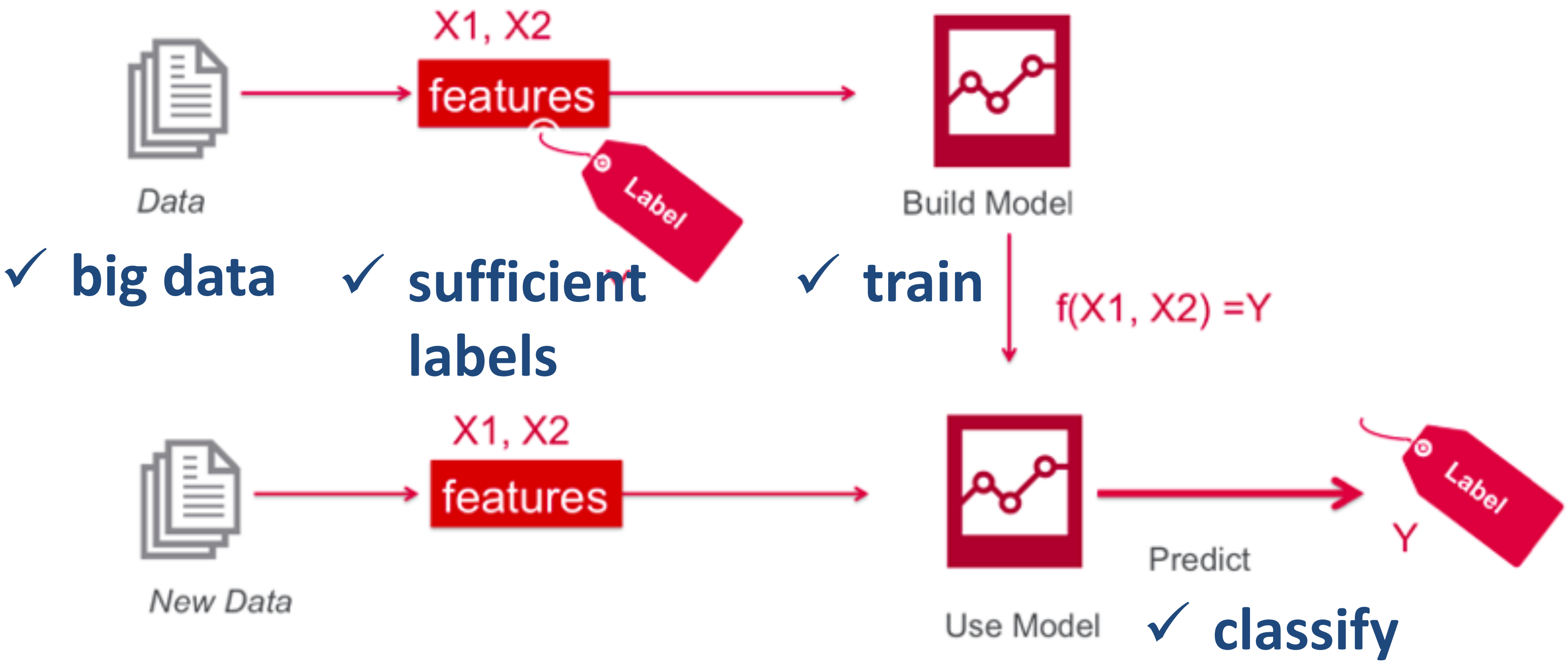
+Institute of Computing Technology ICT, CAS

Outline

- AI healthcare in a nutshell
  - Diagnosis with classification schema
- … …

Classification schema is naturally used for diagnosis

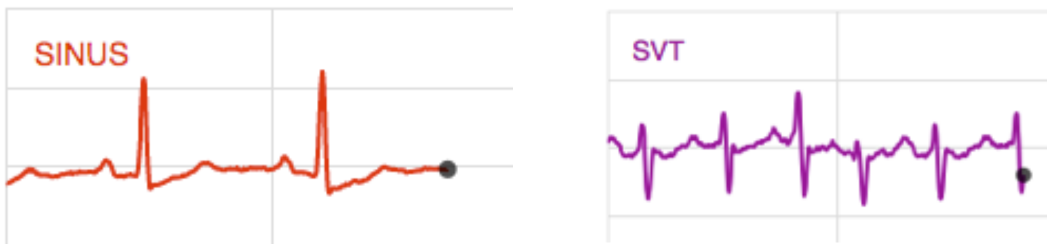
- Key steps

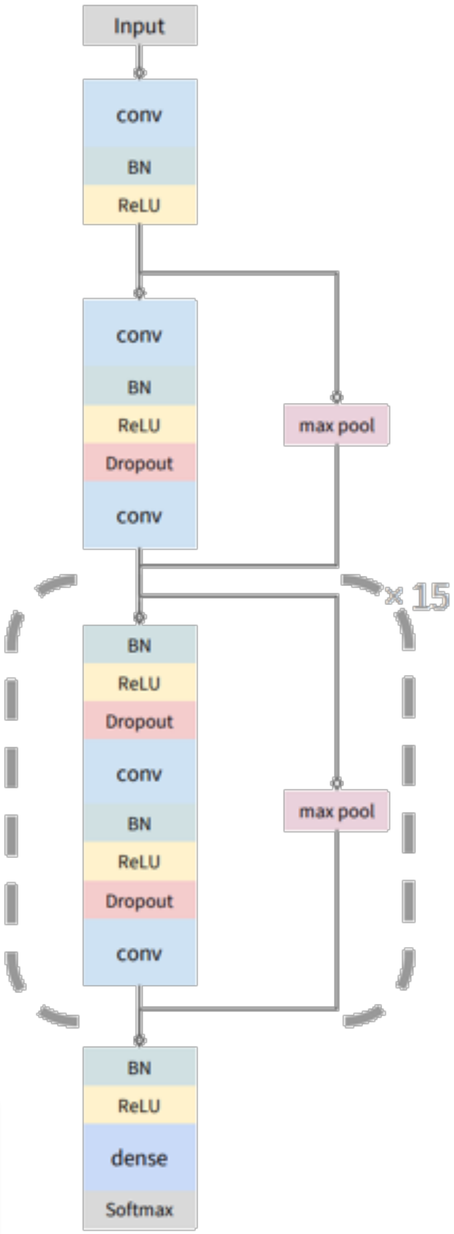


Carol McDonald, Demystifying AI, Machine Learning and Deep Learning, in MAPR, Aug 2017

Diagnosing irregular heart rhythms

- 34-layer CNN model
- classify 10 arrhythmias
  - map a sequence of ECG to a sequence of rhythm class
  - require large annotated dataset

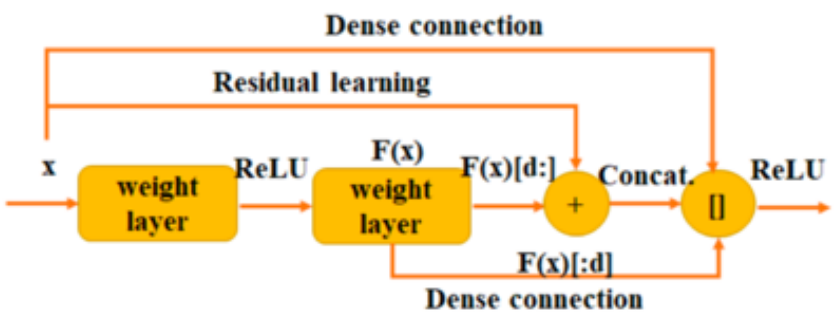




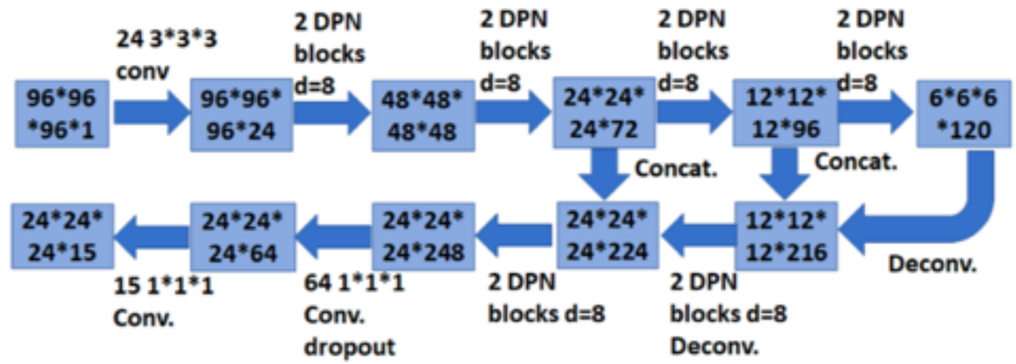
Pranav Rajpurkar+, Cardiologist-level arrhythmia detection with convolutional neural networks. Nature Medicine, 25:65–69, 2019

Diagnosing pulmonary nodule

- 3D dual path network
- Luna16 dataset



Dual path net (DPN)

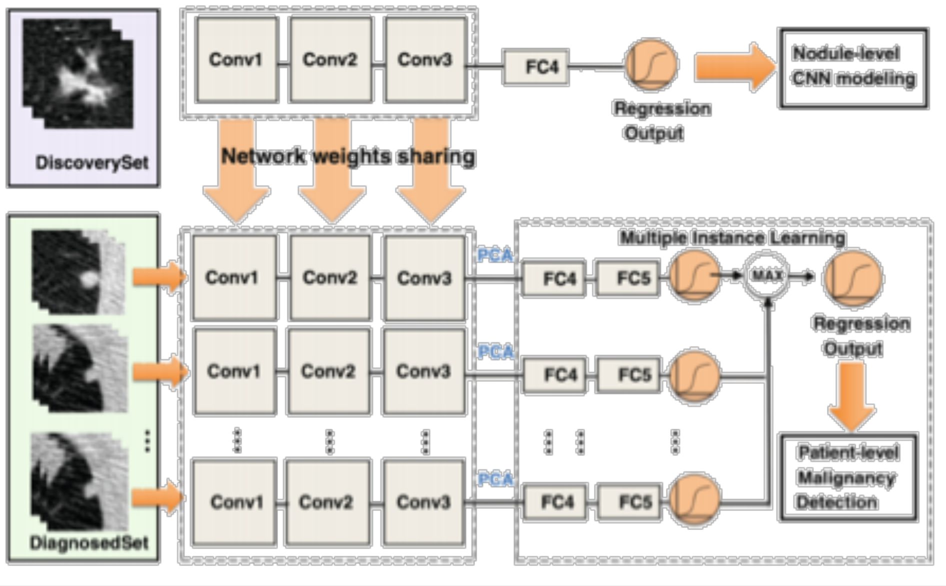


Network framework

Zhu, W.+, Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification, WACV. IEEE, 2018: 673-681.

