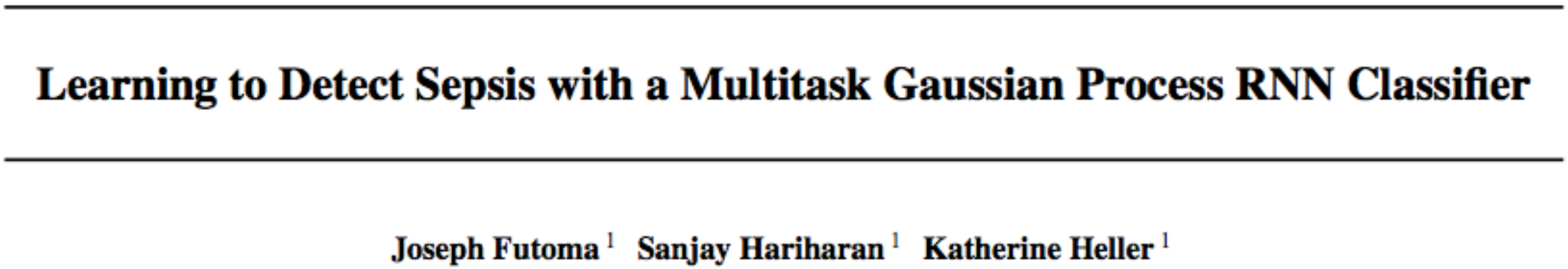


Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier

Joseph Futoma [1] Sanjay Hariharan [1] Katherine Heller [1]

Proceedings of the 34th International
Conference on Machine Learning

Presented by William Hang for CS273B

What is sepsis?

Body’s own immune response to infection causes huge inflammatory response in the rest of the body

Usually caused by bacterial or fungal infections

Sepsis is totally treatable with IV and antibiotics

Causes millions of deaths worldwide annually

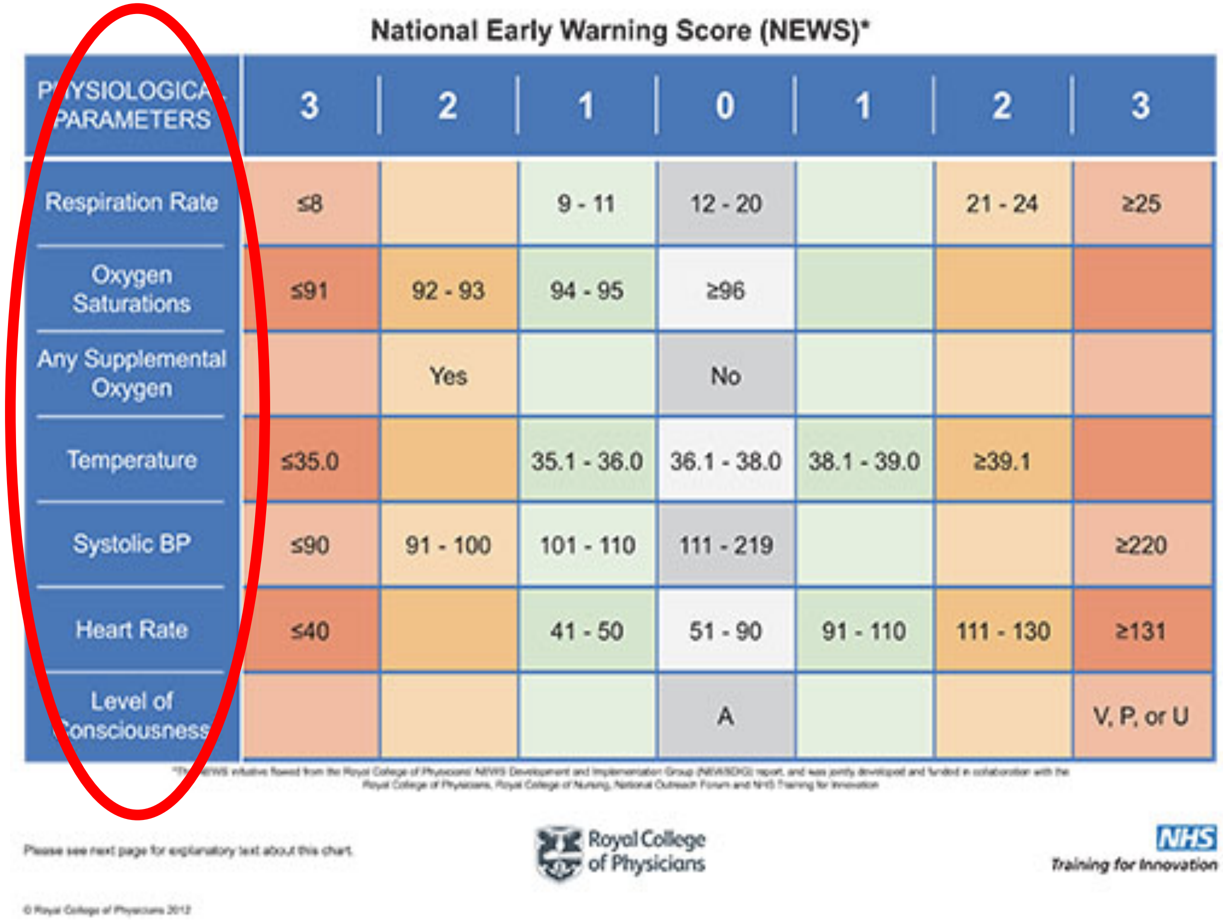
Detecting and then treating this early is key

