

Machine Learning for Healthcare

6.871, HST.956

Lecture 9 – Small to big Data: Case Studies with Physiological Time-series

David Sontag



Course announcements

- We decided to have only 4 problem sets
- PS3 released by Mon 3/9, due Fri 3/20
 - Interpretability, learning to defer, dataset shift
- PS4 released Fri 3/20 and due Wed 4/8
 - Fairness, causal inference
- Project proposals due Wed 4/1
 - Guidelines to be released early next week
 - Teams should be formed by 3/17

Roadmap

- **Module 1: Overview of clinical care & data (3 lectures)**
- **Module 2: Using ML for risk stratification and diagnosis (9 lectures)**
 - Learning with noisy labels, Interpretability, Learning to Defer, Uncertainty
 - **Applications to** Clinical NLP and **Physiological Time-Series**
 - Dataset Shift, Fairness
- **Module 3: Suggesting treatments (4 lectures)**
 - Causal inference; Off-policy reinforcement learning

QUIZ

- **Module 4: Understanding disease and its progression (3 lectures)**
 - Unsupervised learning on censored time series with substantial missing data
 - Discovery of disease subtypes; Precision medicine
- **Module 5: Human factors (3 lectures)**
 - Differential diagnosis; Utility-theoretic trade-offs
 - Automating clinical workflows
 - Translating technology into the clinic

Physiological time-series

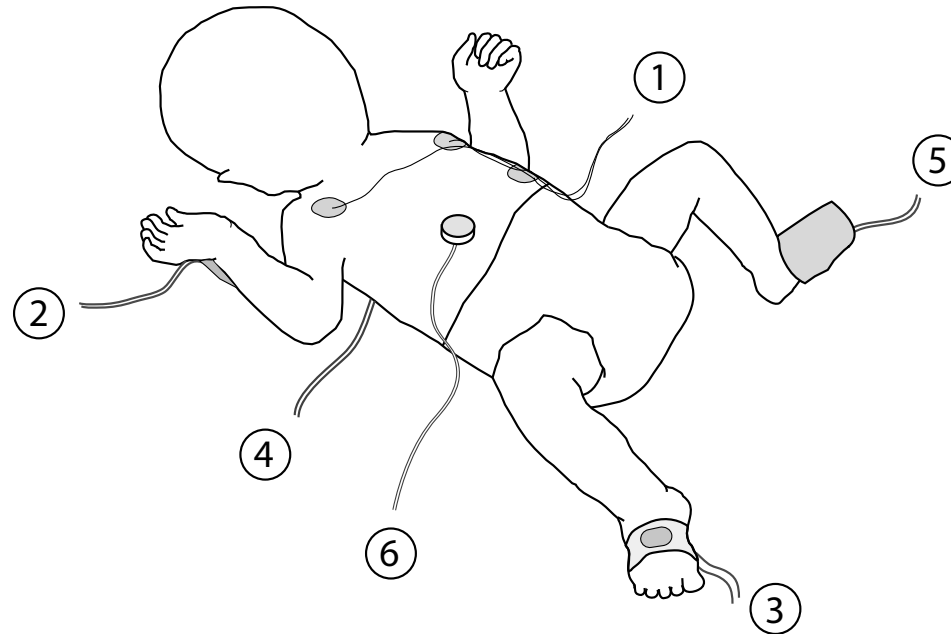
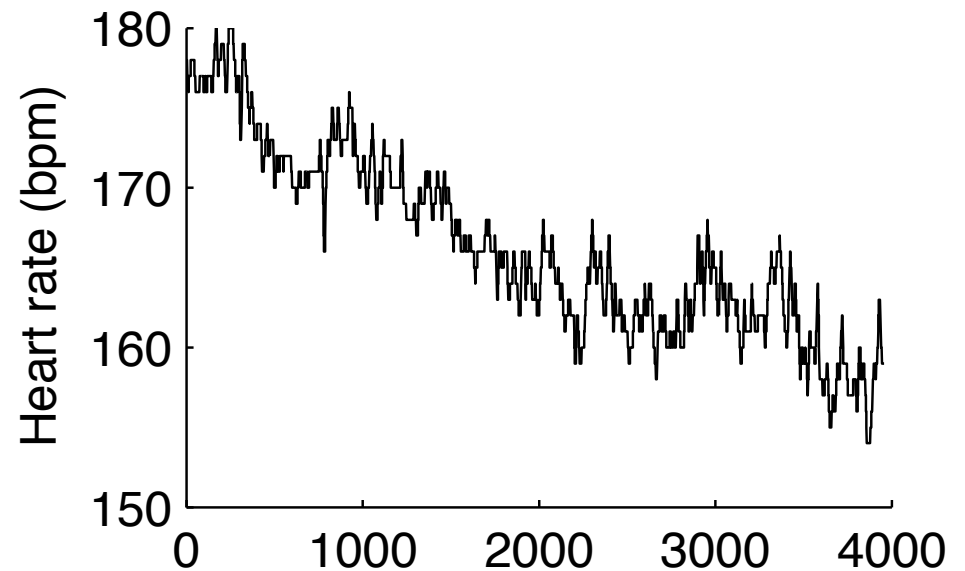
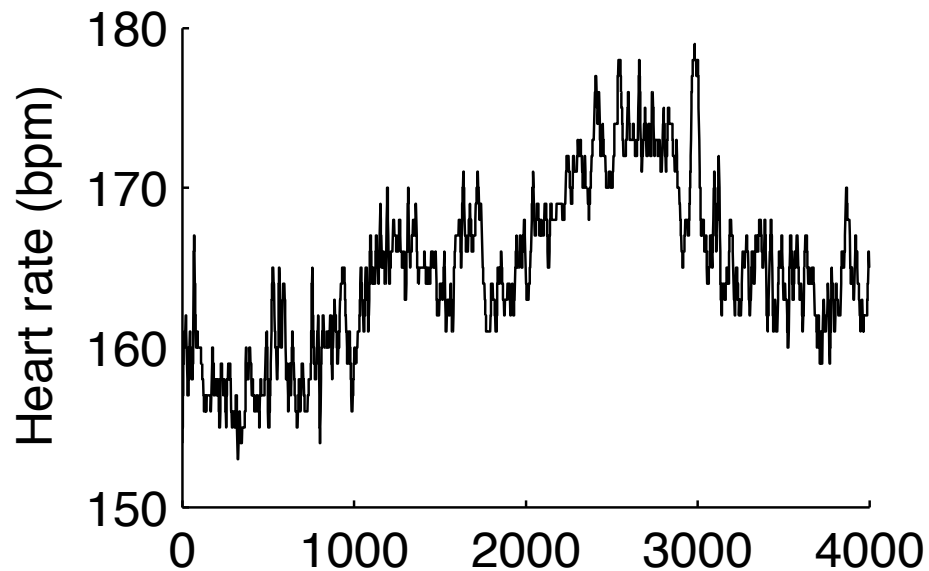


Fig. 4. Probes used to collect vital signs data from an infant in intensive care. 1) Three-lead ECG, 2) arterial line (connected to blood pressure transducer), 3) pulse oximeter, 4) core temperature probe (underneath shoulder blades), 5) peripheral temperature probe, 6) transcutaneous probe.

Physiological time-series

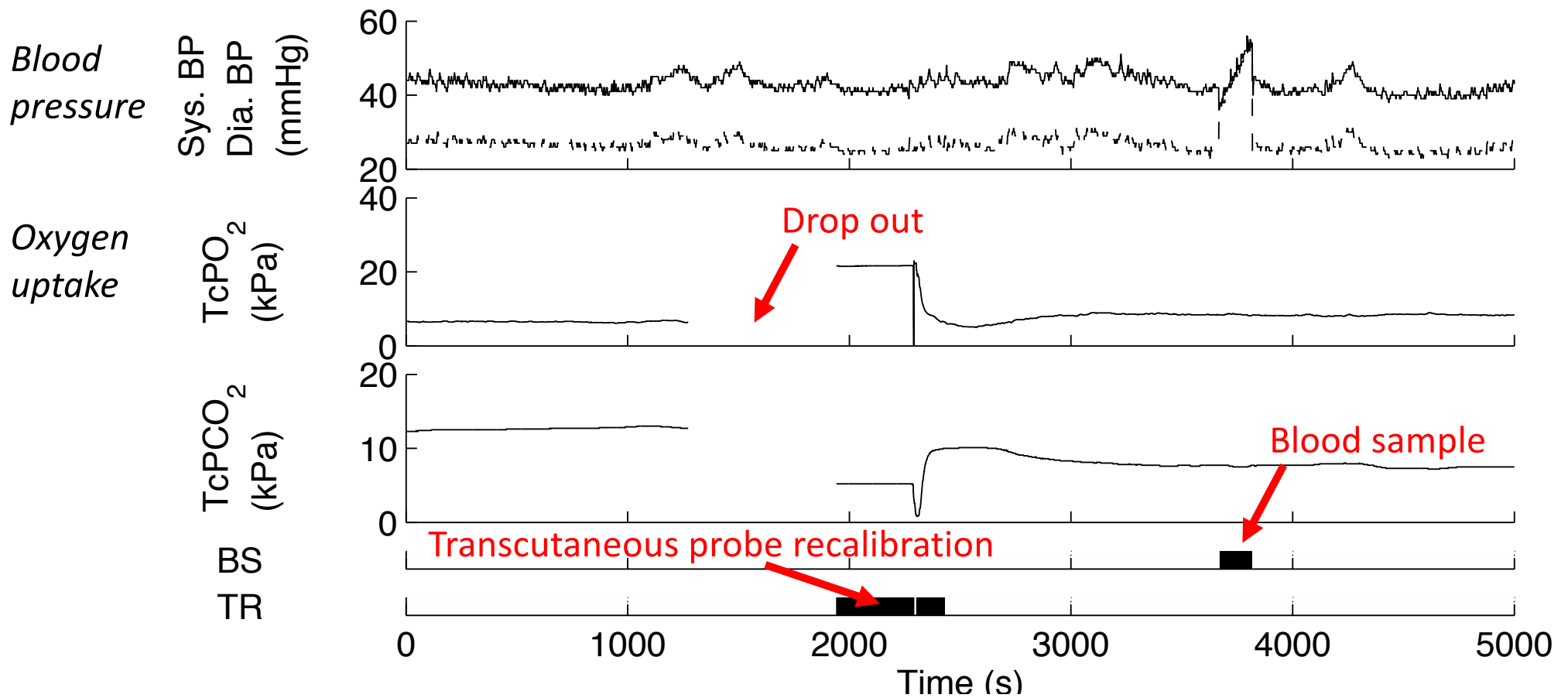
- Typical use cases:
 1. Infer true physiological signal from noisy observations
 2. Risk stratification, e.g. predict clinical deterioration, or diagnosis
- Approach taken depends on:
 - Is labeled data available?
 - Do we have a good mechanistic/statistical model?
 - How much training data is there?

