Learning time series representations through contextualized LSTMs

Madalina Fiterau, Stanford University
mfiterau@cs.stanford.edu

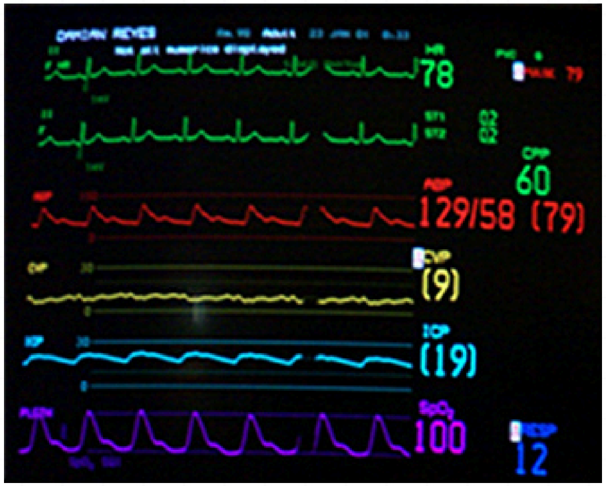
Advisors: Christopher Ré, Scott Delp

WiML Workshop
Dec 5th 2016
Barcelona, Spain

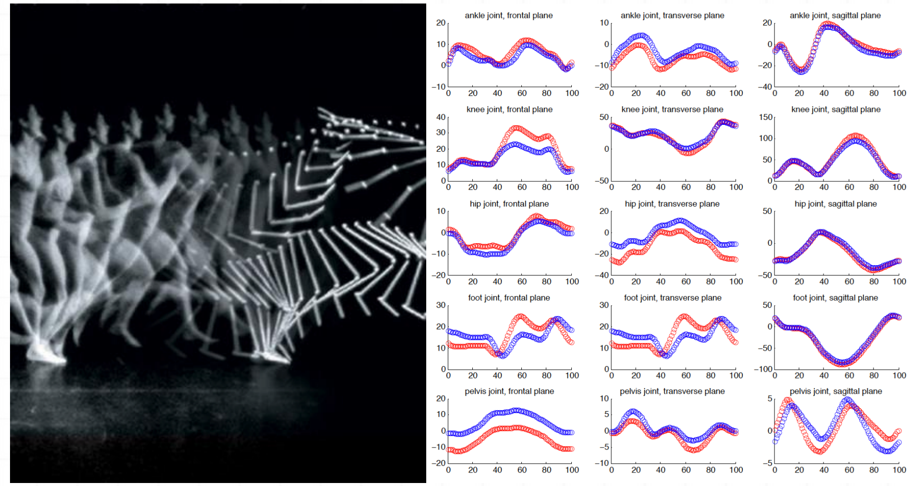
We acknowledge support from the NIH (U54 EB020405).