How do you detect it early right now?

Duke Hospital uses NEWS

Looks at 6 physiological variables to determine a score

Really low specificity, almost 63.4% false positive rate in Duke Hospital



National Early Warning Score (NEWS)*

PHYSIOLOGICAL PARAMETERS	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51 - 90	91 - 110	111 - 130	≥131
Level of Consciousness				A			V, P, or U

*The NEWS initiative flowed from the Royal College of Physicians' NEWS Development and Implementation Group (NEWSDIG) report, and was jointly developed and funded in collaboration with the Royal College of Physicians, Royal College of Nursing, National Outreach Forum and NHS Training for Innovation

Please see next page for explanatory text about this chart.

Royal College of Physicians

NHS Training for Innovation

Challenges to early detection

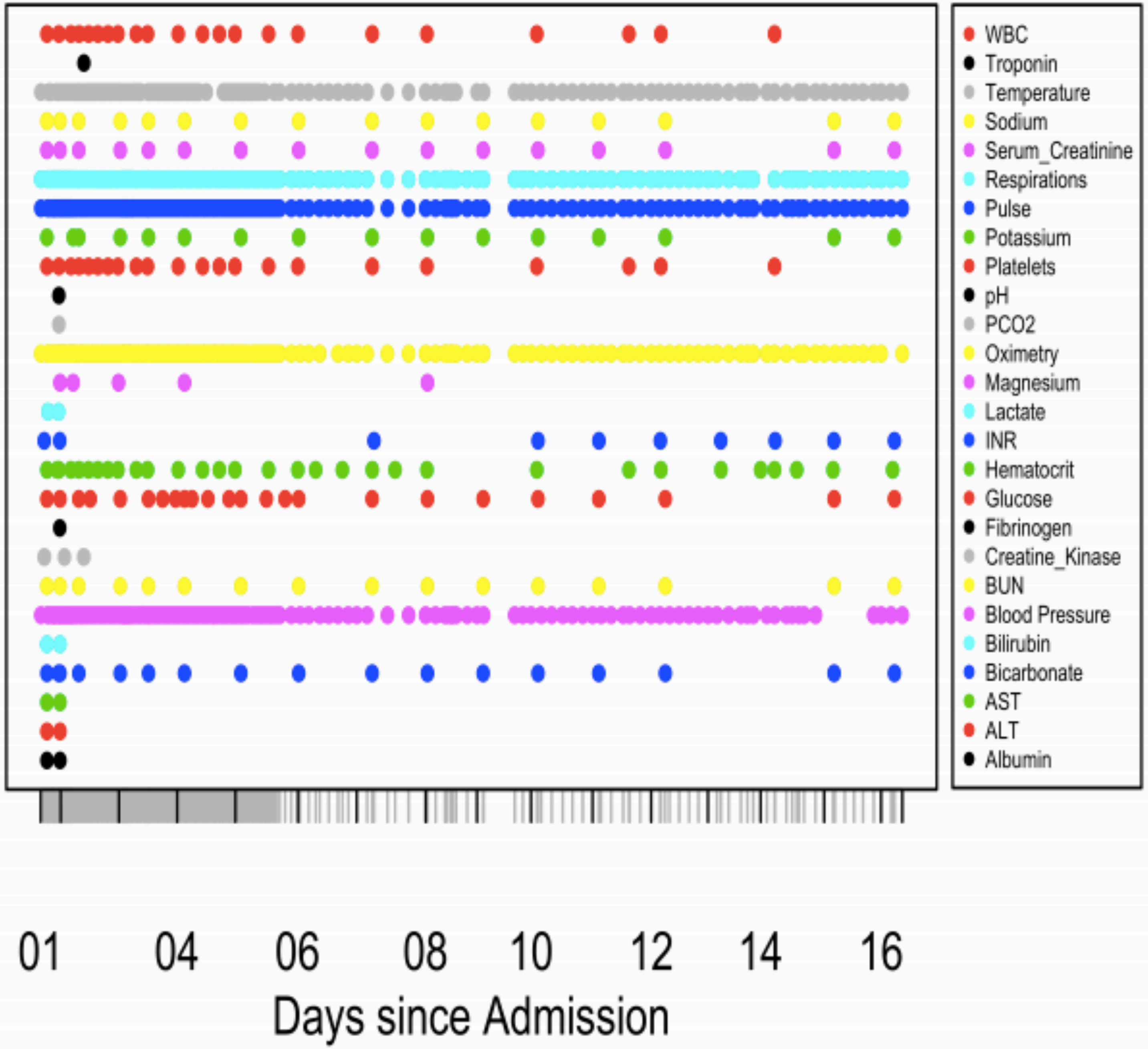
Every patient is different!