Why is using it hard? High-dimensional, noisy, trajectories



(Quinn et al., TPAMI 2008)

Problem: measurements confounded by interventions & measurement errors



(Quinn et al., TPAMI 2008)

Can we identify the artifactual processes?

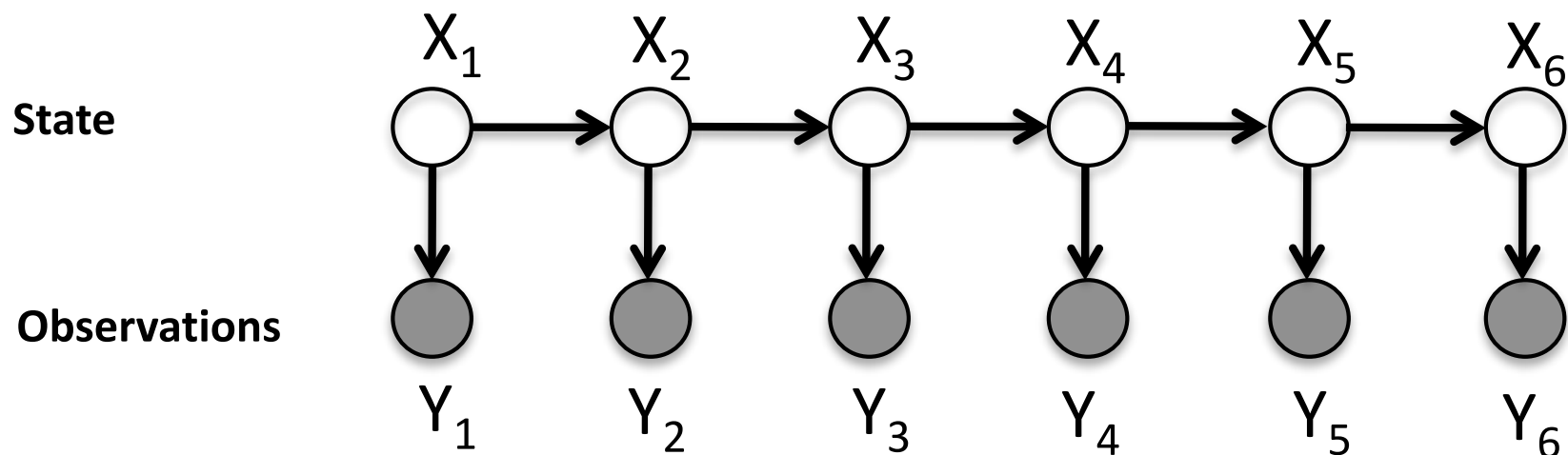
- Once identified, can remove for use in downstream predictive tasks (must deal with missing data)
- Can help mitigate **alarm fatigue** by not alerting the clinicians when unnecessary
- More broadly, can we maintain beliefs about the true physiological values of a patient?

(Switching) linear dynamical systems

- Conditioned on s_t , linear Gaussian state-space models (Kalman filters):

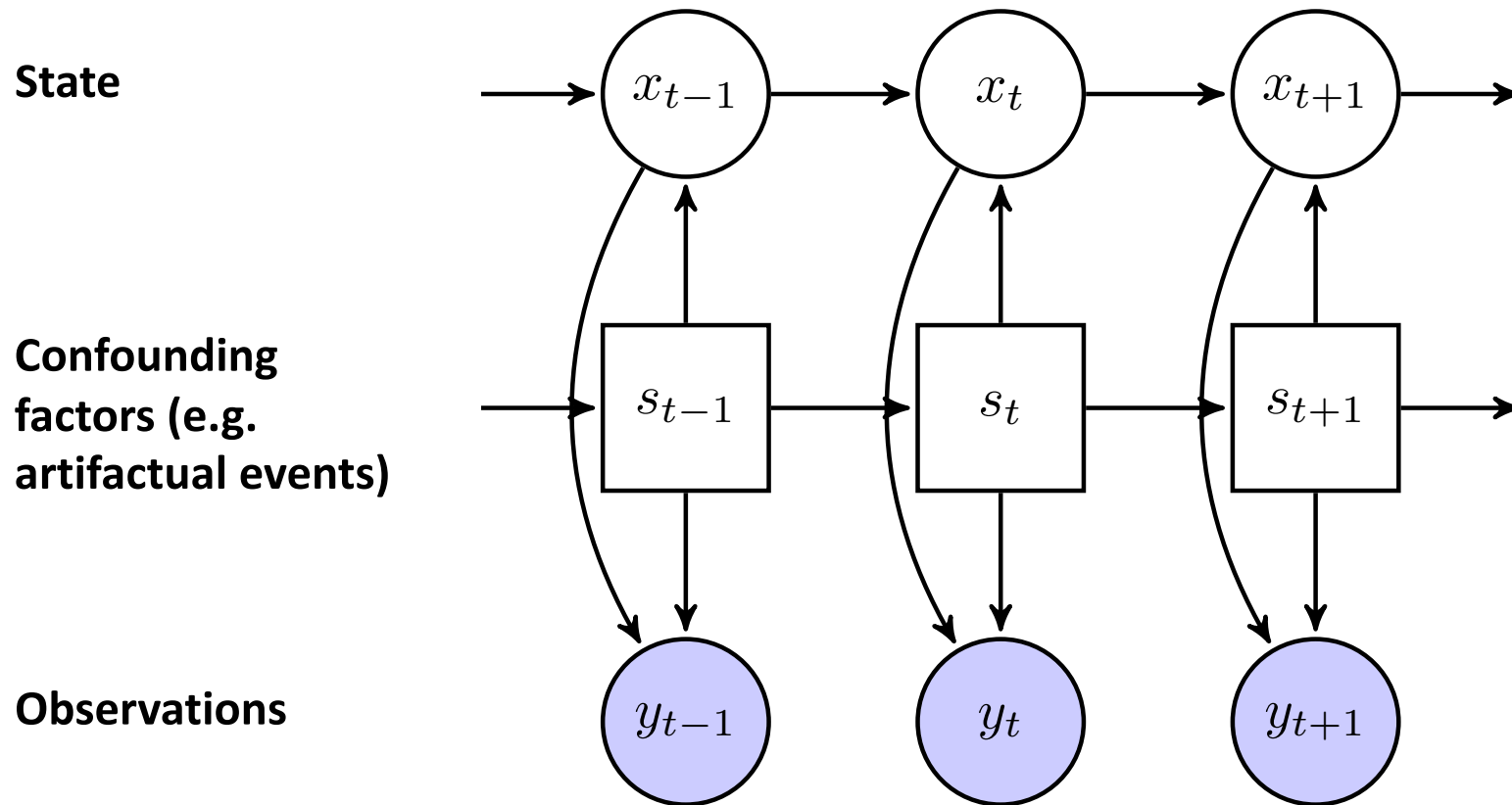
$$\mathbf{x}_t \sim \mathcal{N} \left(\mathbf{A}^{(s_t)} \mathbf{x}_{t-1} + \mathbf{d}^{(s_t)}, \mathbf{Q}^{(s_t)} \right)$$

$$\mathbf{y}_t \sim \mathcal{N} \left(\mathbf{C}^{(s_t)} \mathbf{x}_t, \mathbf{R}^{(s_t)} \right)$$