Prevalence of time series data

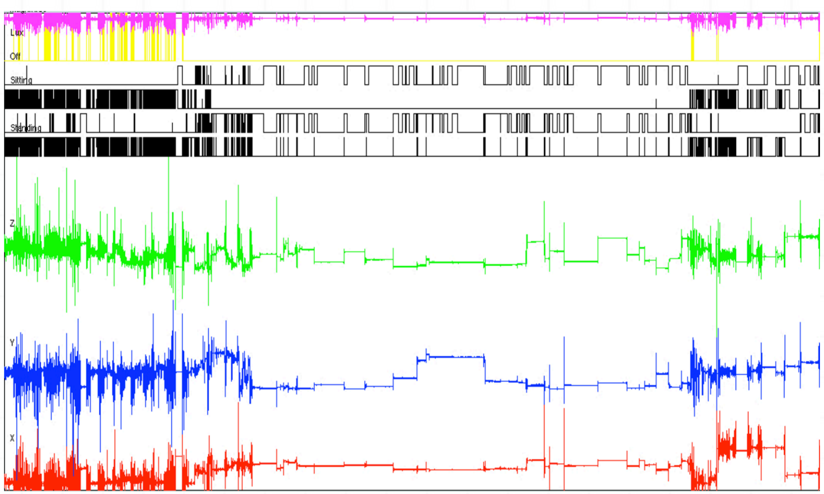
Vital Signs



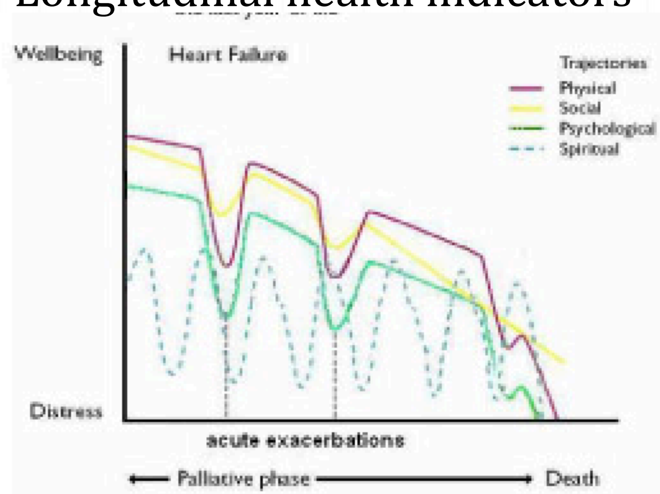
Gait Kinematics



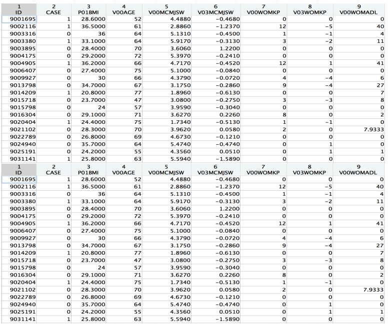
Accelerometer data



Longitudinal health indicators



Prevalence of time series data

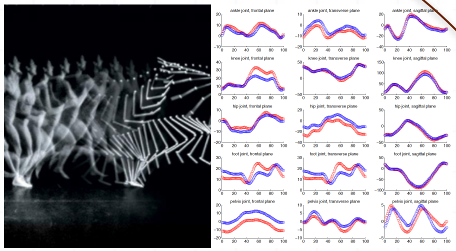


Clinical tests, medication (structured)

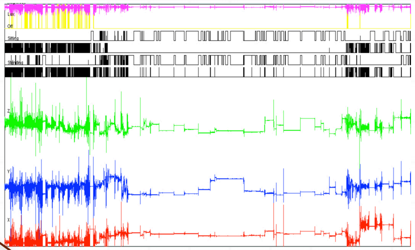
Stereo recordings (video)