Diagnosing lung cancer

- Learns transferable deep features
- CNN-MIL model
  - multi-instance learning



W. Shen+, Learning from experts: Developing transferable deep features for patient-level lung cancer prediction. In MICCAI, 2016.

More diagnosis with classification schema

- 2,032 categories of skin cancer,
  - Inception v3 network [A. Esteva+, Nature17]
- Eye glaucoma (fundus image),
  - texton and local configuration pattern features [Acharya U R, CBM17]
- Liver metastases (CT),
  - fully convolutional network [A. Ben-Cohen+, DLMIA16]
- Breast masses (X-rays),
  - classifying with deep features[Dhungel N+, MICCAI16]
- Brain tumor (MRI),
  - feature extraction with co-occurrence matrix [Praveen G.B.+, CCIS15]
- … …

Summary of classification-based diagnosis

- Classification is powered by ML/Deep Learning
  - Provide data-driven, evidence-based clinical intelligence
- Not an all-purpose solution
  - superficial features, patterns, and correlation
  - lack explainability and hard to adjust certain identified problems
  - as good as the training data (biases, noises, errors often exist in real-world data)
  - difficulty to generalize the finding beyond its training dataset
  - specifying the context of its proper use

M Chen, M Decary, AI in Healthcare: From Hype to Impact, workshop on TCH 2019: Improving Usability, Safety and Patient Outcomes with Health Information Technology, 2019.

Outline

- Diagnosis with classification schema
- Diagnosis as anomaly detection
  - … …

How does human detect anomalies?

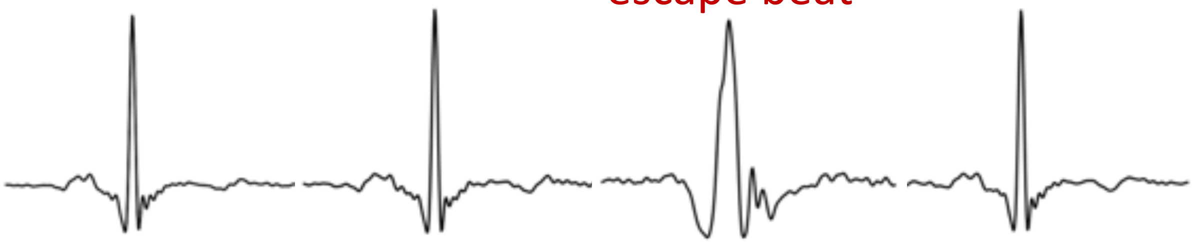
Human detects anomalies without recognizing their features.

- stand out penguin

blue fur
black head
yellow cheek
yellow lower-beak

Detect medical anomalies

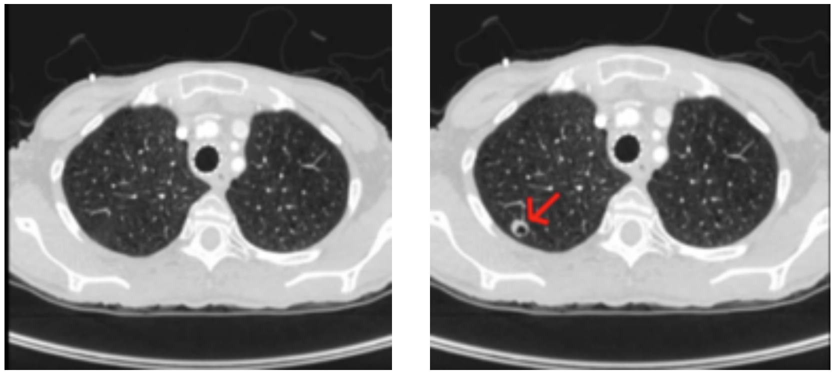
Ventricular
escape beat



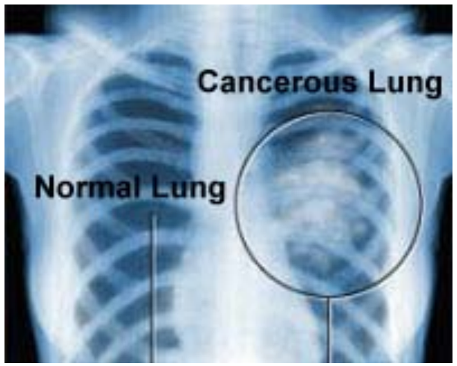
ECG heart beats

understand robustly
normality/majority

liver
Cancer / Tumor



pulmonary CT scan
nodule



X-rays
cancer

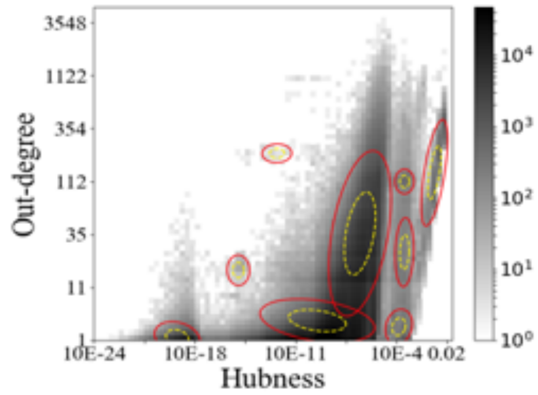
Ed Susman UPI Correspondent Philadelphia, SPACE MEDICINE, Nov 06, 2006

How Can We Diagnose in Anomaly Detection Perspective?

Outline

- Diagnosis with classification schema

- Diagnosis as anomaly detection
  - Feature-based methods
    - ✓One-class SVM [Schölkopf, B.+, 2001],
    - ✓Elliptic Envelope [Rousseeuw, P.J.+, 1999],
    - ✓Isolation Forest [Liu, F.+, 2008]
    - ✓Local Outlier Factor (LOF) [Breunig, K. N.+ 2000]
    - ✓Robust Random Cut Forest (RCF) [S Guha+, 2016]
    - ✓**EagleMine** [Feng W.+, 2019]
    - ✓… …
  - … …



EagleMine

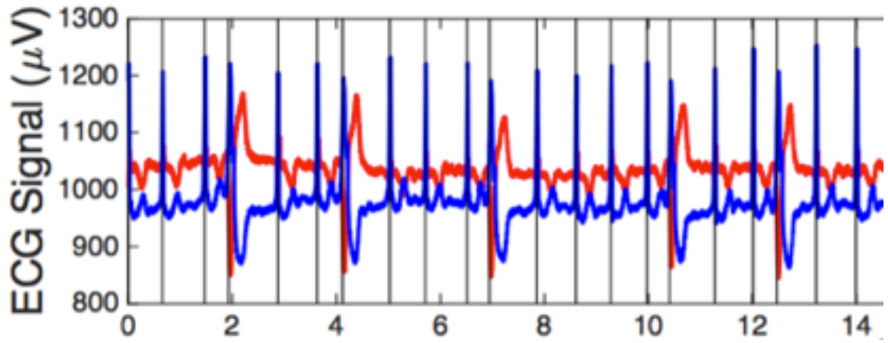
Outline

- Diagnosis with classification schema

- Diagnosis as anomaly detection
  - Feature-based methods
  - Summarization
    - ✓ e.g. Time series, ECG data
  - … …
  -

Diagnosis unusual beats with summarization

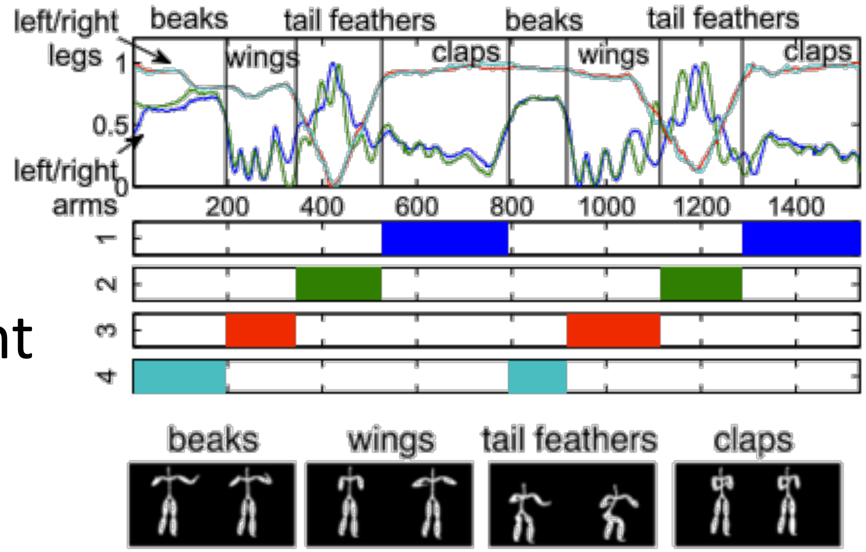
- Given
  - multivariate time series X, containing patterns



- Find
  - **Summarization**: description of patterns in X
  - **Anomalies**: periods of unusual segment
  - **Forecast**: predict future data