Time series prediction problem

Lots of missing data per patient at various points in time

Sepsis start time is generally unknown

When do you serve timely alarms?

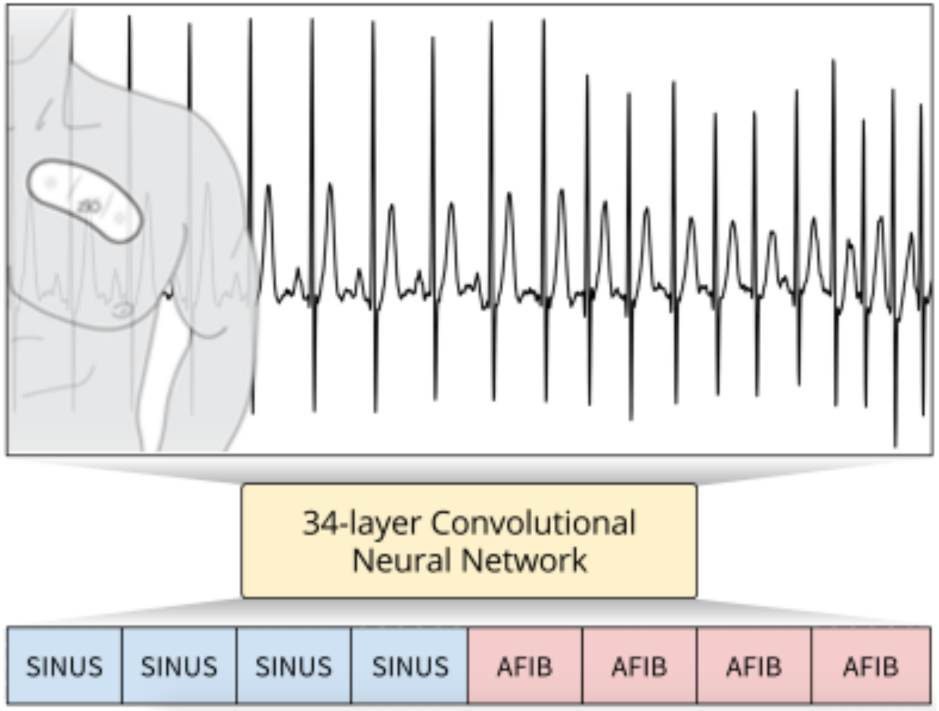

WBC
Troponin
Temperature
Sodium
Serum_Creatinine
Respirations
Pulse
Potassium
Platelets
pH
PCO2
Oximetry
Magnesium
Lactate
INR
Hematocrit
Glucose
Fibrinogen
Creatine_Kinase
BUN
Blood Pressure
Bilirubin
Bicarbonate
AST
ALT
Albumin
01
04
06
08
10
12
14
16
Days since Admission