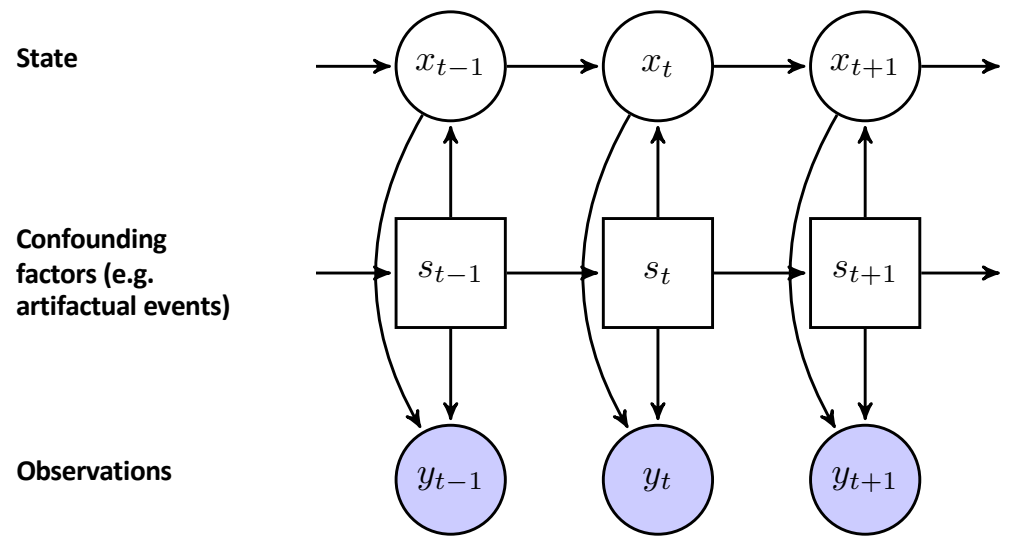
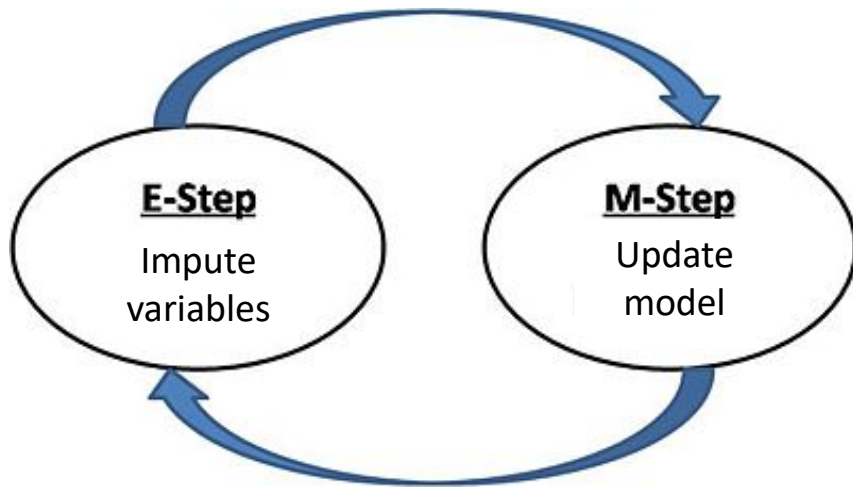
(Switching) linear dynamical systems

- Full model:



Learning SLDS models

- Assume some labeled training data $\{\mathbf{s}, \mathbf{y}\}$
- *True state \mathbf{x} assumed to never be observed*
- Learn using expectation maximization



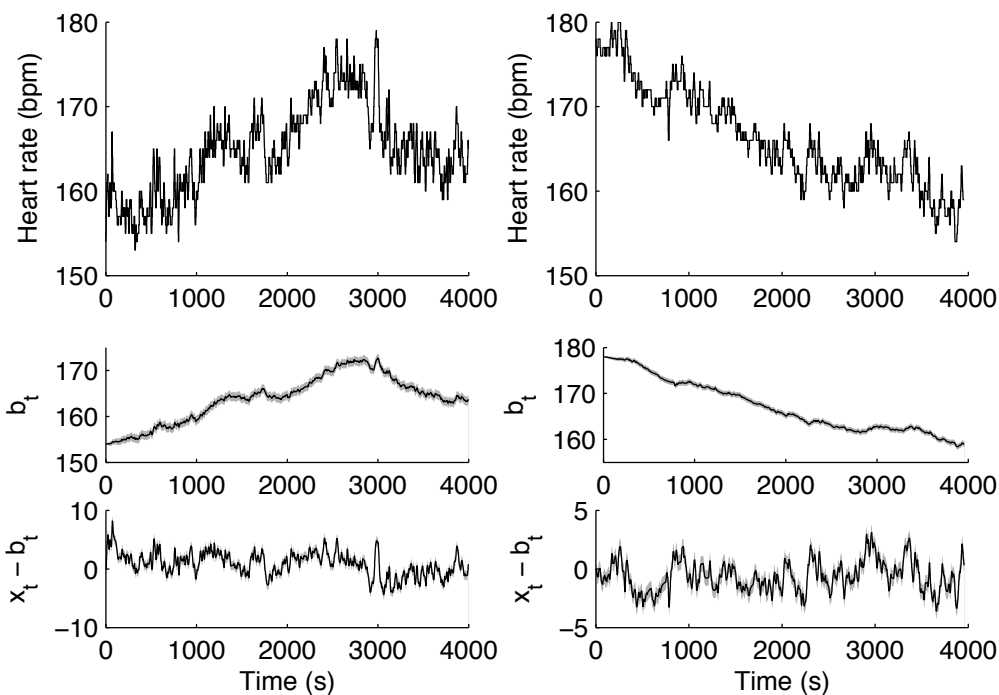
Parameterizing model

- Normal heart rate dynamics are well-modeled using an autoregressive process, e.g.

$$x_t - b_t \sim \mathcal{N} \left(\sum_{k=1}^{p_1} \alpha_k (x_{t-k} - b_{t-k}), \eta_1 \right)$$
$$b_t \sim \mathcal{N} \left(\sum_{k=1}^{p_2} \beta_k b_{t-k}, \eta_2 \right),$$

Baseline process (smooth) \longrightarrow

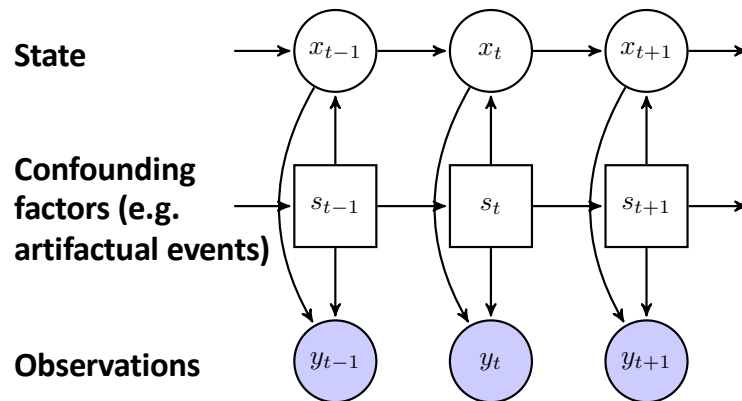
Zero-mean, high frequency \longrightarrow



(Quinn et al., TPAMI 2008)

Parameterizing model

- One can use domain knowledge to specify parts of the artifacts model
 - Probe dropouts modeled by removing dependence of observation y_t on patient state x_t
 - Temperature probe disconnection: exponential decay to room temperature



(Quinn et al., TPAMI 2008)

Evaluation

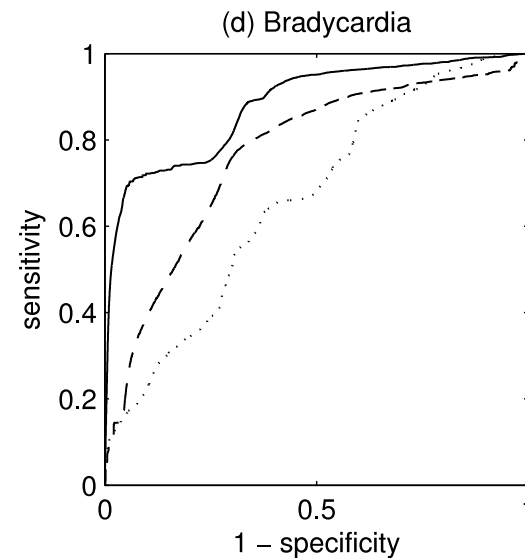
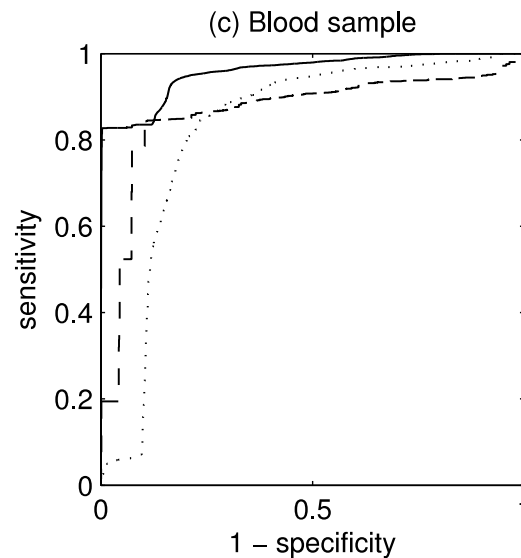
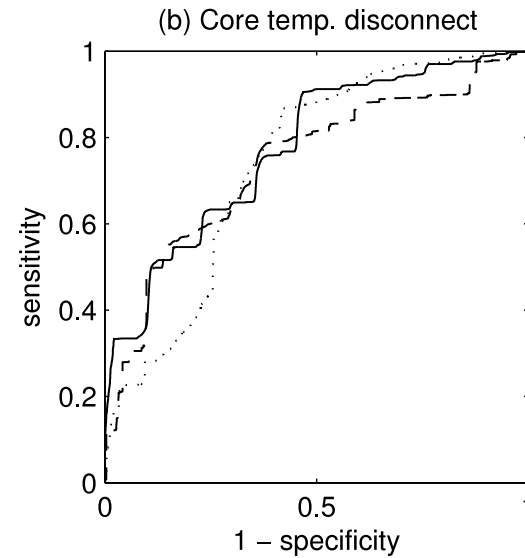
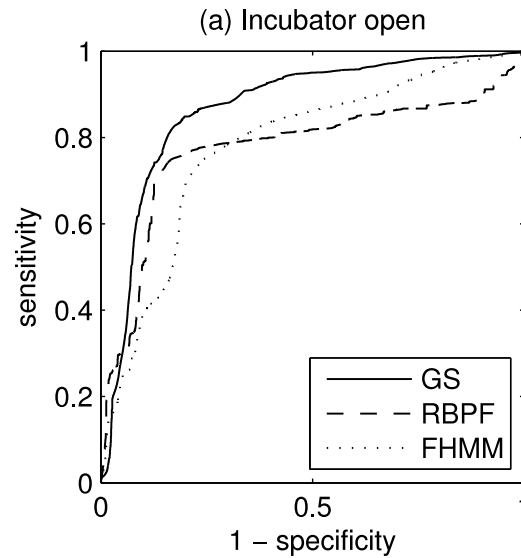
- 3-fold cross validation, where for each fold train on 10 babies and test on 5
- 24-hours of data for each baby
- Normal dynamics refit for test babies using a 30-minute section near the start

