What is Machine Learning and How Can It Be Leveraged in Anesthesiology?

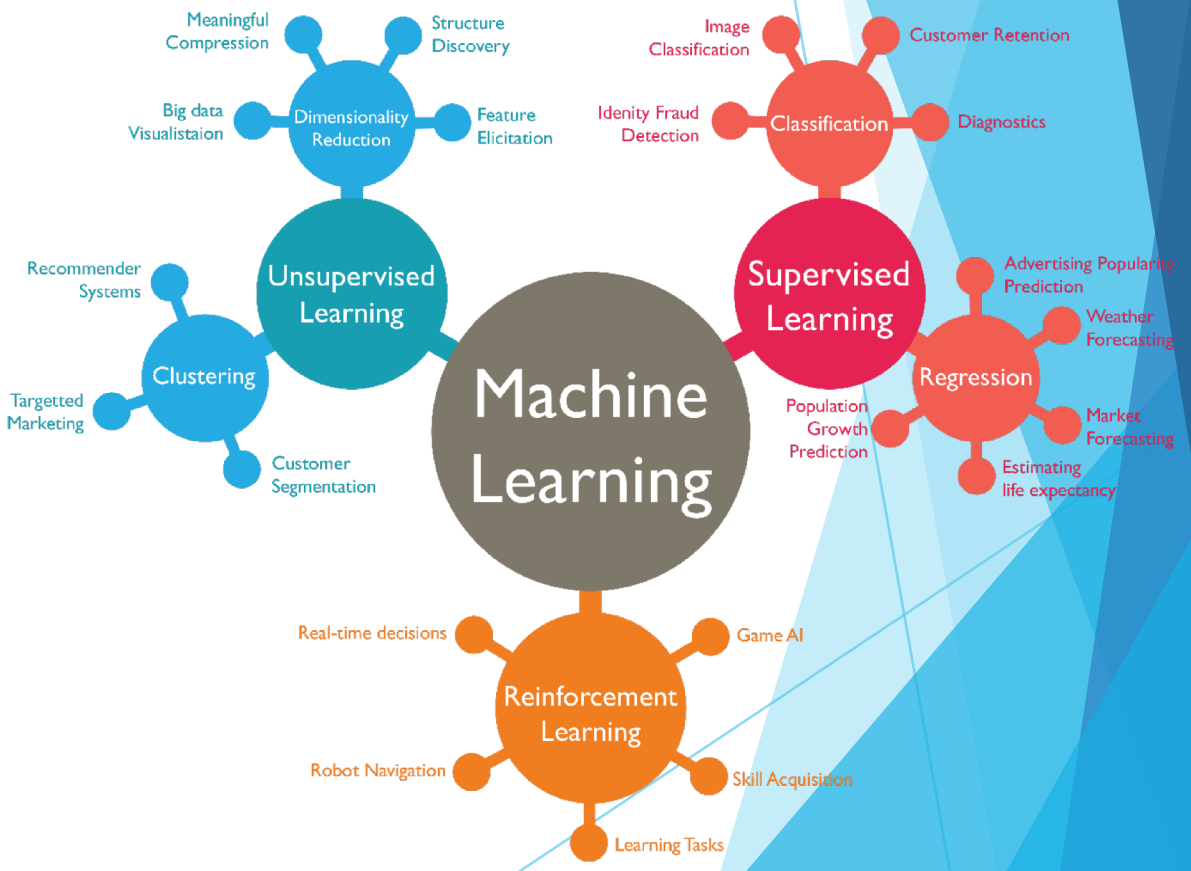
Christine Lee, MS

PhD Candidate, Department of Biomedical Engineering, UC Irvine

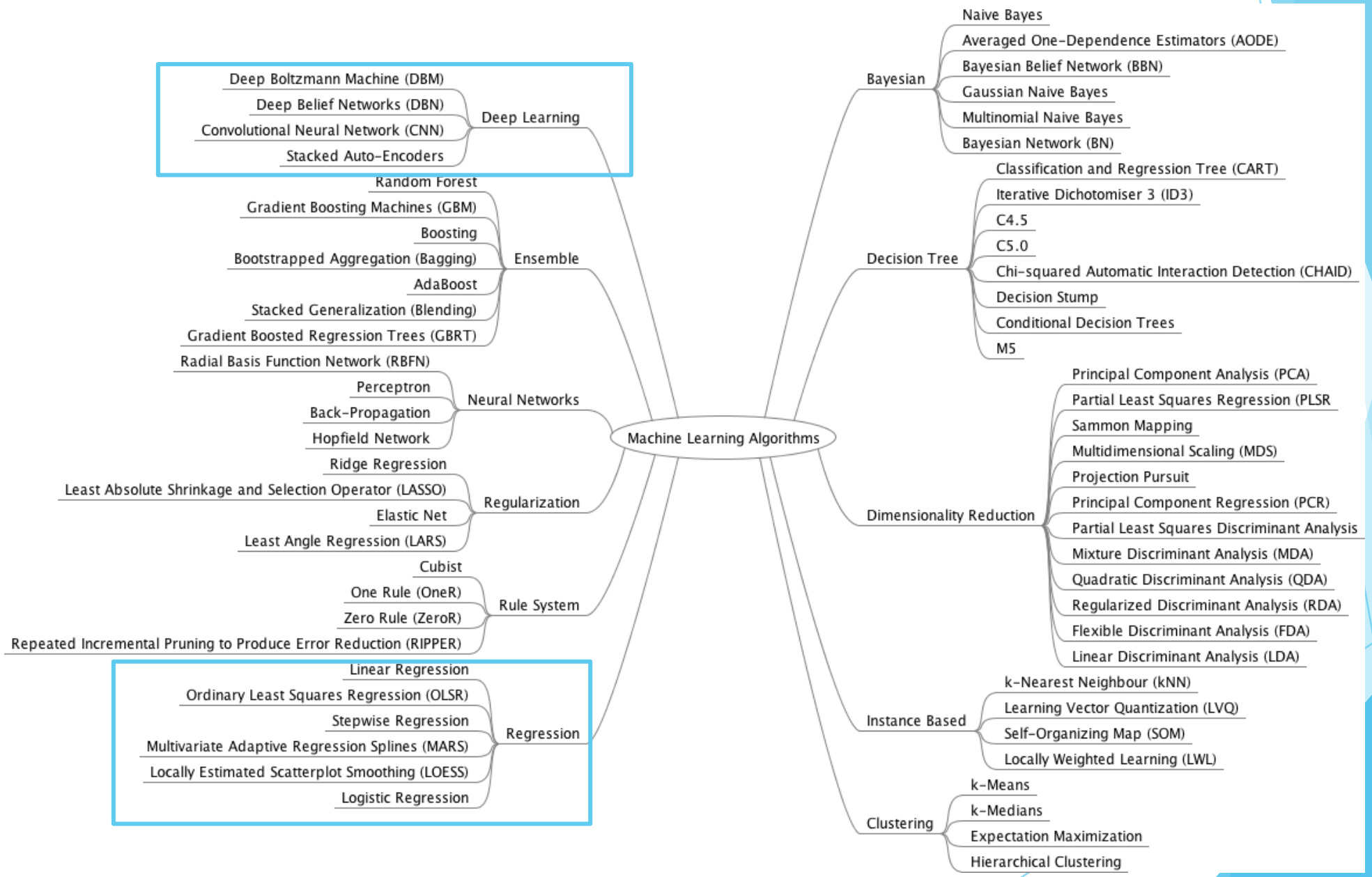
Senior Engineer, Algorithms-Applied ML Group, Edwards Lifesciences

What is Machine Learning?

- Sub-field of computer science
- Creating algorithms that can learn from data and make predictions on new data
- Supervised vs Unsupervised Learning
- “Machine learning is the science of getting computers to act without being explicitly programmed.” – Andrew Ng, VP & Chief Scientist of Baidu; Co-Chairman and Co-Founder of Coursera; and an Adjunct Professor at Stanford University



https://lh3.googleusercontent.com/OooN8vYsq5JVT6i7sbyX1JOURxzBpmrQ6mGc0lmNoakLz84ZJWOYnR8TtdaUoJlWnOkoCDeBvsjdNdyRsLEVgUsUoaWFwCeThpwOSmD7oj01F_6MTJcV1cm5r0dZXf2VAntEI-x6



http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

ML in Healthcare

JAMA | Original Investigation
December 12, 2017
Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer

Babak Ehteshami Bejnordi, MS; Mitko Veta, PhD; Paul Johannes van Diest, MD, PhD; Bram van Ginneken, PhD; Nico Karssemeijer, PhD; Geert Litjens, PhD; Jeroen A. W. M. van der Laak, PhD; and the CAMELYON16 Consortium

Abstract | Full Text

JAMA. 2017; 318(22):2199-2210. doi: 10.1001/jama.2017.14585

This diagnostic accuracy study compares the ability of machine learning algorithms vs clinical pathologists to detect cancer metastases in whole-slide images of axillary lymph nodes dissected from women with breast cancer.

JAMA | Original Investigation
December 12, 2017
Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes

Daniel Shu Wei Ting, MD, P
Haslina Hamzah, BSc; Ren
banayagam, MD, PhD; Man
L. Lamoureux, PhD; Ian Y. L
Guang He, MD, PhD; Ching
Yin Wong, MD, PhD

Abstract | Full Text

JAMA. 2017; 318(22):2211-;

This diagnostic accuracy st
ing diabetic retinopathy, gl

MIT News Browse or Search

Machine-learning system could aid critical decisions in sepsis care

Model predicts whether ER patients suffering from sepsis urgently need a change in therapy.

Rob Matheson | MIT News Office
November 7, 2018

Perioperative Medicine | October 2018
Supervised Machine-learning Predictive Analytics for Prediction of Postinduction Hypotension

Samir Kendale, M.D.; Prathamesh Kulkarni, Ph.D.; Andrew D. Rosenberg, M.D.; Jing Wang, M.D., Ph.D.