Vital Signs

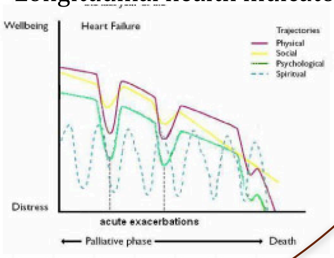
Gait Kinematics



Accelerometer data



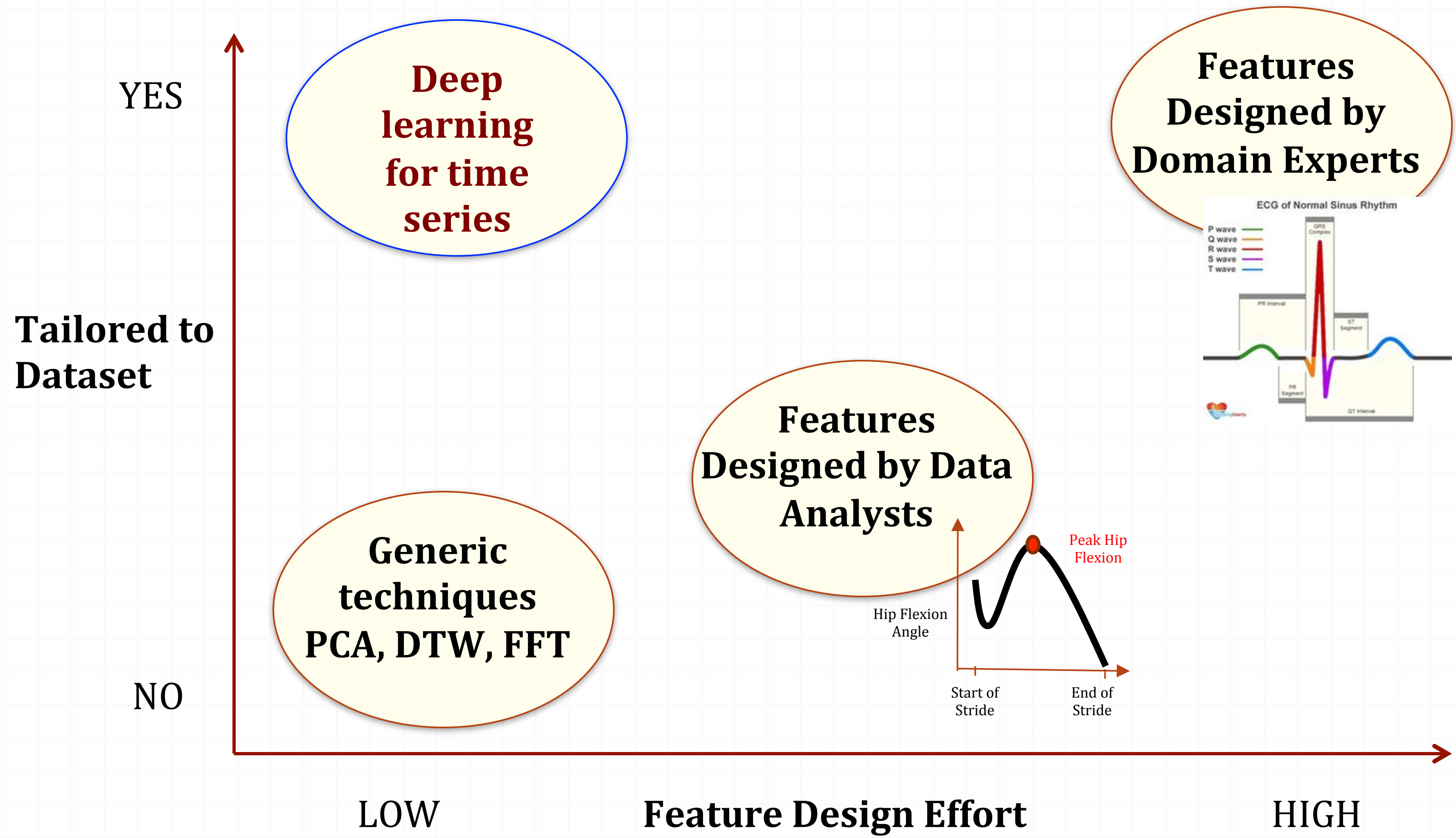