Other research trying to tackle this problem

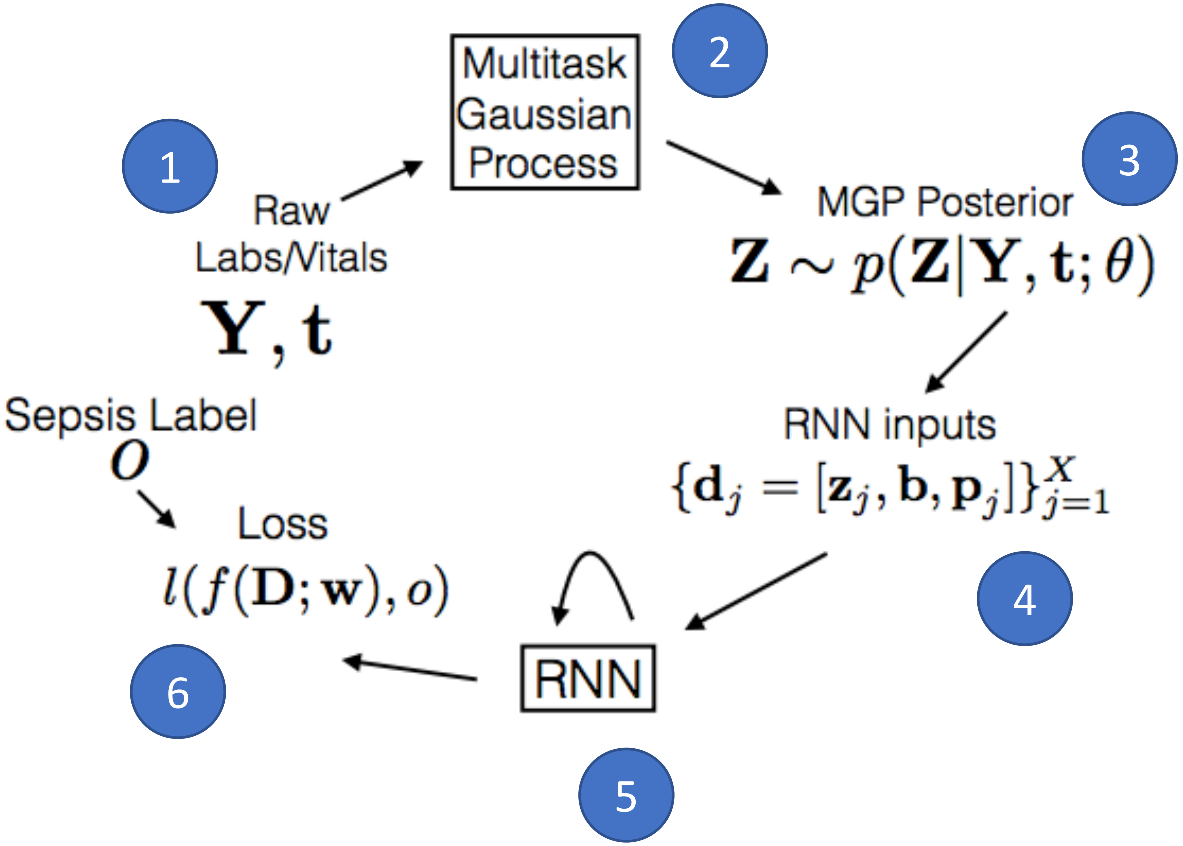
Rothman Index + Cox regression don’t operate on time series data

Works on chronic kidney disease and glaucoma prediction operate on months and years timescale; sepsis is important on the order of hours

