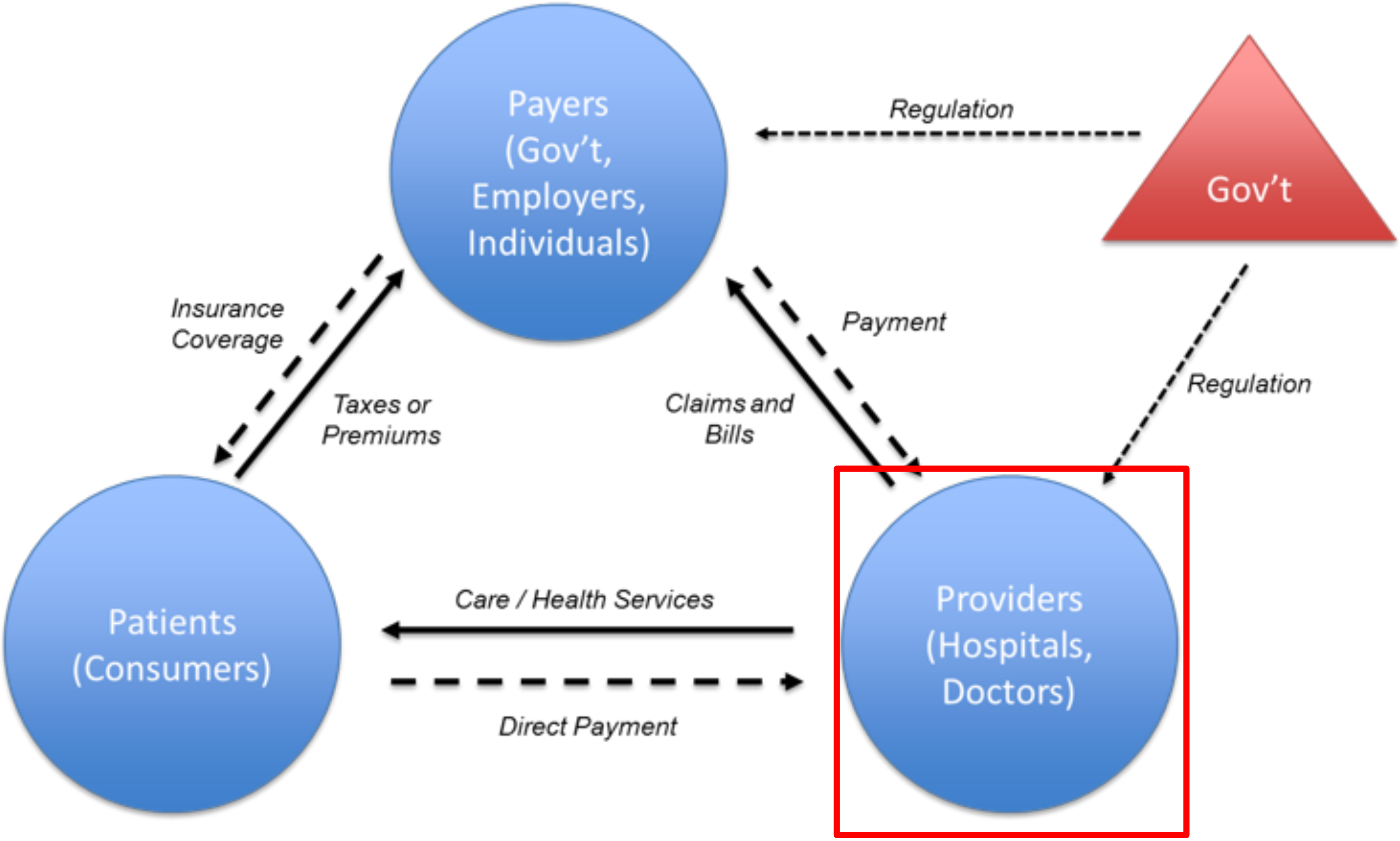
Machine Learning for Healthcare: Clinical text, vital signs, imaging

David Sontag
Clinical Machine Learning Group
MIT

CSAIL

INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

Health stakeholders



Source for figure:
http://www.mahesh-vc.com/blog/understanding-whos-paying-for-what-in-the-healthcare-industry

Outline

1. Clinical text
   – Case study: Prediction of sepsis (severe infection) from electronic health records
2. Physiological time-series
   – Case study: Monitoring babies in neonatal ICUs
   – Case study: Detecting atrial fibrillation
3. Imaging
   – Cardiology, pathology, radiology

Bulk of valuable data is in narrative text

orange=demographics
blue=patient condition, diseases, etc.
brown=procedures, tests
magenta=results of measurements
purple=time

Mr. Blind is a 79-year-old white white male with a history of diabetes mellitus, inferior myocardial infarction, who underwent open repair of his increased diverticulum November 13th at Sephsandpot Center.

The patient developed hematemesis November 15th and was intubated for respiratory distress. He was transferred to the Valtawnprinceel Community Memorial Hospital for endoscopy and esophagoscopy on the 16th of November which showed a 2 cm linear tear of the esophagus at 30 to 32 cm. The patient's hematocrit was stable and he was given no further intervention.

The patient attempted a gastrografin swallow on the 21st, but was unable to cooperate with probable aspiration. The patient also had been receiving generous intravenous hydration during the period for which he was NPO for his esophageal tear and intravenous Lasix for a question of pulmonary congestion.

On the morning of the 22nd the patient developed tachypnea with a chest X-ray showing a question of congestive heart failure. A medical consult was obtained at the Valtawnprinceel Community Memorial Hospital. The patient was given intravenous Lasix.

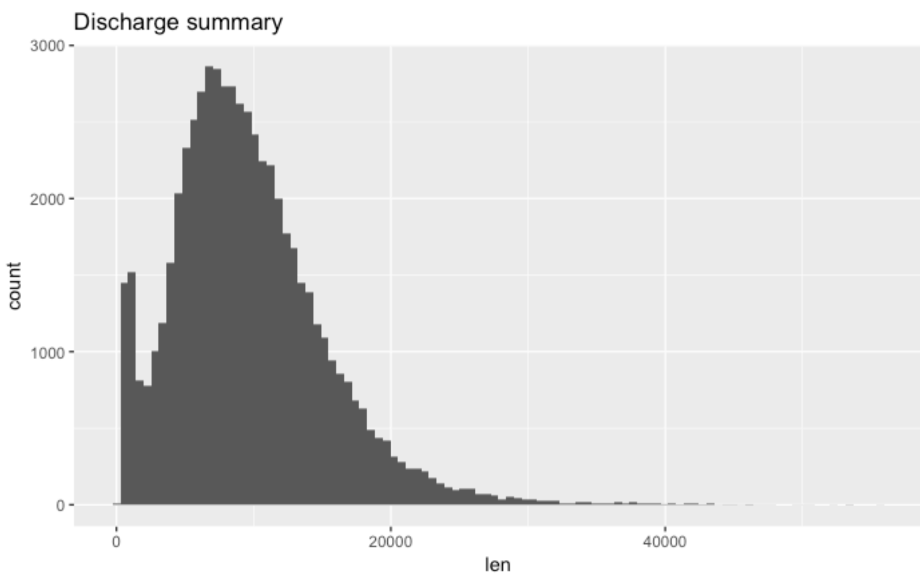
[Slide credit: Pete Szolovits]

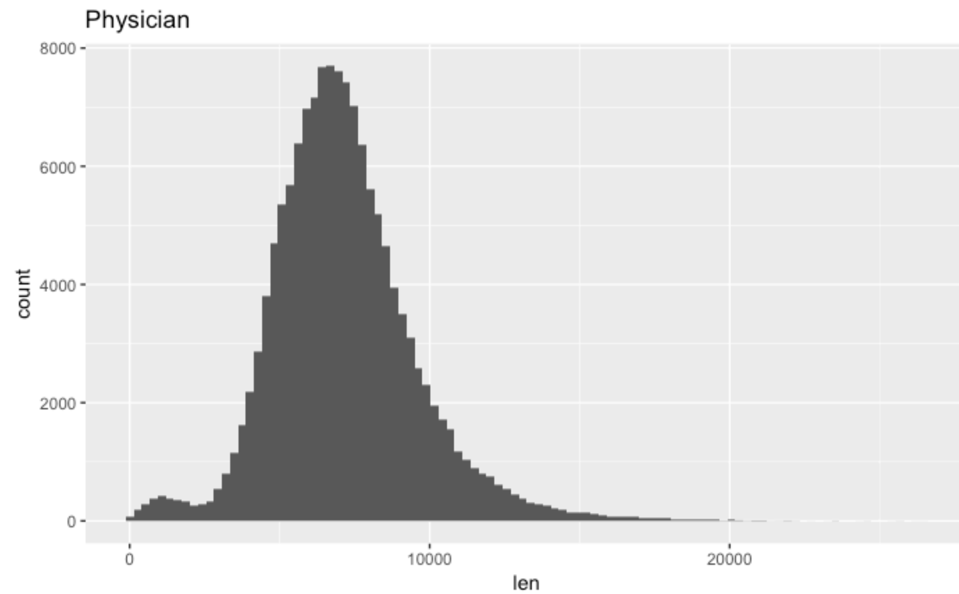
Clinical notes in MIMIC

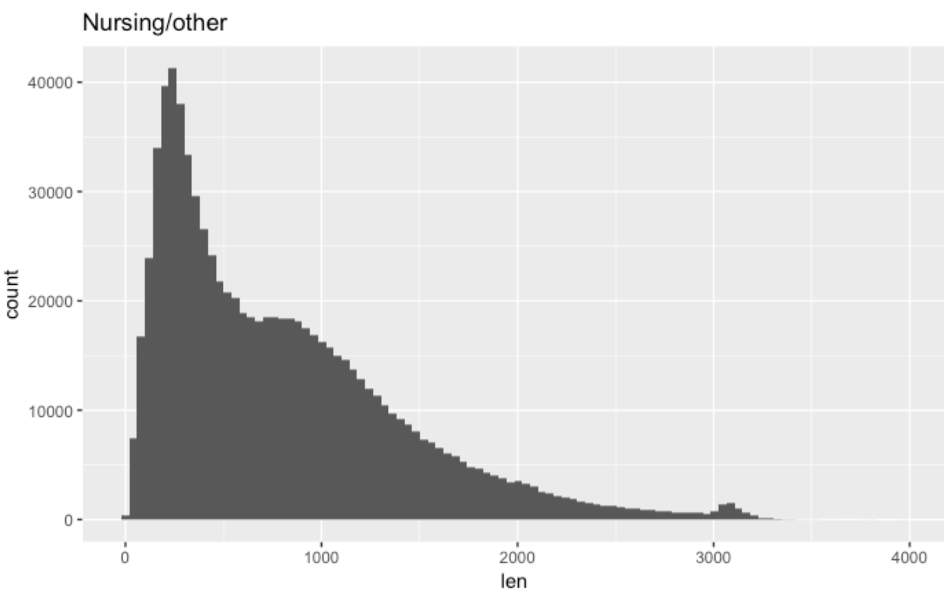
Nursing/other	822497
Radiology	522279
Nursing	223556
ECG	209051
Physician	141624
Discharge summary	59652
Echo	45794
Respiratory	31739
Nutrition	9418
General	8301
Rehab Services	5431
Social Work	2670
Case Management	967
Pharmacy	103
Consult	98

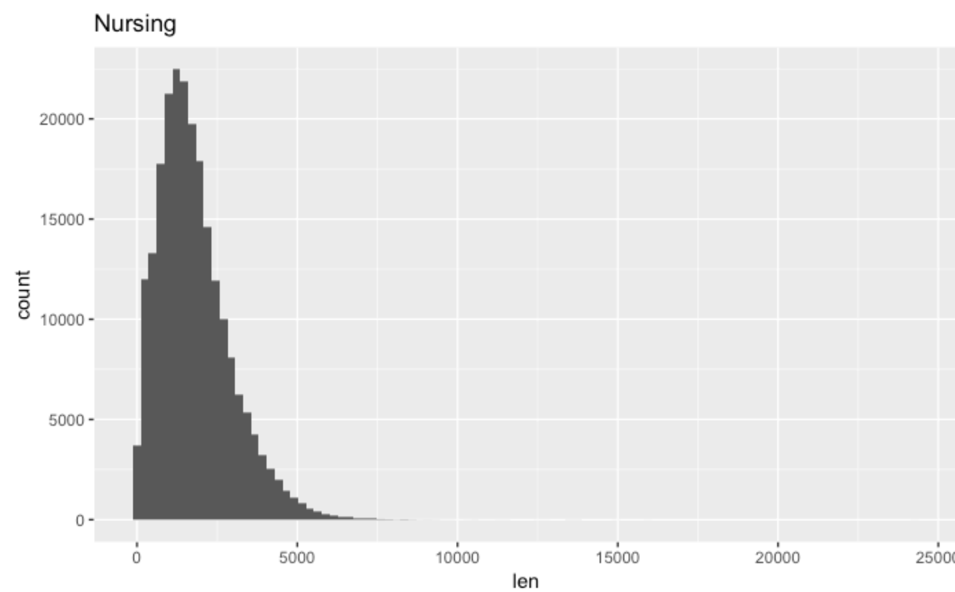
[Slide credit: Pete Szolovits]

Lengths of different note types









[Slide credit: Pete Szolovits]

Nursing note

Hypotension (not Shock)
Assessment:
Pt remains on phenylephrine drip at   0.75 mcg/kg/min
Action:
No titration needed at this time
Response:
BP stable at > 100, MAP >65
Plan:
Wean Neo if tolerated
Wound infection
Assessment:
Anterior groin area open and oozing mod amts thin pink tinged serous fluid
Pt stooling, with small amts stool on dsg and dangerously close to open wound
Action:
Urology resident in to change dressing
Propofol increased to 100 mcg nad fentanyl 100 mcg given for comfort during dsg change
Flexiseal inserted to help contain bowel movements
Stool sent for c diff.
Response:
Pt comfortable during proceedure
Plan:
Continue sedation as needed, increasing Propofol to 100 mcg for sedation during dsg changes.
Keep wound area as clean as possible, check for incontinence of stool as needed

Admission Date: [**2198-7-16**] Discharge Date: [**2198-7-28**]

Date of Birth: [**2153-5-26**] Sex: F

Service: SURGERY

Discharge Summary

Allergies:
No Known Allergies / Adverse Drug Reactions

Attending:[**First Name3 (LF) 1234**]
Chief Complaint:
Leg pain, erythema and swelling secondary to infection of left femoral-poplital bypass

Major Surgical or Invasive Procedure:
1. Incision and drainage and pulse irrigation of left groin and left above-knee popliteal site incisions with xxploration of bypass graft ([**2198-7-16**])
2. Excision of entire left common femoral artery-to-above-knee popliteal artery bypass graft; Repair of common femoral artery and above-knee popliteal artery with harvested left arm cephalic vein ([**2198-7-18**])
3. I and D/washout of left groin with complex wound closure over 2 drains

History of Present Illness:
Ms. [**Known lastname **] is a 45 y/o F who underwent a left fem-AK [**Doctor Last Name **] BPG with PTFE over one month ago on [**2198-6-11**]. She had been doing well postoperatively, and was seen in the clinic 6 days prior to presentation. At this time, she acutely developed nausea/vomiting, fevers, and progressive redness/swelling/pain of her left thigh
directly at the surgical incision. She has been unable to keep down food or liquids. At the time, she denied any ischemic-type pain in her lower leg, and denied any chest pain or shortness of breath.

Example NLP pipeline (cTAKEs)

An example of a sentence discovered by the sentence boundary detector:

Fx of obesity but no fx of coronary artery diseases.

Tokenizer output – 11 tokens found:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  diseases  .
```

Normalizer output:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  disease  .
```

Part-of-speech tagger output:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  diseases  .
NN  IN  NN       CC   DT  NN  IN  JJ        NN      NNS       .
```

Shallow parser output:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  diseases  .
NP  PP  NP            NP      PP              NP
```

Named Entity Recognition – 5 Named Entities found:

```
Fx of obesity but no fx of coronary artery diseases .
      obesity  (type=diseases/disorders, UMLS CUI=C0028754, SNOMED-CT codes=308124008 and 5476005)
                           coronary artery diseases (type=diseases/disorders, CUI=C0010054, SNOMED-CT=8957000)
                           coronary artery (type=anatomy, CUI(s) and SNOMED-CT codes assigned)
                                    artery (type=anatomy, CUI(s) and SNOMED-CT codes assigned)
                                           diseases (type=diseases/disorders, CUI = C0010054)
```

Status and Negation attributes assigned to Named Entities:

```
Fx of obesity but no fx of coronary artery diseases .
      obesity (status = family_history_of; negation = not_negated)
                           coronary artery diseases (status = family_history_of, negation = is_negated)
```

Figure 1 Example sentence processed through cTAKES components 'family history of obesity but no family history of coronary artery diseases.' Fx, family history.