Evaluation

GS = Gaussian-sum approximation (used for inference)

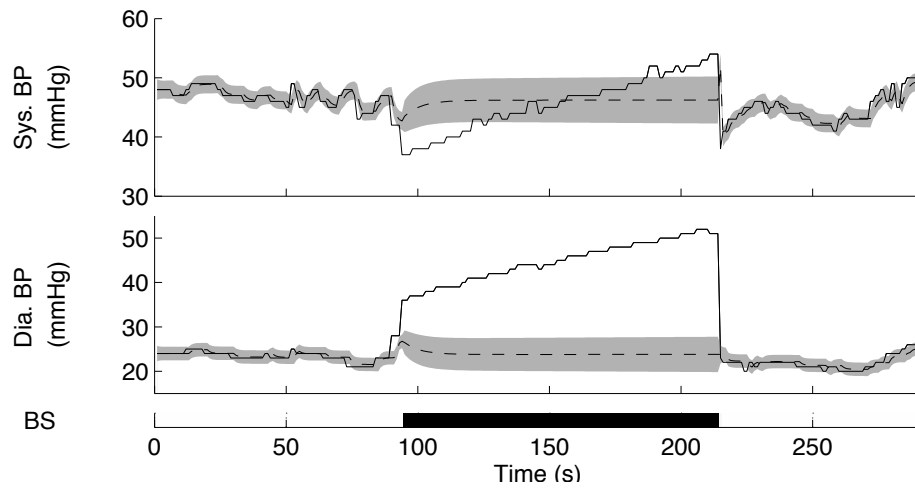
RBPF = Rao-Blackwellized particle filtering approximation (used for inference)

FHMM = Factorial HMM (simpler model which does not model normal physiological dynamics)

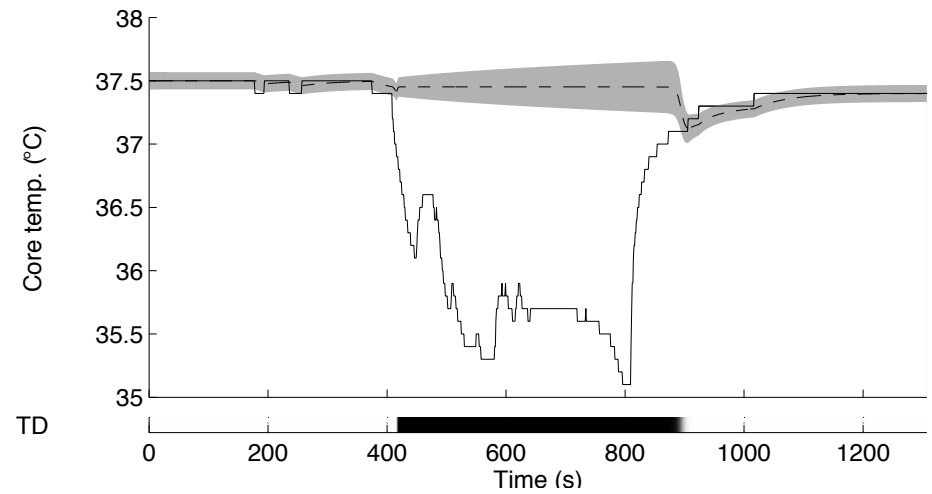


(Quinn et al., TPAMI 2008)

Inference of physiological state

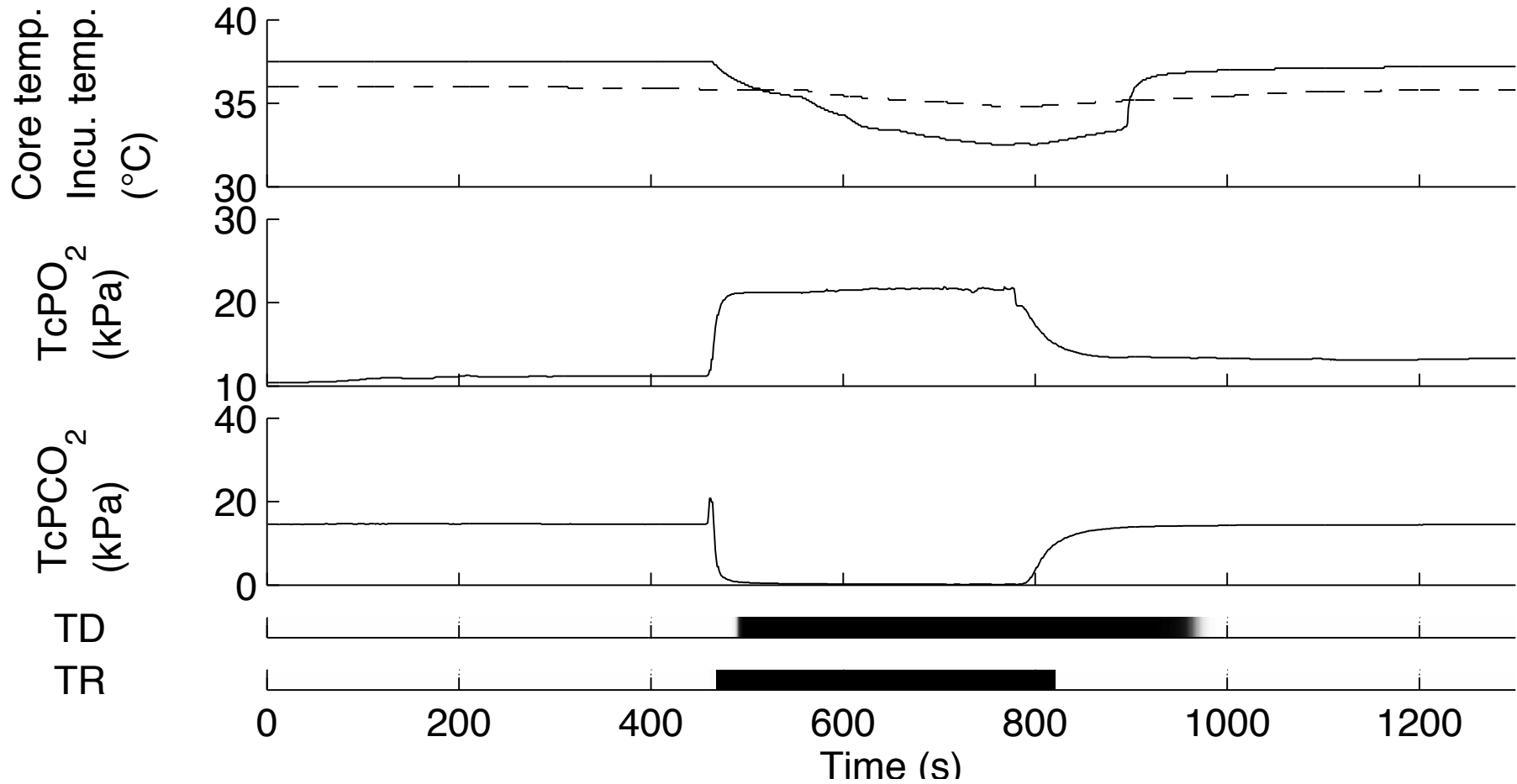


Blood sample draw



Temperature probe disconnection

Inferred switch settings



TD= core temperature probe disconnection

TR = recalibration

Outline of today's lecture

- Monitoring babies in neonatal ICUs
- **Detecting irregular heart arrhythmias**
 - Small data approach
 - Big data approach

Detecting atrial fibrillation



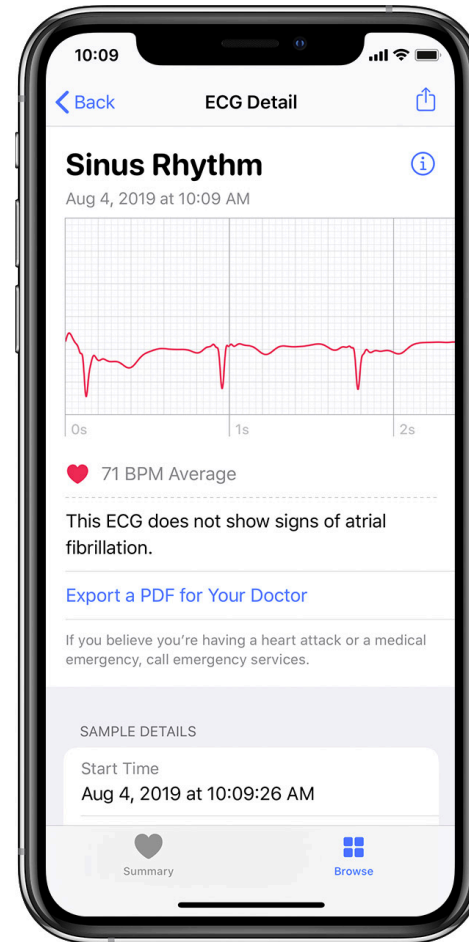
AliveCore ECG
device

ECG = electrocardiogram

Detecting atrial fibrillation

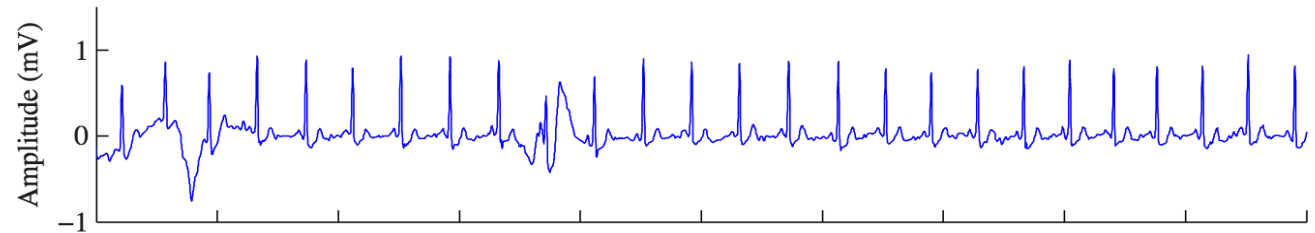


Apple Watch

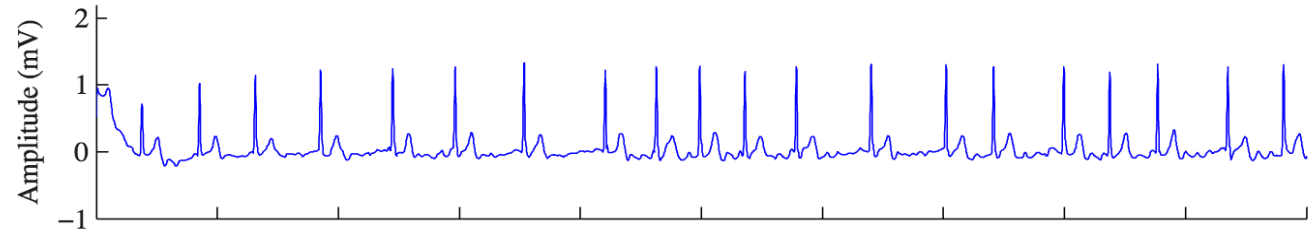


What type of heart rhythm?