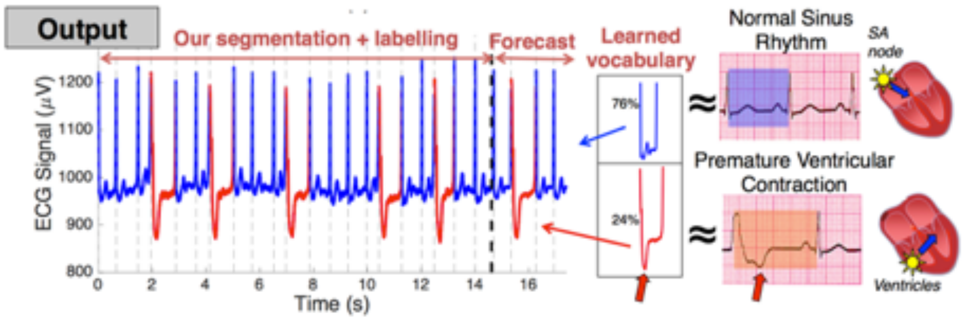
Related works

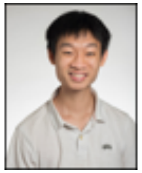
- Time Series Summarization
  - Autoplait [Matsubara+, 2014]
  - segmentation, and
  - high-level patterns assignment



- Distance-based Methods
  - Dynamic Time Warping [Berndt+, 1994]
  - distance of a given time series segment from others

* Carnegie Mellon University, † Chinese Academy of Sciences





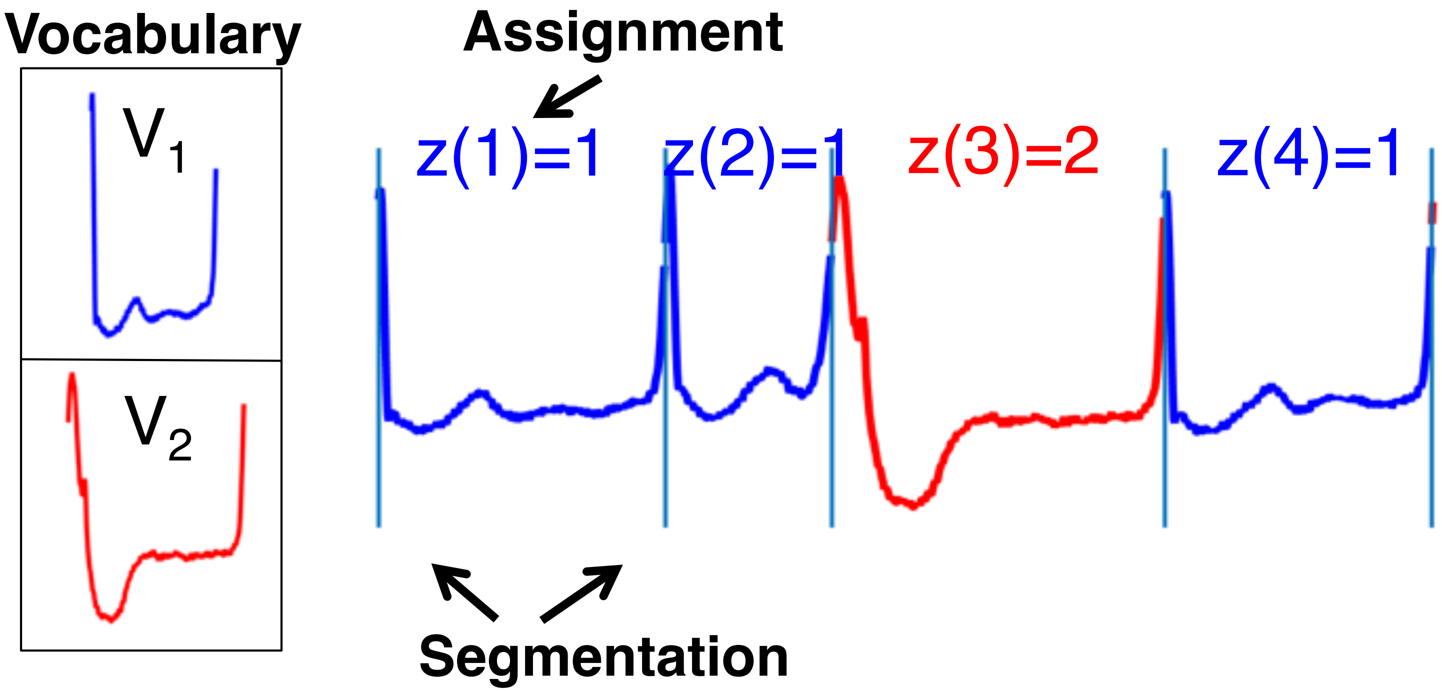

BEATLEX: Summarizing and Forecasting Time Series with Patterns

B Hooi*, S Liu†, A Smailagic*, C Faloutsos*

code: http://www.andrew.cmu.edu/user/bhooi/code/beatlex.zip

BeatLex Model

- **Vocabulary**: short time series patterns
- **Segmentation**: how X is split into segments
- **Assignment**: of vocabulary terms to segments

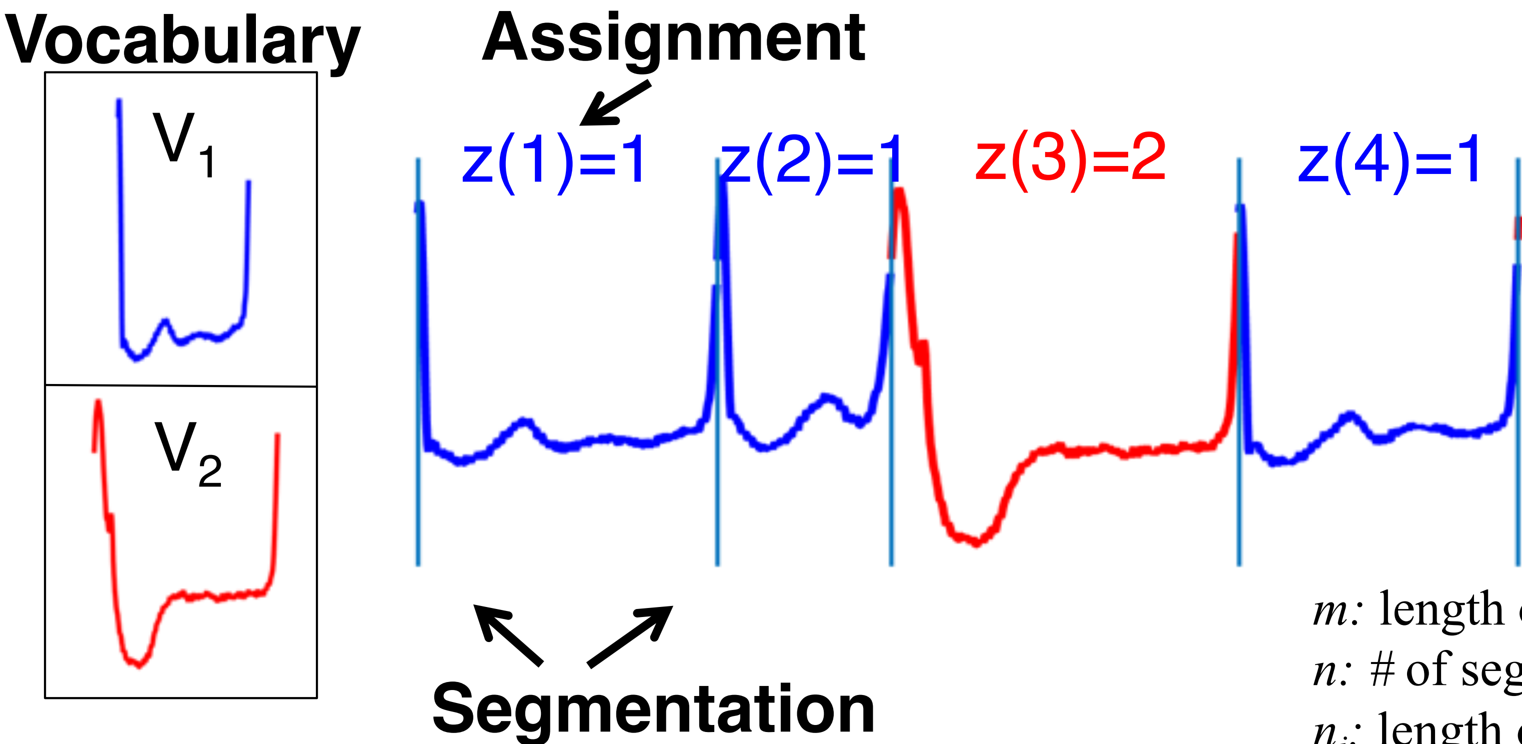


Minimum Description Length (MDL)