(Slide credit: Nigam Shah)

Outline

1. Clinical text
   – **Case study: Prediction of sepsis (severe infection) from electronic health records**
2. Physiological time-series
   – Case study: Monitoring babies in neonatal ICUs
   – Case study: Detecting atrial fibrillation
3. Imaging
   – Cardiology, pathology, radiology

Early identification of sepsis

- Sepsis is a systemic inflammatory response secondary to infection
- Hospital mortality rate reported to be 30-50%
- Estimated 751,000 cases/year in the US, with a cost of care of $16.7 billion
- Reducing the time to administration of antibiotics by one hour has shown to reduce mortality from 33.2% to 19.5%

Early Goal-Directed Therapy improves sepsis outcomes

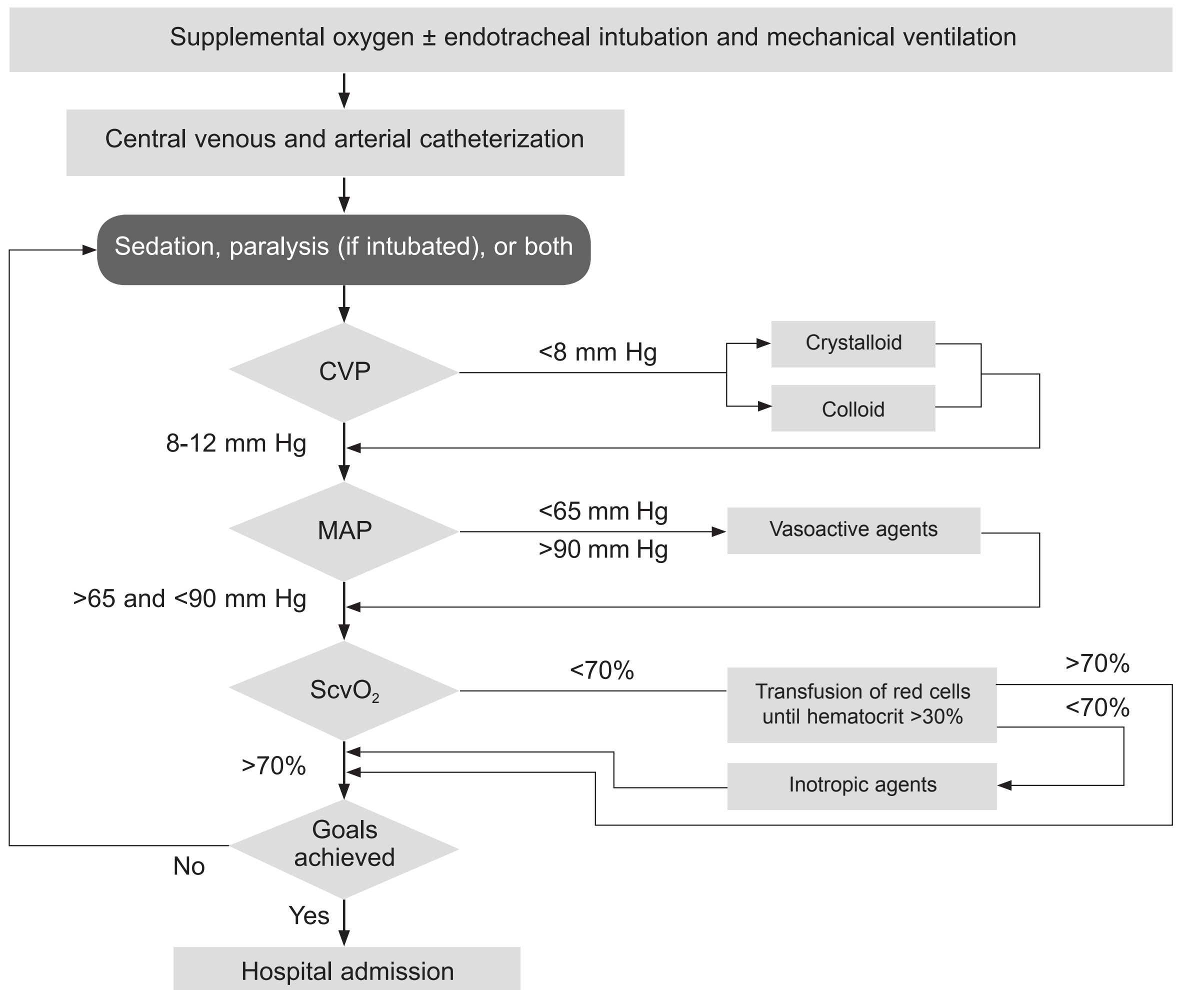
Administer antibiotics

Give fluids

Maintain blood pressure

Maintain oxygen saturation

Figure 1. Protocol for Early Goal-Directed Therapy from Rivers et al.[5]



CVP=central venous pressure MAP=mean arterial pressure $ScvO_2$=central venous oxygen saturation

Sepsis Triage Criteria

Does the patient have **any three** of the following:

- ☐ Temp > 100.4 or < 96.5 or rigors
- ☐ HR > 90
- ☐ RR > 20
- ☐ O2 Sat < 90%
- ☐ SBP < 90
- ~~☐ Suspected Infection~~
- ~~☐ Any alteration of mental status~~

☐ Yes ☐ No

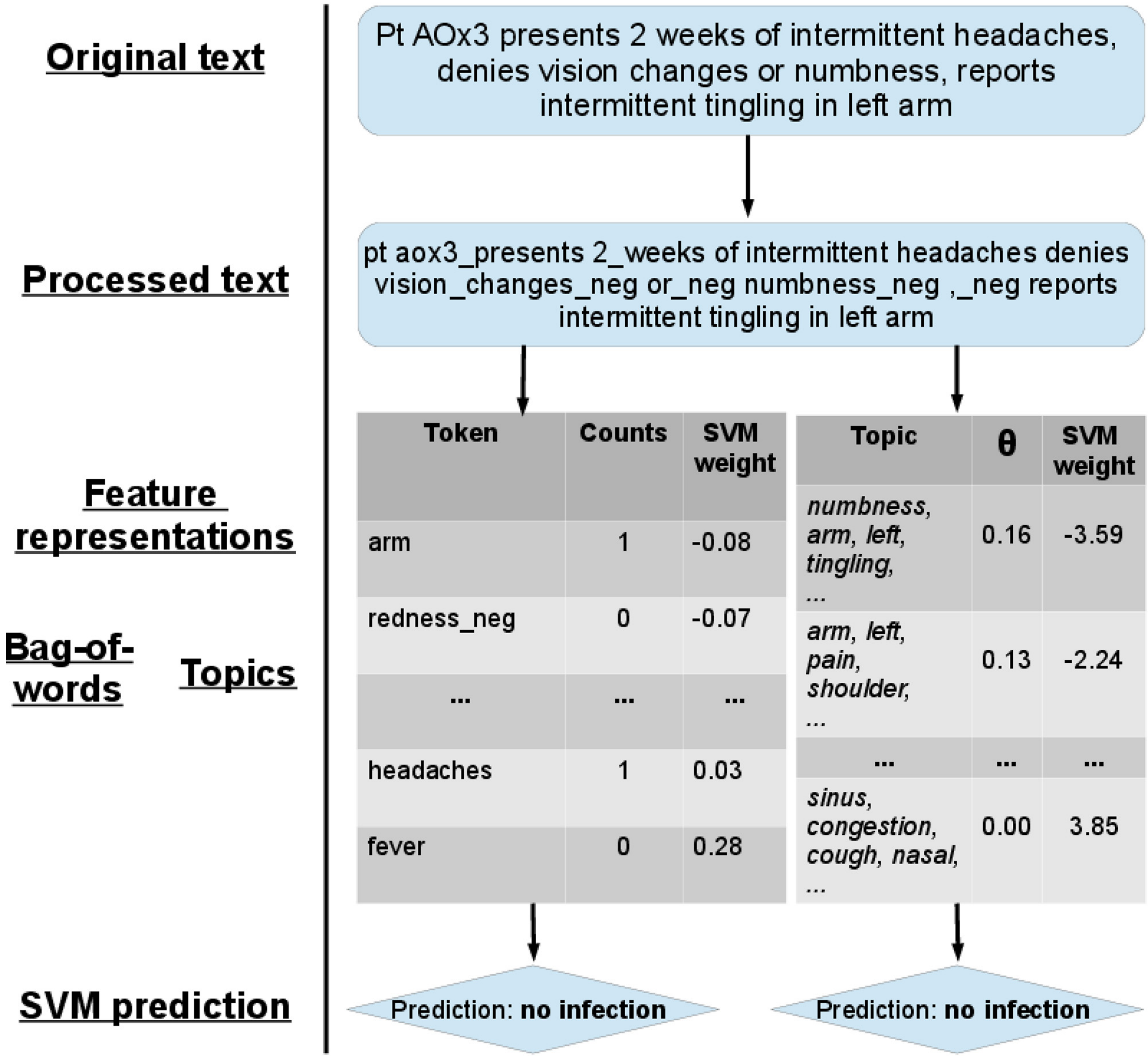
Suspect SEPSIS Save Lives

Never used by sepsis alerts, since not explicitly recorded

Predicting infection at triage

- Use data from 230,936 patients from 12/08 to 2/13 at tertiary academic teaching hospital
- 14% have positive label (infection according to ED ICD9 discharge diagnosis)
- Compare use of only *structured data* versus also using *unstructured data* (text)

[Horng, Sontag, et al. "Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning". PLOS ONE, 2017]

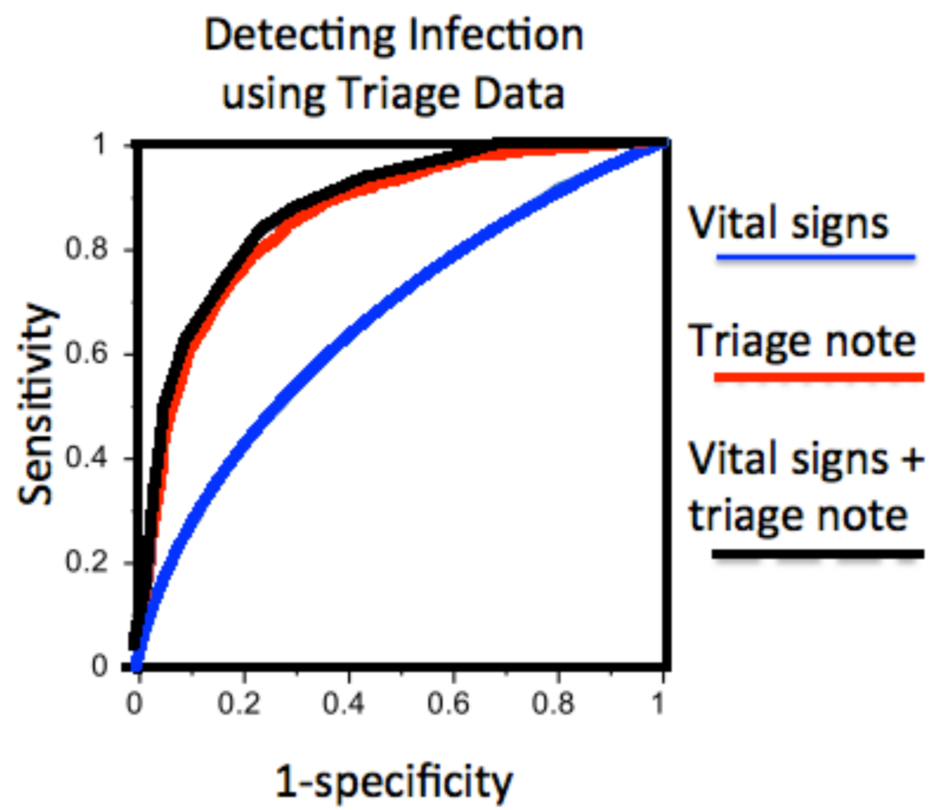

Original text
Pt AOx3 presents 2 weeks of intermittent headaches, denies vision changes or numbness, reports intermittent tingling in left arm
Processed text
pt aox3_presents 2_weeks of intermittent headaches denies vision_changes_neg or_neg numbness_neg ,_neg reports intermittent tingling in left arm
Feature representations
Bag-of-words
Topics
Token
Counts
SVM weight
arm
1
-0.08
redness_neg
0
-0.07
...
headaches
1
0.03
fever
0
0.28
Topic
θ
SVM weight
numbness, arm, left, tingling, ...
0.16
-3.59
arm, left, pain, shoulder, ...
0.13
-2.24
...
sinus, congestion, cough, nasal, ...
0.00
3.85
SVM prediction
Prediction: no infection
Prediction: no infection

Text is much more valuable than structured data

Example Triage Notes:

FOOT INFECTION. *"Pt here from ___ hosp.with ?osteomyelitis. Footis pink swollen and warm to the touch on the right foot. Denies fevers at home. hx of multiple infections after a mvc ankle break in ___"*

CHEST PAIN. *"presents with left sided chest pain intermittant described as gas pain today pain reoccured during episode of stress, developed fluttering in left chest with left arm pain. Denies n/v/d or dyspnea"*



Area under the curve (AUC) of .86 versus .67

Table 8. SVM model learned using bag-of-words.

Weight	Word
0.98	cellulitis
0.80	uti
0.79	redness_swelling
0.78	sore_throat
0.77	abscess
0.73	diverticulitis
0.72	abscess
0.70	dysuria
0.66	st
0.65	erythema
0.20	swelling
-0.29	swelling_neg
-0.35	pancreatic
-0.36	eye
-0.36	bleed
-0.37	etoh
-0.37	epistaxis
-0.38	pancreatitis
-0.39	injury
-0.57	mvc

Most positive (indicative of infection) and negative (suggesting no infection) words used by the model built by machine learning using the bag-of-words model on triage notes.

Alternative – topic model

Topic models are powerful tools for exploring large data sets and for making inferences about the content of documents

Documents

Unsupervised learning

Topics

politics	religion	sports
president	hindu	baseball
obama	judiasm	soccer
washington	ethics	basketball
religion	buddhism	football
…	…	…

Many applications in information retrieval, document summarization, and classification

New document

Words $w_1, \ldots, w_N$

Dimensionality reduction

What is this document about?

weather	.50
finance	.49
sports	.01

Distribution of topics θ

Alternative – topic model

- First, learn a topic model over all triage notes
- Then, learn predictive model on the topics instead of the words themselves
- Disadvantage:
  - **Bit worse** predictive performance compared to bag-of-words model (Test AUC of 0.85)
- Advantages:
  - Easier to interpret, may transfer better

Latent Dirichlet allocation

- Generative model for documents (patient's triage text)
- Assume there are T topics (**for us, T=500**), and the variable z_i denotes the assignment of a topic to the i'th word
- Generative model for single patient's triage text:
  - $\theta \sim \mathrm{Dir}(\alpha)$ (θ is a distribution over the T topics)
  - For each word i,

$$z_i \sim \mathrm{Multinomial}(\theta)$$ (choose a topic for i'th word)

$$w_i \sim \Pr(w \mid z = z_i)$$ (sample a word)

- We learn the distributions Pr(w | z = t) and the "priors" α_t

[Blei, Ng, Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 2003]

Table 9. SVM model learned using topics.

Weight	Topic (described by most frequent words)
11.00	redness, cellulitis, left, leg, swelling, area, rle, arm, lle, increased, erythema
8.38	abcess, buttock, area, drainage, axilla, groin, painful, thigh, left, hx, abcesses, red, boil
8.15	cellulitis, abx, pt, iv, infection, po, keflex, antibiotics, leg, treated, started, yesterday
7.13	red, swollen, touch, warm, painful, area, left, infection, swelling, tender, slightly, hot
6.65	abscess, left, area, fevers_neg, axilla, cyst, size, i&d, lesion, lump, swelling, mass, thigh
6.60	pna, pneumonia, cxr, wbc, dec_num, transfer, rll, anon_1140, rehab, fever, lll, recent
6.40	sore_throat, throat, st, voice, secretions, swallowing, pain, swallow, difficulty_swallowing
5.90	uti, pt, cipro, abx, dx, started, treated, recent, bactrim, fever, c/o, recently, infection
5.69	pna, cough, sob, pneumonia, cxr, recent, dx, abx, fever, r/o, fevers, bronchitis, recently, tb
5.64	dysuria, hematuria, uti, c/o, urination, pain_neg, burning, denies, frequency, urgency,
2.12	wound, check, eval, pt, abcess, wick, i&d, abscess, drained, removal, returns, fevers_neg
-1.80	pain, ankle, weight, bearing, left, foot, swelling, knee, wt, injury, bear, unable_bear
-3.44	struck, bike, car, ped, accident, bicycle, loc_neg, pain, riding, hit, bicyclist, pt, fell, c/o
-3.59	numbness, arm, left, tingling, facial, hand, leg, weakness, side, sided, c/o, today, resolved
-3.63	epistaxis, bleeding, nose, pt, bleed, pressure, bleeding_neg, blood, on_coumadin, stopped
-3.64	status_post_mvc, mvc, car, restrained_driver, loc_neg, passenger, neck, driver, front, side
-3.89	fall, status_post_fall, fell, ladder, feet, pain, landed, ft, 10, loc_neg, back, approx, foot, steps
-3.90	gi, bleed, status_post, colonoscopy, endoscopy, procedure, today, esophageal, upper, scope
-4.26	playing, injury, ball, soccer, pt, game, football, hit, hockey, player, struck, baseball, loc_neg
-4.29	mvc, trauma, gsw, basic, mcc, 21, status_post_mvc, transfer, rollover, rm, room, stabbing
-4.91	etoh, found, vomiting, apparently, drunk, drinking, denies, friends, trauma_neg, triage,
-5.18	watching, tv, sitting, sudden_onset, movie, television, smoked, couch, pt, pot, 5pm, theater

Most positive (indicative of infection) and negative (suggesting no infection) topics from the model built by machine learning using features derived from the topic model on triage notes.