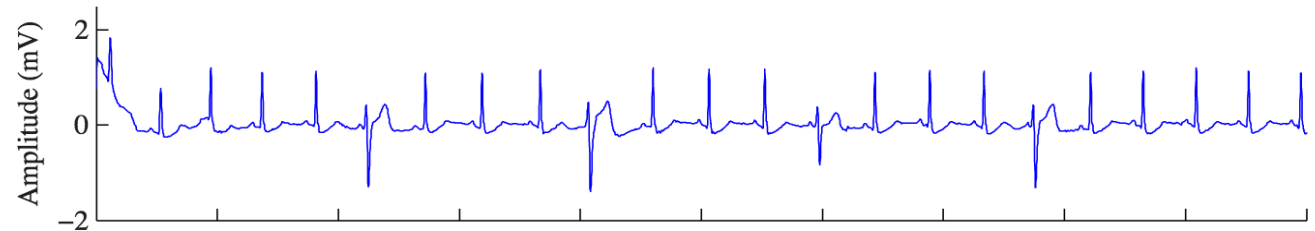
Normal rhythm



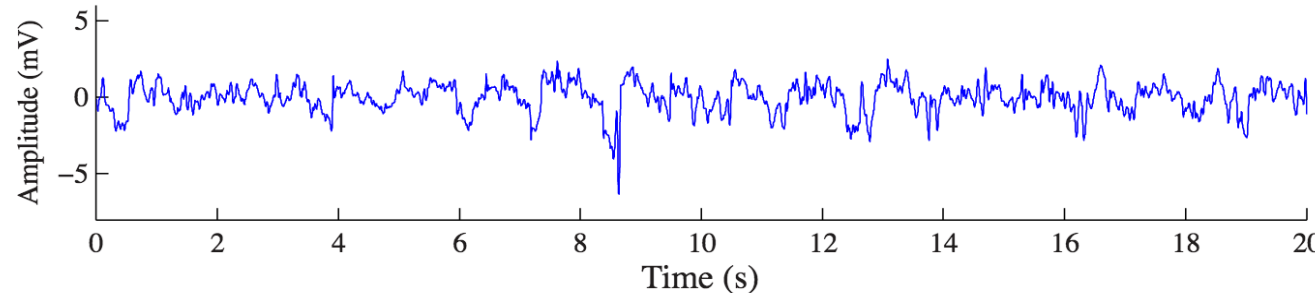
AF rhythm

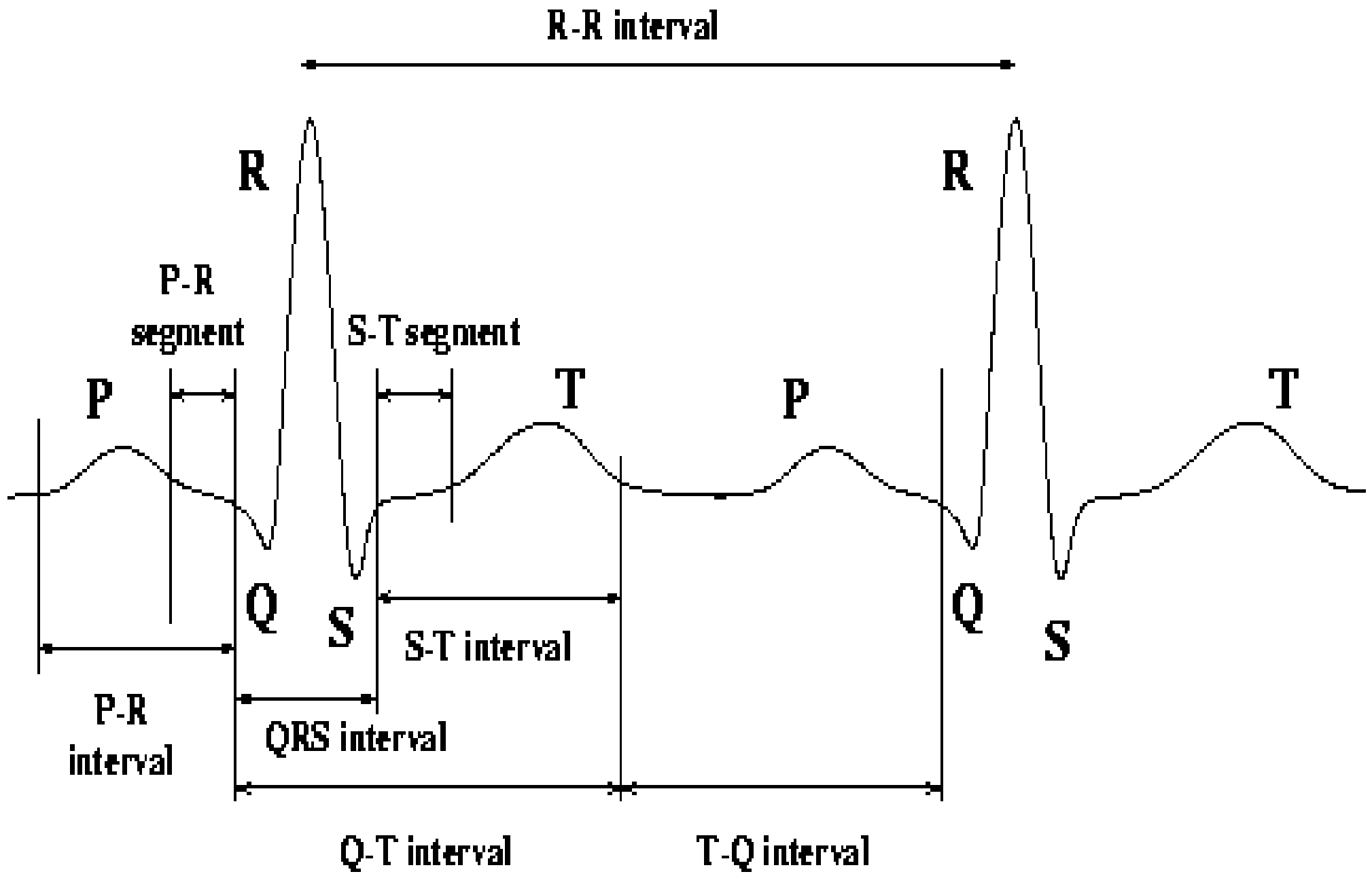


Other rhythm

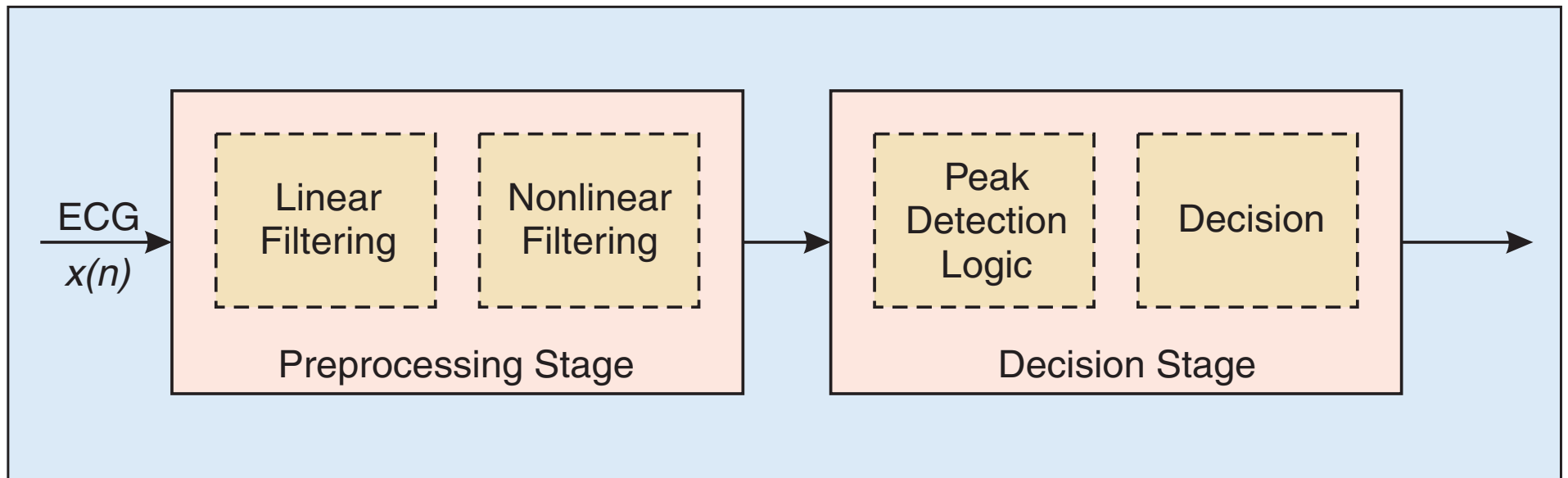


Noisy recording



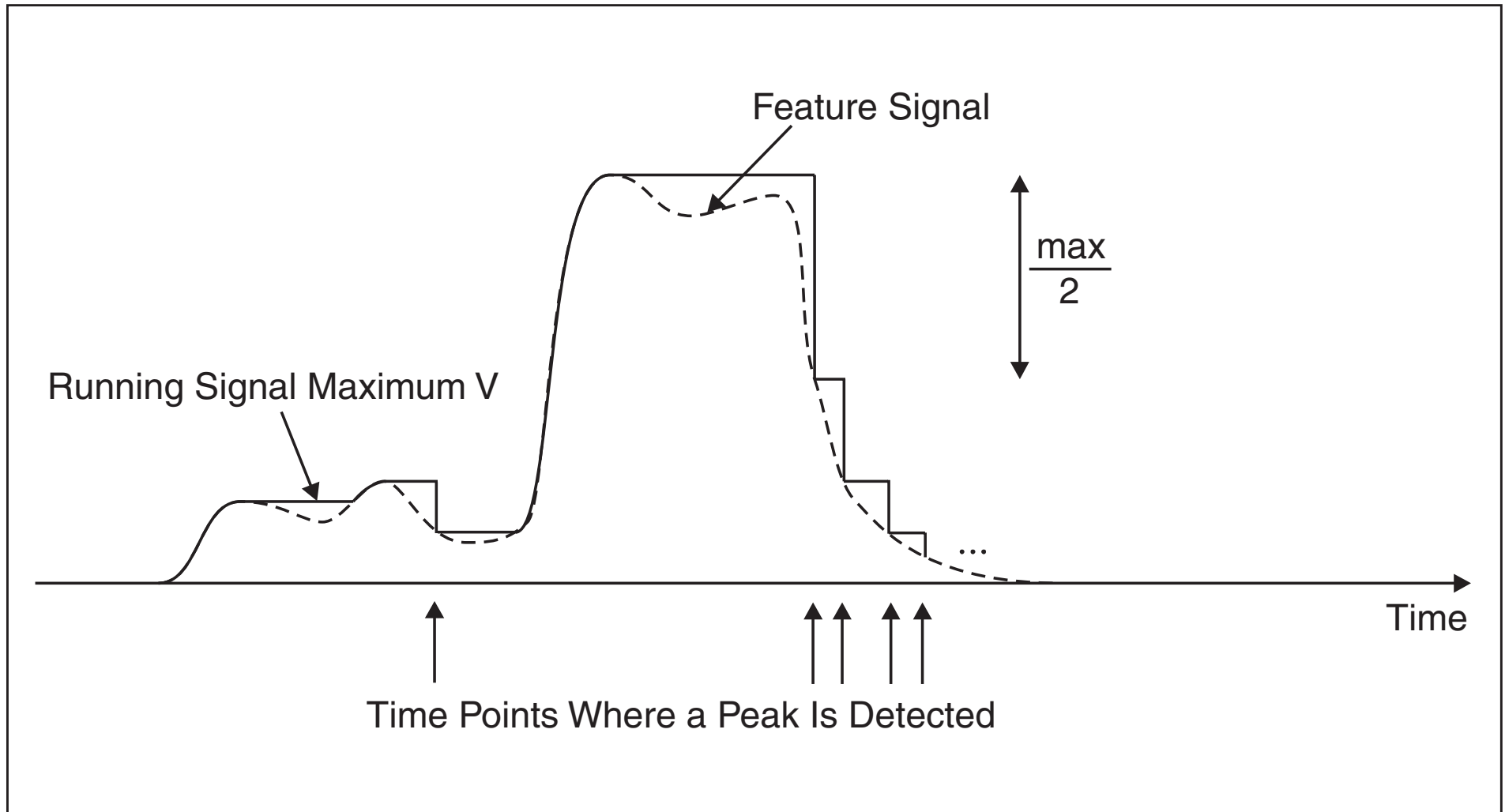


Traditional approach



2. Common structure of the QRS detectors.

[Kohler, Hennig, Orglmeister. The Principles of Software QRS Detection, IEEE Engineering in Medicine & Biology, 2002]



3. Peak detector proposed in [41].

[Kohler, Hennig, Orglmeister. The Principles of Software QRS Detection, IEEE Engineering in Medicine & Biology, 2002]

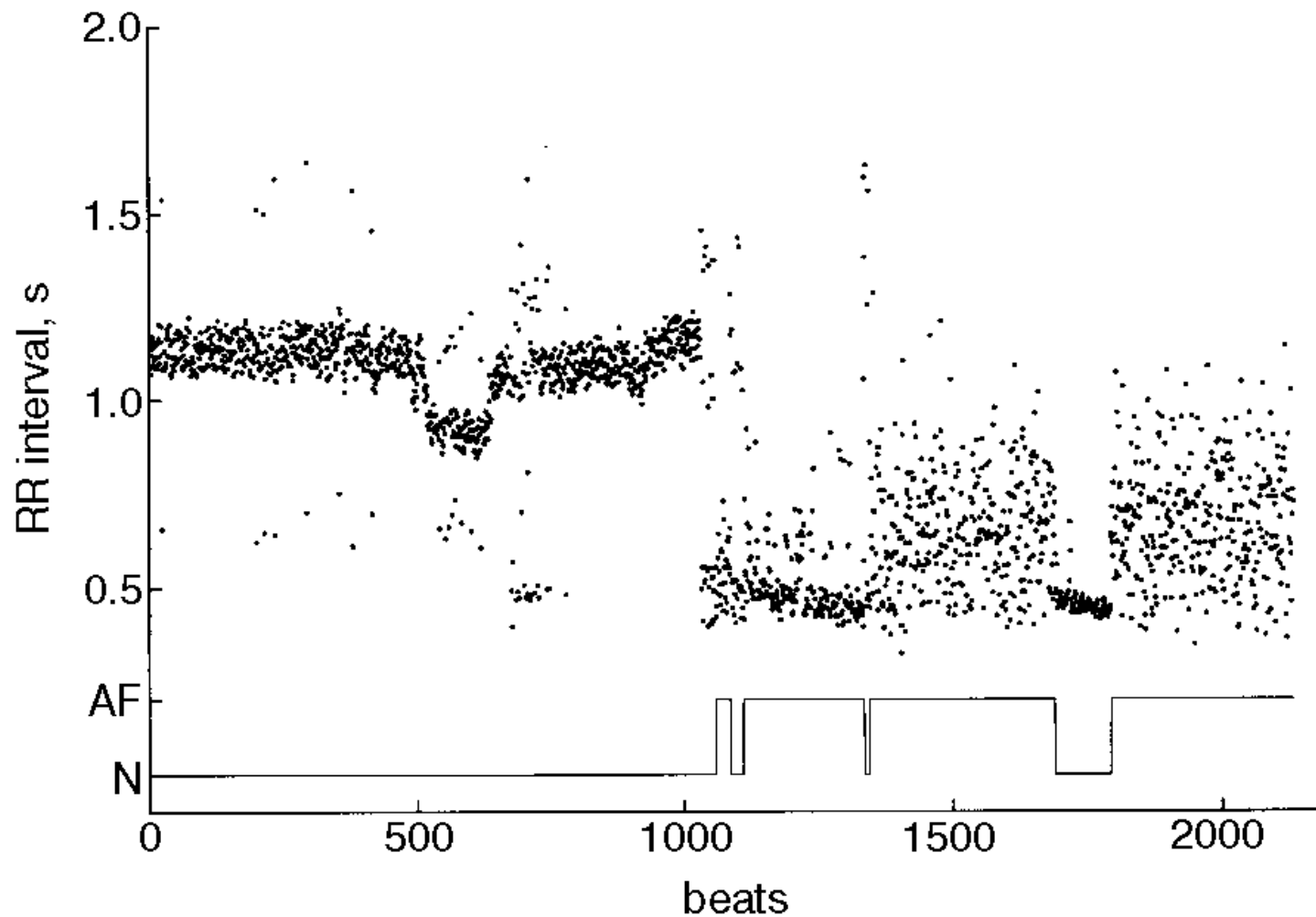


Fig. 1 *Time series showing RR intervals from subject 202 from MIT-BIH arrhythmia database. (—) Assessment of atrial fibrillation (AF) or non-atrial fibrillation (N) as reported in database*

[Tateno & Glass, Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and Δ RR intervals. MBEC, 2001]

Cardiac **Arrhythmia Classification:**
A Heart-Beat Interval-Markov Chain Approach *

WILL GERSCH,† DAVID M. EDDY,‡ AND EUGENE DONG, JR.§

*Division of Cardiovascular Surgery, Department of Surgery, Stanford University
Medical Center, Stanford, California 94305*

Received March 2, 1970

A sequence of heart-beat intervals (R-R wave intervals) is automatically transformed into a three-symbol Markov chain sequence. For convenience the symbols used may be thought of as S-R-L for short, regular, and long heart-beat intervals, respectively. The **probability** that the observed sequence was generated by each of a set of prototype models characteristic of different cardiac disorders is computed. That prototype corresponding to the largest probability of observed sequence generation is designated as the disorder. This procedure is the equivalent of **Kullback's** classification by the minimization of directed divergence procedure.

In a **preliminary** experiment **primarily** using data sequences of 100 heart-beat intervals, 35 different known cases were automatically classified into six cardiac disorders without error. The disorders considered were **atrial fibrillation, APC** and VPC, bigeminy, sinus tachycardia with occasional bigeminy, sinus tachycardia, and ventricular tachycardia.