+ Author Notes

Anesthesiology 10 2018, Vol.129, 675-688. doi:10.1097/ALN.000

Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks

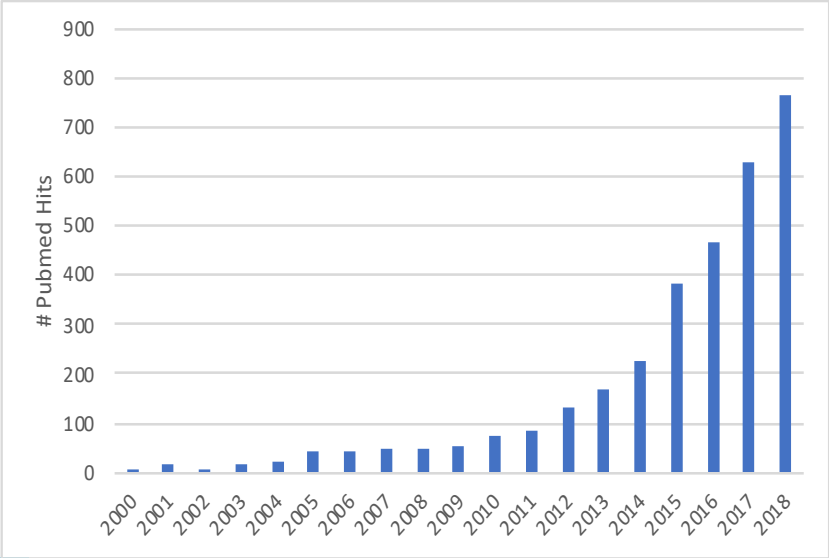
PRANAVSR@CS.STANFORD.EDU
AWNI@CS.STANFORD.EDU
MHAGHPANAHI@IRHYTHMTECH.COM
CBOURN@IRHYTHMTECH.COM
ANG@CS.STANFORD.EDU

Andrew Y. Ng

34-layer Convolutional Neural Network

SINUS SINUS SINUS SINUS AFIB AFIB AFIB AFIB

NCBI Resources How To
PubMed.gov US National Library of Medicine National Institutes of Health
PubMed "machine learning" & "diagnosis" Search
Create RSS Create alert Advanced



IBM
IBM Watson Health About Products Services Thought leadership Support

Today's decisions deserve tomorrow's insights.

At Watson Health we have impacted over 13,000 clients and partners, 80,000 professionals and 184,000 patients and consumers with our cognitive solutions.

Learn more Explore our products

verily Projects Leadership Partners Careers

Verily Projects

We are developing tools and platforms to enable more continuous health data collection for timely decision-making and effective interventions. We are running longitudinal studies to better understand ways to predict and prevent disease onset and progression. And, we are undertaking significant joint efforts with partners to radically transform the way healthcare is delivered.

FDA Nod for Edwards' Acumen HPI Software Widens Portfolio

March 20, 2018, 12:10:00 PM EDT By Zacks Equity Research, Zacks.com

Shutterstock photo

In a bid to strengthen its position in predictive analytics, **Edwards Lifesciences Corporation** EW , has recently announced the approval of De Novo request from the FDA for the Acumen Hypotension Prediction Index (HPI) software. Notably, the company plans a targeted launch of the product in the United States.

DeepMind Health

lumiata PRODUCTS NEWS BLOG ABOUT

We are making healthcare smarter.

AI Powered Health Analytics
We are Lumiata. We are delivering health analytics for managing risk and cost to make healthcare more efficient.

LEARN MORE

enlitic HOME NEWS JOBS

Enlitic uses deep learning to make doctors faster and more accurate

ML Methods Aren't New

Handwritten Digit Recognition with a Back-Propagation Network

Y. Le Cun, B. Boser, J. S. Denker, D. Henderson,
R. E. Howard, W. Hubbard, and L. D. Jackel
AT&T Bell Laboratories, Holmdel, N. J. 07733

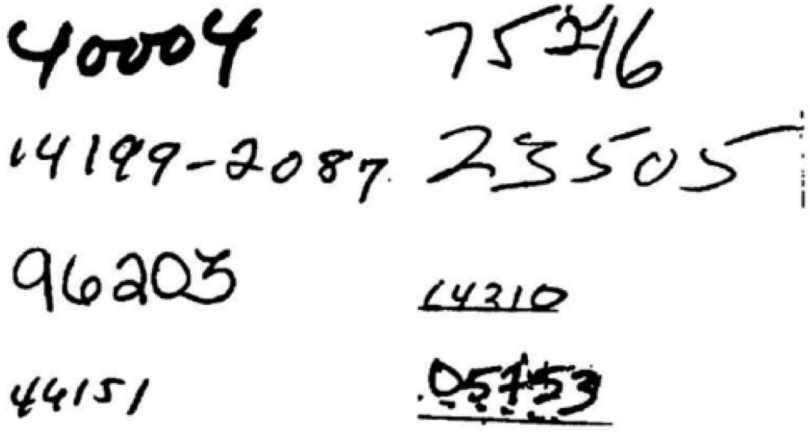


Figure 1: Examples of original zipcodes from the testing set.

LONG SHORT-TERM MEMORY

Neural Computation 9(8):1735–1780, 1997

Sepp Hochreiter
Fakultät für Informatik
Technische Universität München
80290 München, Germany
hochreit@informatik.tu-muenchen.de
http://www7.informatik.tu-muenchen.de/~hochreit

Jürgen Schmidhuber
IDSIA
Corso Elvezia 36
6900 Lugano, Switzerland
juergen@idsia.ch
http://www.idsia.ch/~juergen

Why now?

1. The Hardware Exists

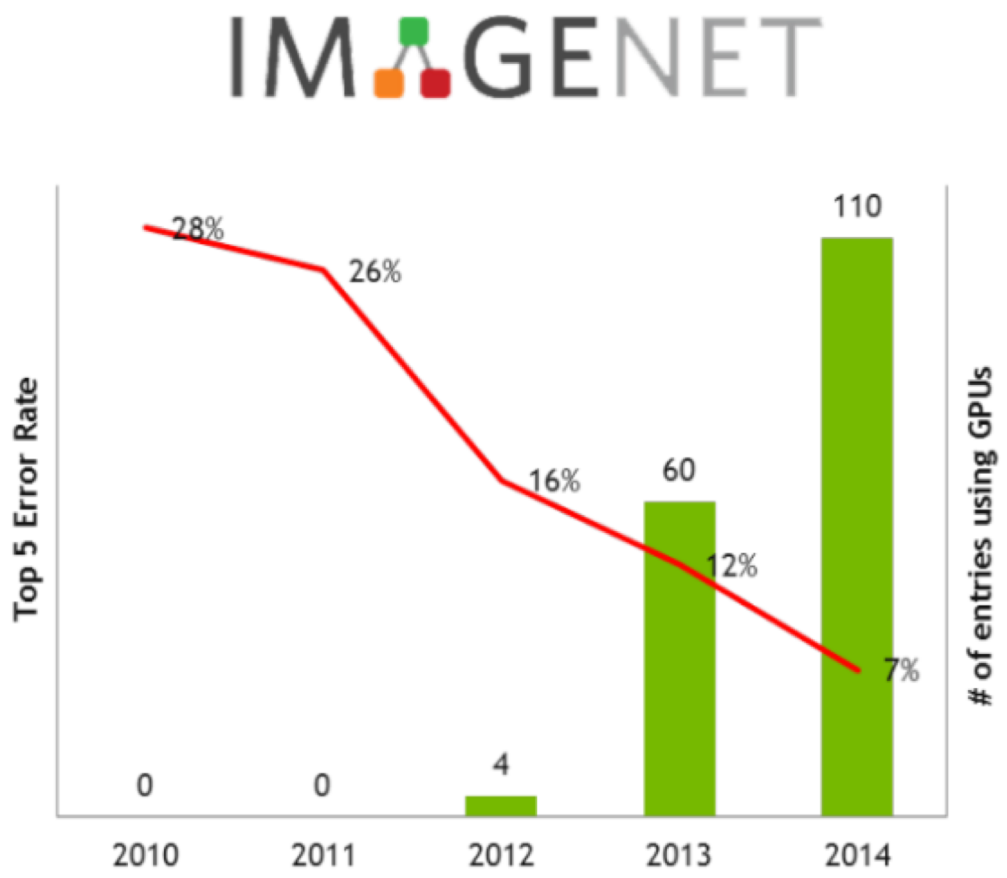
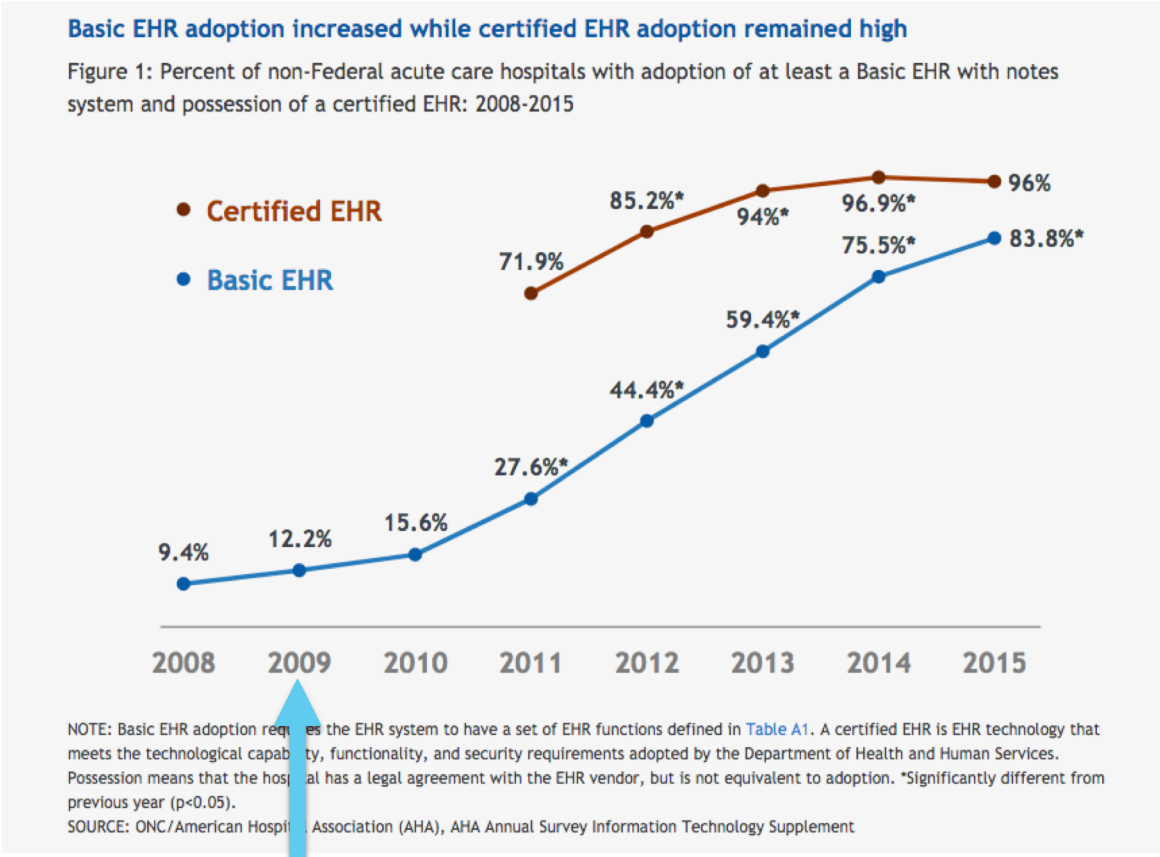


Figure 2. The rapidly falling ILSVRC winning entry classification error rate (red line) and the growing number of entries using GPUs each year (green bars).

https://devblogs.nvidia.com/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/

2. The Data Exists



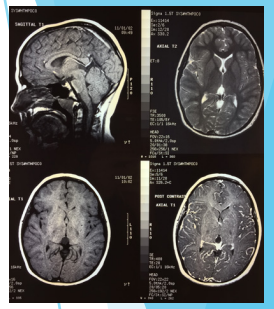
2009 Health Information Technology for Economic and Clinical Health Act

> 30 million patients visit the hospital in the US alone [1]

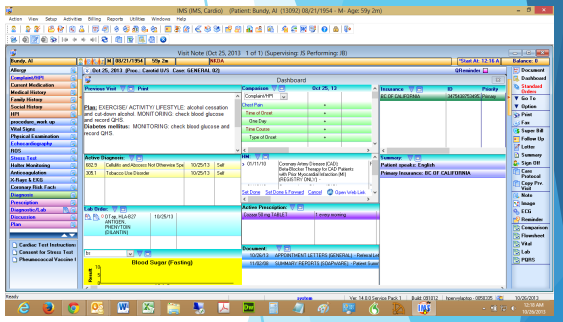
"Medical data is estimated to grow over 50-fold this decade, to 25,000 petabytes worldwide by 2020"[3]

What is Medical Data?