Other works use RNNs to classify clinical time series



This paper’s approach



1

Raw Labs/Vitals

$\mathbf{Y}, \mathbf{t}$

2

Multitask Gaussian Process

3

MGP Posterior

$\mathbf{Z} \sim p(\mathbf{Z}|\mathbf{Y}, \mathbf{t}; \theta)$

4

RNN inputs

$\{\mathbf{d}_j = [\mathbf{z}_j, \mathbf{b}, \mathbf{p}_j]\}_{j=1}^{X}$

5

RNN

6

Sepsis Label

o

Loss

$l(f(\mathbf{D}; \mathbf{w}), o)$

Data for a single patient:

$$\mathcal{D}_i = \{\mathbf{b}_i, \mathbf{t}_i, \mathbf{Y}_i, \mathcal{P}_i, o_i\}$$

b = baseline state

t = time

Y = physiological variables: vitals and lab tests

P = administered drugs at various time steps

o = whether the patient got sepsis

Predict **o**!

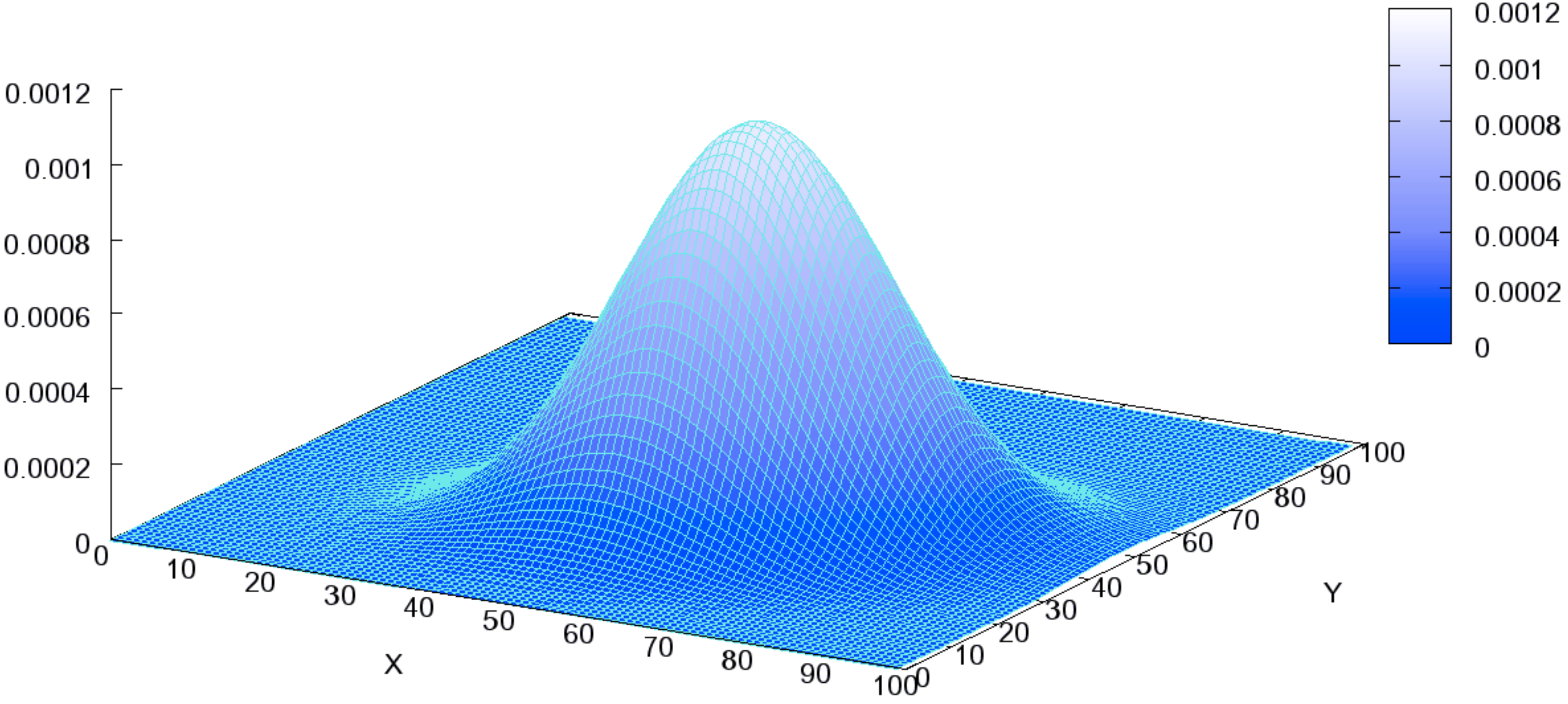
Okay, what's a Gaussian Process?