Longitudinal health indicators



X-rays (images)

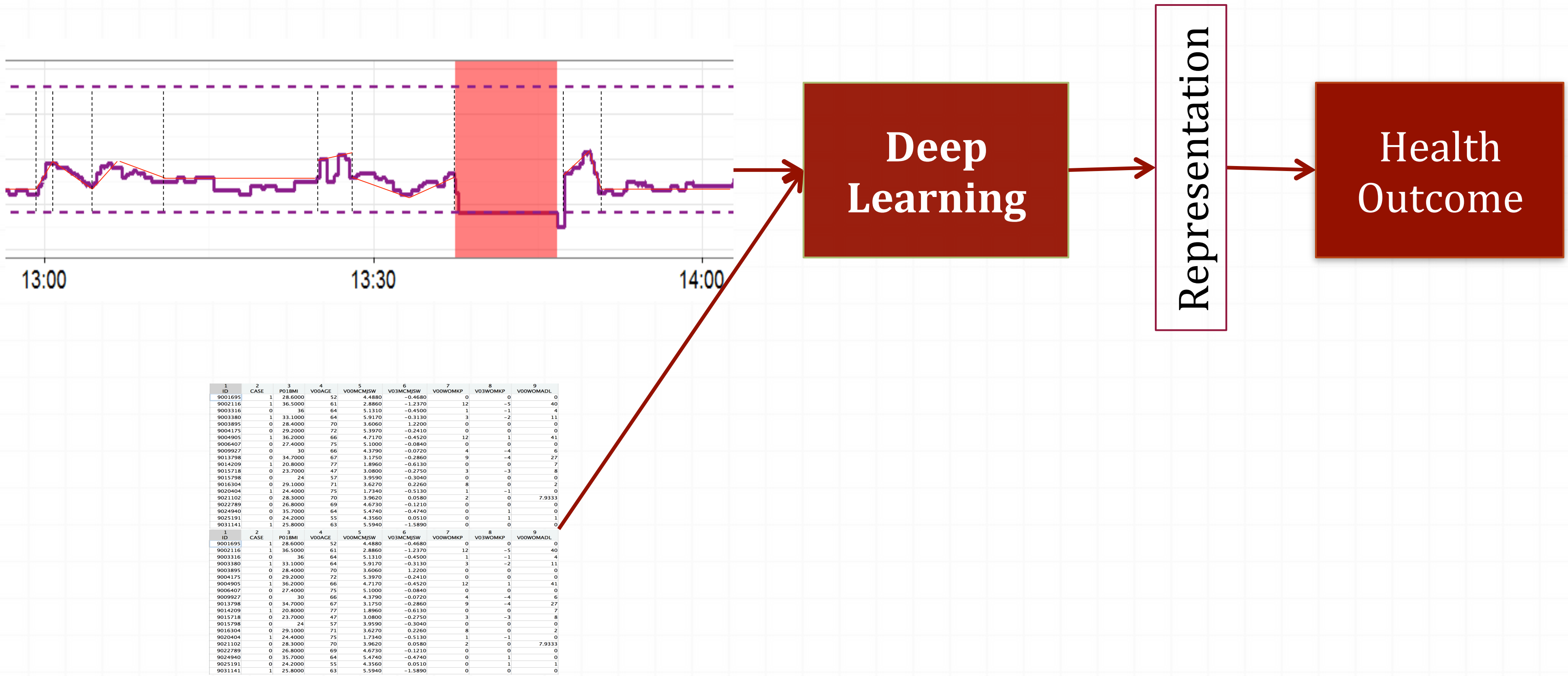
Clinician notes (text)