Images → MRI, CT Scans, X-rays

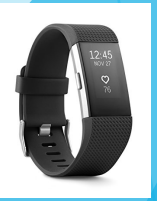


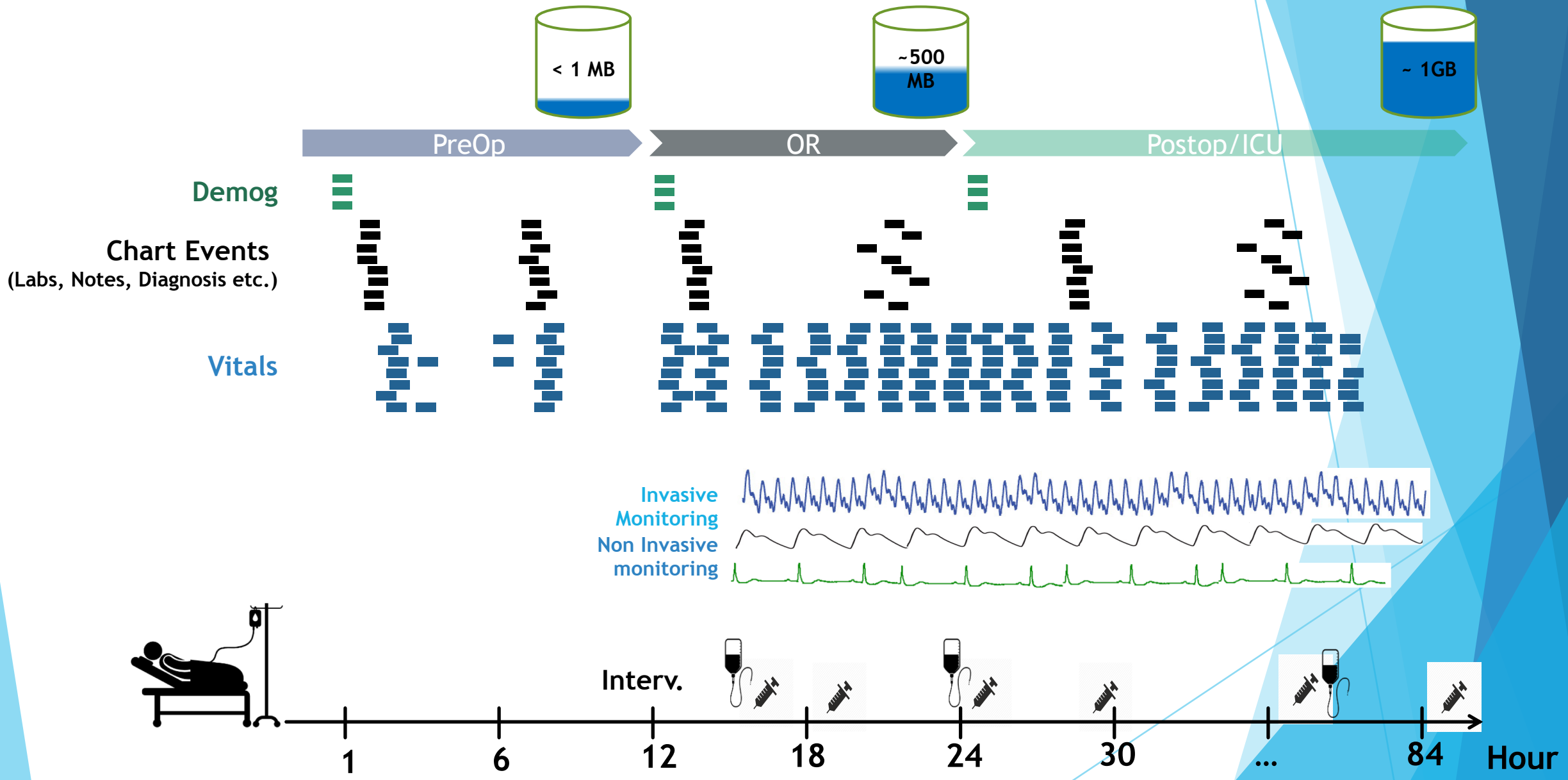
Sparse Manual Data → Clinical Notes, Labs, Test results, Diagnosis, Patient history, Interventions



Higher Resolution Data → Physiological Waveforms, Intermittent Vitals, Genomics

Personal Health Data → *Sociological/Behavioral, Fitness trackers, Health apps*




< 1 MB
~500 MB
~ 1GB
PreOp
OR
Postop/ICU
Demog
Chart Events
(Labs, Notes, Diagnosis etc.)
Vitals
Invasive Monitoring
Non Invasive monitoring
Interv.
1
6
12
18
24
30
...
84
Hour

ML Basics

Developing a machine learning model

1. Define the classification problem
   - Predict postoperative reintubation using intraoperative data
2. Label data [0, 1]
   - Reintubation event annotated in the EMR
3. Data preparation/preprocessing
   - Feature extraction and selection (if necessary)
   - Statistical, existing, or domain expert features
   - Feature selection like PCA or sequential forward
4. Train the model
5. Finalize the model and test performance

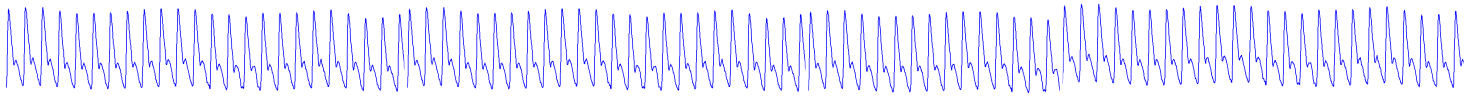
Developing a machine learning model

1. Define the classification problem
   - Predict postoperative reintubation using intraoperative data
2. Label data [0, 1]
   - Reintubation event annotated in the EMR
3. **Data preparation/preprocessing**
   - **Feature extraction and selection (if necessary)**
   - Statistical, existing, or domain expert features
   - Reduce dimensionality (PCA, Fourier, etc.)
4. Train the model
5. Finalize the model and test performance

Featurizing The Data

Simple 1 Patient Example

Raw ABP Signal



Sampled at 100 Hz over 30 minutes
= 180,000 datapoints

HR, SYS, DIA, MAP, TV, RR, ETCO2, SpO2

Sampled every 1 minute over 30 minutes
= 30 datapoints * 8

Hgb

Manual
2 datapoints

Vasopressor Administration

Manual
5 datapoints

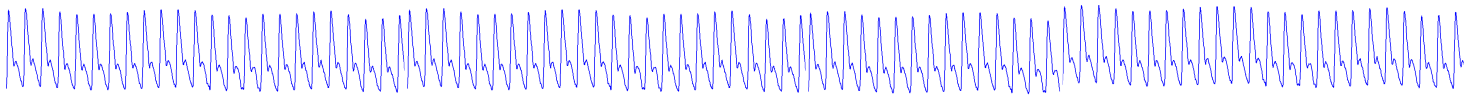
Age, Height, Weight, Gender

1 Time
4 datapoints

> 180,000 datapoints TOTAL per patient!

Preprocessing and Data Limitations

Raw ABP Signal



Needs to be filtered for noise
Need to reduce dimensionality → overfitting

HR, SYS, DIA, MAP, TV, RR, ETCO2, SpO2

Remove outliers

Hgb

Vasopressor Administration

Non uniformly sampled

Sparse

Timing and/or dosage may be unreliable

Age, Height, Weight, Gender

Incorrect or empty entries

Preprocessing and Data Limitations

Raw ABP Signal

HR, SYS, DIA, MAP, TV, RR, ETCO2, SpO2

Hgb

Vasopressor Administration

Age, Height, Weight, Gender

How to time align non-uniformly sampled data with uniformly sampled?

How to scale everything to N patients?

Needs to be filtered for noise
Need to reduce dimensionality → overfitting

Remove outliers

Non uniformly sampled

Sparse

Timing and/or dosage may be unreliable

Incorrect or empty entries

Alignment and Scaling Data to N Patients

Take 1 representative value per feature per patient

	Minimum, Max, Mean, Std Values												
	HR	**SYS**	**DIA**	**MAP**	**TV**	**RR**	**SpO2**	**Hgb**	**# Vasopressors Boluses**	**Age**	**Height**	**Weight**	**Gender**
PT 1	x	x	x	x	x	x	x	x	x	x	x	x	x
PT 2	x	x	x	x	x	x	x	x	x	x	x	x	x
PT 3	x	x	x	x	x	x	x	x	x	x	x	x	x
…	x	x	x	x	x	x	x	x	x	x	x	x	x
PT N	x	x	x	x	x	x	x	x	x	x	x	x	x