- Occam's razor ("simplest explanation is best")
- Define
  - **Cost(*M*)**: number of bits to describe model *M*
  - **Cost(*X* | *M*)**: number of bits to describe data *X* given *M*
- **MDL principle:**

$$\underset{M}{\text{minimize}}\ \text{Cost}(M) + \text{Cost}(X|M)$$

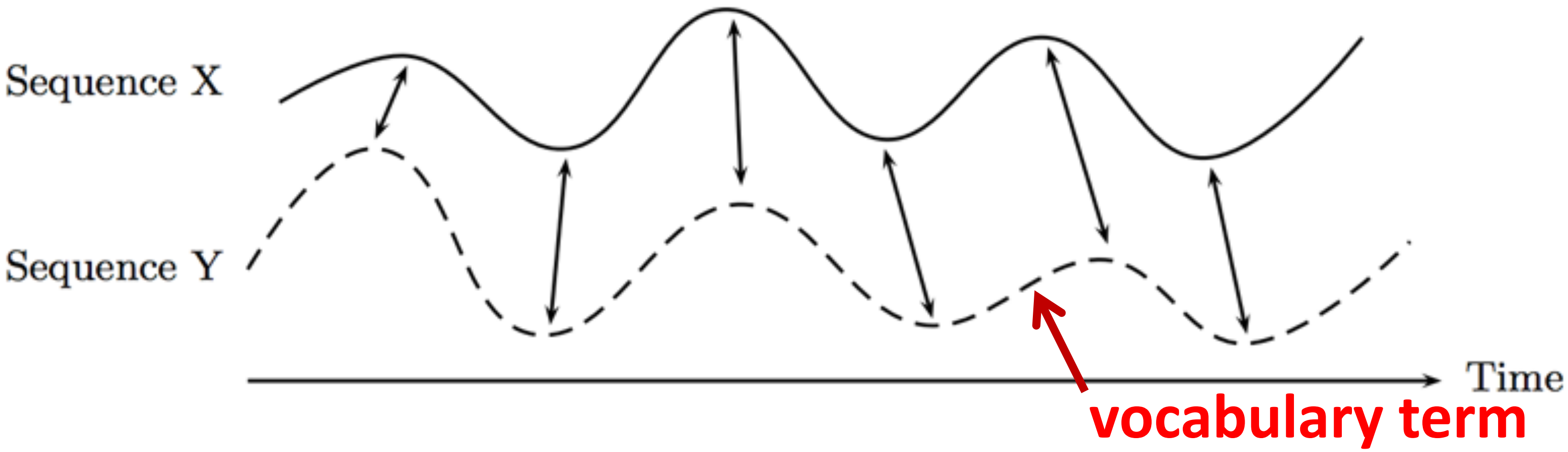
Model Cost



m: length of ts
n: # of segments
n_i: length of vocabulary
k: # of vocabulary

$$\mathrm{Cost}(M) = \underbrace{\log^*(k)}_{\text{cost of } k} + \underbrace{C_F \sum_{i=1}^{k} n_i}_{\text{vocab. cost}} + \underbrace{(n-1)\log(m)}_{\text{segmentation cost}} + \underbrace{n\log(k)}_{\text{assign. cost}}$$

Data Cost



Dynamic Time Warping measures the distance between X and Y by aligning them, then computing squared distance

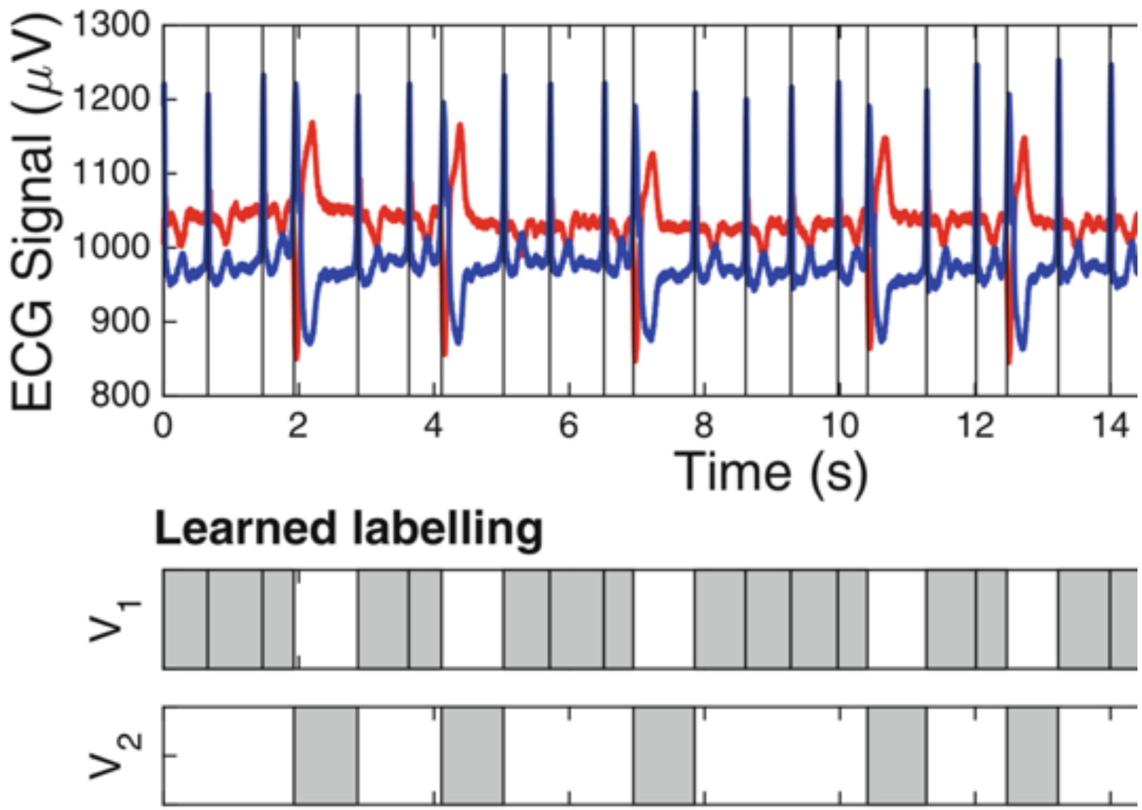
Modified DTW distance: number of bits needed to encode X given Y, using this alignment (adding penalty of length of segment and vocabulary)

$$\mathrm{Cost}(X|M) = \sum_{i=1}^{n} \mathrm{MDTW}(X^i, V_{z(i)}))$$

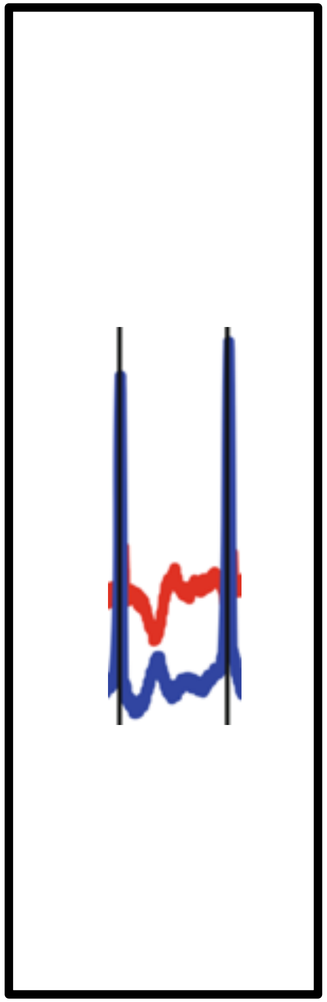
i **th segment of X** **Assigned vocab. term**

Algorithm

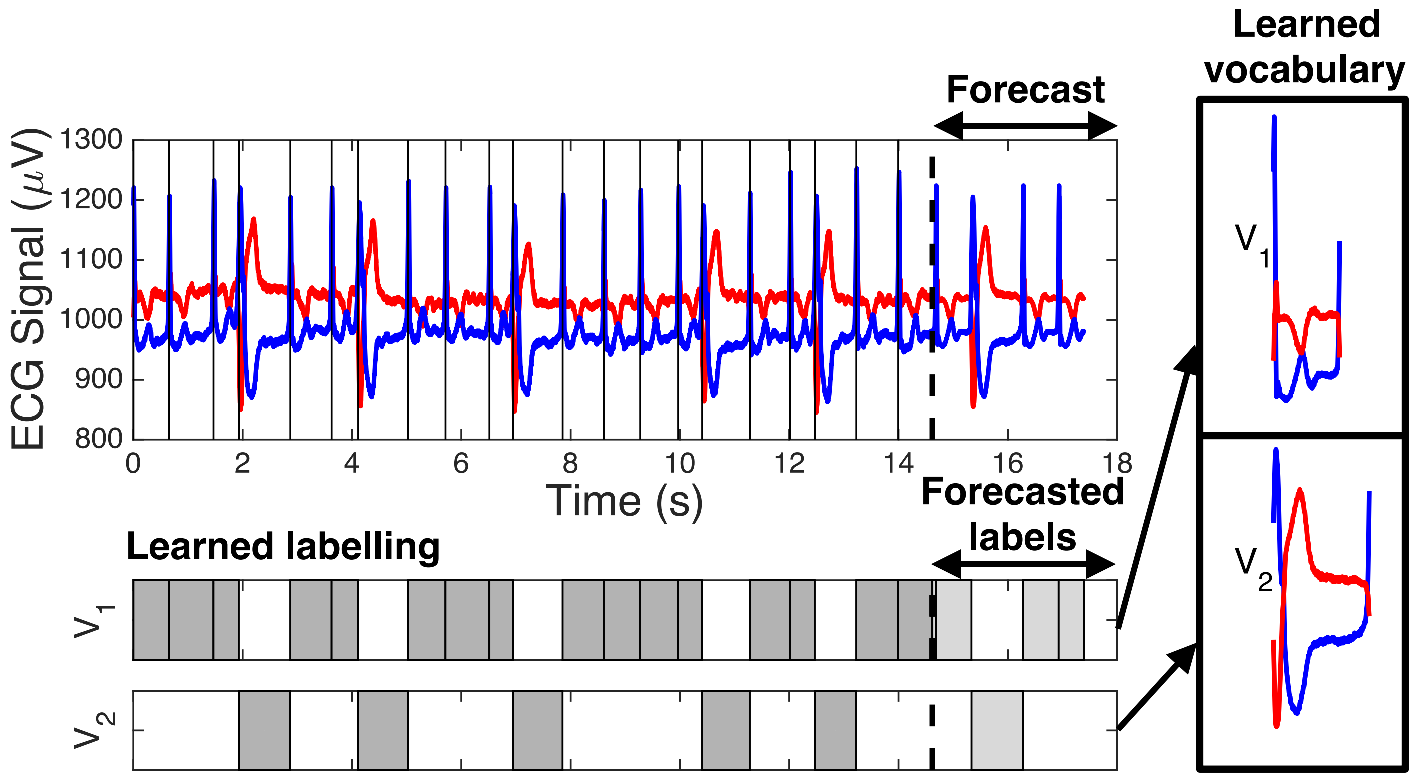
- Greedy streaming algorithm guided by MDL objective