An automatic procedure to classify cardiac **arrhythmias** using a Markov chain interpretation of heart-beat interval **data** is reported. A sequence of heart-beat

Detection of Atrial Fibrillation Using Artificial Neural Networks

SG Artis, RG Mark, GB Moody

Harvard-MIT

Division of Health Sciences and Technology, Cambridge, MA

Abstract

Artificial neural networks (ANNs) were used as pattern detectors to detect atrial fibrillation (AF) in the MIT-BIH Arrhythmia Database. ECG data was represented using generalized interval transition matrices, as in Markov model AF detectors[1]. A training file was developed, using these transition matrices, for a backpropagation ANN. This file consisted of approximately 15 minutes each of AF and non-AF data. The ANN was successfully trained using this data. Three standard databases were used to test network performance. Post-processing of the ANN output yielded an AF sensitivity of 92.86% and an AF positive predictive accuracy of 92.34%.

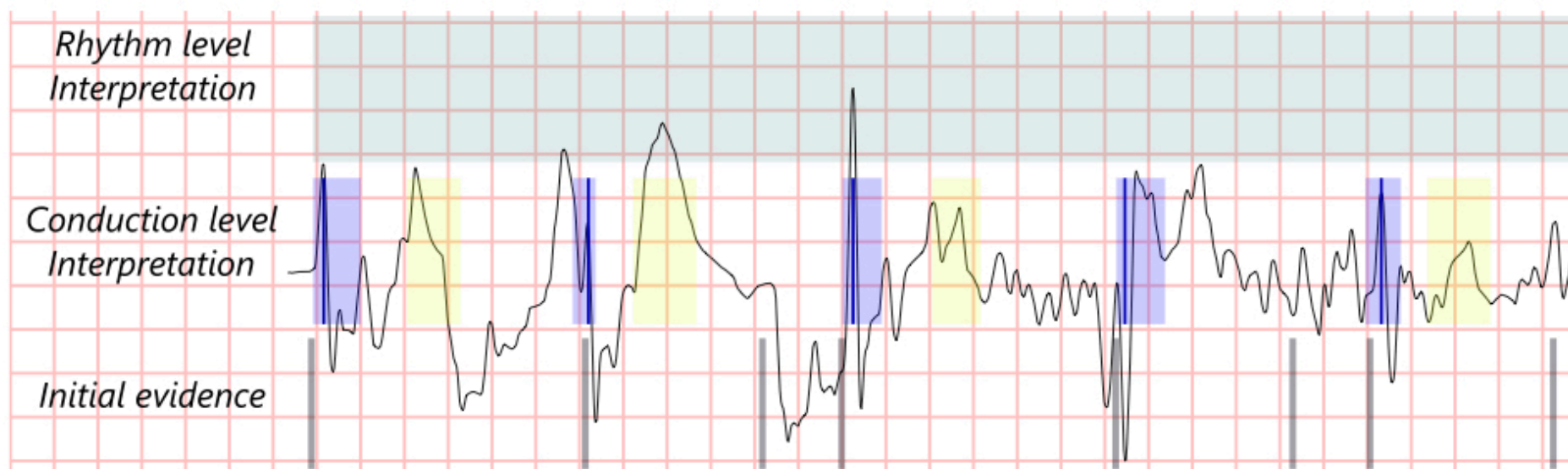
1 Introduction

on R-R interval sequences using a variety of statistical methods [1] but there is room for improvement in these techniques.

Pattern classifiers exist in many forms, and artificial neural networks (ANNs) represent an important subset of these classifiers. ANNs are attractive for solving pattern recognition problems because few assumptions about the underlying data need to be made. The task of the operator of an ANN is to separate the data into subsets. The network will be able classify these subsets according to type as long as they are distinct. Neural network training requires appropriate training data, pre-processing and post-processing algorithms, an appropriate network topology, and a training algorithm, as well as evaluation databases. This document will present the design and evaluation of a technique which detects AF in the presence of other cardiac arrhythmias using a backpropagation artificial neural network.

Winning approach

- Training data in 2017 Physionet challenge: ~8500 ECGs
- Best algorithms use a combination of expert-derived features and machine learning



[Teijeiro, Garcia, Castro, Felix. arXiv:1802.05998, 2018]

Table 1: Set of features used to train the global classifier

tSR: Proportion of the record length interpreted as a regular rhythm (Normal rhythm, tachycardia or bradycardia).	t1b: Number of milliseconds from the beginning of the record to the first interpreted heartbeat.
tOR: Number of milliseconds interpreted as a non-regular rhythm.	longTch: Longest period of time with heart rate over 100bpm.
RR: Median RR interval of regular rhythms.	RRd_std: Standard deviation of the instant RR variation.
RRd: Median Absolute Deviation (MAD) of the RR interval in regular rhythms.	MRRd: Max. absolute variation of the RR interval in regular rhythms.
RR_MIrr: Max. RR irregularity measure.	RR_Irr: Median RR irregularity measure.
PNN{10,50,100}: Global PNNx measures.	o_PNN50: PNN50 of non-regular rhythms.
mRR: Min. RR interval of regular rhythms.	o_mRR: Min. RR interval of non-regular rhythms.
n_nP: Proportion of heartbeats with detected P-wave inside regular rhythms.	n_aT: Median of the amplitude of the T waves inside regular rhythms.
n_PR: Median PR duration inside regular rhythms.	Psmooth: Median of the ratio between the standard deviation and the mean value of P-waves' derivative signal.
Pdist: MAD of the measure given by the P wave delineation method.	MPdist: Max. of the measure given by the P wave delineation method.
prof: Profile of the full signal.	pw_prof: MAD of pw_prof .
xcorr: Median of the maximum cross-correlation between QRS complexes interpreted in regular rhythms.	o_xcorr: Median of the maximum cross-correlation between QRS complexes interpreted in non-regular rhythms.
PRd: Global MAD of the PR durations.	QT: Median of the corrected QT measure.
TP: Median of the prevailing frequency in the TP intervals.	TPfreq: Median of the frequency entropy in the TP intervals.
pw_prof: Profile measure of the signal in the P-wave area.	nT: Proportion of QRS complexes with detected T waves.
n_Txcorr: Median of the maximum cross-correlation between T-waves inside regular rhythms.	n_Pxcorr: Median of the maximum cross-correlation between P-waves inside regular rhythms.
baseline: Profile of the baseline in regular rhythms.	o_baseline: Profile of the baseline in non-regular rhythms.
wQRS: Proportion of wide QRS complexes (duration longer than 110ms).	wQRS_xc: Median of the maximum cross-correlation between wide QRS complexes.
wQRS_prof: Median of the signal profile in the 300ms before each wide QRS complex.	w_PR: Proportion of heartbeats with long PR interval (longer than 210 ms).
x_xc: Median of the maximum cross-correlation between ectopic beats.	x_rrel: Median of the ratio between the previous and next RR intervals for each ectopic beat.

[Teijeiro, Garcia, Castro, Felix. arXiv:1802.05998, 2018]

Not enough data for deep learning? Wrong architectures?

“However, the fact that a standard random forest with well chosen features performed as well as more complex approaches, indicates that perhaps a set of 8,528 training patterns was not enough to give the more complex approaches an advantage. With so many parameters and hyperparameters to tune, the search space can be enormous and significant overtraining was seen...”

[Clifford et al. AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge, Computing in Cardiology 2017]

Secure | <https://stanfordmlgroup.github.io/projects/ecg/>

Stanford ML Group

Cardiologist-Level Arrhythmia Detection With Convolutional Neural Networks

Pranav Rajpurkar*, Awni Hannun*, Masoumeh Haghpanahi, Codie Bourn, and Andrew Ng

A collaboration between Stanford University and iRhythm Technologies

We develop a model which can diagnose irregular heart rhythms, also known as arrhythmias, from single-lead ECG signals better than a cardiologist.

Key to exceeding expert performance is a deep convolutional network which can map a sequence of ECG samples to a sequence of arrhythmia annotations along with a novel dataset two orders of magnitude larger than previous datasets of its kind.



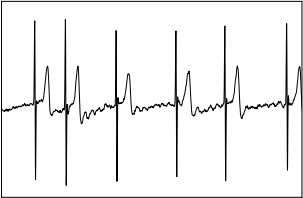
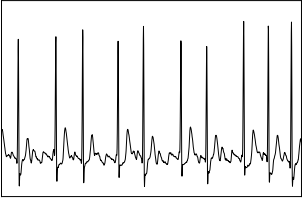
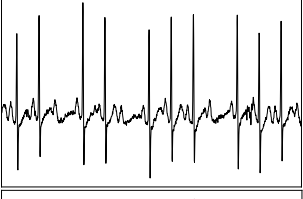


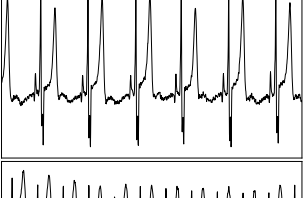
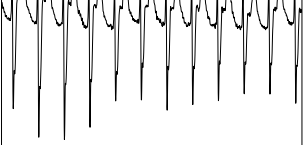
[Rajpurkar et al., arXiv:1707.01836, 2017; Nature Medicine '19]

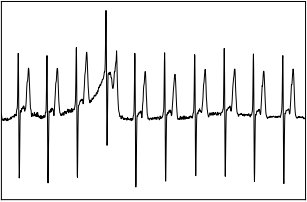
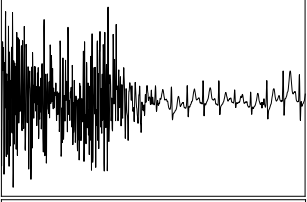
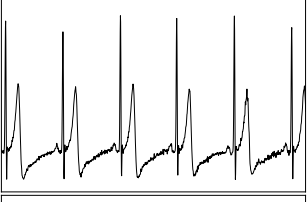
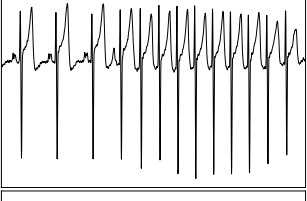
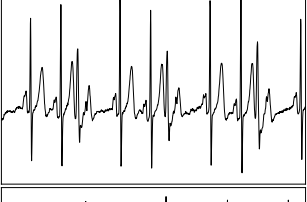
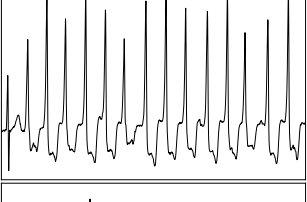
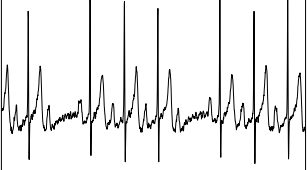
Differences with previous work