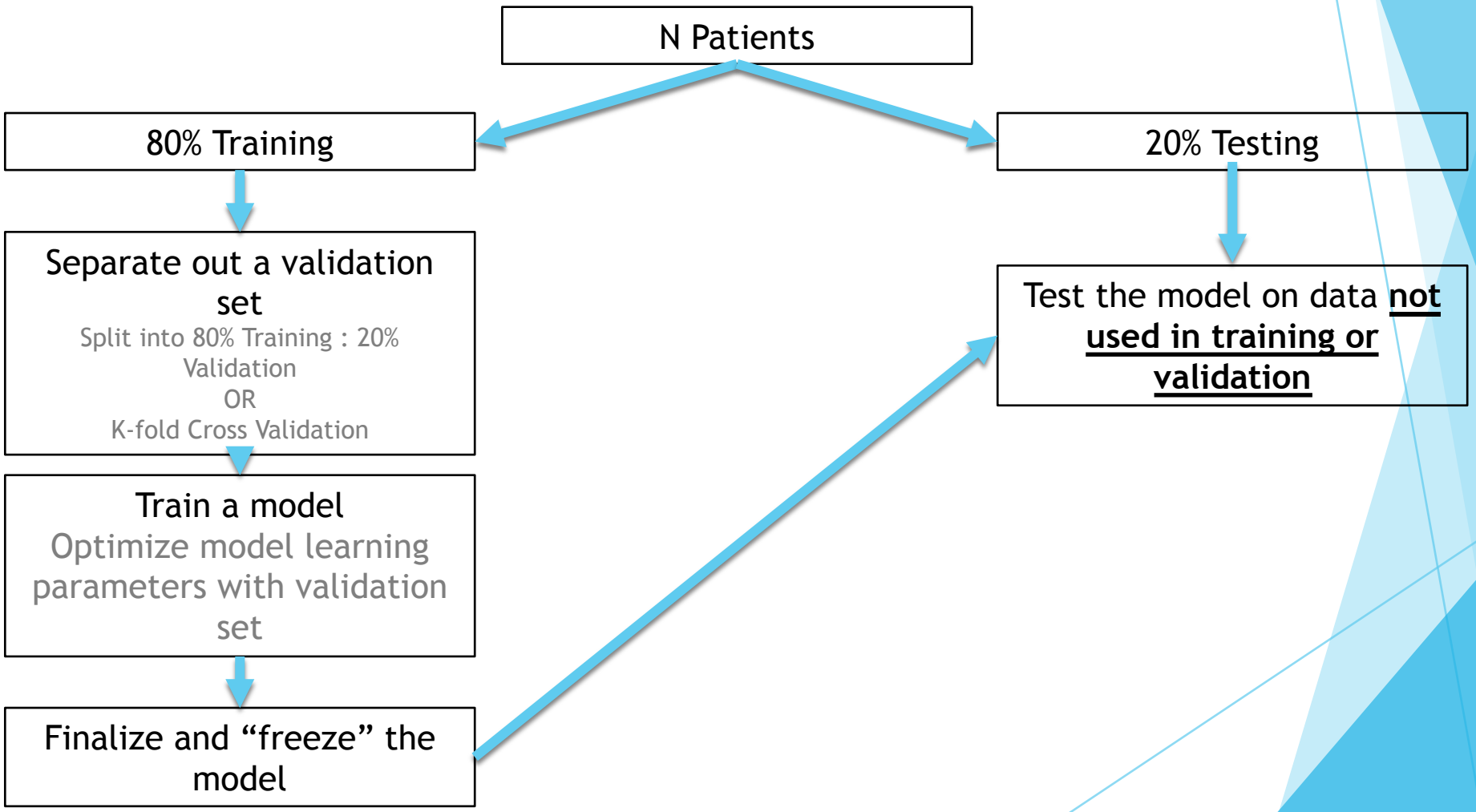
Size N patients x 34 features

Ideal for models like logistic regression

Developing a machine learning model

1. Define the classification problem
   - Predict postoperative reintubation using intraoperative data
2. Label data [0, 1]
   - Reintubation event annotated in the EMR
3. Data preparation/preprocessing
   - Feature extraction and selection (if necessary)
   - Statistical, existing, or domain expert features
   - Feature selection like PCA or sequential forward
4. Train the model
5. Finalize the model and test performance

Training the model



What does it mean to train a model?

The goal is to tune the model and learn $\boldsymbol{\beta}$ such that we minimize the **error** between **predicted y (0,1)** and the **true label [0,1]**

Logistic Regression

$$Predicted\ y = f(x) = \frac{1}{1 + e^{-(\boldsymbol{\beta}_0 + \sum_1^i \boldsymbol{\beta}_i \boldsymbol{x})}}$$

Probability between 0 and 1

Bias

Weight Coefficients

Values of each feature in the model

Learnable Parameters

# of features + 1

What does it mean to train a model?

The goal is to tune the model and learn $\boldsymbol{\beta}$ such that we minimize the **error** between **predicted y (0,1)** and the **true label [0,1]**

Logistic Regression

$$Predicted\ y = f(x) = \frac{1}{1 + e^{-(\boldsymbol{\beta}_0 + \sum_1^i \boldsymbol{\beta}_i \boldsymbol{x})}}$$

Bias

Weight Coefficients

Values of each feature in the model

Learnable Parameters
# of features + 1

Probability between 0 and 1

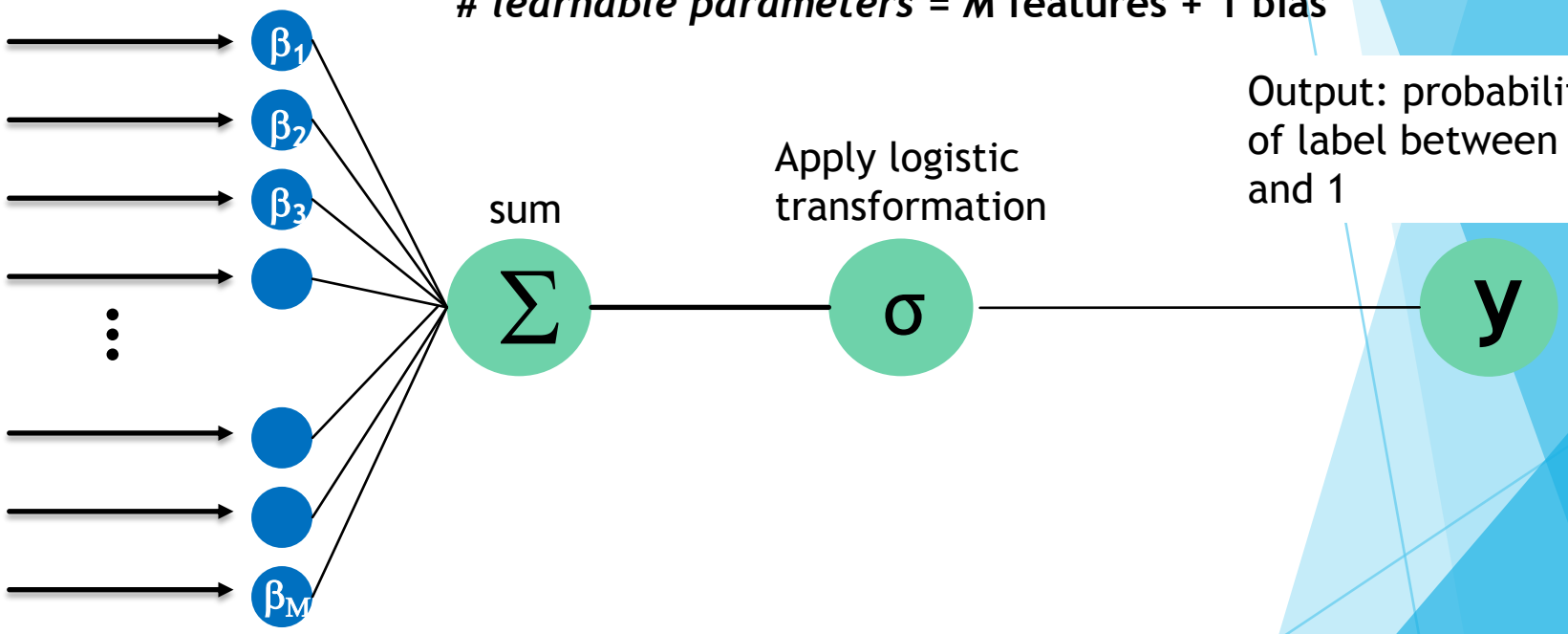
Error = Predicted y - True y

Iteratively update the values of $\boldsymbol{\beta}$ in the direction of minimizing the error

Logistic Regression

Input: *N* patients by *M* features

PT 1	PT 2	PT 3	…	PT N	
x	x	x	x	x	HR
x	x	x	x	x	SYS
x	x	x	x	x	DIA
x	x	x	x	x	MAP
x	x	x	x	x	TV
x	x	x	x	x	RR
x	x	x	x	x	SpO2
x	x	x	x	x	Hgb
x	x	x	x	x	# Vasopressors Boluses
x	x	x	x	x	Age
x	x	x	x	x	Height
x	x	x	x	x	Weight
x	x	x	x	x	Gender



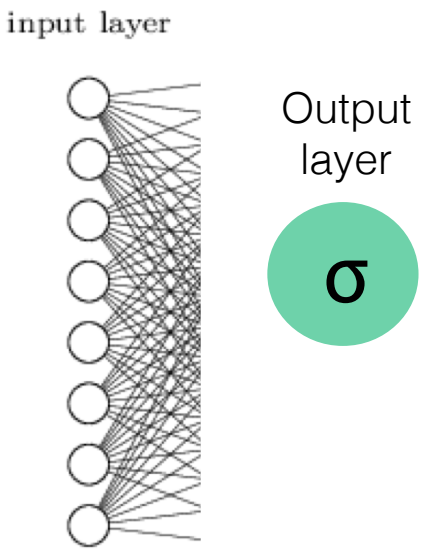
***# learnable parameters* = *M* features + 1 bias**

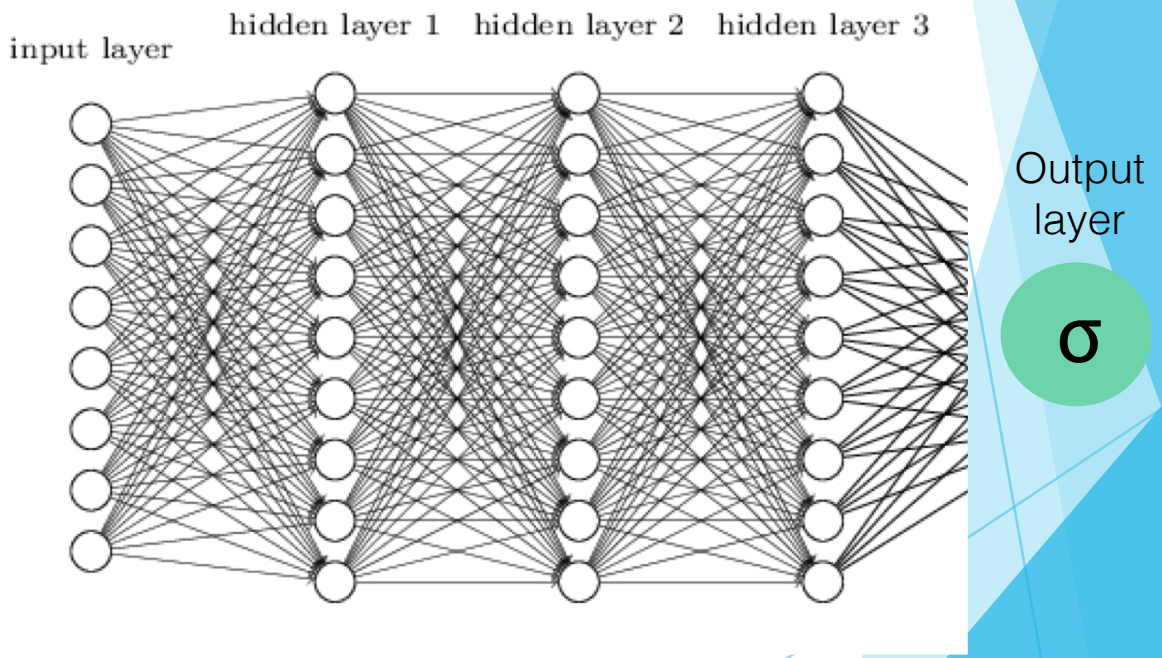
Output: probability of label between 0 and 1