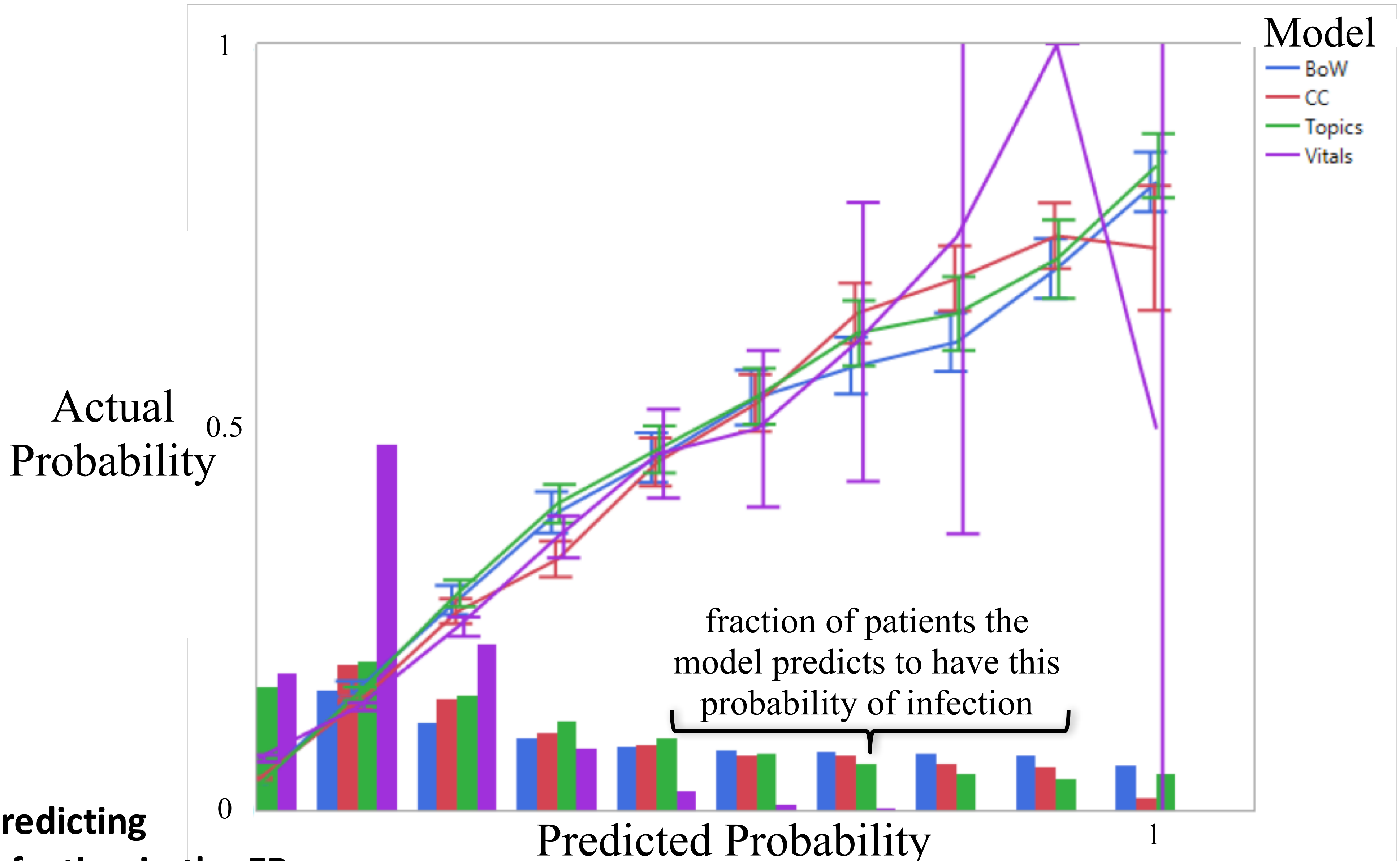
Synonyms are grouped together

More likely

Infection

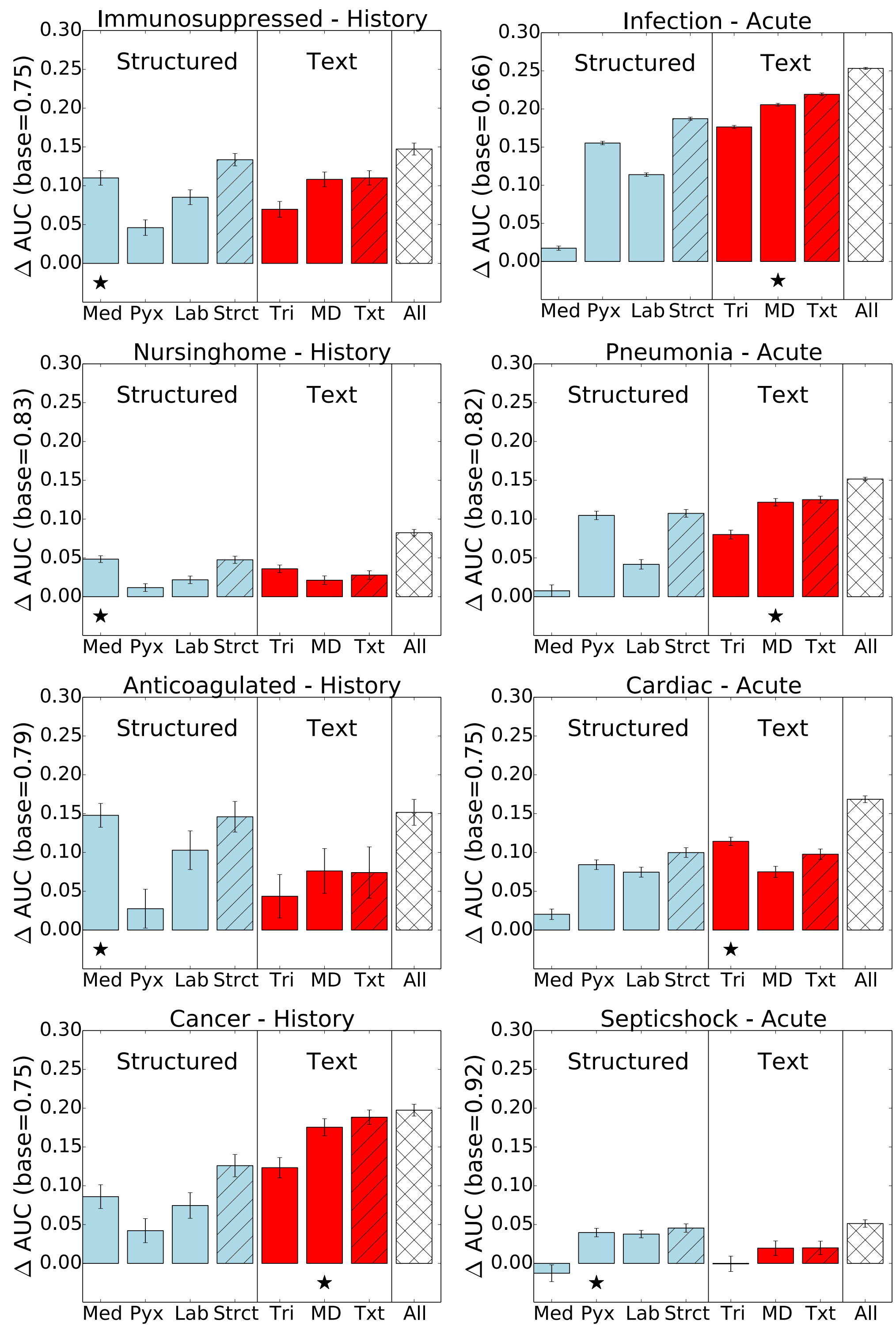
Less likely

Evaluating model *calibration*



Predicting infection in the ER

Value of data types across prediction tasks



Med Medication history (prior to visit)
Pyx Medication dispensing record (during visit)
Lab Laboratory values
Strct All Structured data (Med + Pyx + Labs)
Tri Triage Nursing Text
MD Physician Comments
Txt All Text (Tri + MD)
All All features (Structured + Text).

[Halpern, Horng, Choi, Sontag, JAMIA ‘16]

Outline

1. Clinical text
   – Case study: Prediction of sepsis (severe infection) from electronic health records
2. Physiological time-series
   – Case study: Monitoring babies in neonatal ICUs
   – Case study: Detecting atrial fibrillation
3. Imaging
   – Cardiology, pathology, radiology

Physiological time-series

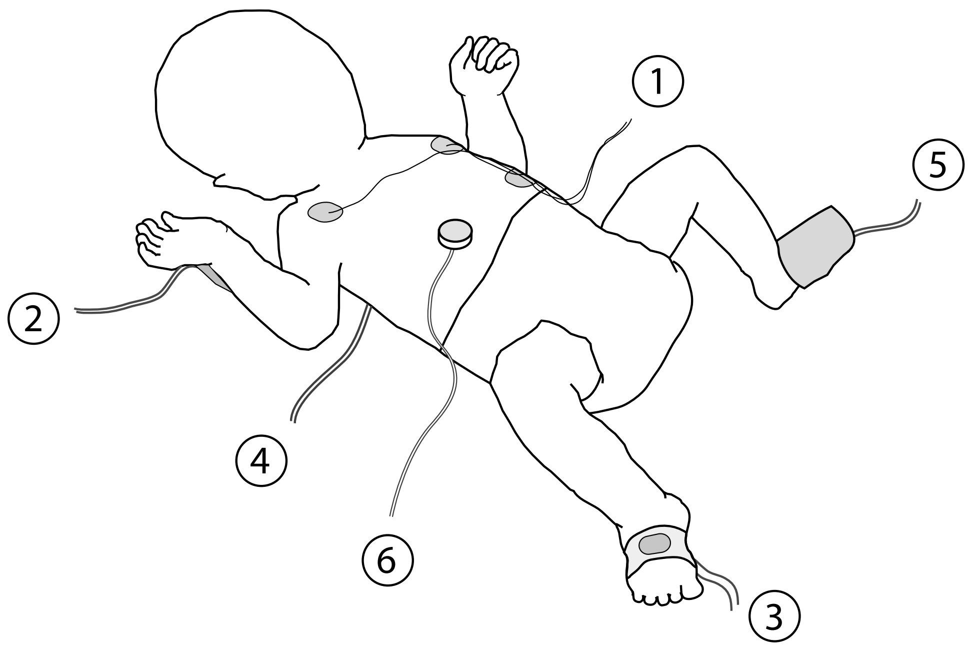
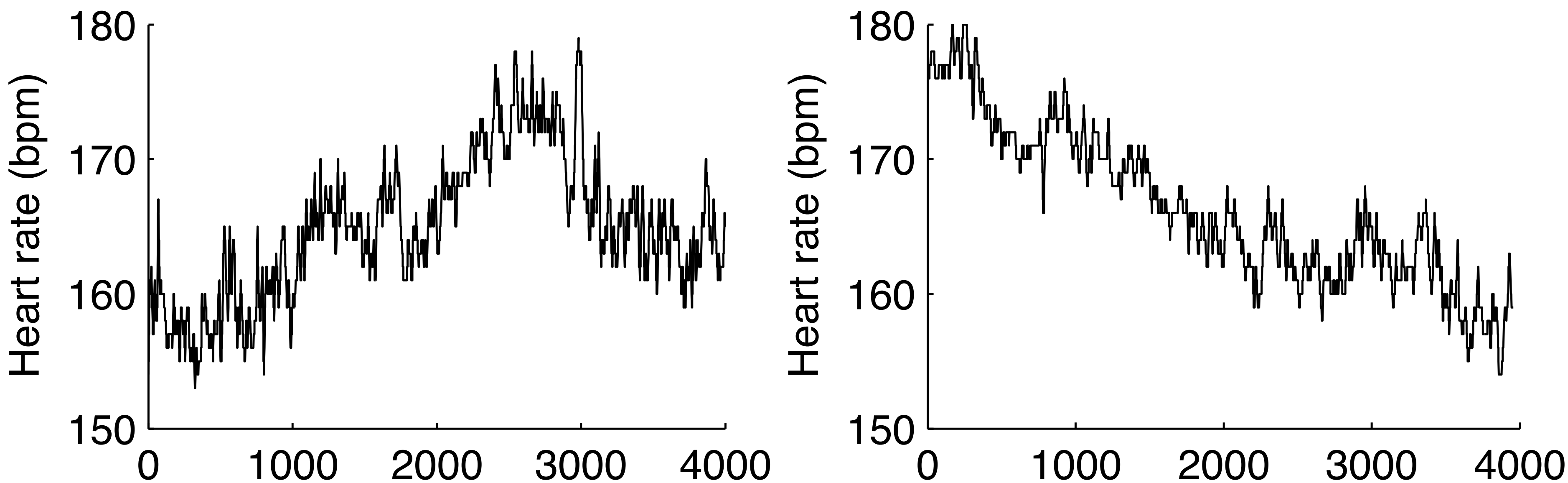


Fig. 4. Probes used to collect vital signs data from an infant in intensive care. 1) Three-lead ECG, 2) arterial line (connected to blood pressure transducer), 3) pulse oximeter, 4) core temperature probe (underneath shoulder blades), 5) peripheral temperature probe, 6) transcutaneous probe.

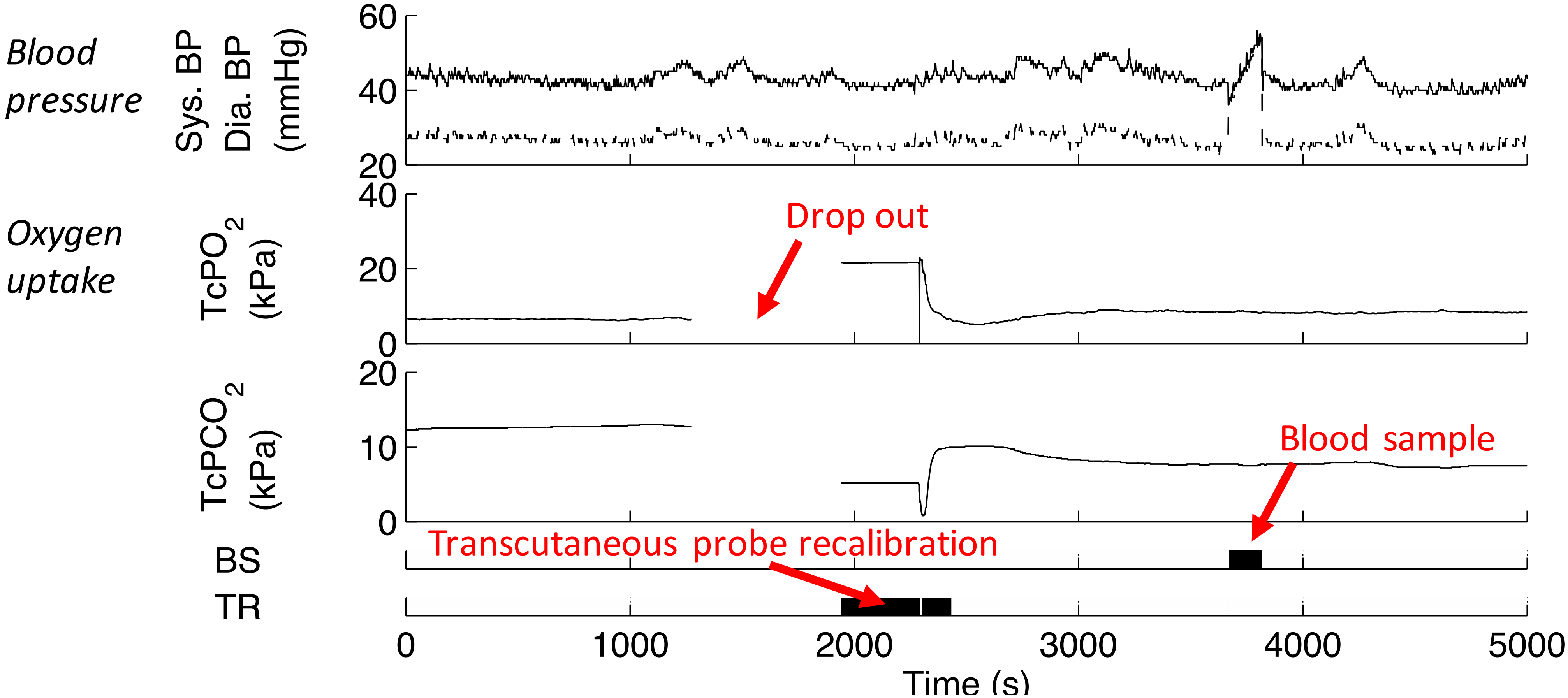
(Quinn et al., TPAMI 2008)

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



(Quinn et al., TPAMI 2008)

Measurements confounded by interventions & measurement errors

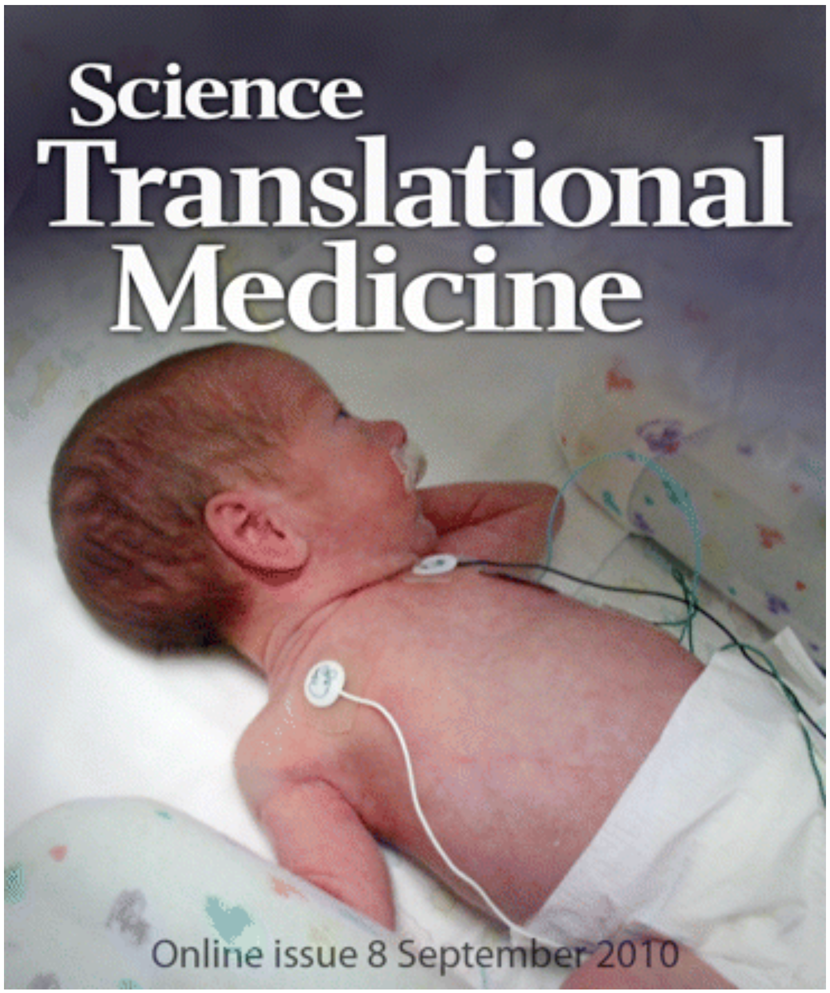


(Quinn et al., TPAMI 2008)

Outline

1. Clinical text
   - Case study: Prediction of sepsis (severe infection) from electronic health records
2. Physiological time-series
   - **Case study: Monitoring babies in neonatal ICUs**
   - Case study: Detecting atrial fibrillation
3. Imaging
   - Cardiology, pathology, radiology

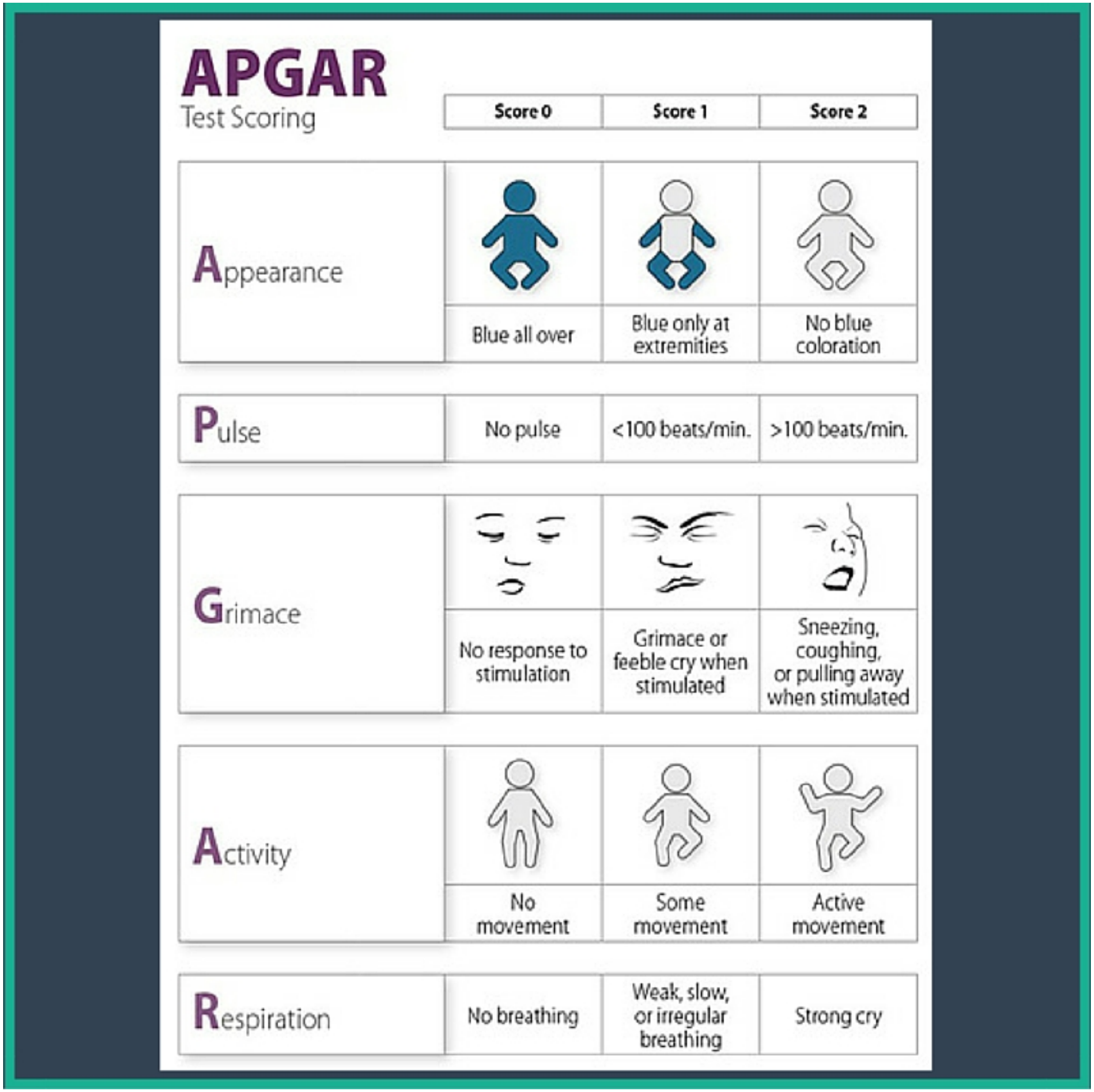
Predicting morbidity in preterm newborns



Saria et al.,
Science Translational
Medicine 2010

Can we predict major complications?