Learned vocabulary



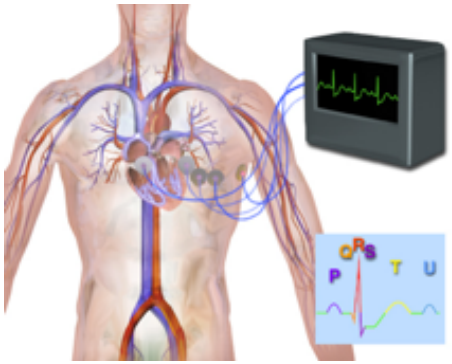
Forecasting

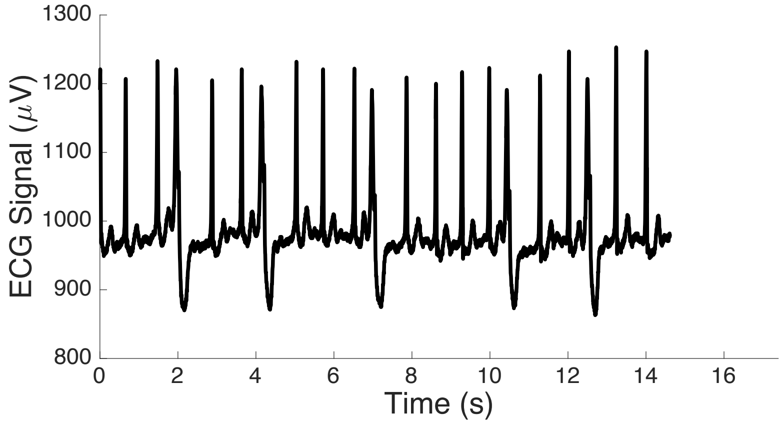


1. **Forecast labels** using Markov model based approach
2. **Forecast data** by replacing each label by its vocabulary term

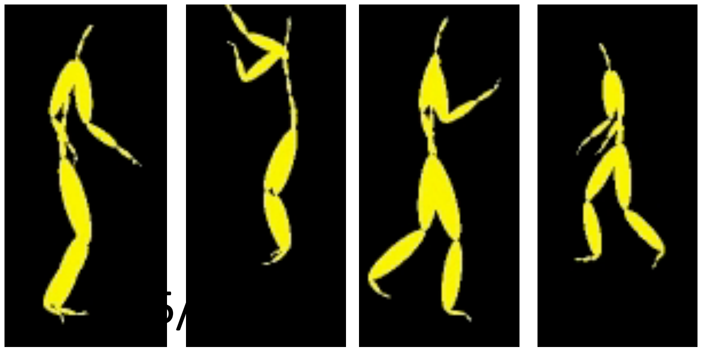
Datasets

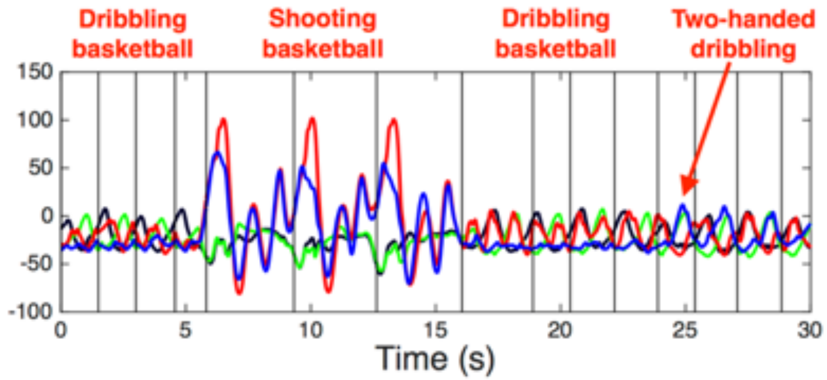
- **MIT-DB:** electrocardiogram recordings



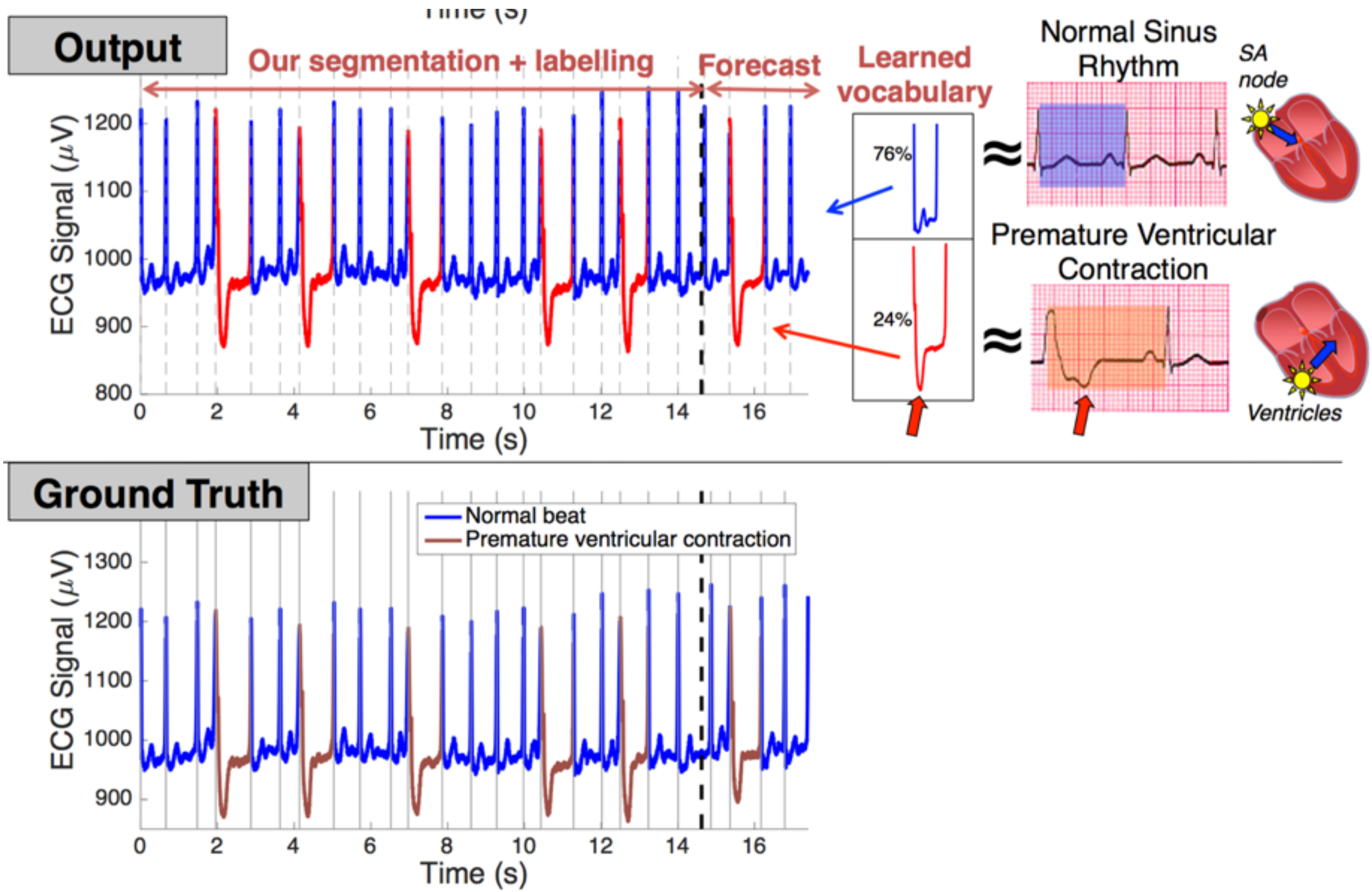


- **CMU Motion Capture dataset** motion-captured subjects performing various actions