Deep Learning

Self learns a complex representation of input data

Logistic regression is a neural network with NO hidden layers

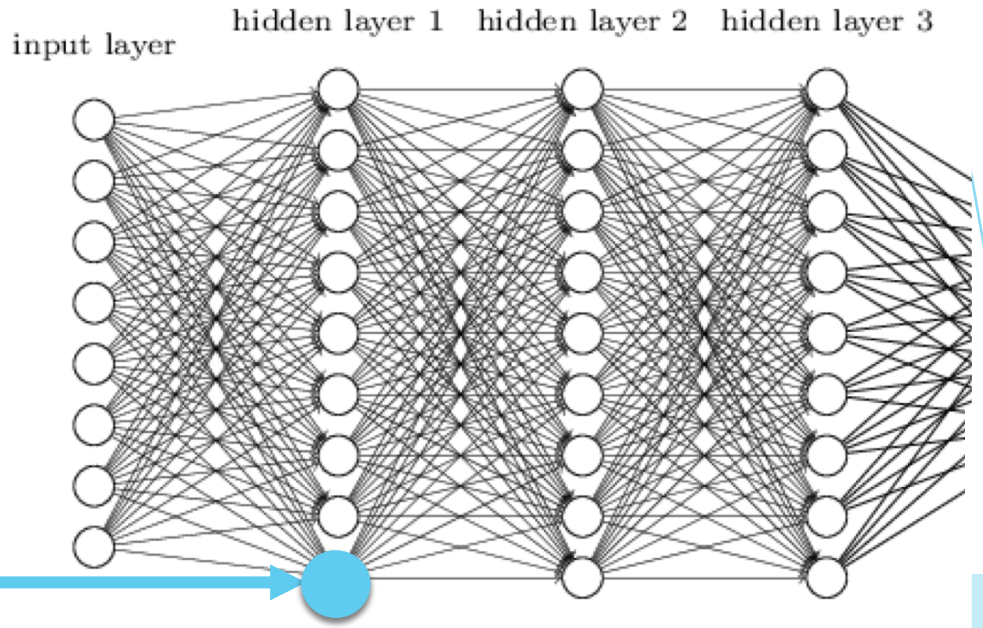




Deep Learning-Fully Connected Feed Forward Neural Network

Self learns a complex representation of input data

- Model Hyperparameters
  - Number of neurons
  - Number of hidden layers
  - Batch size and epochs
  - Activation function
  - Learning rate
  - Regularization parameters



Probability of label between 0 and 1

σ → y

Logistic Output Layer

Neuron:
$f(\sum_0^i \beta\, x)$,

f : **activation** function ReLu or tanh

β : learnable parameters

x : output of the previous layer
Output layer = logistic function

Recall…Logistic Regression
f : **logistic function**

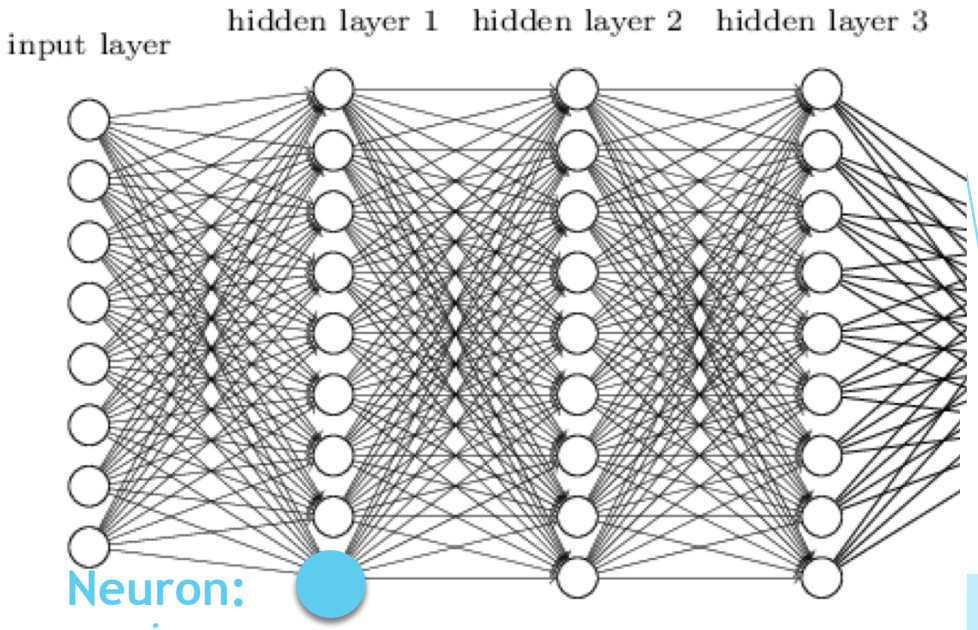
β : learnable parameters

x : input features (input layer only)

Deep Learning-Fully Connected Feed Forward Neural Network

Self learns a complex representation of input data

- Model Hyperparameters
  - Number of neurons
  - Number of hidden layers
  - Batch size and epochs
  - Activation function
  - Learning rate
  - Regularization parameters



Neuron:
$f(\sum_0^i \beta\, x)$,

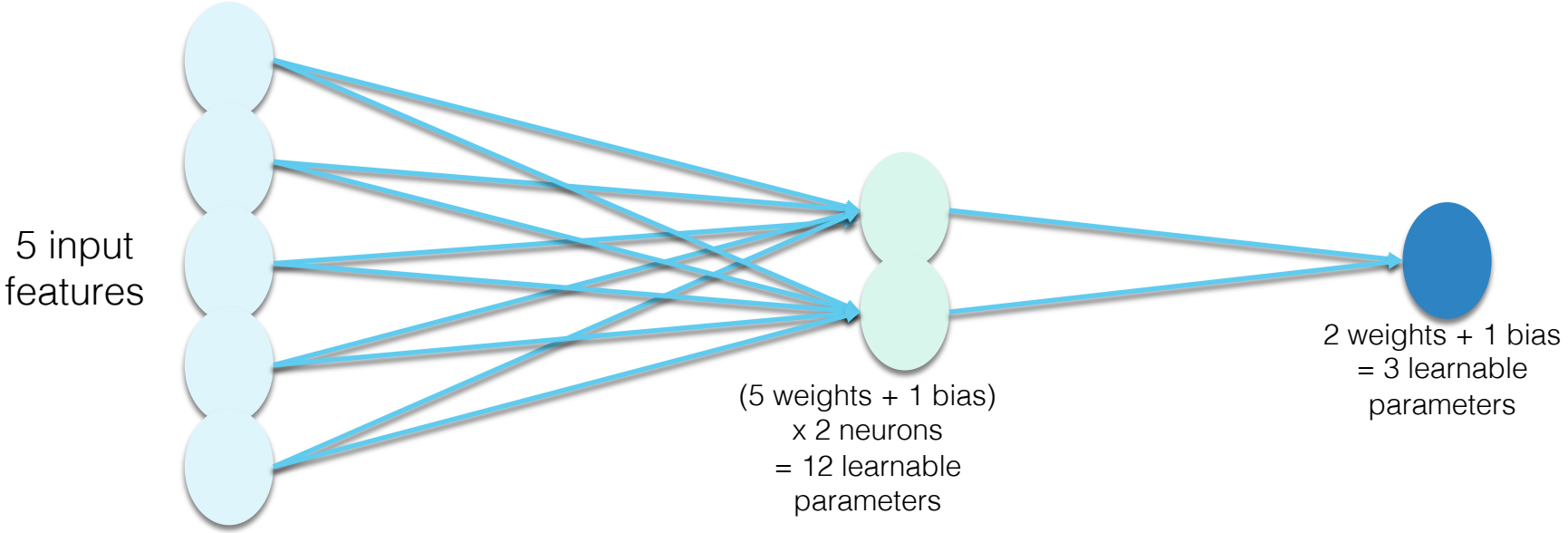
Probability of label between 0 and 1

σ → y

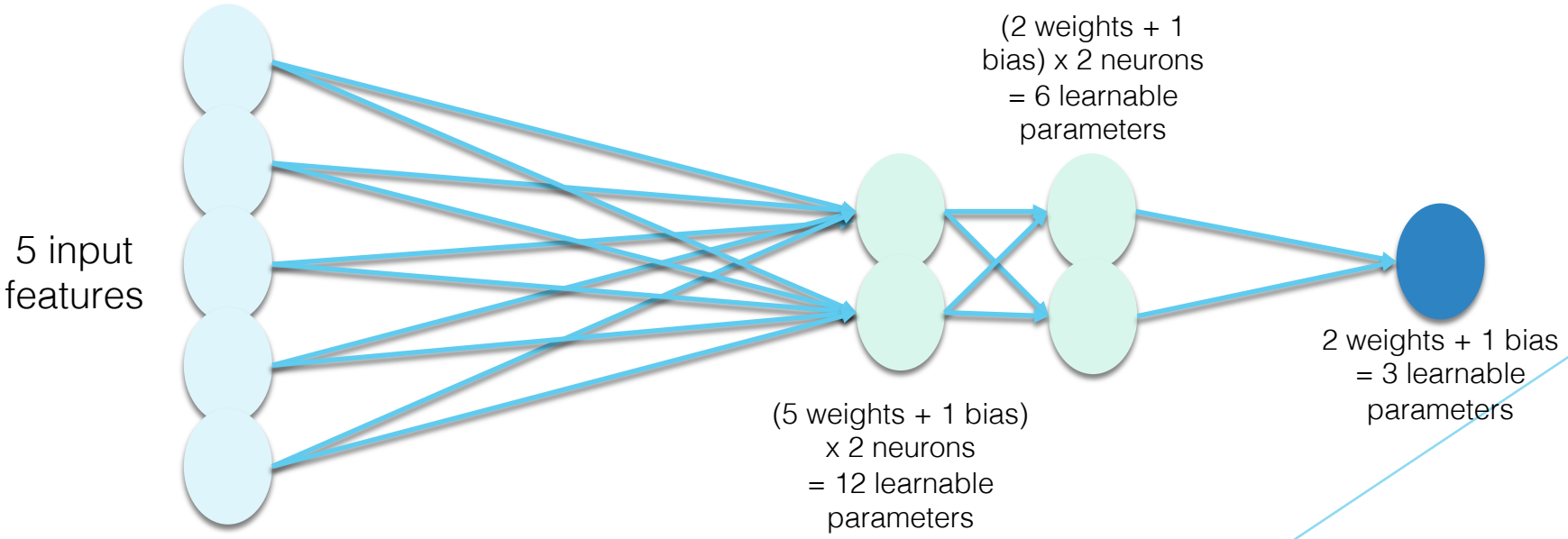
Logistic Output Layer

Iteratively update all the values of β in **each neuron** in the direction of **minimizing the error**

↑ the complexity = ↑ number of learnable parameters (β).