- Preterm neonates 34 weeks gestational age or less and <2000 g in weight
- Goal: estimate probability infant would have high morbidity (HM), *using data in first 3 hours of life*
  - Includes death, sepsis, hemorrhage, pulmonary hypertension, acute hemodynamic instability, and retinopathy of prematurity
  - Outcomes can manifest days or weeks later
- A benefit of using only first 3 hours is that data not typically confounded by medical intervention
  - Models may generalize better across NICUs



APGAR
Test Scoring

	Score 0	Score 1	Score 2
Appearance	Blue all over	Blue only at extremities	No blue coloration
Pulse	No pulse	<100 beats/min.	>100 beats/min.
Grimace	No response to stimulation	Grimace or feeble cry when stimulated	Sneezing, coughing, or pulling away when stimulated
Activity	No movement	Some movement	Active movement
Respiration	No breathing	Weak, slow, or irregular breathing	Strong cry

Figure from: http://www.medicinehack.com/2010/05/apgar-scoring.html

Goal of study

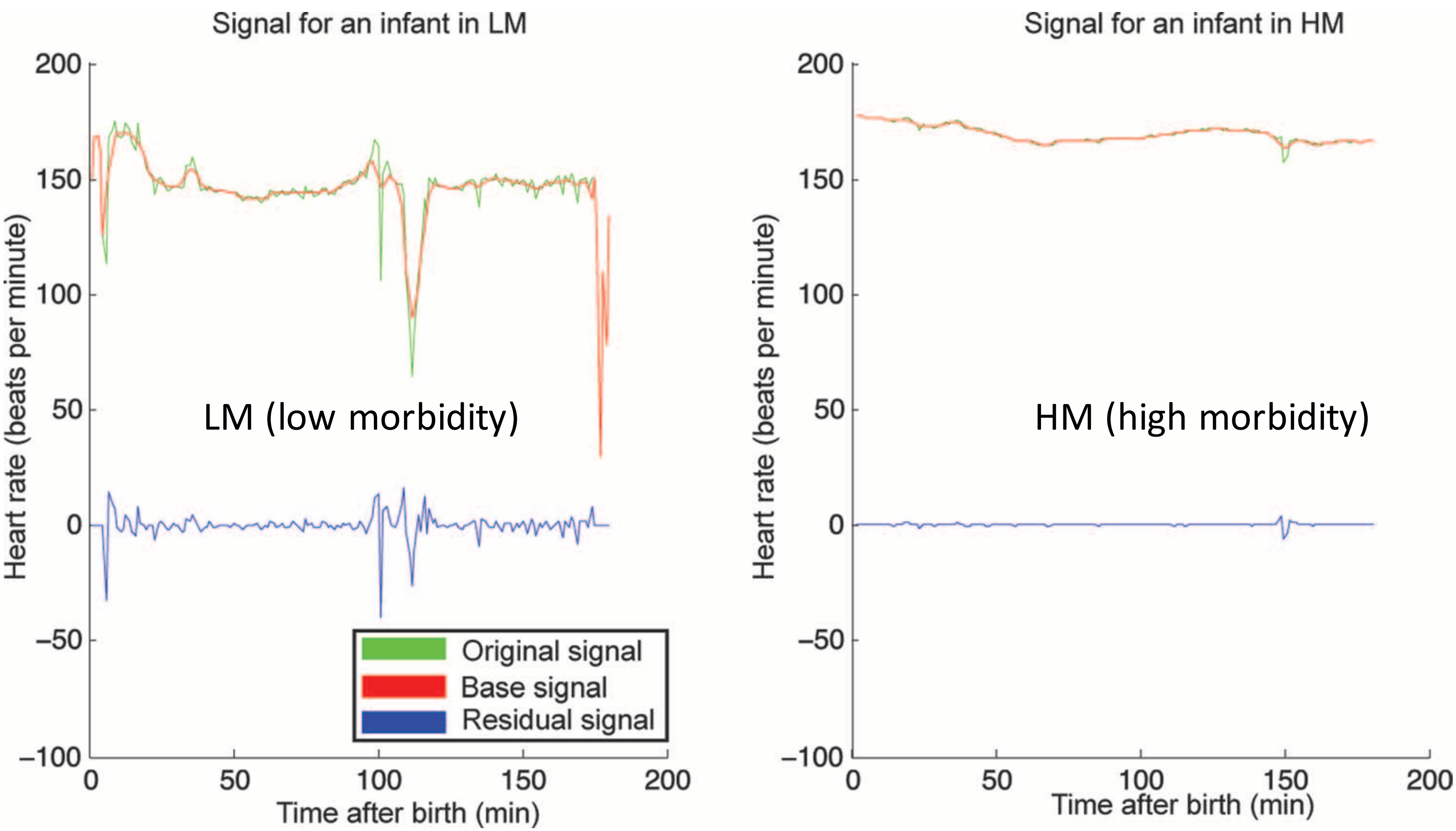
- “Electronic” Apgar score
- Better inform decisions regarding
  - Aggressive use of intensive care
  - Need for transport to tertiary centers
  - Resource allocation (currently $26 billion per year in US spent because of preterm birth)

Machine learning setup

- Binary classification
- Features:
  - Mean heart rate (+ base and residual variability); mean respiratory rate (+base and residual variability); mean oxygen saturation and cumulative hypoxia time
  - Gestational age and birth weight
- 138 preterm neonates (35 with HM complications)
- Leave-one-out cross-validation – no need for nested cross-validation since no hyperparameter tuning

HM = high morbidity
LM = low morbidity

Deriving the features: variability



(Saria et al., Science Translational Medicine 2010)

Prediction using probabilistic model

- L2-regularized logistic regression used to learn predict whether baby will be "high morbidity" (HM):

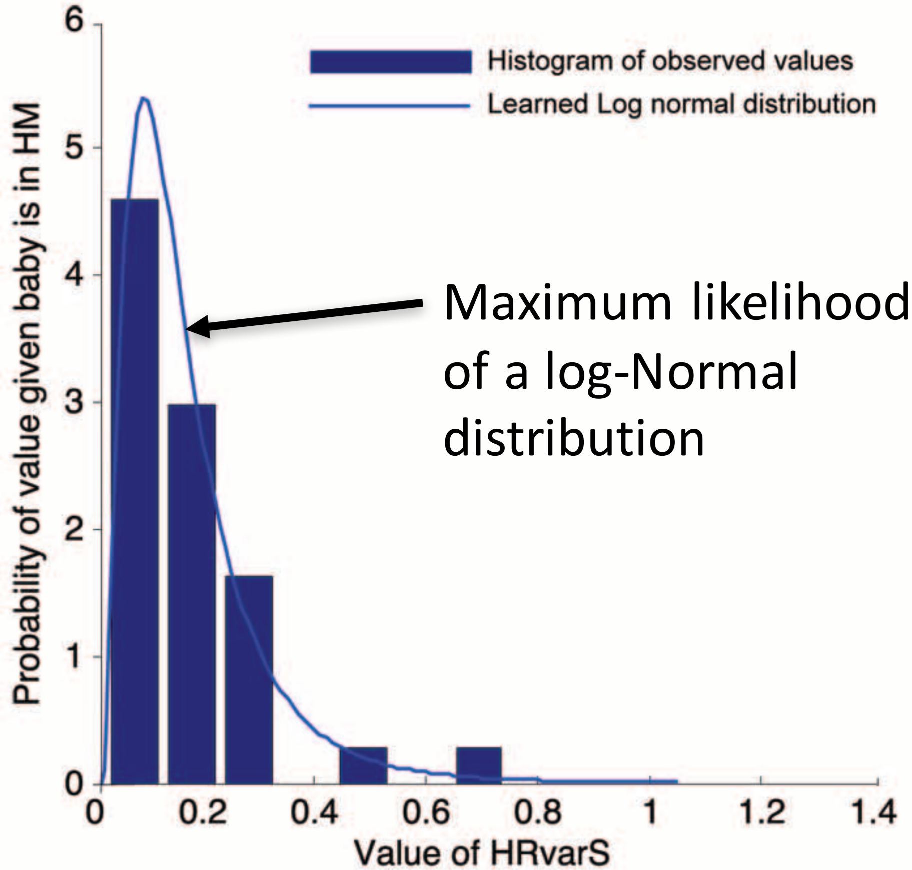
$$P(\mathrm{HM}|v_1,v_2,\ldots,v_n) = \left(1+\exp\left(b+w_0 {*} c+\sum_{i=1}^{n} w_i {*} f(v_i)\right)\right)^{-1}$$

- Non-linear transformation applied to the features:
  - Estimate $\Pr(v_i \mid C)$ for each class of patient C={HM or LM}) using parametric models: exponential, Weibull, lognormal, gamma
  - Use log odds ratio of observed value as feature if observed, 0 if the value is missing:

$$f(v_i) = \log \frac{\Pr(v_i \mid HM)}{\Pr(v_i \mid LM)}$$

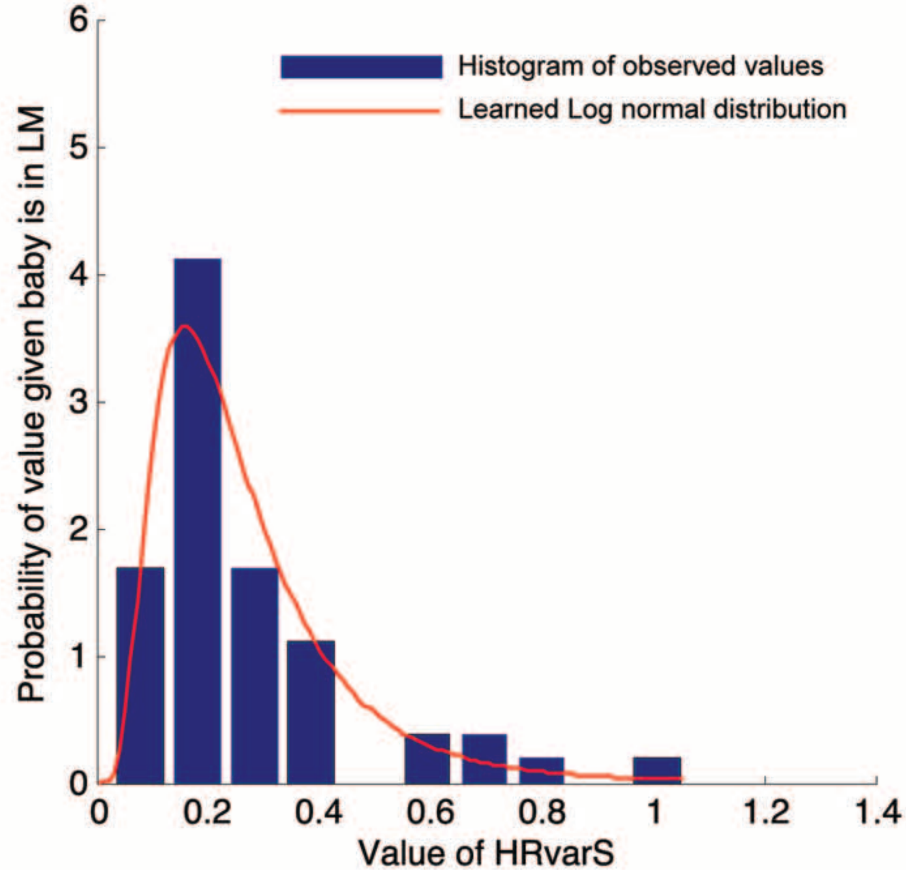
  - No need to do imputation with this approach!
  - Also use missingness indicators given that it is often informative

Prediction using probabilistic model



Maximum likelihood fit of a log-Normal distribution

Distribution of heart rate variability for patients with HM (high morbidity)



Distribution of heart rate variability for patients with LM (low morbidity)

Mean heart rate

Short-term variability of heart rate

Long-term variability of heart rate

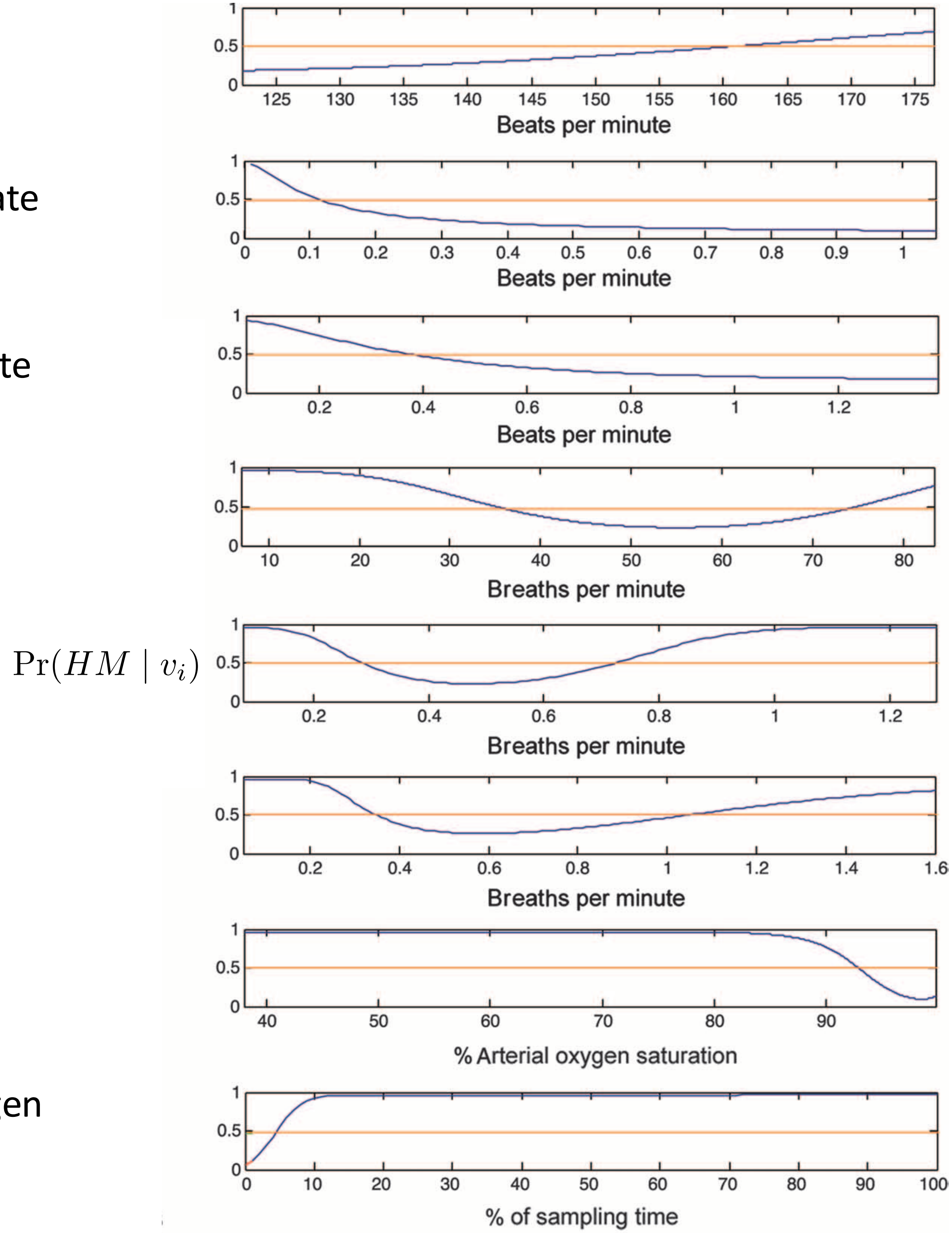
Mean respiratory rate

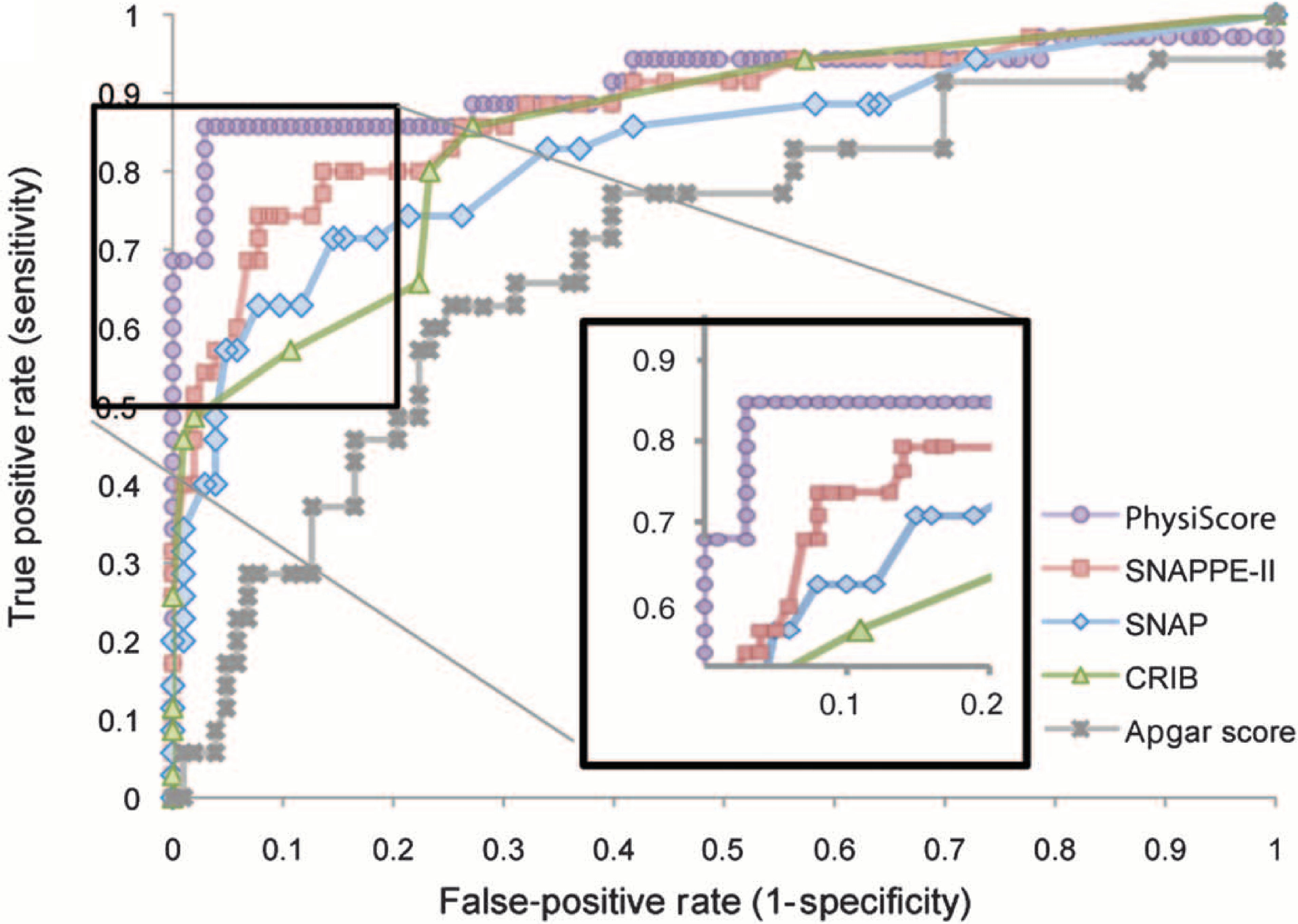
Short-term variability of respiratory rate