Learning time series representations

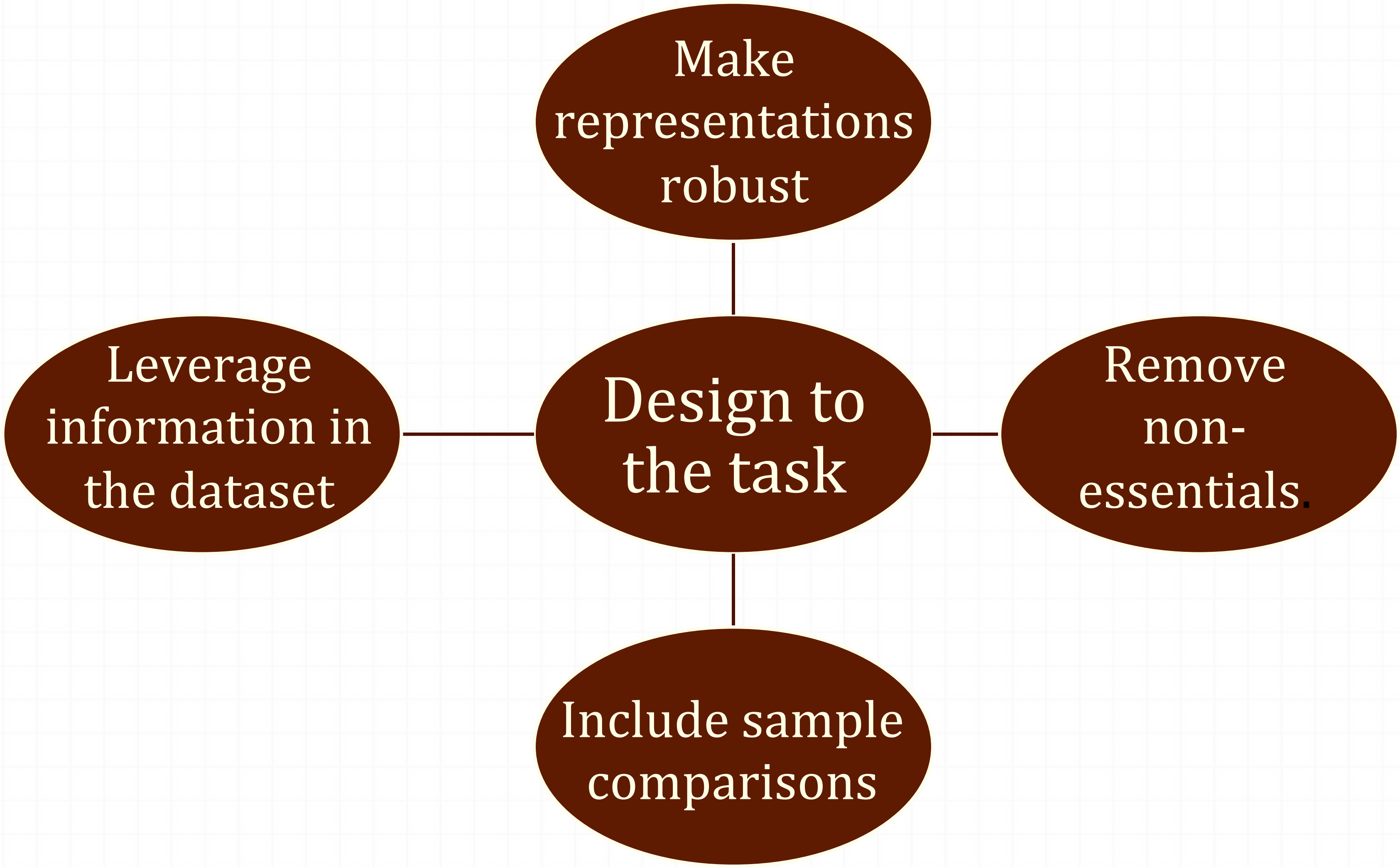


Learning setup

- Classification or regression of health outcome