15 learnable parameters



21 learnable parameters

Adding Complexity

High Resolution Data over Time over N Patients

How to time align non-uniformly sampled data with uniformly sampled?

PT 1

Time	HR	SYS	DIA	MAP	TV	RR	SpO2	Hgb	Vasopressor Administered [0 or 1]	Age	Height	Weight	Gender
0	x	x	x	x	x	x	x	9	0	x	x	x	x
1	x	x	x	x	x	x	x	9	1	x	x	x	x
2	x	x	x	x	x	x	x	9	0	x	x	x	x
…	x	x	x	x	x	x	x	10	0	x	x	x	x
30 minutes	x	x	x	x	x	x	x	10	0	x	x	x	x

Backfilling/ Forwardfilling

Binary

Repeated values

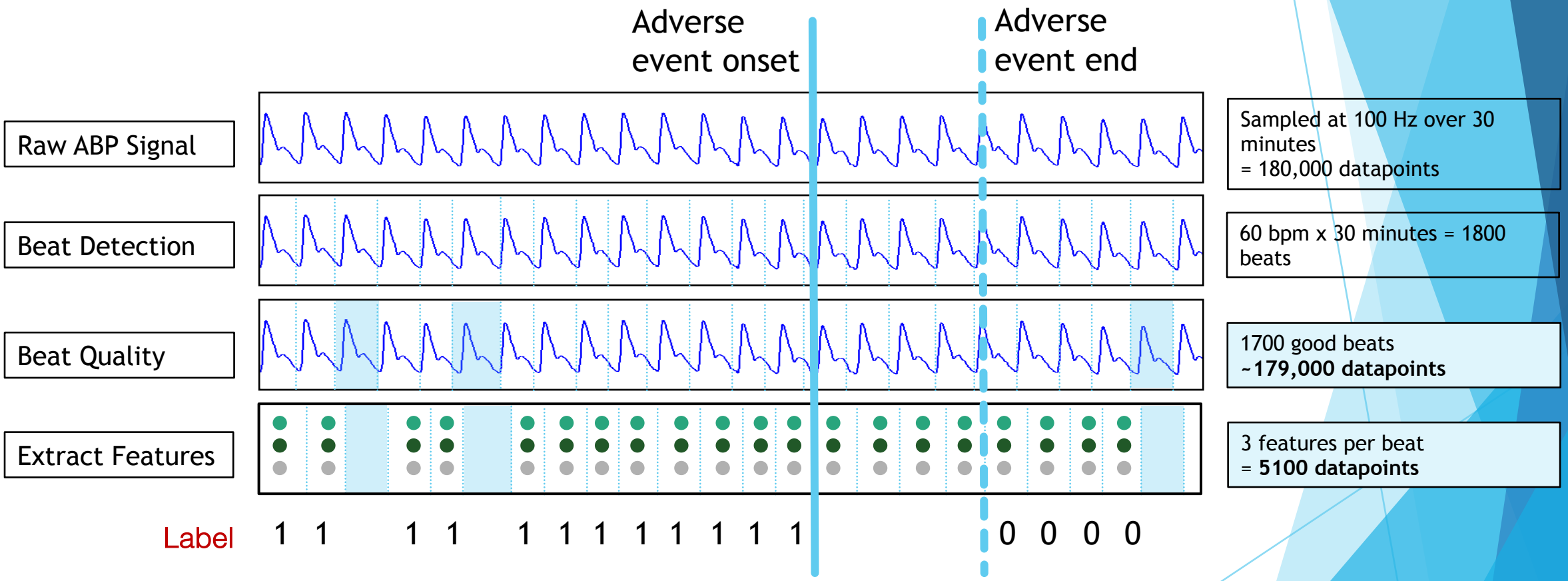
PT N

Time	HR	SYS	DIA	MAP	TV	RR	SpO2	Hgb	Vasopressor Administered [0 or 1]	Age	Height	Weight	Gender
0	x	x	x	x	x	x	x	8	0	x	x	x	x
1	x	x	x	x	x	x	x	8	1	x	x	x	x
2	x	x	x	x	x	x	x	8	0	x	x	x	x
…	x	x	x	x	x	x	x	8	0	x	x	x	x
30 minutes	x	x	x	x	x	x	x	10	0	x	x	x	x

N patients x 30 timepoints x 13 features

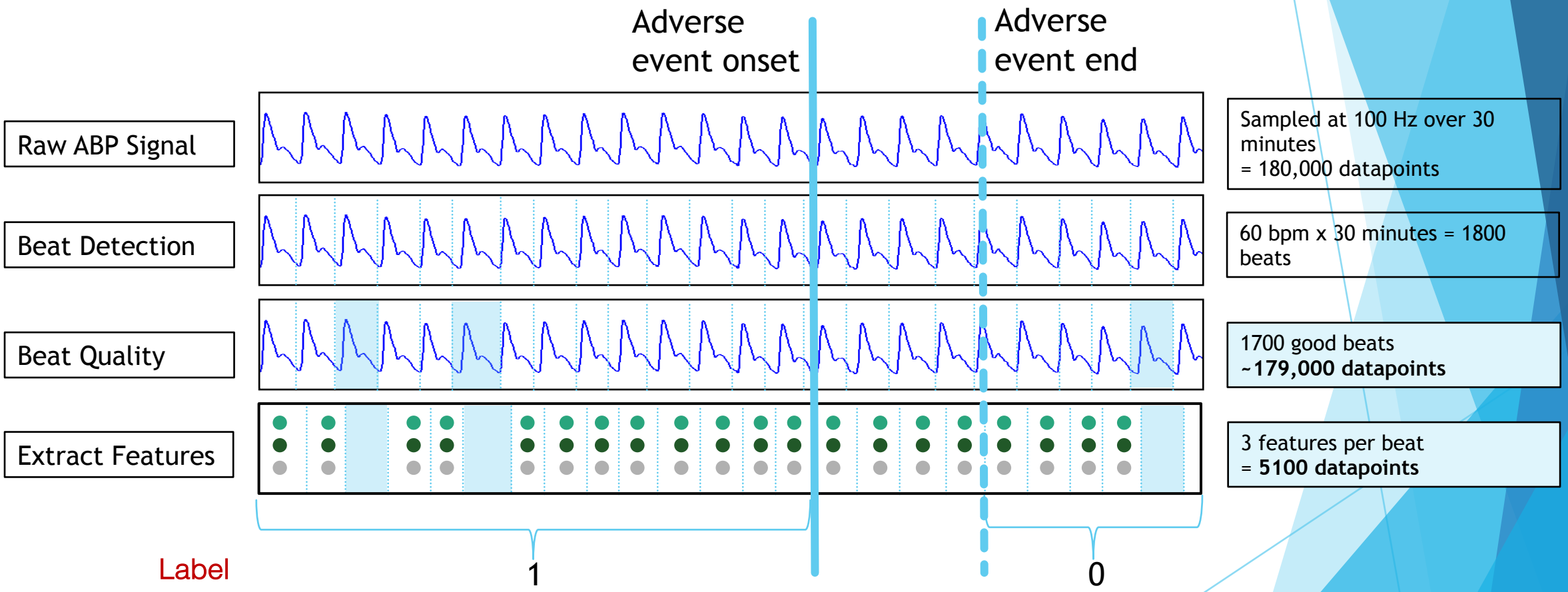
Additional dimension of time

High Resolution Waveforms



1. Take the features at each timepoint and treat as a static input (i.e. N patients by M features) → Lose time series relationship
2. Take the features as a time series as an input (i.e. N patients by T time by M channels)

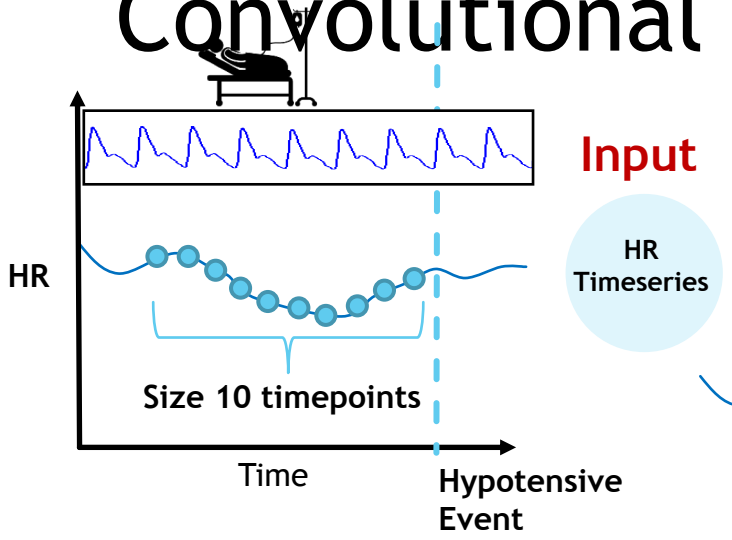
High Resolution Waveforms



1. Take the features at each timepoint and treat as a static input (i.e. N patients by M features)
2. Take the features as a time series as an input (i.e. N patients by T time by M channels)

Convolutional Neural Networks

Deep Neural Networks with a Convolutional Layer



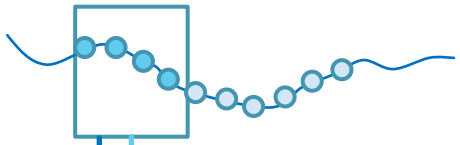
Input

HR Timeseries

1 Convolutional Layer

2 filters (kernels) of length 4 timepoints

And stride of 2 timepoints



Filter 1

Weights

(Size 4 x 1)

β_1	β_2	β_3	β_4

Filter 2

Weights

(Size 4 x 1)

β_5	β_6	β_7	β_8

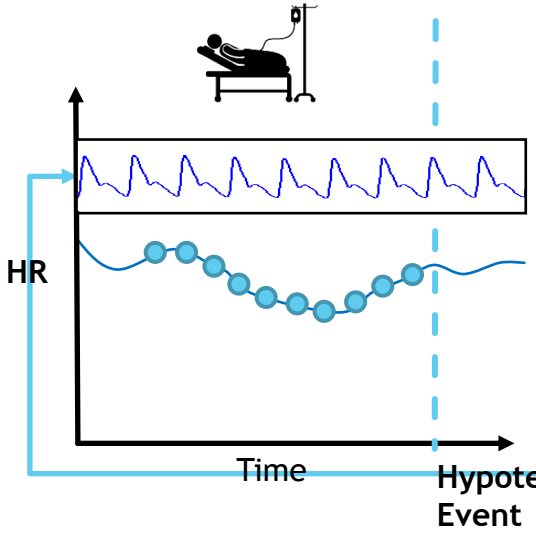
Convolutional Layer Output

Activation Map

Size 4 x 2

$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$
$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$f_{2,4}$