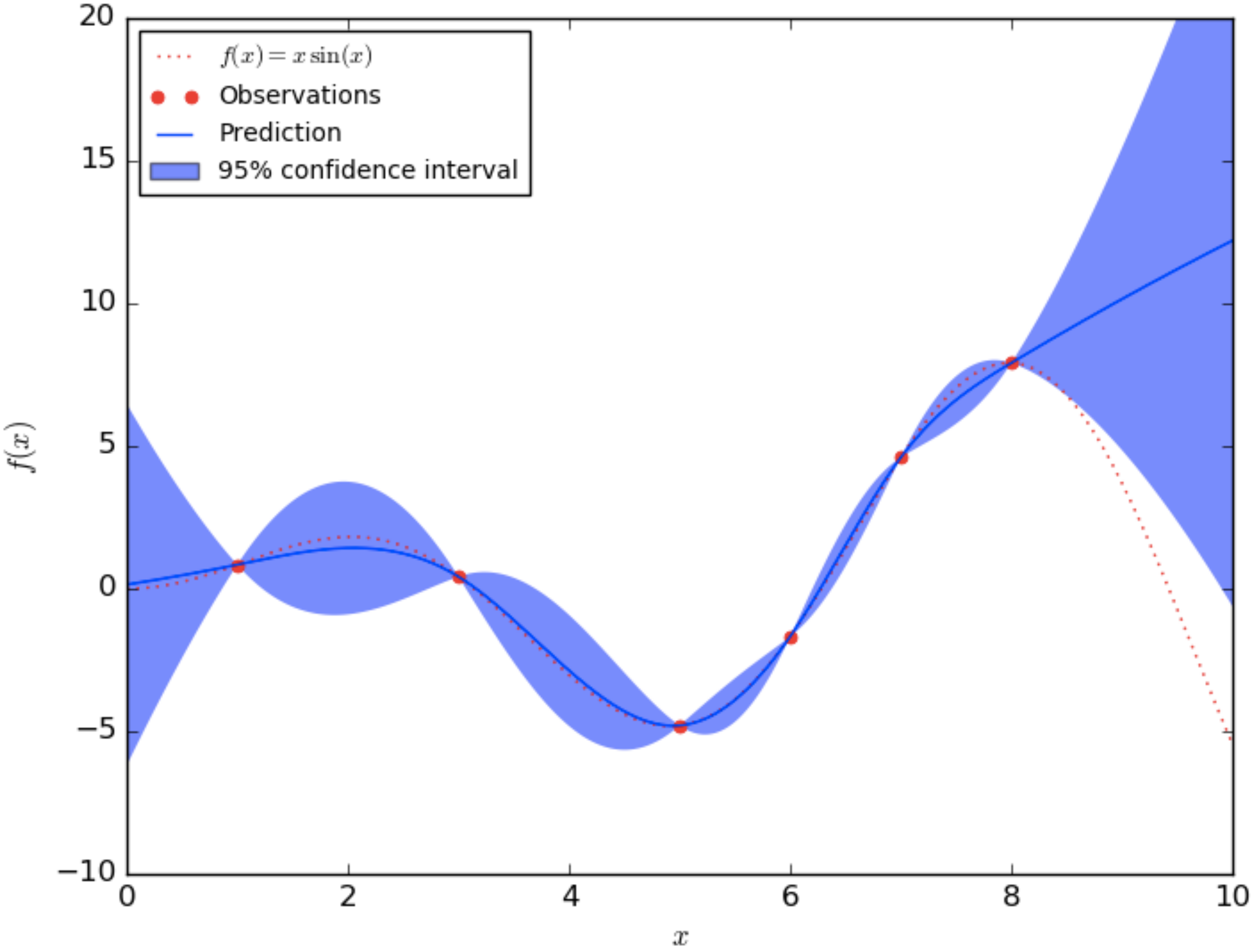
Estimates the probability distribution over *f(***X***)* given training data X_{train}

We do this by trying to calculate a posterior over possible *f*(**X**) by comparing how similar **X** is to X_{train} through the use of covariance matrices. This covariance matrix is computed with a kernel function

If you have something in **X** that looks kind of like stuff you've seen in X_{train}, you could say that its mean and variance is distributed like things you've already seen

Multivariate Normal Distribution




$f(x) = x\sin(x)$
Observations
Prediction
95% confidence interval
20
15
10
5
0
−5
−10
$f(x)$
0
2
4
6
8
10
x

So what?