Long-term variability of respiratory rate

Mean oxygen saturation

% of time spent below 85% oxygen saturation

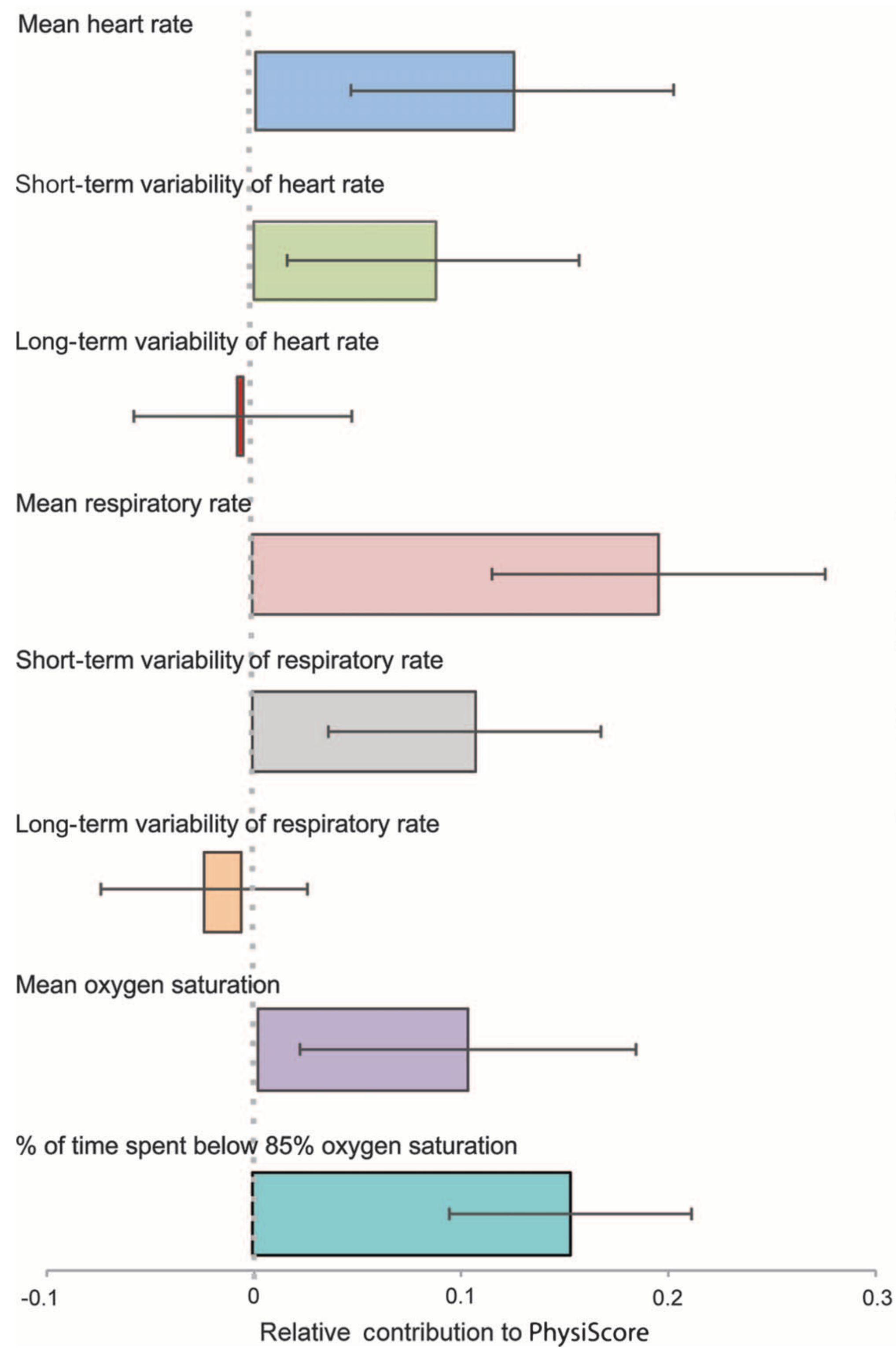




True positive rate (sensitivity)
False-positive rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
PhysiScore
SNAPPE-II
SNAP
CRIB
Apgar score

Feature importance

Error bars = variation over folds in cross-validation



Outline

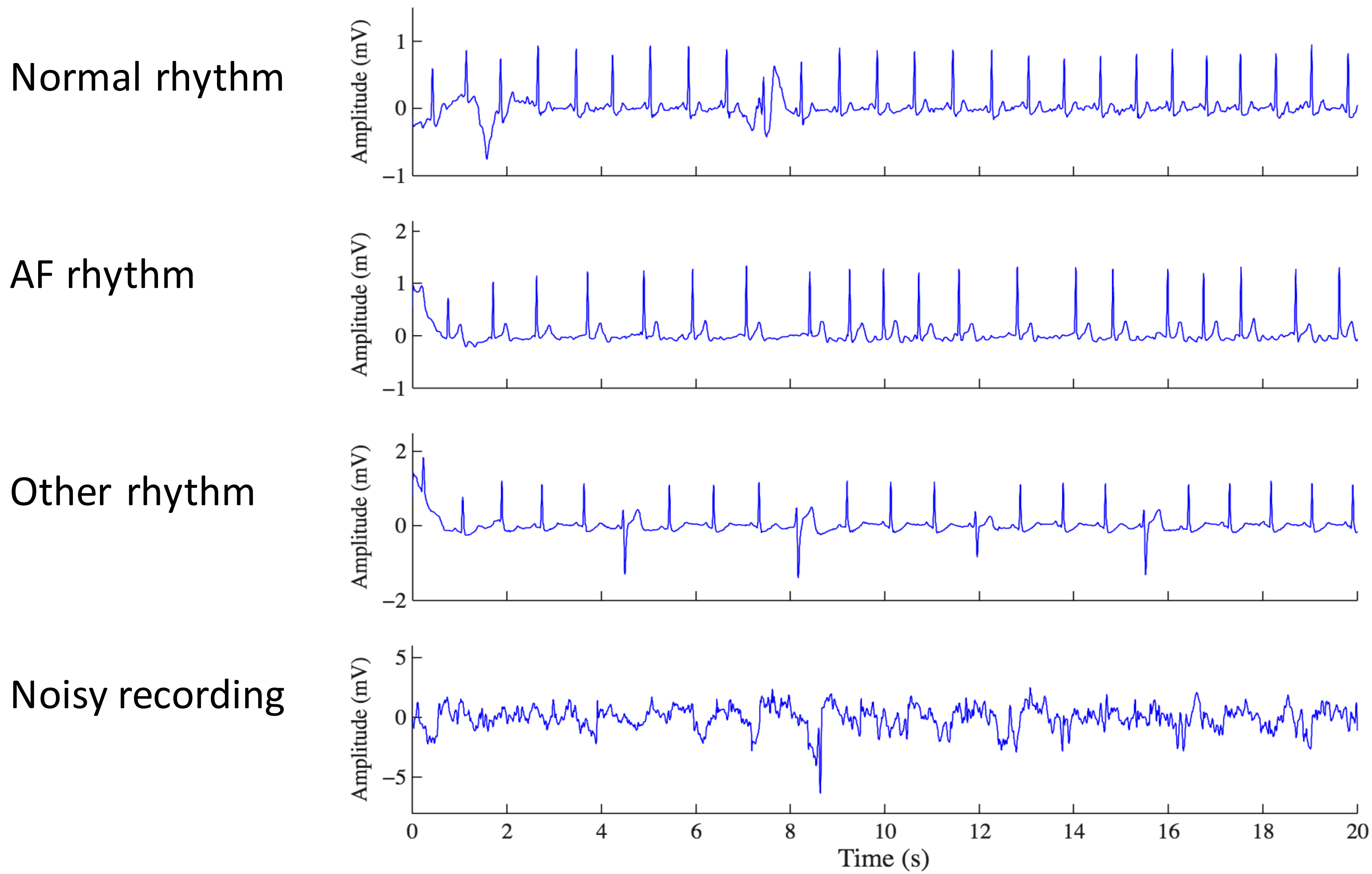
1. Clinical text
   - Case study: Prediction of sepsis (severe infection) from electronic health records
2. Physiological time-series
   - Case study: Monitoring babies in neonatal ICUs
   - **Case study: Detecting atrial fibrillation**
3. Imaging
   - Cardiology, pathology, radiology

Detecting atrial fibrillation

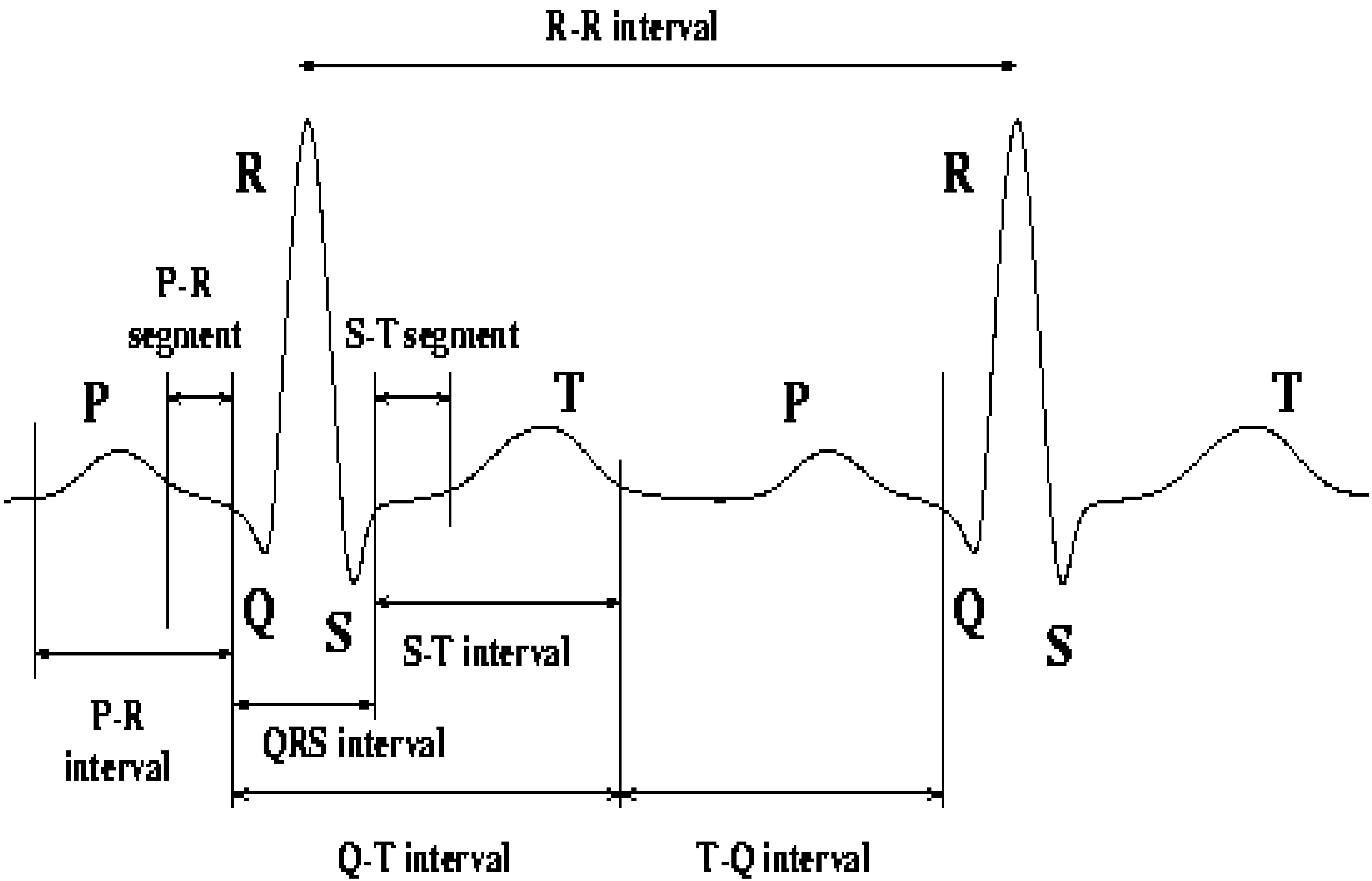
AliveCore ECG device

ECG = electrocardiogram

What type of heart rhythm?

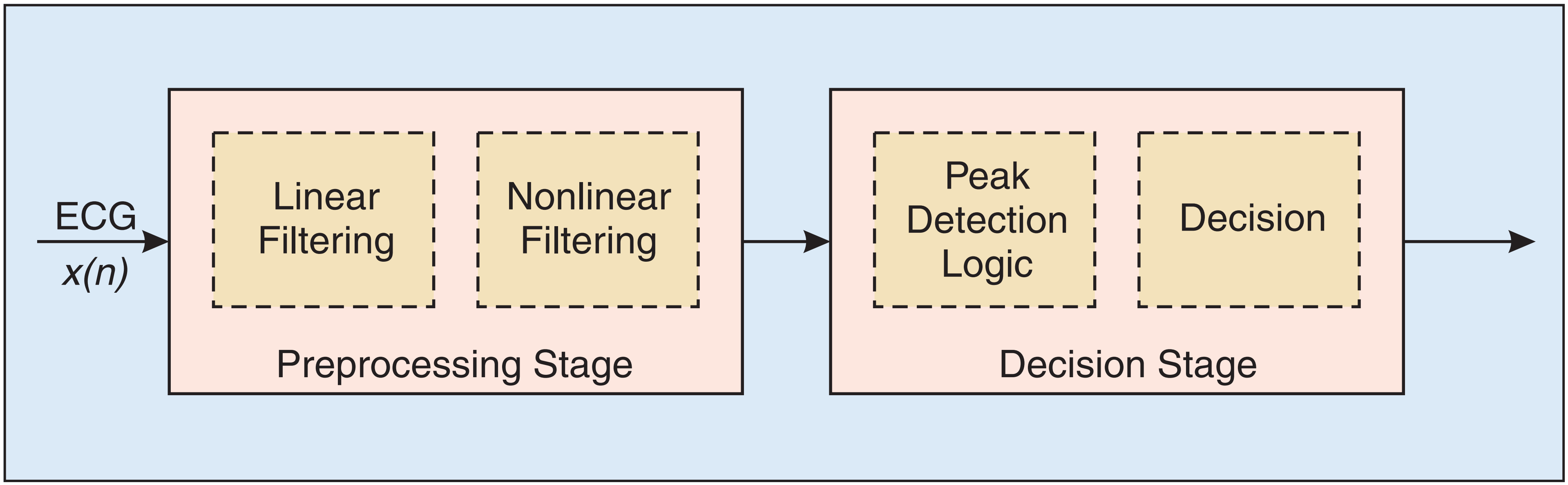


[Clifford, Liu, Moody, Mark. PhysioNet Computing in Cardiology Challenge 2017]



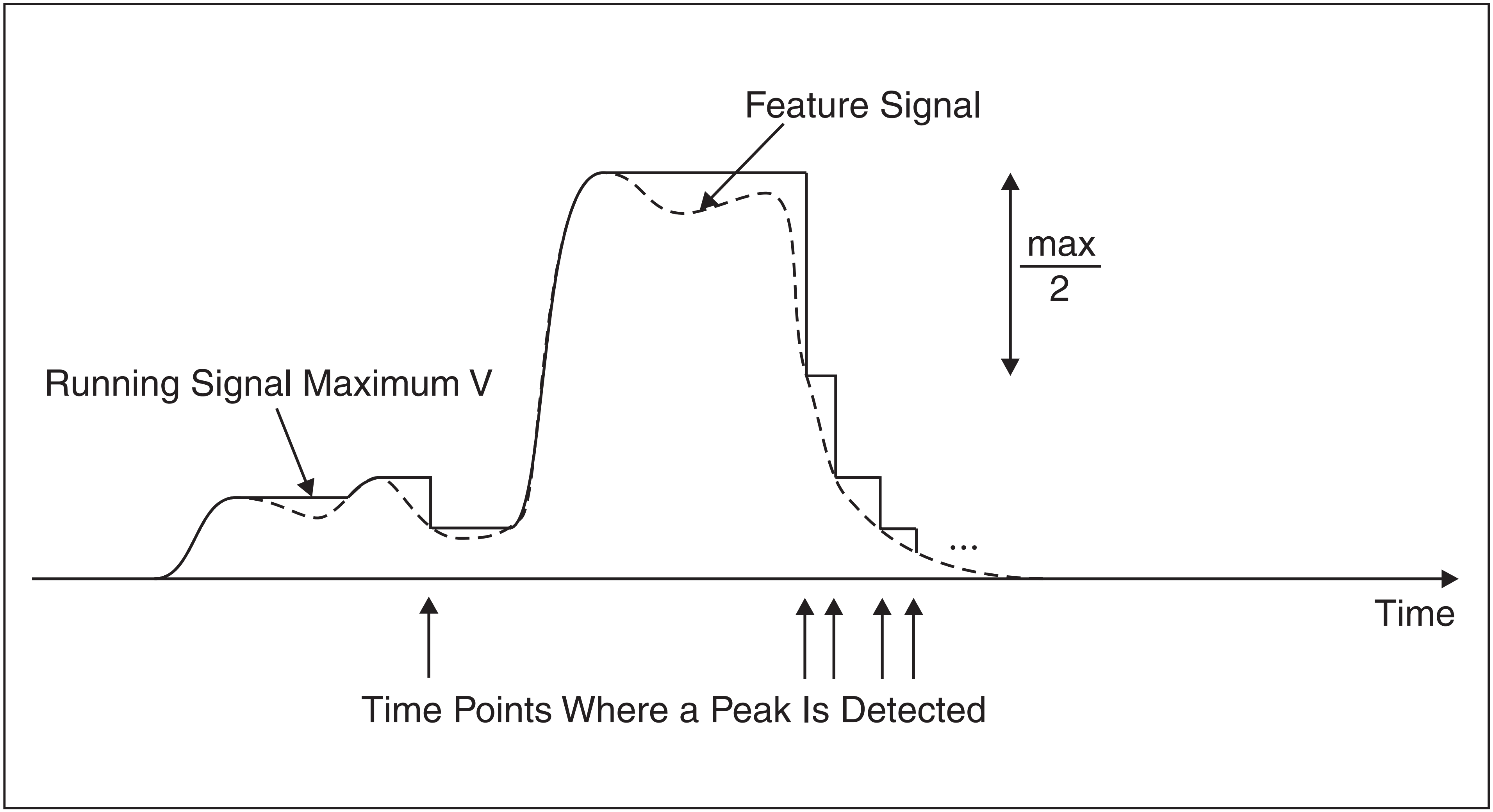
R-R interval
R
R
P-R
segment
S-T segment
P
T
P
T
Q
S
S-T interval
Q
S
P-R
interval
QRS interval
Q-T interval
T-Q interval

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]

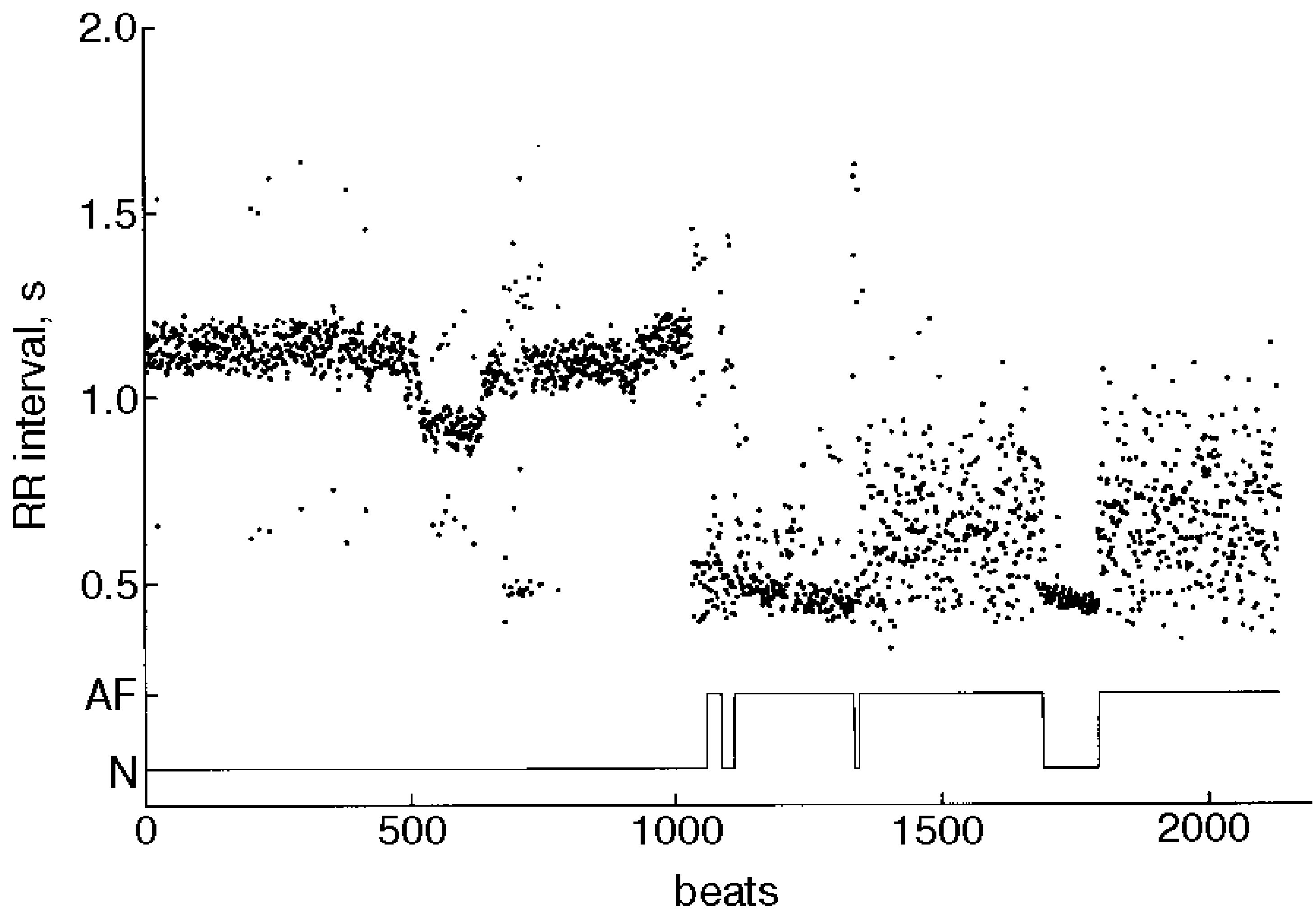


Fig. 1 *Time series showing RR intervals from subject 202 from MIT-BIH arrhythmia database. (——) Assessment of atrial fibrillation (AF) or non-atrial fibrillation (N) as reported in database*

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

COMPUTERS AND BIOMEDICAL RESEARCH 4, 385-392 (1970)

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

Proceedings Computers in Cardiology (1991)

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 succesfully trained using this data. Three standard databases were used to test network performance. Postprocessing of the ANN output yielded an AF sensitivity of 92.86% *and an AF positive predictive accuracy of* 92.34%.