Deep Neural Networks with a Convolutional Layer



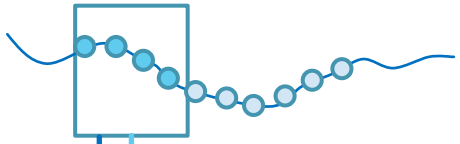
Input

HR Timeseries

1 Convolutional Layer

2 filters (kernels) of length 4 timepoints

And stride of 2 timepoints



"Convolving" the time series by filters…multiplying by $\boldsymbol{\beta}$

Filter 1

Weights

(Size 4 x 1)

β_1	β_2	β_3	β_4

$$f(\beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + \beta_4 * t_4)$$

Filter 2

Weights

(Size 4 x 1)

β_5	β_6	β_7	β_8

$$f(\beta_5 * t_1 + \beta_6 * t_2 + \beta_7 * t_3 + \beta_8 * t_4)$$

Apply Activation Function

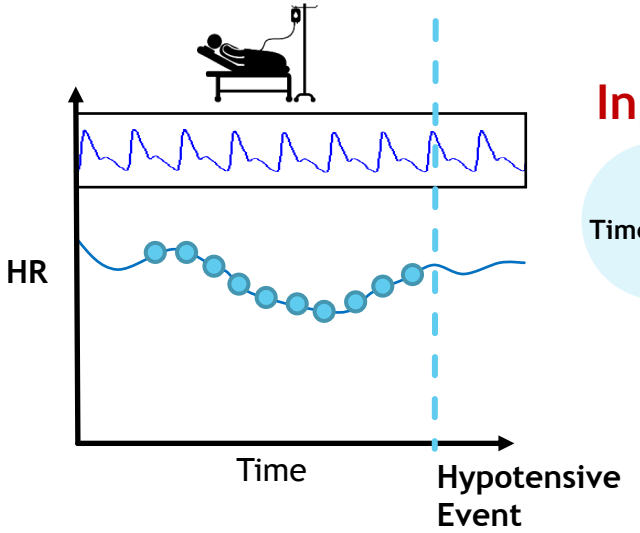
Convolutional Layer Output

Activation Map

Size 4 x 2

$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$
$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$f_{2,4}$

Deep Neural Networks with a Convolutional Layer



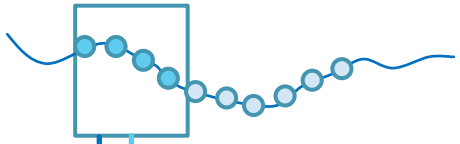
Input

HR Timeseries

1 Convolutional Layer

2 filters (kernels) of length 4 timepoints

And stride of 2 timepoints



Filter 1

Weights

(Size 4 x 1)

β_1 β_2 β_3 β_4

Apply Activation Function

$$f(\beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + \beta_4 * t_4)$$

Filter 2

Weights

(Size 4 x 1)

β_5 β_6 β_7 β_8

$$f(\beta_5 * t_1 + \beta_6 * t_2 + \beta_7 * t_3 + \beta_8 * t_4)$$

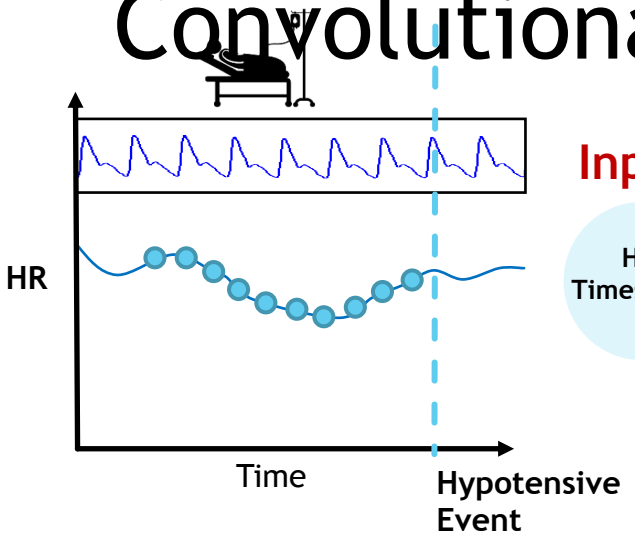
Convolutional Layer Output

Activation Map

Size 4 x 2

$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$
$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$f_{2,4}$

Deep Neural Networks with a Convolutional Layer



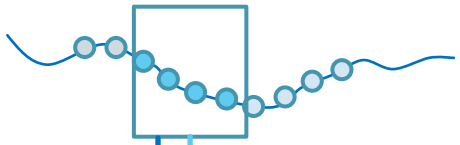
Input

HR Timeseries

1 Convolutional Layer

2 filters (kernels) of length 4 timepoints

And stride of 2 timepoints



Keep moving forward in time…

Filter 1

Weights

(Size 4 x 1)

β_1 β_2 β_3 β_4

Filter 2

Weights

(Size 4 x 1)

β_5 β_6 β_7 β_8

Apply Activation Function

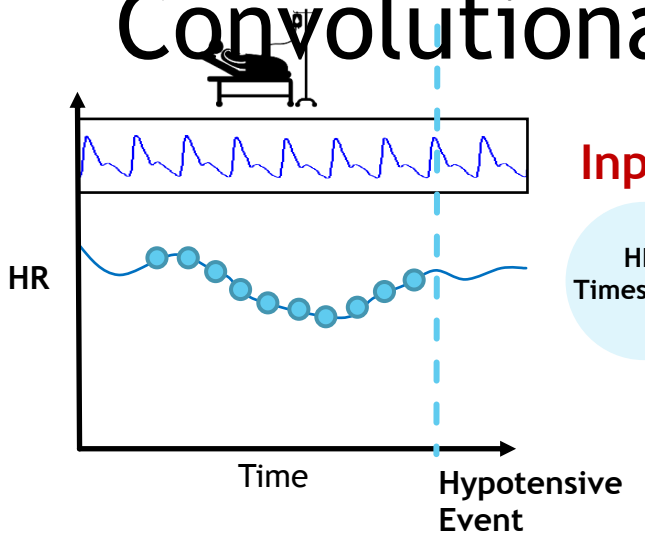
Convolutional Layer Output

Activation Map

Size 4 x 2

$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$
$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$f_{2,4}$

Deep Neural Networks with a Convolutional Layer



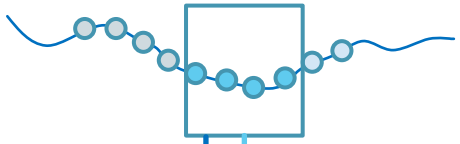
Input

HR Timeseries

1 Convolutional Layer

2 filters (kernels) of length 4 timepoints

And stride of 2 timepoints



Filter 1

Weights

(Size 4 x 1)

β_1 β_2 β_3 β_4

Filter 2

Weights

(Size 4 x 1)

β_5 β_6 β_7 β_8

Apply Activation Function

Convolutional Layer Output

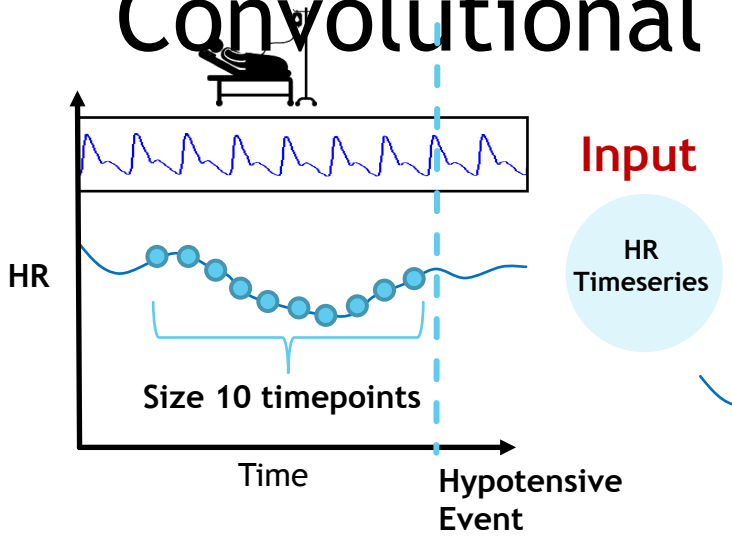
Activation Map

Size 4 x 2

$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$
$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$f_{2,4}$

Keep moving forward in time…

Deep Neural Networks with a Convolutional Layer



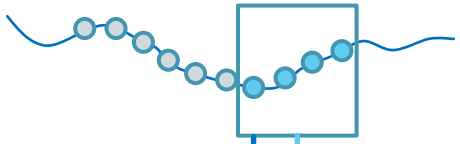
Input

HR Timeseries

1 Convolutional Layer

2 filters (kernels) of length 4 timepoints

And stride of 2 timepoints



Filter 1

Weights

(Size 4 x 1)

β_1 β_2 β_3 β_4

Filter 2

Weights

(Size 4 x 1)

β_5 β_6 β_7 β_8

Apply Activation Function

Convolutional Layer Output

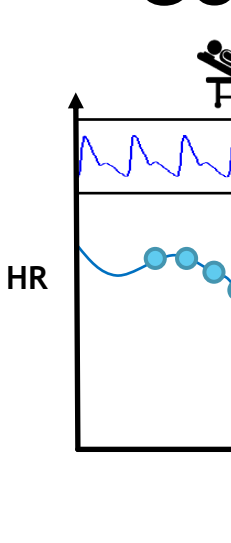
Activation Map

Size 4 x 2

$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$
$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$f_{2,4}$

Keep moving forward in time…

Deep Neural Networks with a Convolutional Layer



Convolutional Layer
Creates Multi-dimensional and "activation maps"
Size 4x2

$f_{1,1}$ $f_{1,2}$ $f_{1,3}$ $f_{1,4}$
$f_{2,1}$ $f_{2,2}$ $f_{2,3}$ $f_{2,4}$

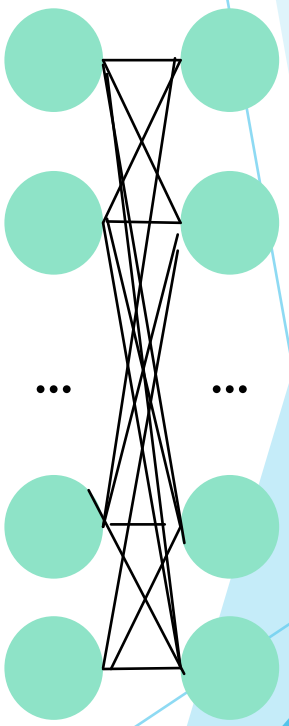
Max Pooling Layer
Halves the output
Size 2x2