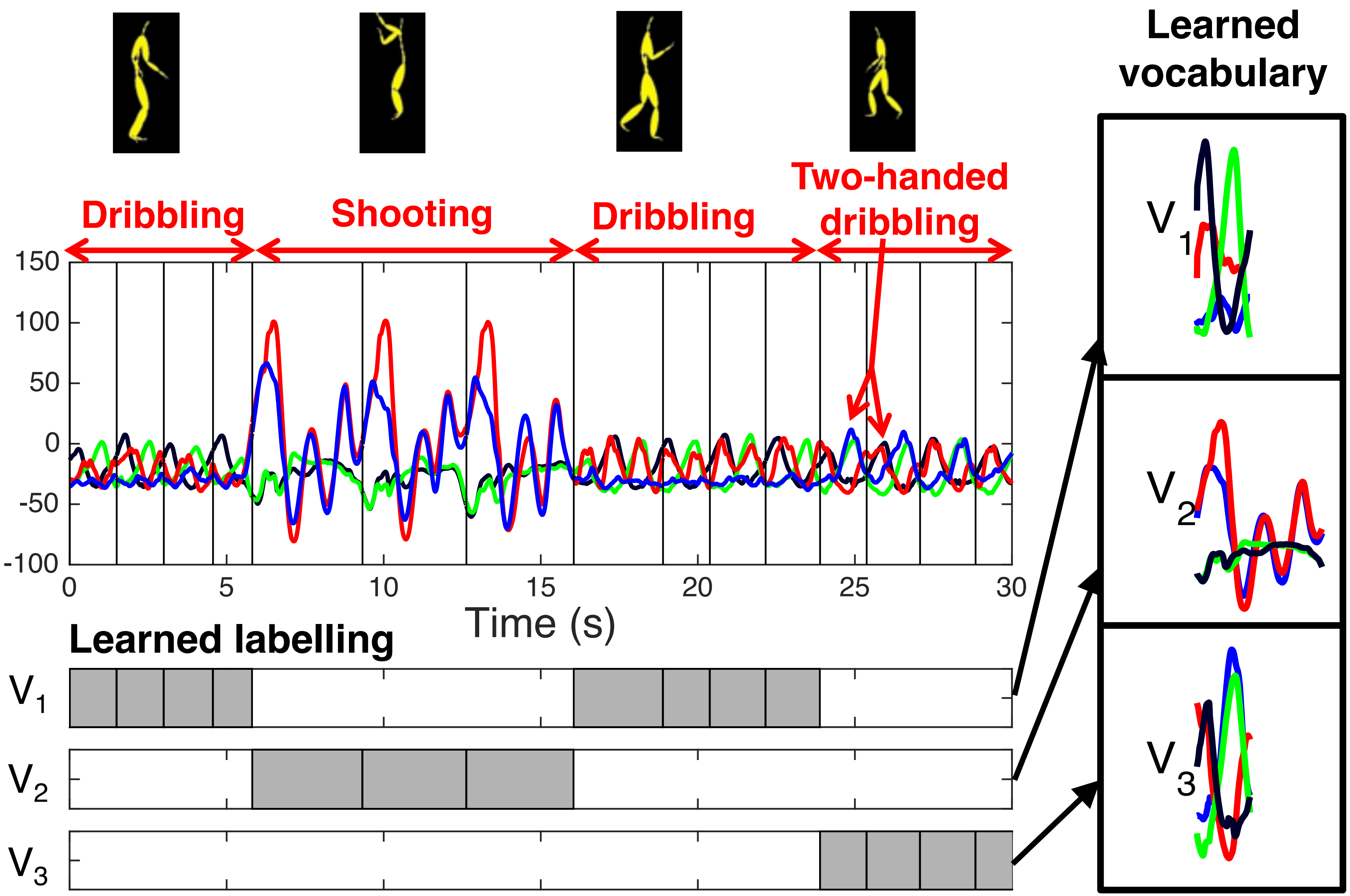




Results: ECG data



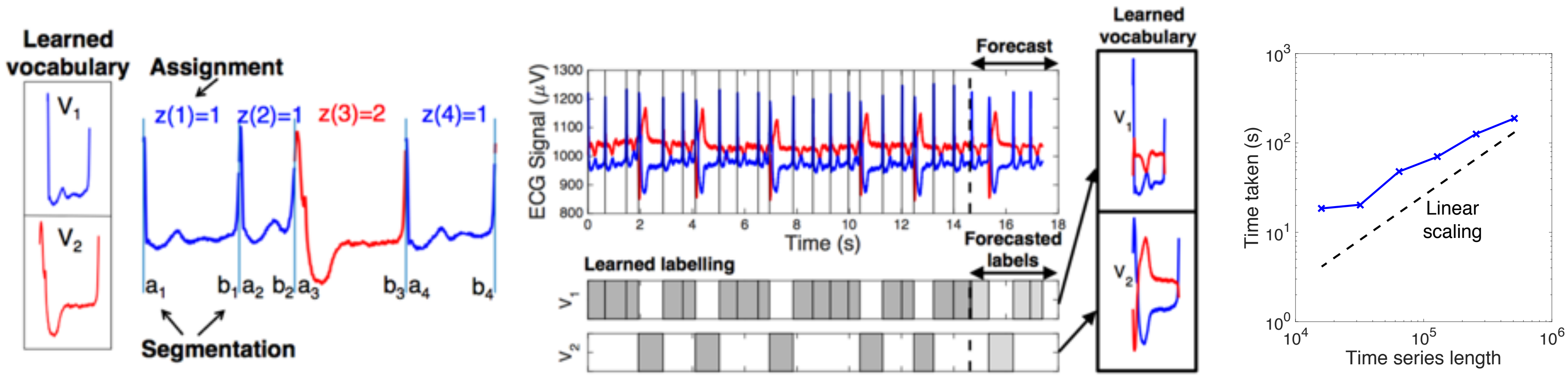
Results: CMU Motion Data



BeatLex: detect unusual from summarization

- **Model**: vocabulary model + modified DTW
- **Objective**: based on MDL
- **Algorithm**: greedy streaming algorithm for segmentation, summarization and forecasting
- **Scalable**

$$\mathrm{Cost}(M) + \mathrm{Cost}(X|M)$$



Outline

- Diagnosis with classification schema

- Diagnosis as anomaly detection
  - Feature-based methods
  - Summarization
    - ✓e.g. Time series, ECG data
  - Reconstruction-based methods

Reconstruction-based framework

- Reconstruction Loss for optimization

$$
\begin{aligned}
L &= ||X - G(X)||_2 + R(G) \\
&= \sum ||x - G(x)||_2 + R(G)
\end{aligned}
$$

- Anomaly detection

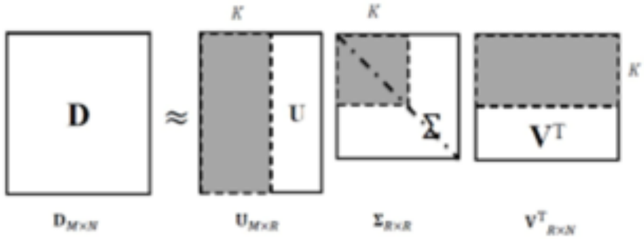
$$A(x) = \|x - G(\cdot)\|$$

Reconstruction-based methods

- SVD/Fbox [N Shah+, 2014]

$$G(X) = \sum_{k=1}^{p} \sigma_k u_k v_k^T$$

$$R(G) = \lambda_1 \left\| I - U^T U \right\|_2 + \lambda_2 \left\| I - V^T V \right\|_2$$

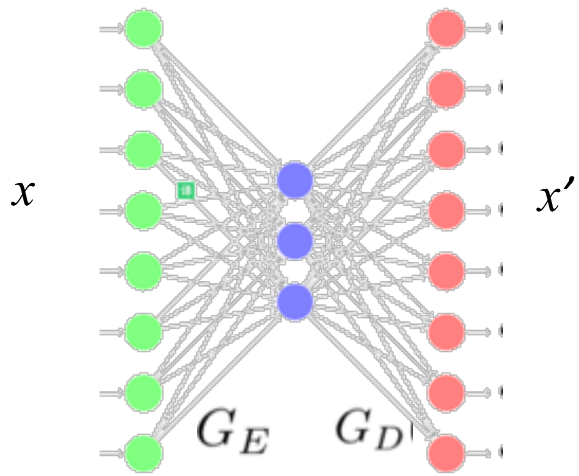


- Auto Encoder networks

$$G(x) = G_D(G_E(x))$$

  - AE [Vincent, P.+, 2010] $R(G) = 0$
  - Variational AE [Kingma, D. P.+, 2013]

$$R(G) = \lambda D_{KL}[Q(z|x)||P(z)]$$



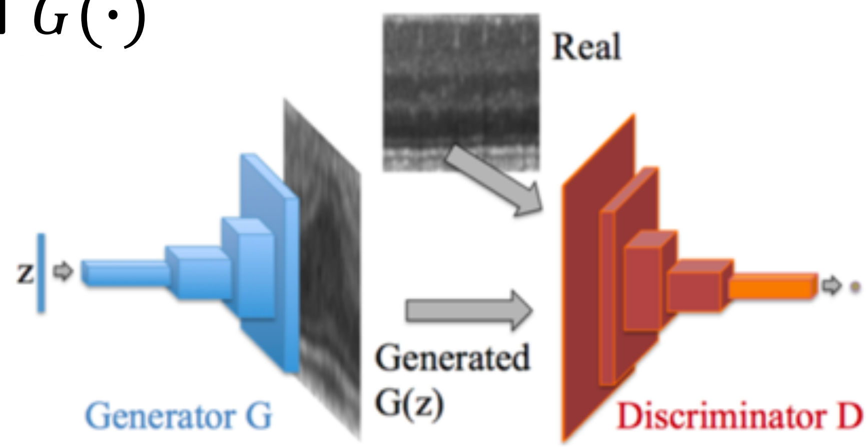
An, Jinwon+, Variational autoencoder based anomaly detection using reconstruction probability, Special Lecture on IE 2 (2015): 1-18.

AnoGAN for anomaly detection

- Train a GAN to generate normal data

$$z \rightarrow G(z) \qquad R(G) = D(x, G(z))$$

- For test sample x
  - fix parameters of the GAN $G(\cdot)$
  - learn z from $G(z\,|x)$
  - output anomaly score
  - $A(x) = ||x - G(z\,|x)||$
- Tune z for each test sample

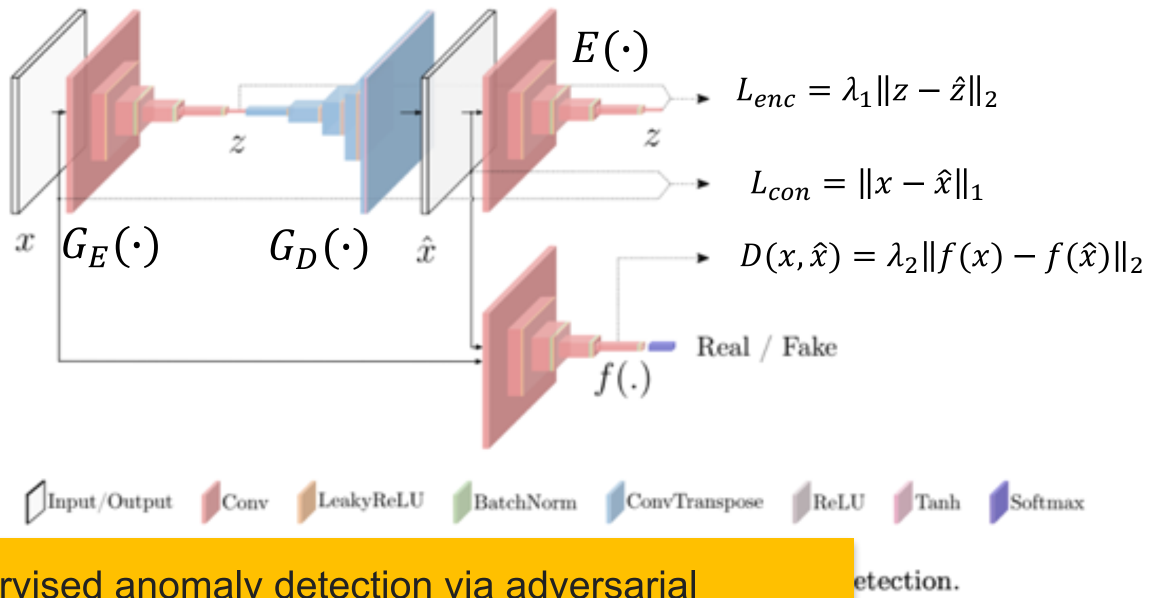


Thomas Schlegl+, Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. IPMI, pp 146–157. Springer, 2017.