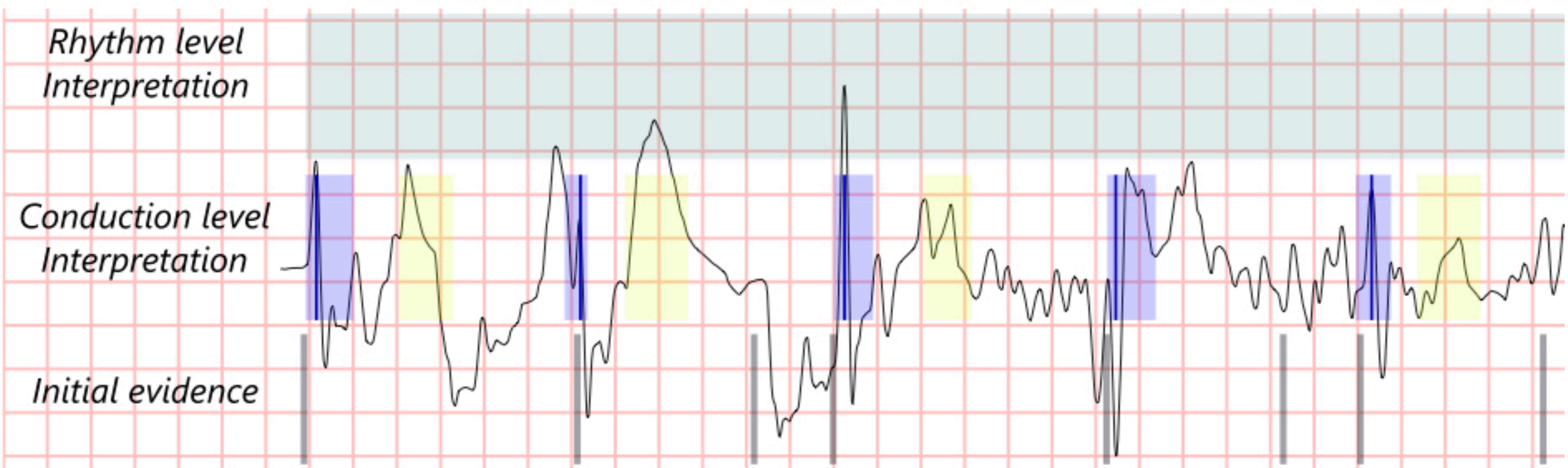
1 Introduction

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.
`Pdistd`: 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_profd`: 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.

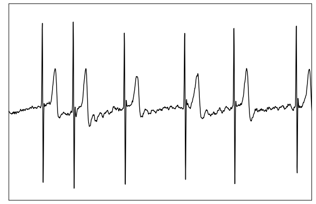
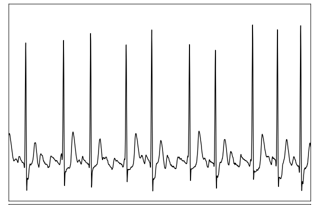
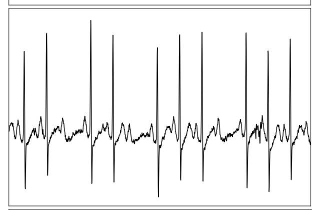
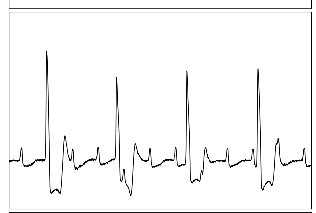
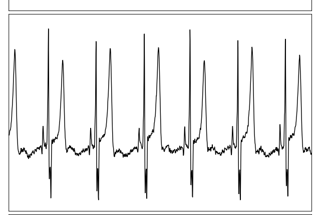
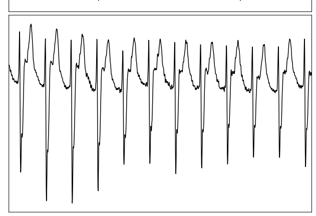
Key to exceeding expert performance is a deep convolutional network which can map a sequence of ECG samples to a sequence of arrhythmia annotations along with a novel dataset two orders of magnitude larger than previous datasets of its kind.

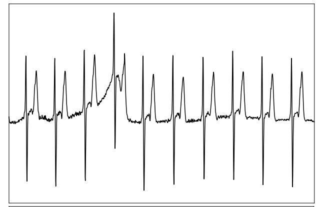
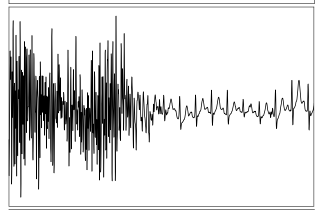
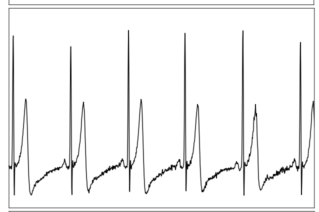
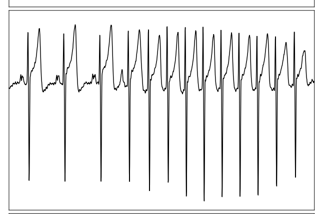
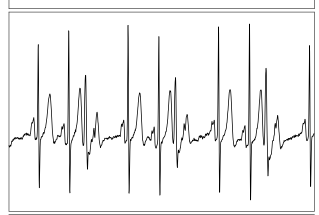
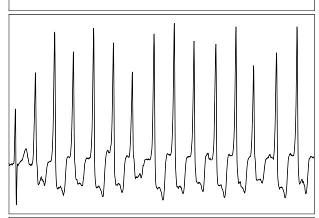
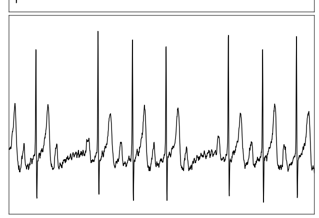
[Rajpurkar et al., arXiv:1707.01836, 2017; Nature Medicine '19]

Differences with previous work

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

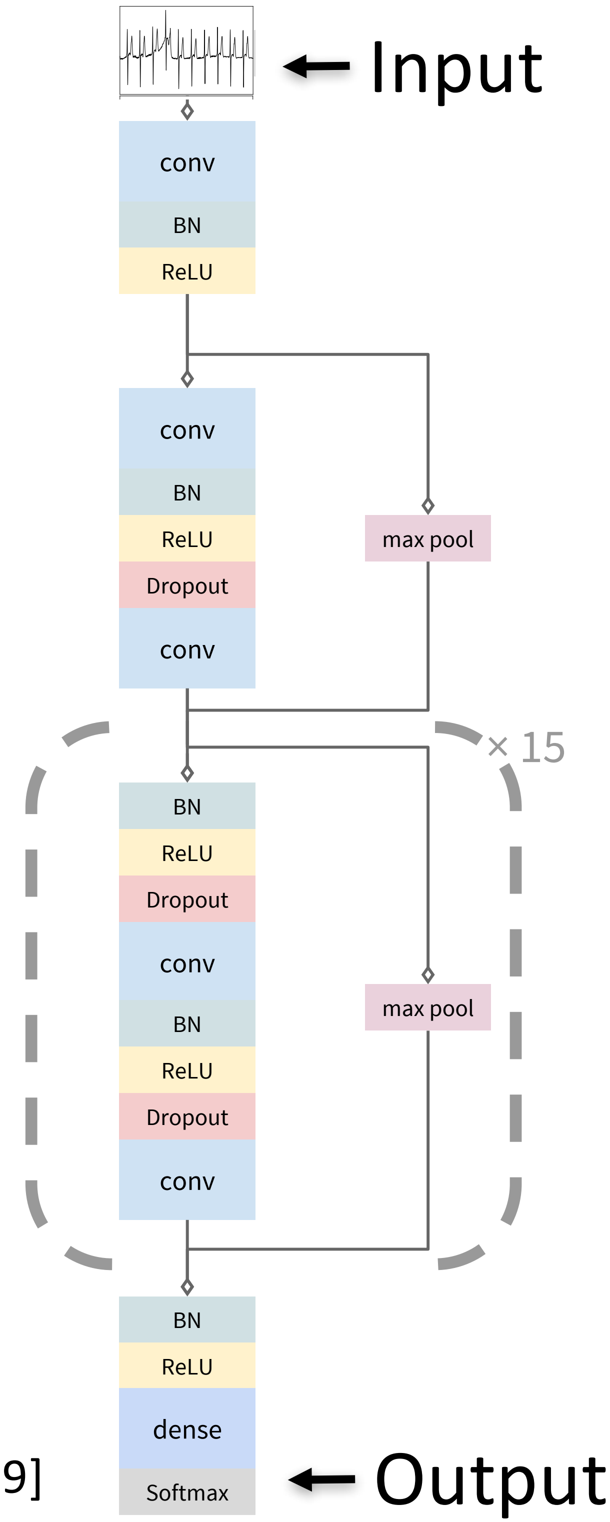
- ~90K ECG records annotated (from ~50K patients)
- Identify 12 heart arrhythmias, sinus rhythm and noise for a total of 14 output classes

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

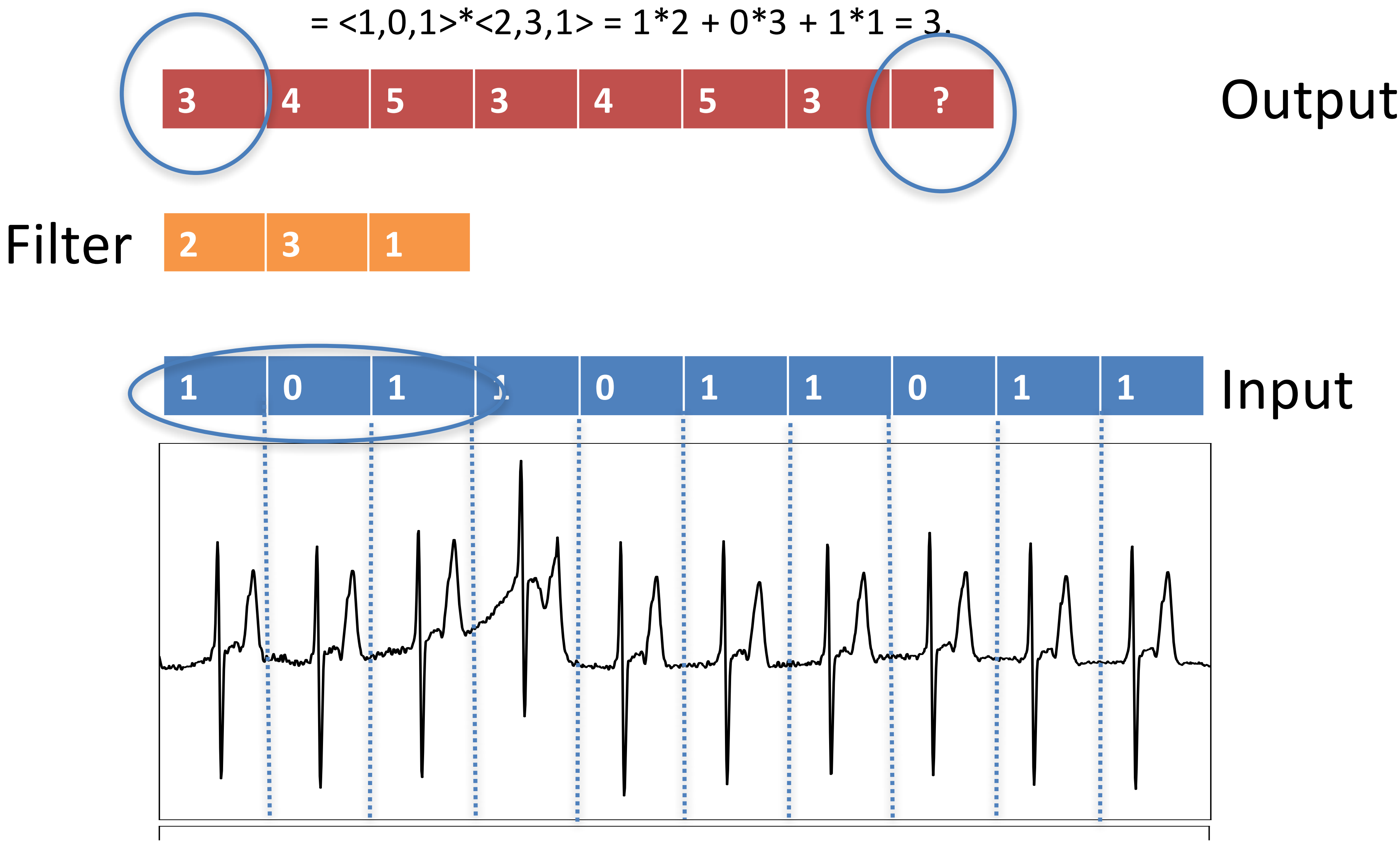
Deep convolutional network

- 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)



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

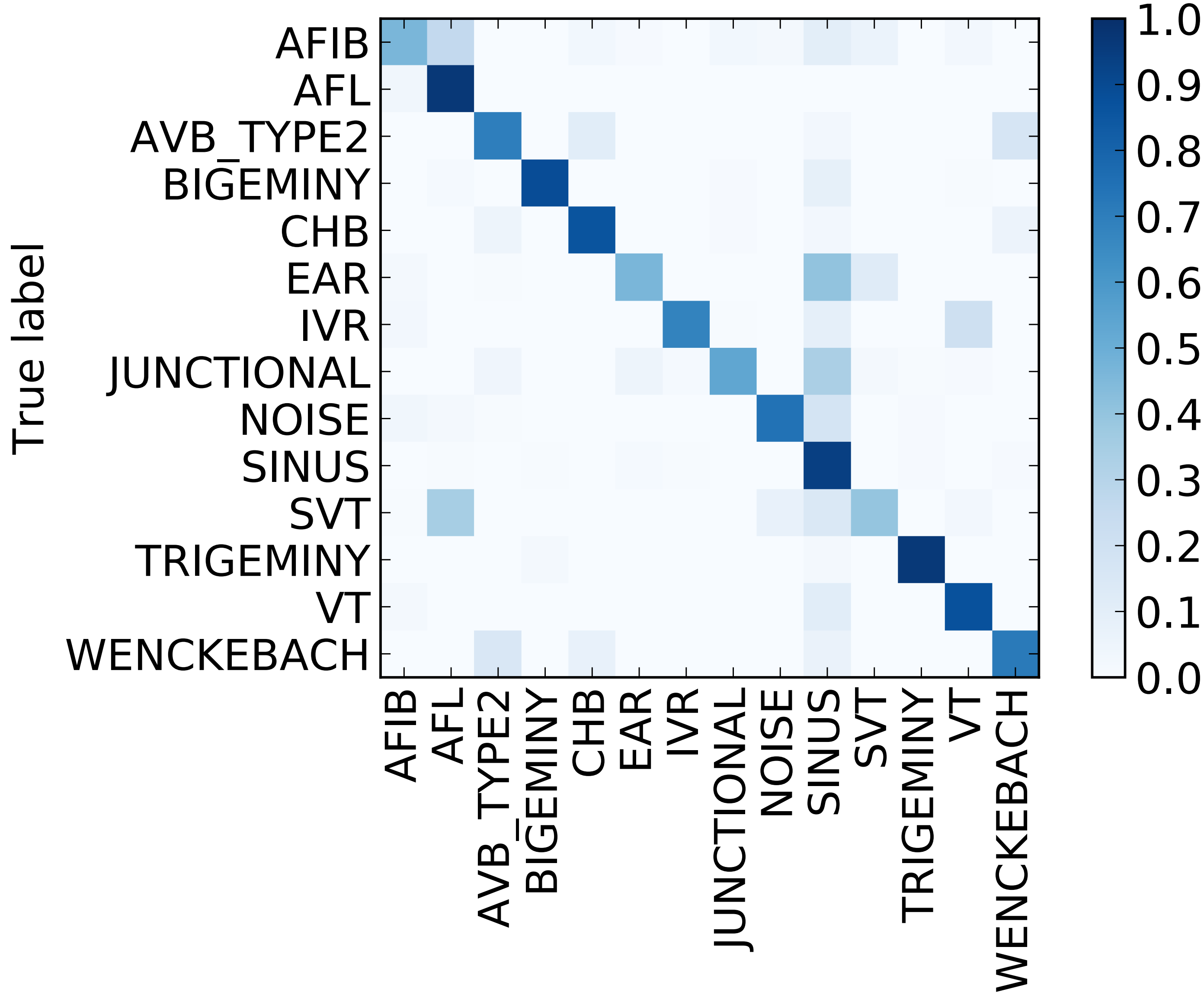
Example of 1D convolution



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

Evaluation



AFIB
AFL
AVB_TYPE2
BIGEMINY
CHB
EAR
IVR
JUNCTIONAL
NOISE
SINUS
SVT
TRIGEMINY
VT
WENCKEBACH
True label
Predicted label
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Summary so far

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

Outline

1. Clinical text
   – Case study: Prediction of sepsis (severe infection) from electronic health records
2. Physiological time-series
   – Case study: Monitoring babies in neonatal ICUs
   – Case study: Detecting atrial fibrillation
3. **Imaging**
   – Cardiology, pathology, radiology