$f_{1,2}$ $f_{1,3}$
$f_{2,2}$ $f_{2,3}$

Flatten
Into 1 Dimension
Size 4

$f_{1,2}$
$f_{1,3}$
$f_{2,2}$
$f_{2,3}$

Deep Neural Network



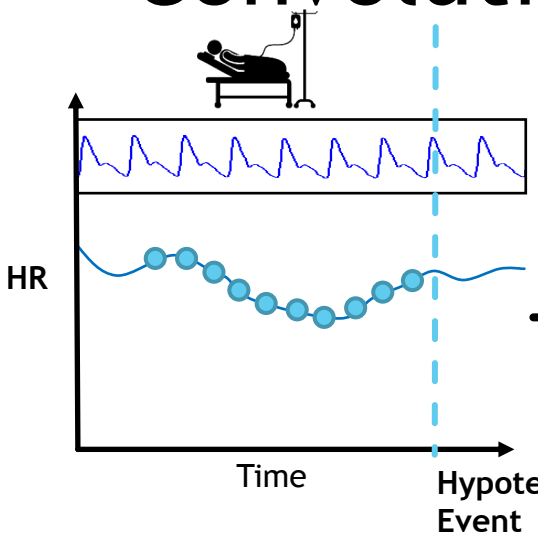
Logistic function

σ

Probability of Event

$\hat{y}$

Deep Neural Networks with a Convolutional Layer



Convolutional Layer
Creates Multi-dimensional and "activation maps"
Size 4x2

$f_{1,1}$ $f_{1,2}$ $f_{1,3}$ $f_{1,4}$
$f_{2,1}$ $f_{2,2}$ $f_{2,3}$ $f_{2,4}$

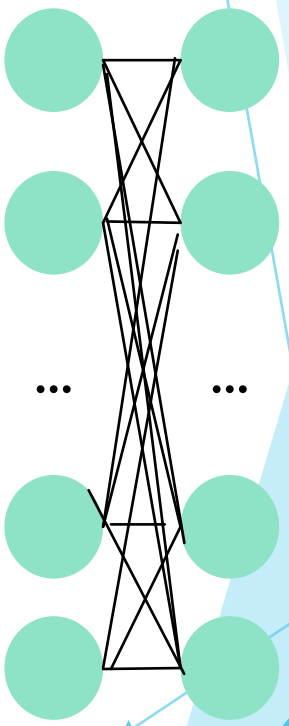
Max Pooling Layer
Halves the output
Size 2x2

$f_{1,2}$ $f_{1,3}$
$f_{2,2}$ $f_{2,3}$

Flatten
Into 1 Dimension
Size 4

$f_{1,2}$
$f_{1,3}$
$f_{2,2}$
$f_{2,3}$

Deep Neural Network



Logistic function

σ

Probability of Event

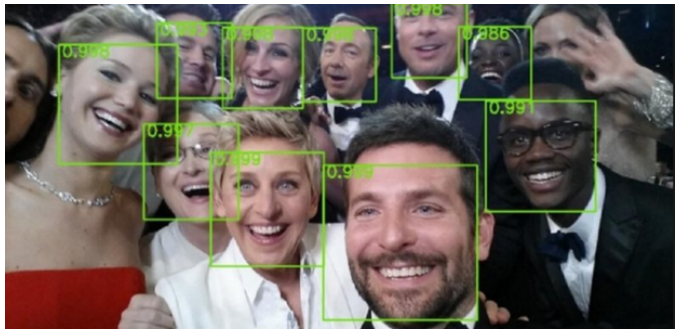
$\hat{y}$

Error =
Predicted Probability of Event - True Label

Update ALL the βs in the direction of minimizing the error

Convolutional Neural Networks in Use Today

Facial Recognition

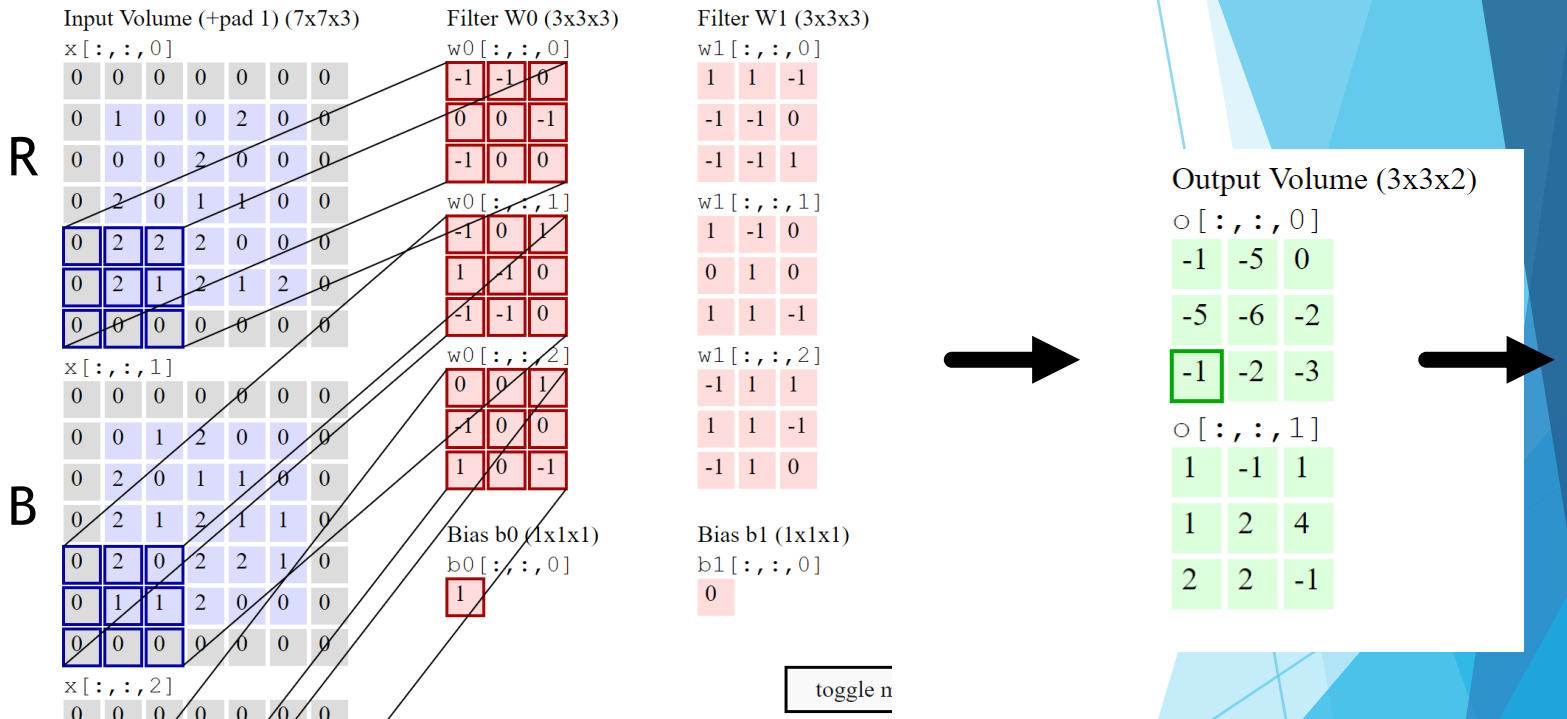


1 Input Image

Size 7x7 with 3 color channels

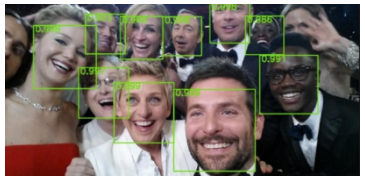
1 Convolutional Layer

2 *filters* of length 3x3 and a *stride* of 2



1 Input Image

Size 7x7x3 color channels



Convolutional Layer

Output Volume (3x3x2)

o[:,:,0]

-1	-5	0
-5	-6	-2
-1	-2	-3

o[:,:,1]

1	-1	1
1	2	4
2	2	-1

Max Pooling layer

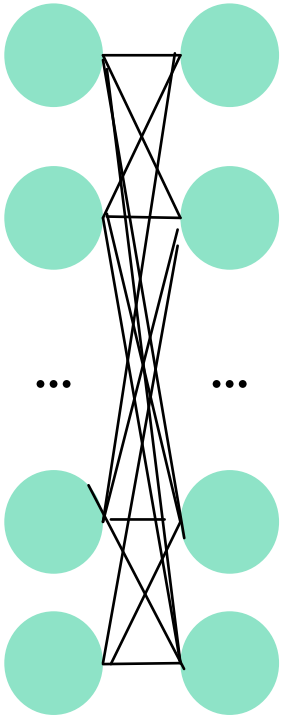
-1	0
-1	-2

2	4
2	4

Flatten

-1
0
-1
-2
2
4
2
4

Deep Neural Network



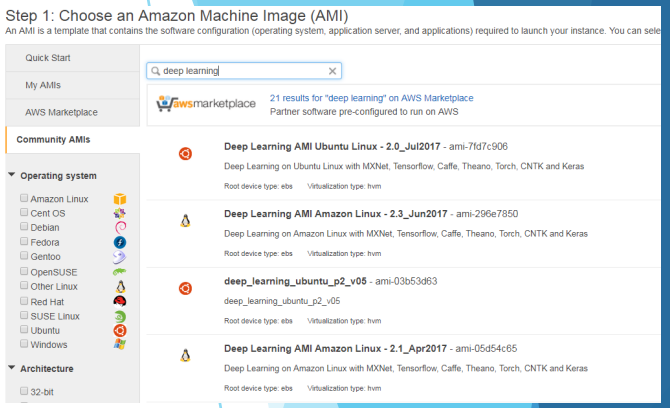
Logistic Output layer

σ

Label Probability of Image

ML Starter Kit

- Pretty much whatever OS you want + ready to use Images with packages pre-downloaded
- **Keras** – Python API for deep learning (runs on top of TensorFlow from Google)
- **Scikit Learn** – statistical models and model analysis
- **Pandas** – pretty data structures
- **Anaconda and Jupyter notebooks** – program to keep all packages easily updated and web app that makes everything pretty



ML Starter Kit

Other Tools:

- Matlab
  - Statistics and Machine Learning Toolbox
  - Neural Network Toolbox
  - Parallel Computing Toolbox
- R

Tutorials:

- Machinelearningmastery.com
- neuralnetworksanddeeplearning.com
- CS231n: Convolutional Neural Networks for Visual Recognition
- Coursera Machine Learning by Andrew Ng

Conclusion

- Medical data can be sparse, high resolution, variable and complex
- Medical data can get BIG…the data is there for the application of ML models
- ML methods are not new
- Clinical application of “traditional methods” like logistic regression tends to focus on “low resolution” data
- Adding the extra dimension of time and also more high resolution data makes the data more complex
- Deep learning can capture that complexity
- As we move forward, the goal is to create models that merge more high resolution data with the low resolution to fully capture relationships in all patient data

Thank you!

linkedin.com/in/christineklee1

chriskl2@uci.edu

Christine_lee@Edwards.com

YOUNG RESEARCHERS WORKSHOP

3:30 = 5:30 PM, IRONWOOD B