If you give a GP input values, you can draw output values from a correlated Gaussian distribution that knows how similar your input values are to things it's seen in the past

This sampling technique is super useful with missing values at various time steps

We will see how we train this GP to give us what we want

Now what’s a Multitask Gaussian Process?

Correlation kernel also operates across time

$$\mathrm{cov}(f_{im}(t), f_{im'}(t')) = K^{M}_{mm'} k^{t}(t, t')$$

You can learn the covariance of each pair of features, across time! This information will allow us to estimate the posterior distribution among all features, across time.

$$\Sigma_i = K^M \otimes K^{T_i} + D \otimes I,$$

This covariance matrix is actually MT × MT, so you know the distribution of feature values at each time step

How are GPs used with an LSTM?

$$\mu_{z_i} = (K^M \otimes K^{X_i T_i})\Sigma_i^{-1}\mathbf{y}_i \tag{5}$$

$$\Sigma_{z_i} = (K^M \otimes K^{X_i}) - (K^M \otimes K^{X_i T_i})\Sigma_i^{-1}(K^M \otimes K^{T_i X_i}) \tag{6}$$

Feed in all the **y**’s and sample a big **Z** that gives you imputed values across all time steps

$$\mathbf{D}_s = [\mathbf{Z}_s^\top, \mathbf{B}^\top, \mathbf{P}^\top]^\top, \quad \mathrm{vec}(\mathbf{Z}_s) \equiv \mathbf{z}_s \sim N(\mu_z, \Sigma_z; \boldsymbol{\theta})$$

The **z**’s go into your LSTM as inputs

How do you even train this thing?

z is a random variable! So you need to optimize the expected loss instead of average loss

$$\boldsymbol{w}^*, \boldsymbol{\theta}^* = \text{argmin}_{w,\theta} \sum_{i=1}^{N} \mathbb{E}_{z_i \sim N(\mu_{z_i}, \Sigma_{z_i}; \theta)}[l(f(\mathbf{D}_i; \boldsymbol{w}), o_i)].$$

According to them, you can’t take the gradient of this function, so you approximate the loss using Monte Carlo

$$\mathbb{E}_{z \sim N(\mu_z, \Sigma_z; \theta)}[l(f(\mathbf{D}; \boldsymbol{w}), o)] \approx \frac{1}{S} \sum_{s=1}^{S} l(f(\mathbf{D}_s; \boldsymbol{w}), o),$$

Everything is good now because you have a loss value whose gradients you can backprop through LSTM **and** the MGP because you can represent

$\mathbf{z} = \mu + Rx$, where x is noise and R is a matrix such that $\Sigma = RR^T$

So if you know the gradients to **z**, you can train the MGP as well!

Dataset

34 physiological variables per patient, some variables only present 2-4% of the time