Ganomaly for anomaly detection

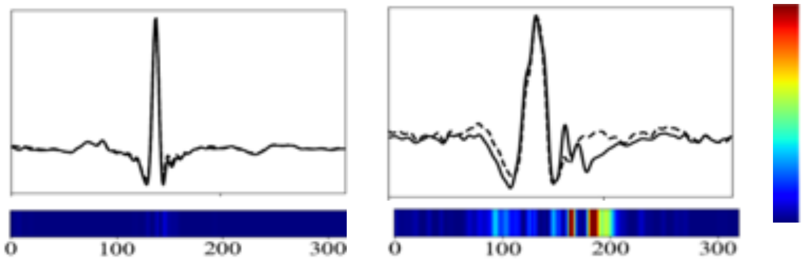
- Two encoder network
  - $L = L_{con} + L_{enc} + R(G)$, where $R(G) = D(x, \hat{x})$ $\quad \hat{x} = G_D(G_E(x))$
- Compute anomaly score in latent space

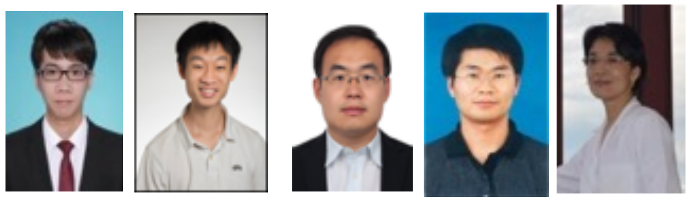
$$A(x) = \|G_E(x) - E(\hat{x})\|_1 = L_{enc}$$



Samet Akcay+, Ganomaly: Semi-supervised anomaly detection via adversarial training. arXiv, 2018.

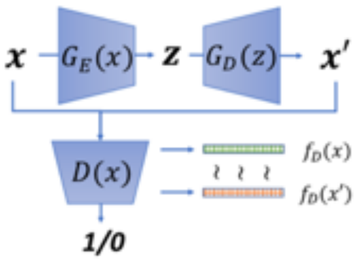
* Carnegie Mellon University, † Chinese Academy of Sciences





BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series

B. Zhou[†], S Liu[†], B Hooi[*], X. Cheng[†], J. Yang

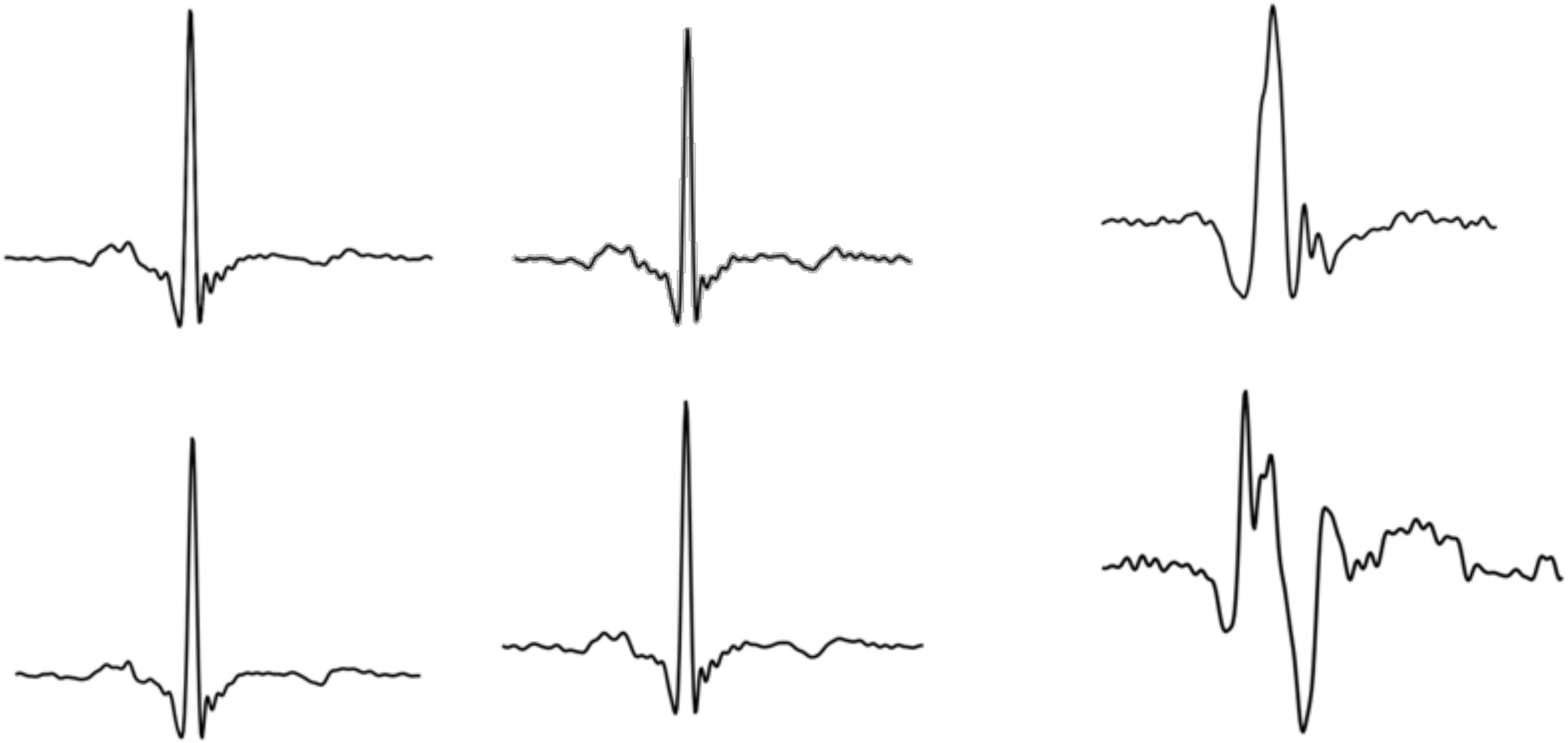


B Zhou+,BEATLEX: BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series, IJCAI 2019 (to be appeared)

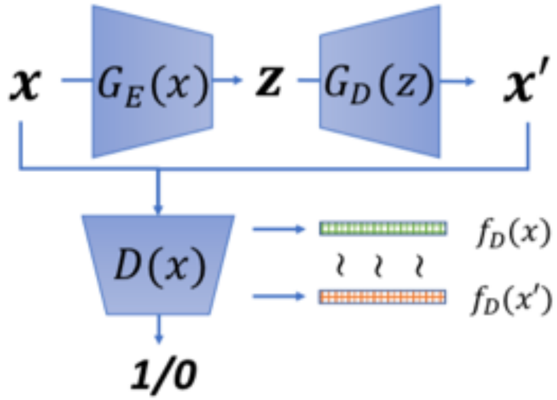
ECG data: approximately segment into beats

Normal Beats

Abnormal Beats



BeatGAN: diagnosing anomalous time series



- Two encoder network and GAN

$$L = \|x - G_D(G_E(x))\| + R(G)$$

- Two regularization terms
  - adversarial (model level) and

$$R_1(G) = D\big(x, G_D(z)\big) = \lambda \|f_D(x) - f_D(x')\|$$

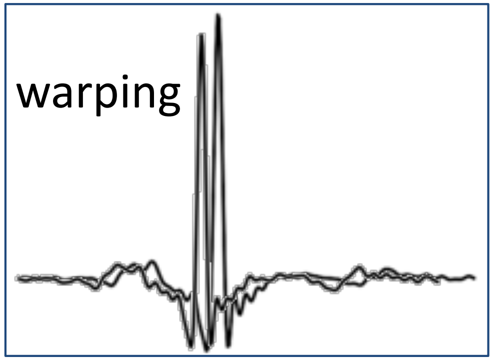
  - time warping (data level)