- Sensor is a Zio patch – conceivably much less noisy:



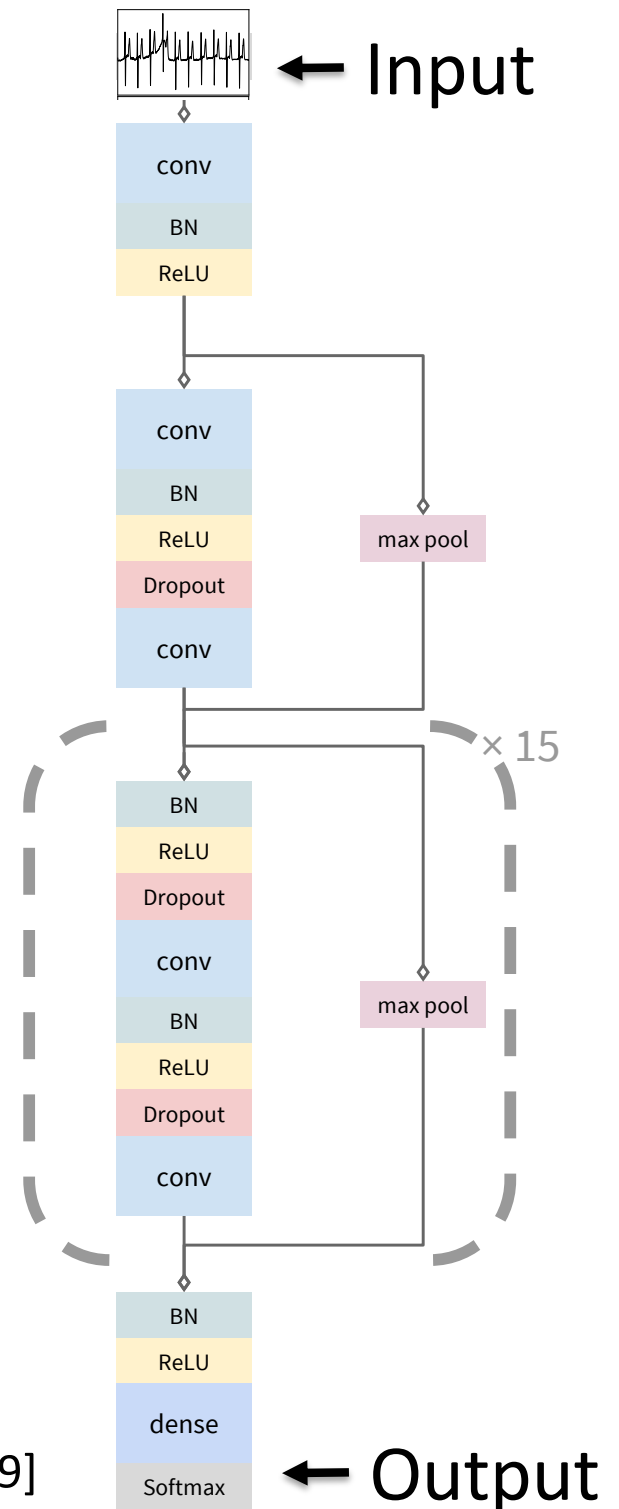
- ~90K ECG records annotated (from ~50K patients)
- Identify 12 heart arrhythmias, sinus rhythm and noise for a total of 14 output classes

Class	Description	Example	Train + Val Patients	Test Patients
AFIB	Atrial Fibrillation		4638	44
AFL	Atrial Flutter		3805	20
AVB_TYPE2	Second degree AV Block Type 2 (Mobitz II)		1905	28
BIGEMINY	Ventricular Bigeminy		2855	22
CHB	Complete Heart Block		843	26
EAR	Ectopic Atrial Rhythm		2623	22
IVR	Idioventricular Rhythm		1962	34

Class	Description	Example	Train + Val Patients	Test Patients
JUNCTIONAL	Junctional Rhythm		2030	36
NOISE	Noise		9940	41
SINUS	Sinus Rhythm		22156	215
SVT	Supraventricular Tachycardia		6301	34
TRIGEMINY	Ventricular Trigeminy		2864	21
VT	Ventricular Tachycardia		4827	17
WENCKEBACH	Wenckebach (Mobitz I)		2051	29

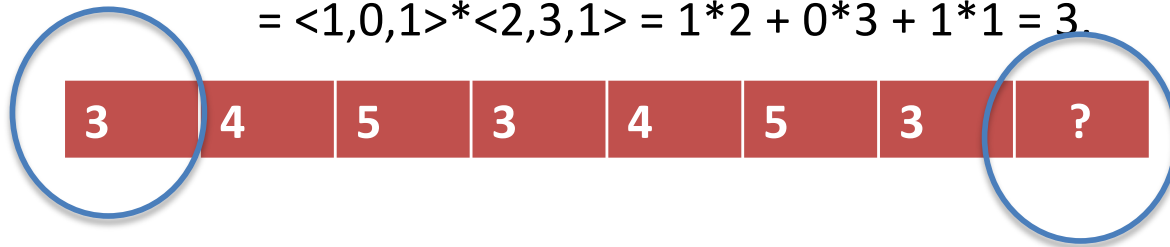
Deep convolutional network

- 1-D signal sampled at 200Hz, labeled at 1 sec intervals
- 34 layers
- Shortcut connections (ala residual networks) with max-pooling
- Subsampled every other layer (2^8 in total)



Example of 1D convolution

$$= \langle 1, 0, 1 \rangle * \langle 2, 3, 1 \rangle = 1*2 + 0*3 + 1*1 = 3.$$



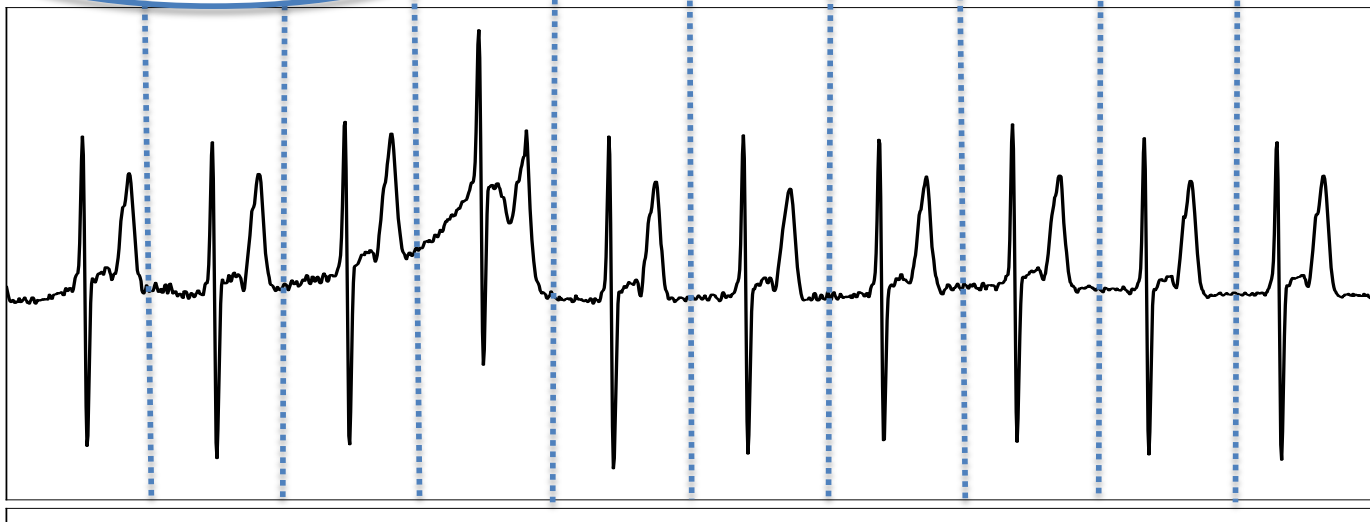
Output

Filter

2	3	1
---	---	---



Input



Evaluation

	Seq		Set	
	Model	Cardiol.	Model	Cardiol.
Class-level F1 Score				
AFIB	0.604	0.515	0.667	0.544
AFL	0.687	0.635	0.679	0.646
AVB_TYPE2	0.689	0.535	0.656	0.529
BIGEMINY	0.897	0.837	0.870	0.849
CHB	0.843	0.701	0.852	0.685
EAR	0.519	0.476	0.571	0.529
IVR	0.761	0.632	0.774	0.720
JUNCTIONAL	0.670	0.684	0.783	0.674
NOISE	0.823	0.768	0.704	0.689
SINUS	0.879	0.847	0.939	0.907
SVT	0.477	0.449	0.658	0.556
TRIGEMINY	0.908	0.843	0.870	0.816
VT	0.506	0.566	0.694	0.769
WENCKEBACH	0.709	0.593	0.806	0.736
Aggregate Results				
Precision (PPV)	0.800	0.723	0.809	0.763
Recall (Sensitivity)	0.784	0.724	0.827	0.744
F1	0.776	0.719	0.809	0.751

Summary

- We are nearly always in realm of “not enough data”
- Modeling and incorporating prior knowledge is critical to good performance
- Design principles
 - Model the distribution of physiological dynamics
 - Derive features using existing clinical knowledge
 - Start from the simplest possible model
 - Share statistical strength across tasks