Image classification

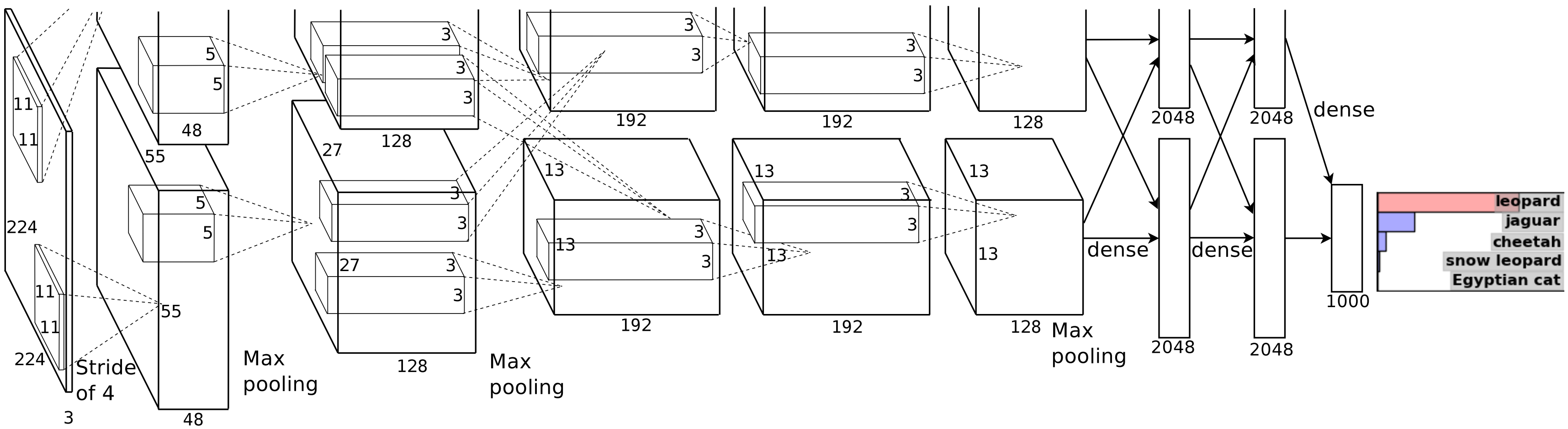
Input:

Output:
label

Image classification using convolutional neural networks

Input:

Output:



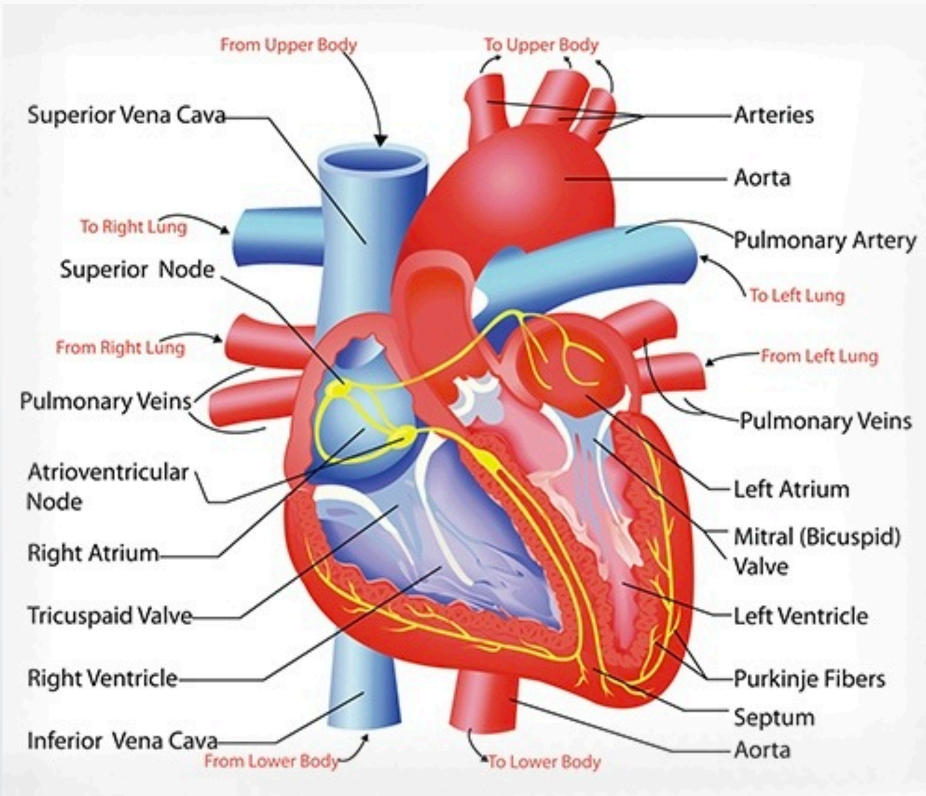
Krizhevsky, Sutskever, Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS ‘12

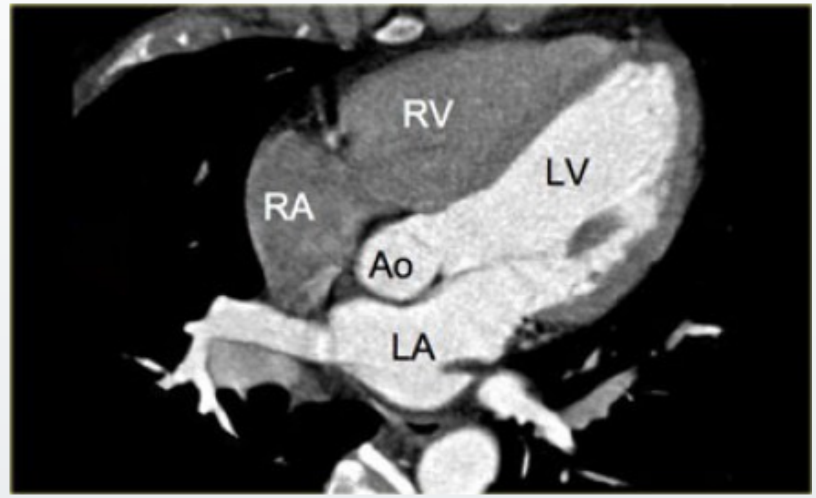
Low hanging fruit: applying image classification to medicine

1. Many simple disease recognition tasks exist in medicine - and can be carried out by an experienced radiologist in 2 minutes or less
    1. e.g. lung cancer or not
    2. pneumonia or not
    3. breast cancer or not
    4. fluid around the heart or not
2. Many of the first successes in medical image classification have involved situations with very large data sets, already labeled in the context of routine clinical care
    1. Chest x-rays
    2. Mammograms
3. Barriers to data export and sharing have limited the size of many other data sets

[Slide credit: Rahul Deo, BWH]

The structure of the heart





4 chambers: RA, RV, LA, LV
4 valves: TV, PV, MR, AV
2 circulations in series: pulmonary and systemic

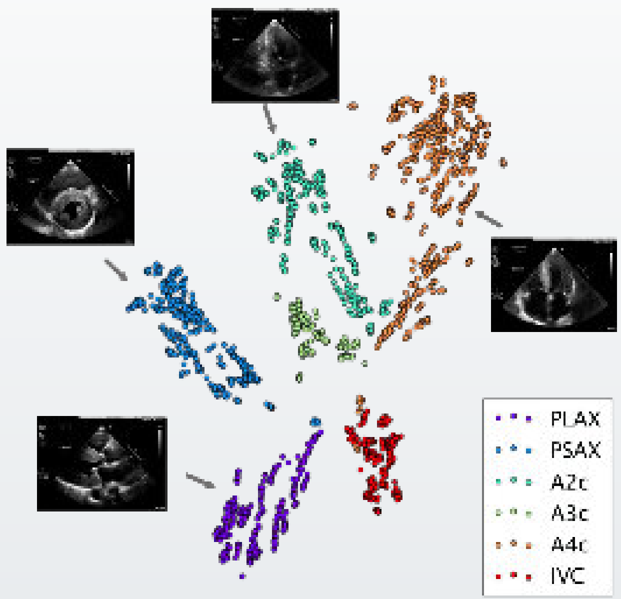
[Slide credit: Rahul Deo, BWH]

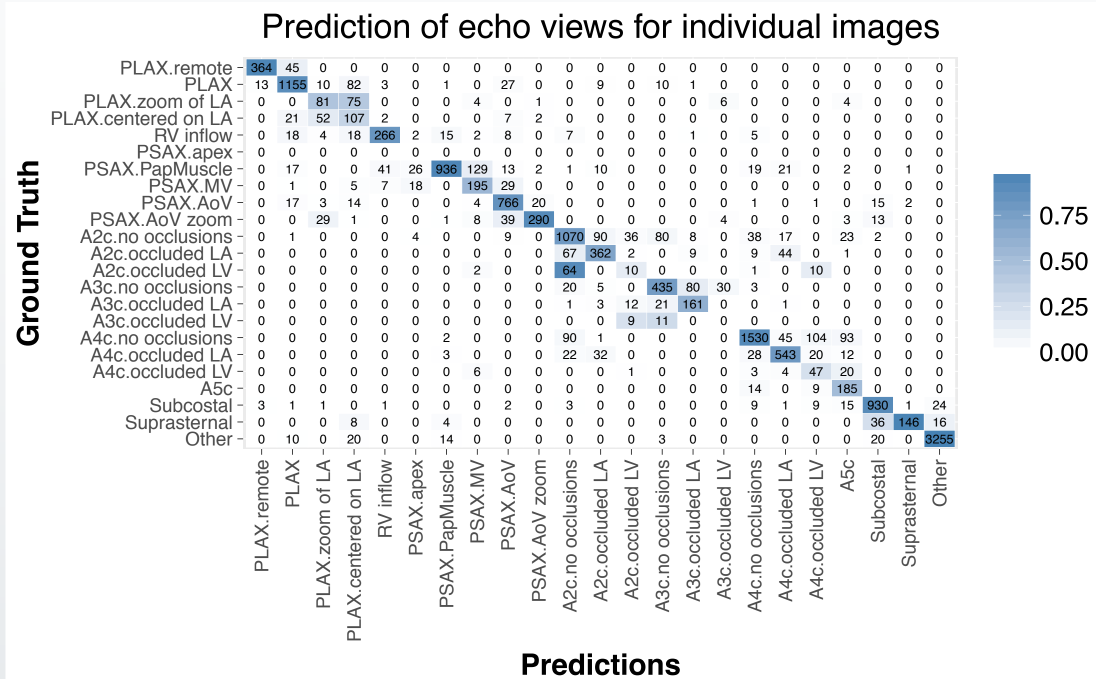
Decisions (sometimes) guided by imaging

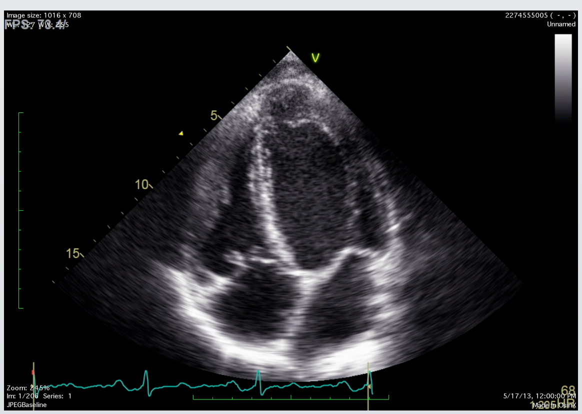
Disease	Decision	Inputs
Heart failure	Decision to implant a defibrillator to prevent sudden death	Symptoms + ejection fraction of the heart <35%
Coronary artery disease	Angioplasty and stenting of a coronary artery	Symptoms + stenosis > 70%
Aortic stenosis	Valve replacement	Symptoms + valve area + enlargement of the heart
Atrial fibrillation	Decision to start anticoagulation to prevent stroke	Age, sex, other diagnoses
Myocardial infarction	Decision to start aspirin and a statin to prevent a future heart attack	A risk model based on age, sex, lab values, blood pressure, diabetes

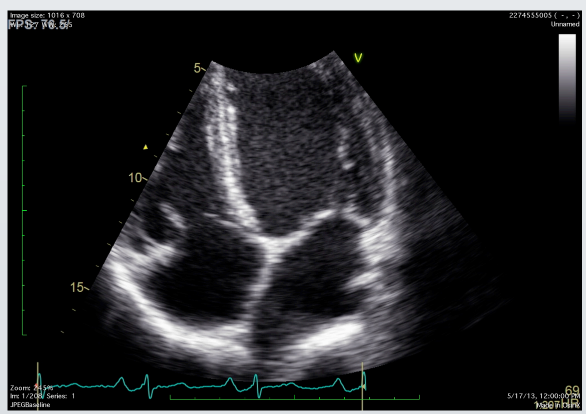
[Slide credit: Rahul Deo, BWH]

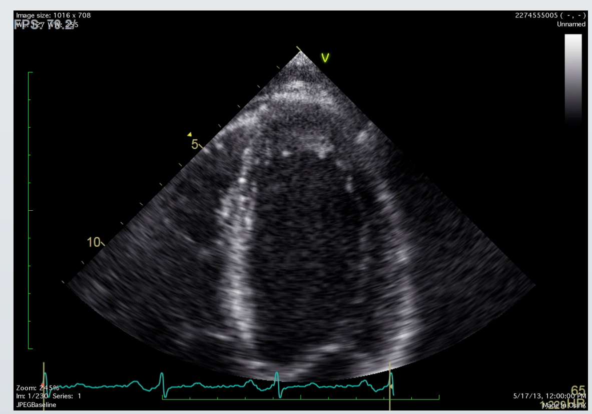
Echocardiography view classification





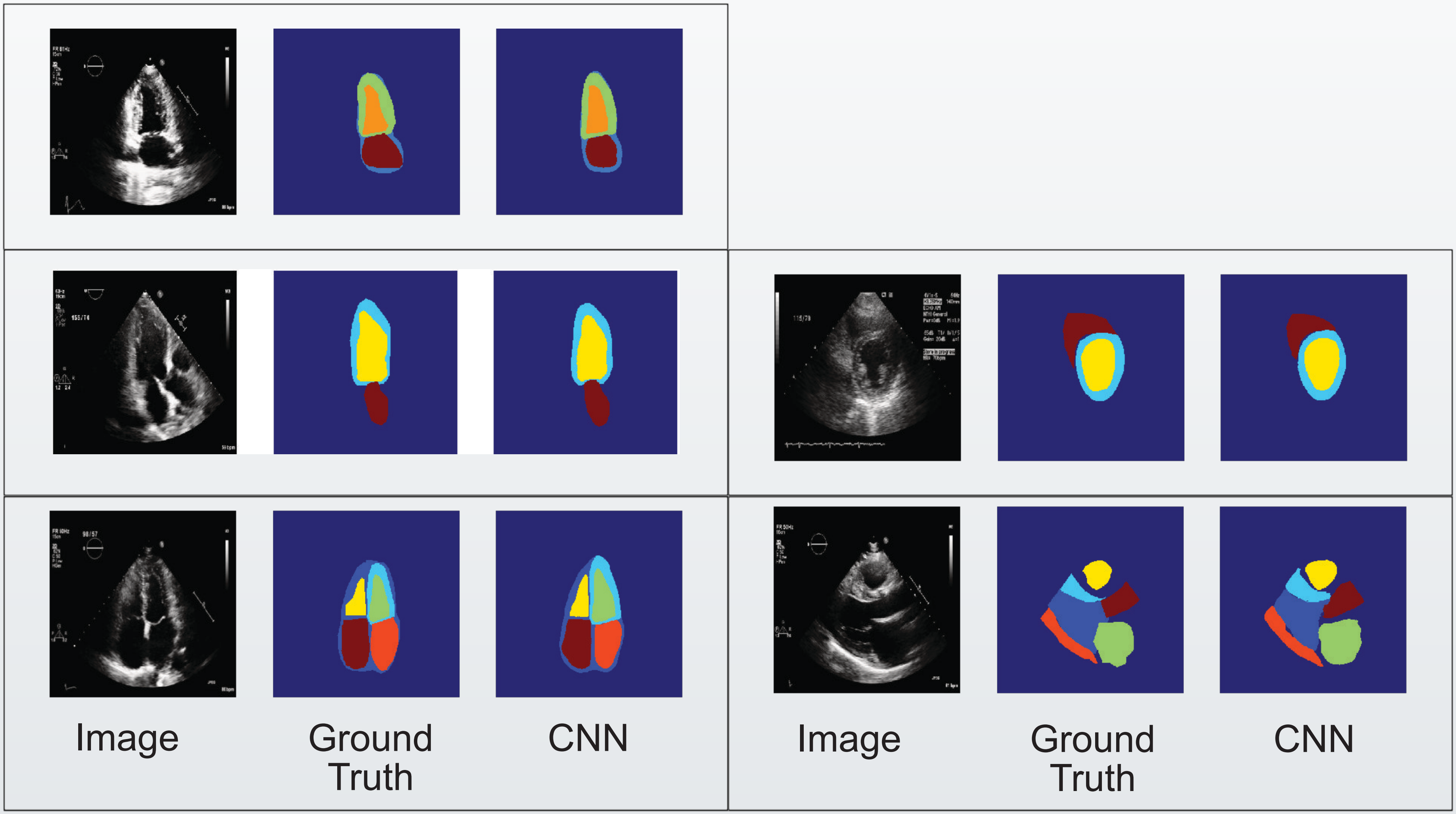






Zhang … Deo, arXiv, 2017

Echocardiography segmentation



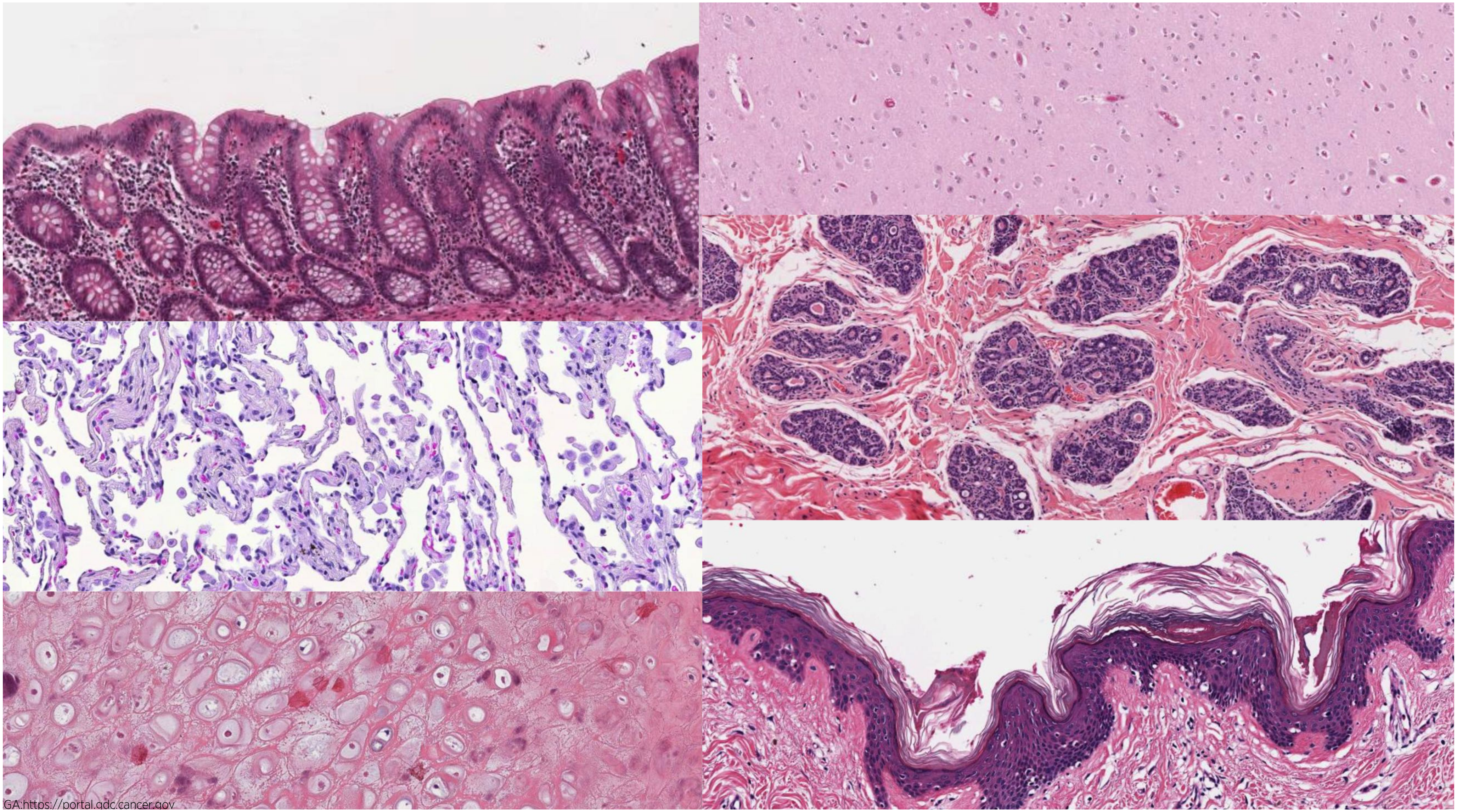
[Slide credit: Rahul Deo, BWH]