$$R_2(G) = \text{DTW(data)}$$

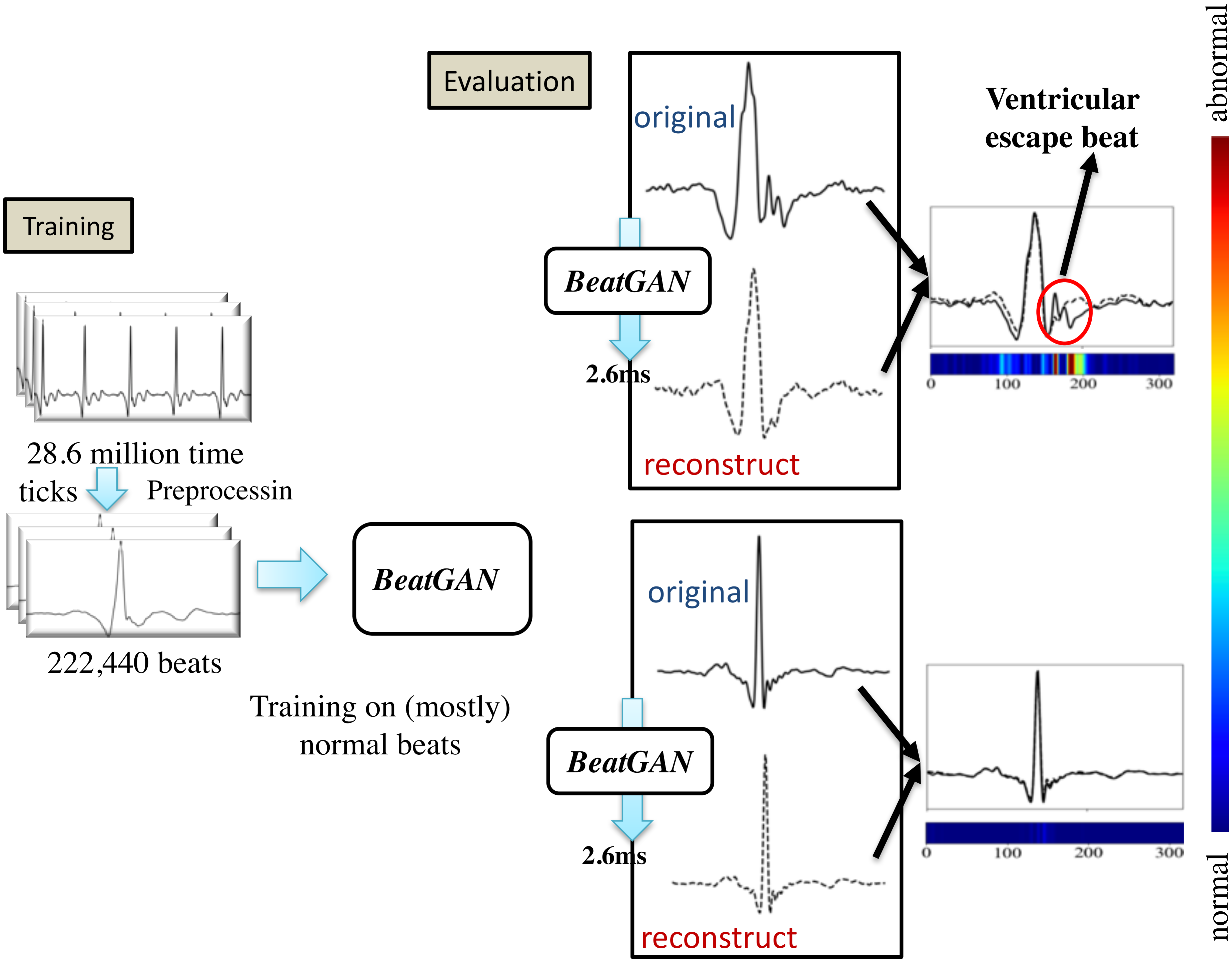
DTW: dynamic time waring to augment data

Why time warping regularization?

- ECG naturally and slightly speed up or slow down
  - DTW



- Augmentation method
  - randomly pick a small number *k*
  - Pick *k* time ticks to “speed up” (delete values)
  - Pick *k* time ticks to “slow down” (insert values)



Evaluation
Training
original
BeatGAN
2.6ms
reconstruct
Ventricular escape beat
0
100
200
300
abnormal
normal
28.6 million time ticks
Preprocessin
222,440 beats
BeatGAN
Training on (mostly) normal beats
original
BeatGAN
2.6ms
reconstruct
0
100
200
300

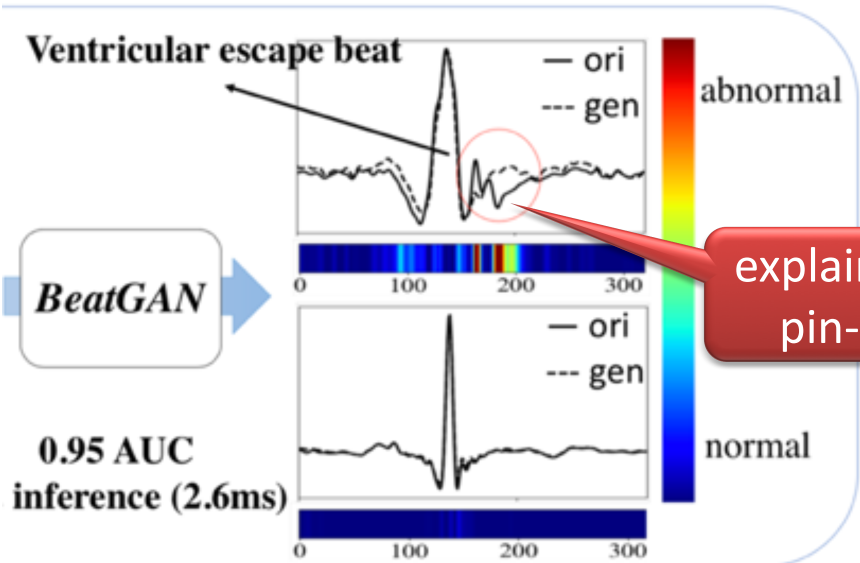
BeatGAN detects anomalous beats and pin-point ticks

Data: MIT-BIH nearly 100,000 beats,
28.6 million time ticks
90% normal beats beats

Method	AUC	AP
PCA	0.8164 ± 0.0037	0.6522 ± 0.0061
OCSVM	0.7917 ± 0.0018	0.7588 ± 0.0027
AE	0.8944 ± 0.0128	0.8415 ± 0.0163
VAE	0.8316 ± 0.0025	0.7882 ± 0.0024
AnoGAN	0.8642 ± 0.0100	0.8035 ± 0.0069
Ganomaly	0.9083 ± 0.0122	0.8701 ± 0.0141
BeatGAN	0.9447 ± 0.0053	0.9108 ± 0.0049
BeatGAN$_{aug-3\times}$	$\mathbf{0.9475 \pm 0.0037}$	$\mathbf{0.9143 \pm 0.0047}$
BeatGAN$^{0.1\%}_{aug-3\times}$	0.9425 ± 0.0022	0.8973 ± 0.0042

Table 3: BeatGAN performs the best for anomalous rhythm detection in ECG data. In BeatGAN$_{aug-3\times}$, we augment the training data size to $3\times$ with time warping. In BeatGAN$^{0.1\%}_{aug-3\times}$, we add the 0.1% anomalous time series to the training data for evaluating robustness. 5-fold cross-validations are run, and averaged metrics and std are given.

Add anomalous beats to training data as noises



explainable results by pin-pointing ticks

BeatGAN is fast

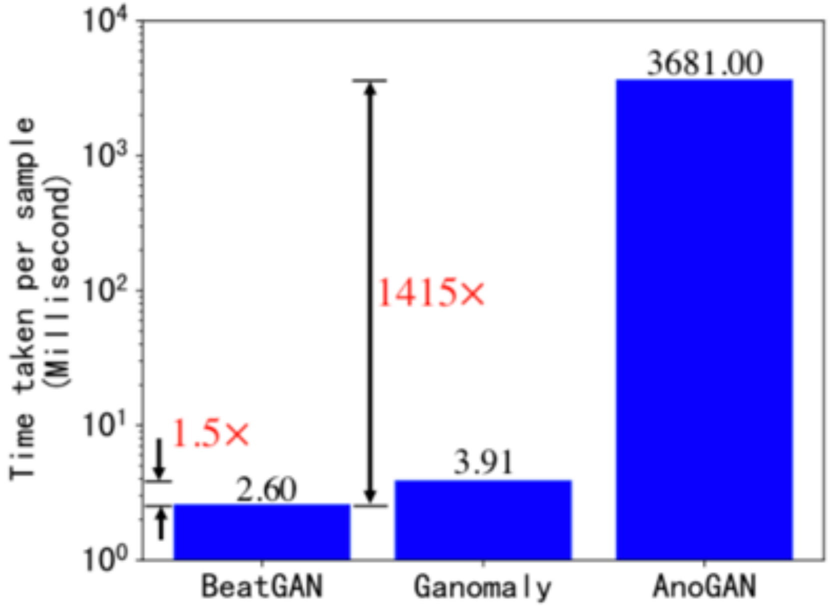


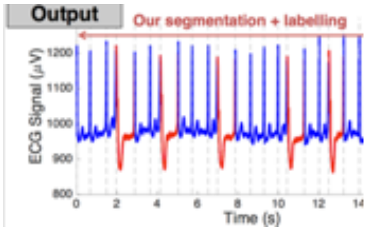
Figure 5: BeatGAN has fast inference (2.6ms) for detecting anomalous beats.

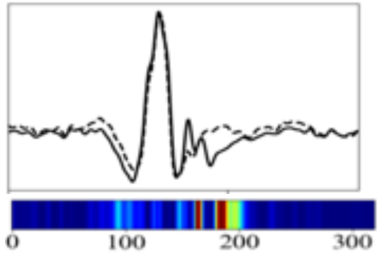
Outline

- Diagnosis with classification schema
- Diagnosis as anomaly detection
  - Feature-based methods
  - Summarization
    - ✓ e.g. Time series, ECG data
  - Reconstruction-based methods
- Conclusion

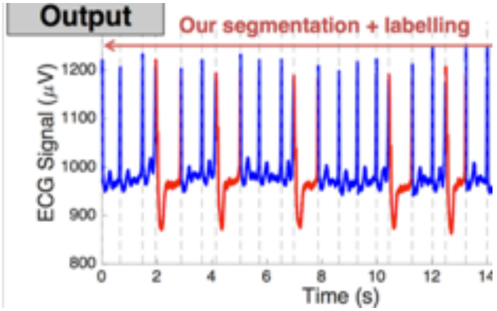
Key requirements for AI success in healthcare

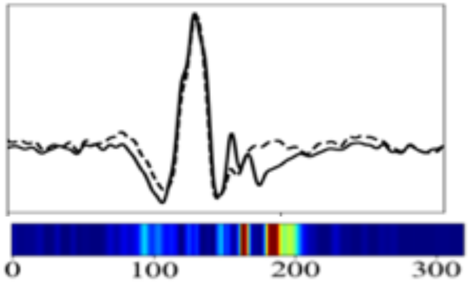
- Understand what a particular types AI technology can or can't do
  - as good as training data
  - specifying the context of its proper use
- **Humans + Machines**
  - Detecting anomaly, and assisting human for diagnosing
- Ensuring privacy, security, and ethics





中国科学院
CHINESE ACADEMY OF SCIENCES





Q&A

THANK YOU

Thanks to