35 baseline variables, 8 drugs, 49k patient admissions

Well defined clinical definition to figure out when a patient gets sepsis

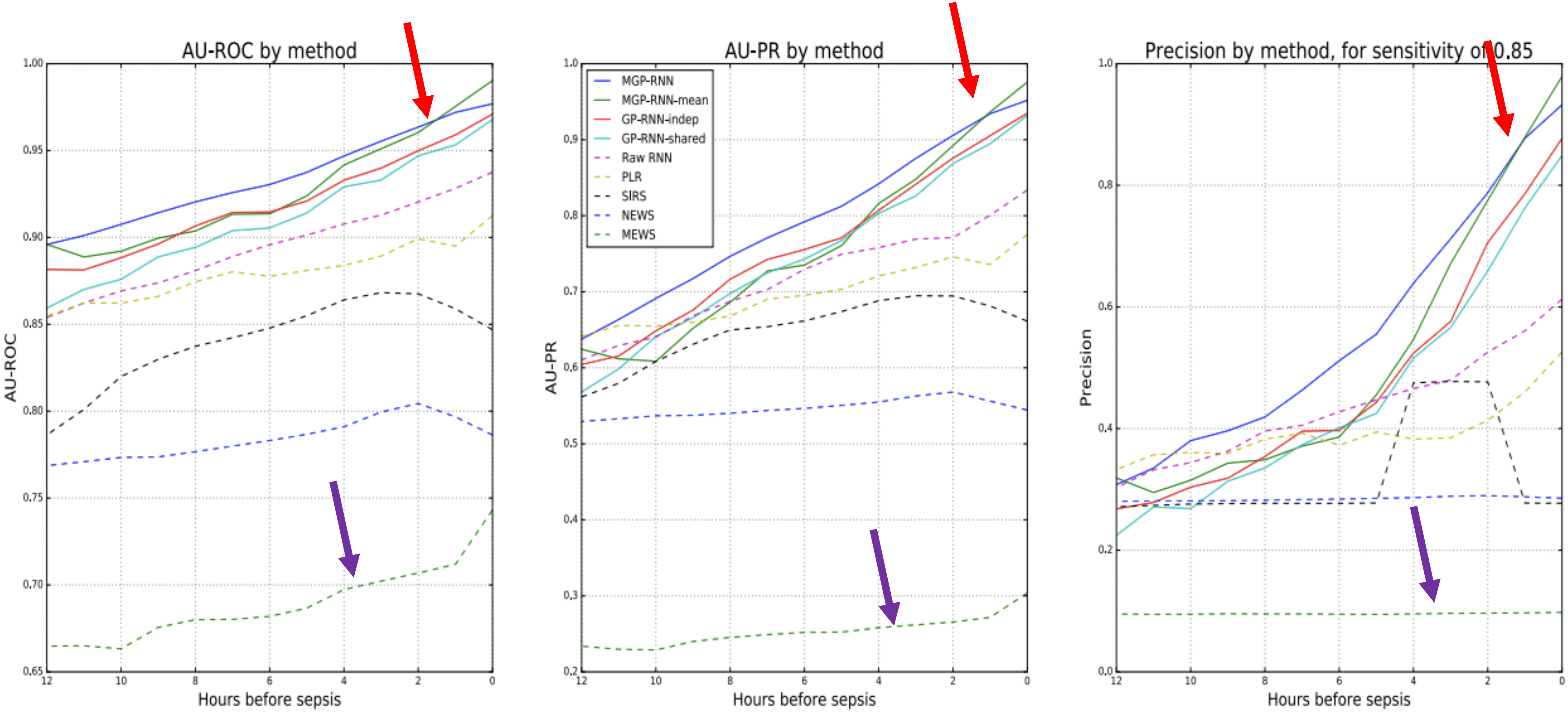
21.4% of patients in this dataset ended up getting sepsis

Training and Testing

80/10/10 split, L2, Adam, Tensorflow

Baselines: crappy hospital methods, baseline LSTMs, and their own modifications to their model

Most notable is MGP-RNN-mean, where they swap out $\mathbf{z}$ with $\boldsymbol{\mu}$


AU-ROC by method
AU-ROC
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
12
10
8
6
4
2
0
Hours before sepsis
AU-PR by method
AU-PR
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
MGP-RNN
MGP-RNN-mean
GP-RNN-indep
GP-RNN-shared
Raw RNN
PLR
SIRS
NEWS
MEWS
Hours before sepsis
Precision by method, for sensitivity of 0.85
Precision
1.0
0.8
0.6
0.4
0.2
0.0
Hours before sepsis

Performance

About 5 times fewer false alarms than NEWS, four hours in advance!

Better than MEWS and NEWS by 19.4% and 55.5% AUCROC

One of the first deep learning papers tackling sepsis detection

Extensions and Limitations

Clustering patients together and training a model for each

Using reinforcement learning or VAEs

Marginalize or subtract away known drug interactions with vital signs (using more domain specific knowledge)

Applying to other datasets

They're actively trying to productize this system to plug into EHRs and make predictions for hospitals (go them!)