Echocardiography automated measurements

Metric	Number of Echo Studies Used for Comparison	Median Value (IQR)	Absolute Deviation - % of Manual (Automated vs. Manual Measurement)		
			50	75	95
Left atrial volume	4800	52.6 (40.0-71.0)	16.1	29.3	66.2
Left ventricular diastolic volume	8457	92.1 (71.8-119.1)	17.2	30.5	68.0
Left ventricular systolic volume	8427	33.2 (24.1-46.8)	26	47	108
Left ventricular mass	5952	148.0 (117.3-159.9)	15.1	27.6	61
Left ventricular ejection fraction	6407	64.8 (58.3-59.41)	9.7	17.2	39.9
Global longitudinal strain	418	19.0 (17.0-21.0)	7.5	13.6	30.8
Global longitudinal strain (Johns Hopkins PKD study)	110	18.0 (16.0-20.0)	9.0	17.1	39.4

[Slide credit: Rahul Deo, BWH]

Pathology



[Slide credit: Andy Beck, PathAI]

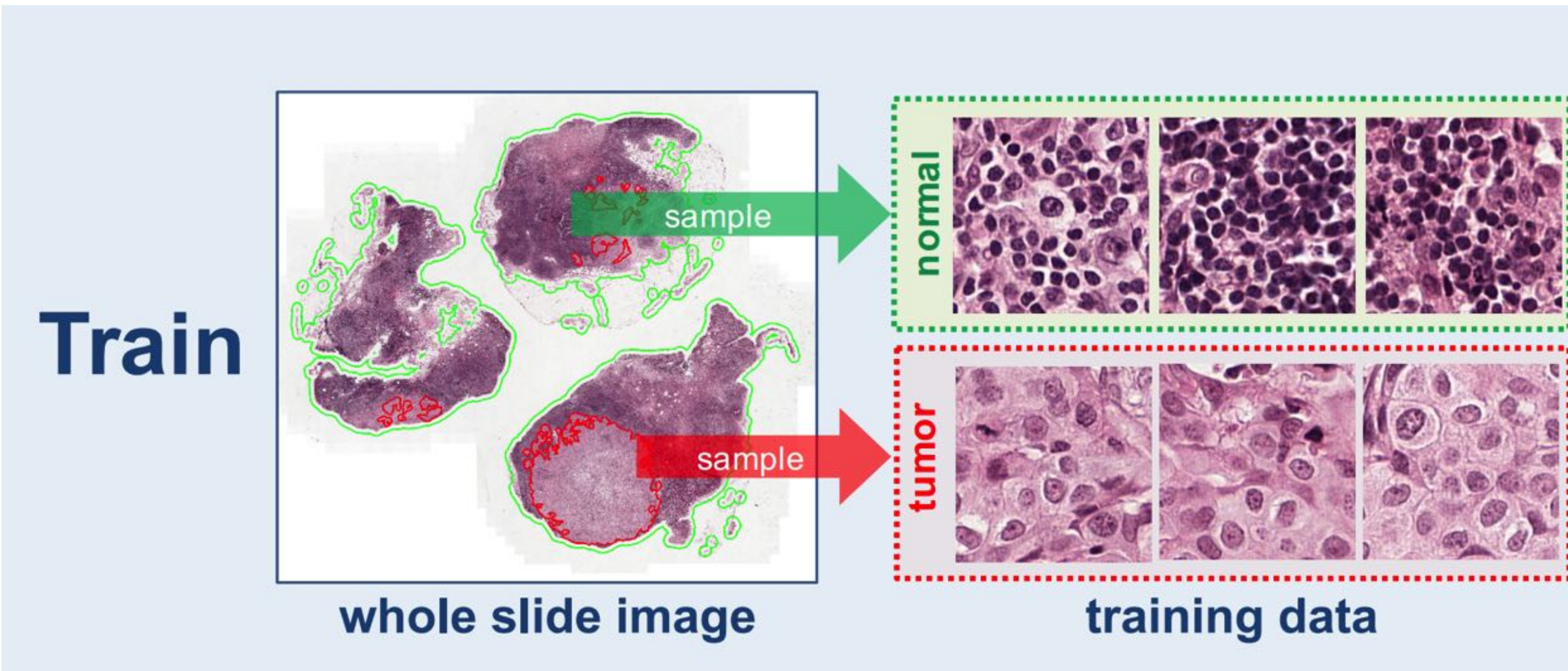
Pathologists aren't consistent – opportunity to increase reliability

	Phase II Interpretation of Same Individual Pathologist					
Phase I Interpretation of Individual pathologist	**Benign without atypia**	**Atypia**	**DCIS**	**Invasive**	**Total**	**Agreement rates of phase I and II interpretations, % (95% CIs)**
Benign without atypia	**947**	137	41	5	1130	84 (81-86)
Atypia	157	**303**	109	2	571	53 (47-59)
Ductal Carcinoma *in situ* (DCIS)	43	94	**792**	14	943	84 (81-87)
Invasive Breast Cancer	8	4	11	**273**	296	92 (88-95)
Total	1155	538	953	294	2940	79 (77-81)
*The same slide was interpreted on two different occasions separated in time by 9 or more months						

Ref: Jackson SL … Elmore JG. Ann Surg Oncol. 2017 May;24(5):1234-1241.

[Slide credit: Andy Beck, PathAI]

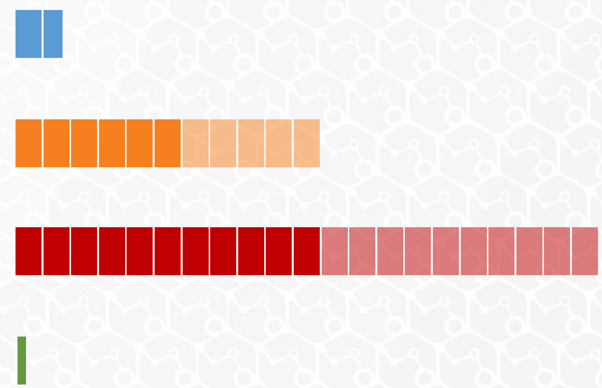
Again, we can apply image classification approaches



[Slide credit: Andy Beck, PathAI]

Deep learning model outperforms human pathologists in the diagnosis of metastatic cancer

	Error Rate (1-AUC)
Pathologists in competition	3.5%
Pathologists in clinical practice[1]	13 – 26%
Pathologists on micro-metastasis[2]	23 – 42%
Deep learning model	0.65%



[1] n=12 [2] Small tumors

References: Wang, Khosla, … Beck (2016) https://arxiv.org/abs/1606.05718 Camelyon16 (JAMA, 2017)

[Slide credit: Andy Beck, PathAI]

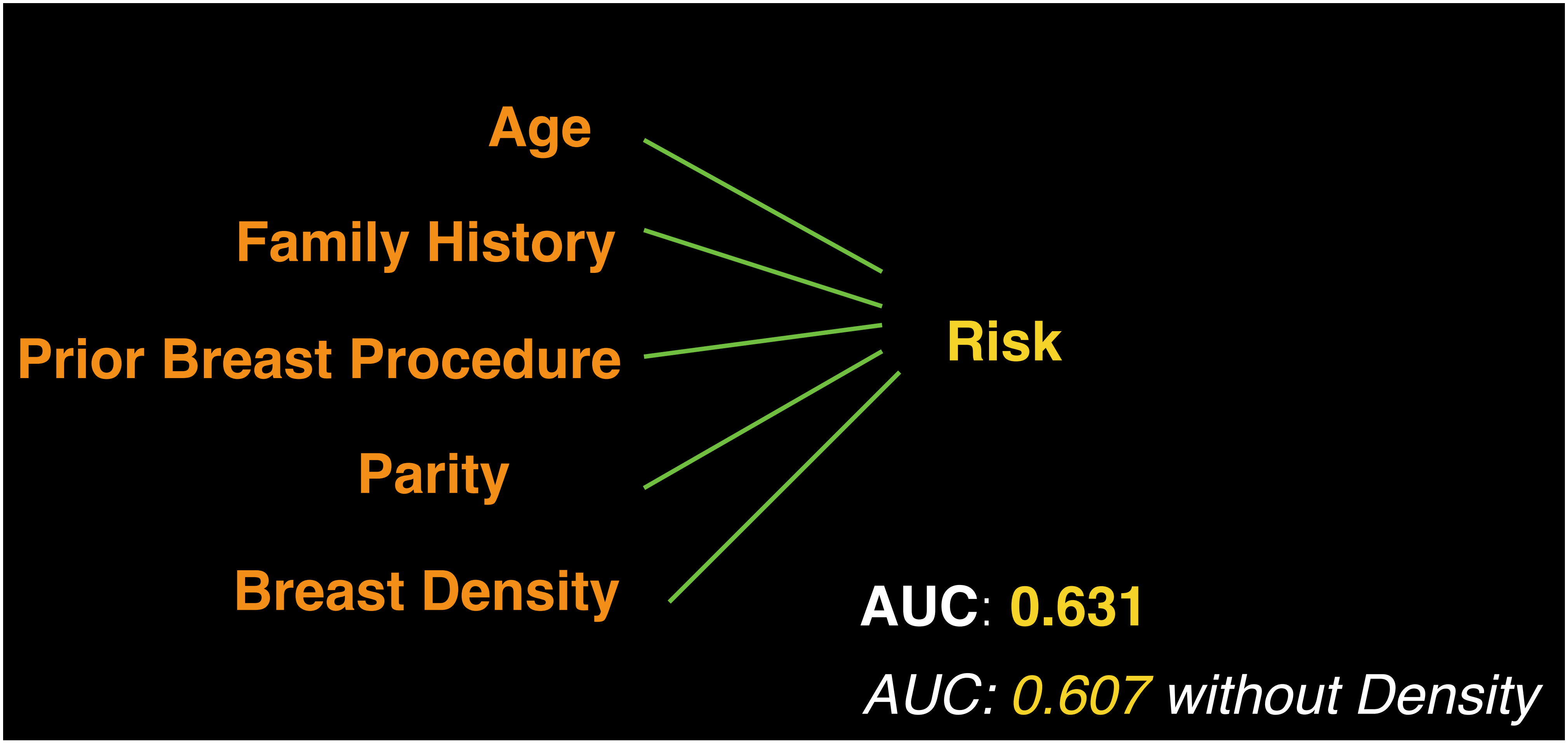
Breast cancer screening

Every Year:

- Of 3.8 billion women in the world, > 2 million diagnosed with breast cancer each year
- > 40,000 deaths in the US alone
- > 600,000 deaths in the world

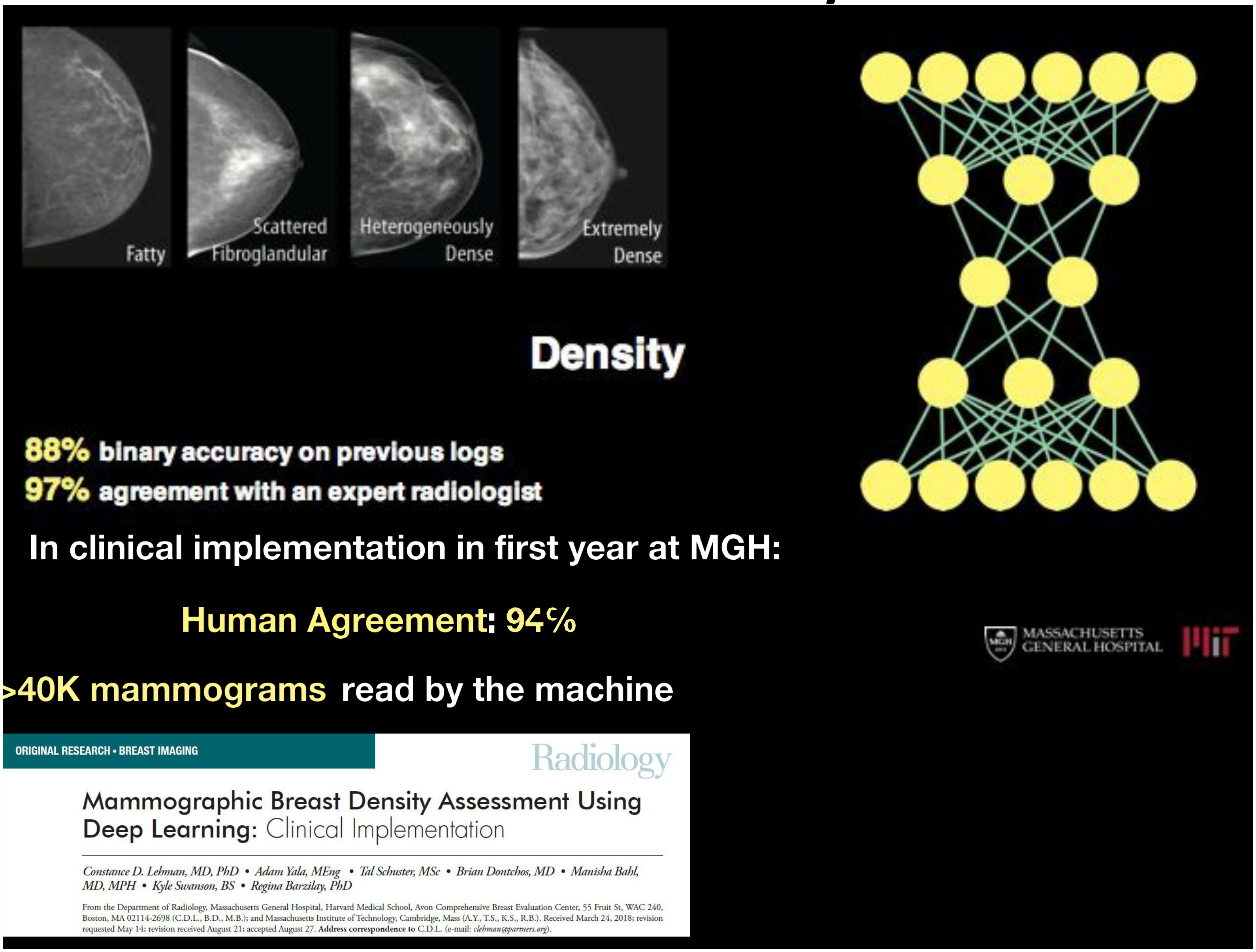
[Slide credit: Connie Lehman, MGH]

Classical risk scores



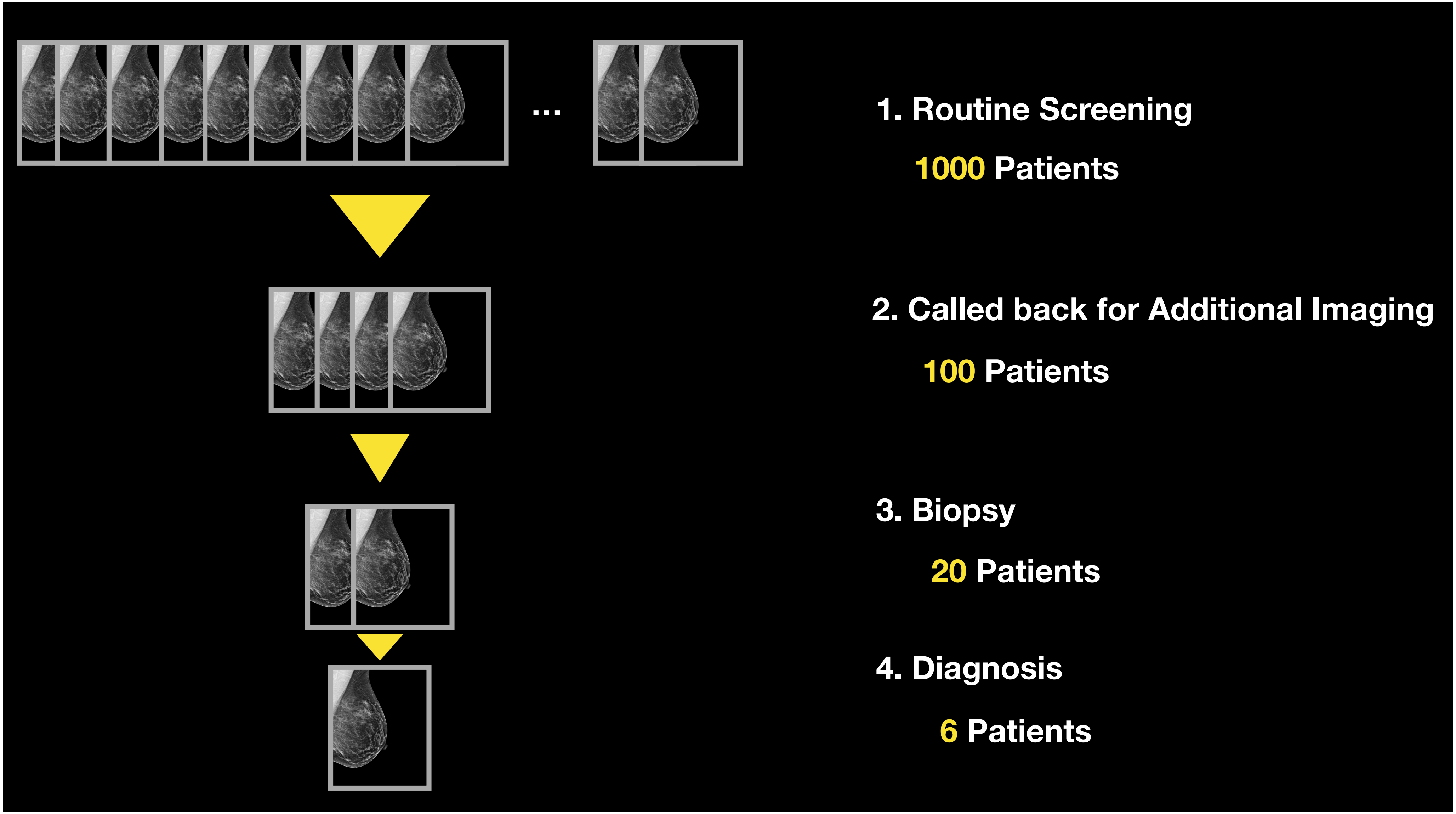
[Slide credit: Connie Lehman, MGH]

Using image classification to predict breast density



[Slide credit: Connie Lehman, MGH]

Triaging mammograms



[Slide credit: Adam Yala, MIT]

Triaging mammograms – estimated to reduce mammograms needing to be ready by 20%

Setting	Sensitivity (95% CI)	Specificity (95% CI)	% Mammograms Read (95% CI)
Original Interpreting Radiologist	90.6% (86.7, 94.8)	93.0% (92.7, 93.3)	100% (100, 100)
Original Interpreting Radiologist + Triage	**90.1% (86.1, 94.5)**	**93.7% (93.0, 94.4)**	**80.7% (80.0, 81.5)**

[Slide credit: Adam Yala, MIT]