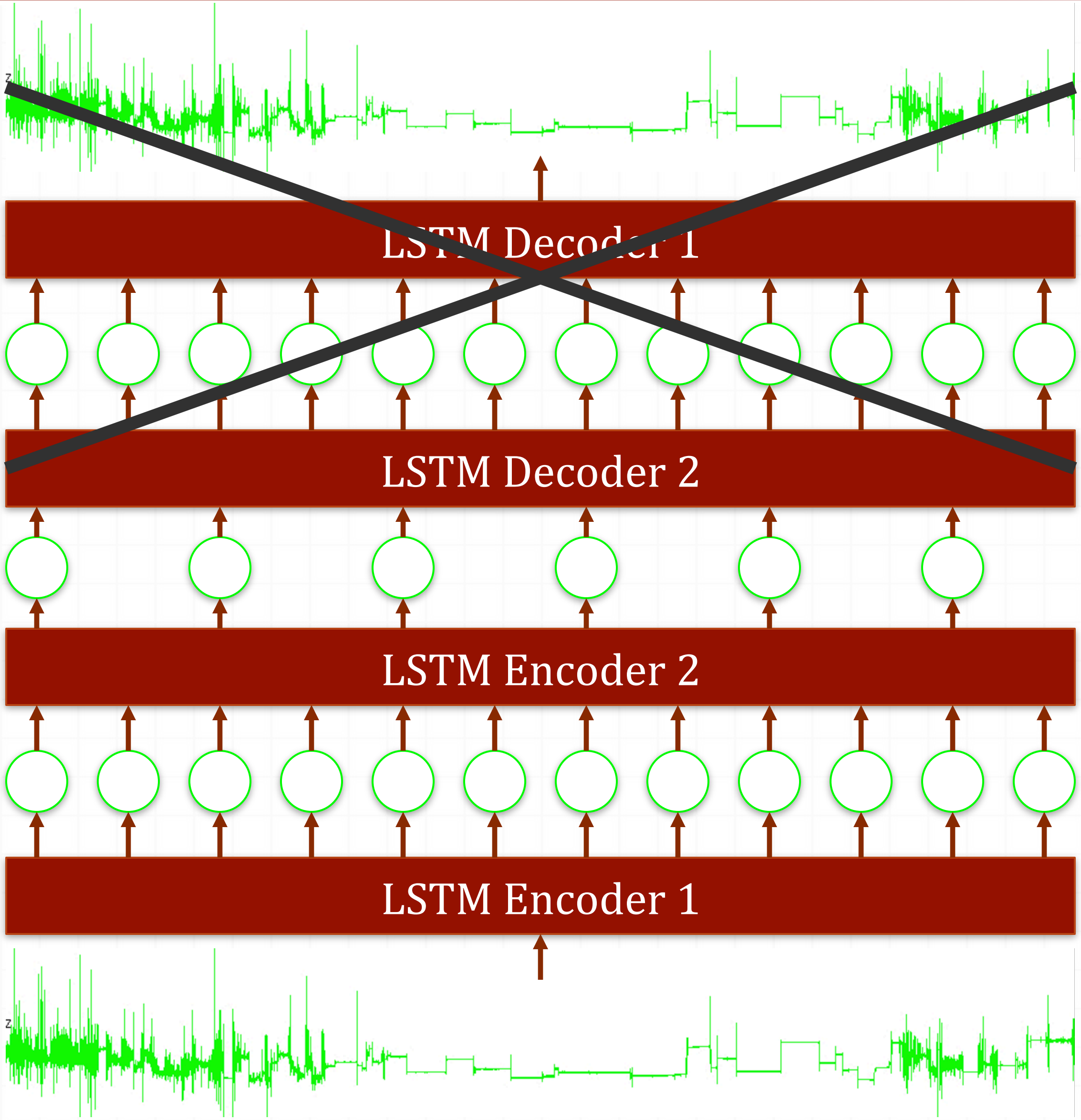


Learning better representations

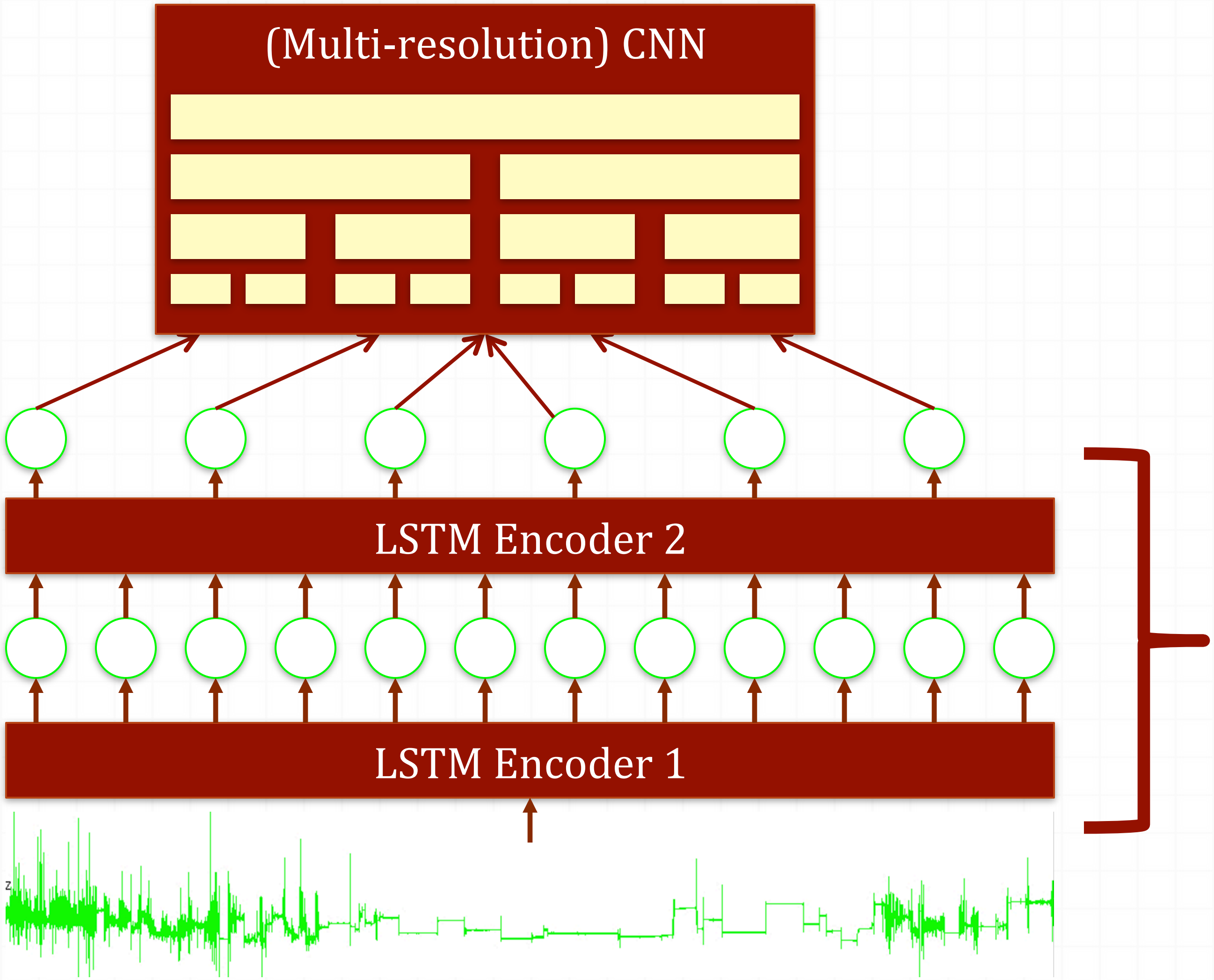


●●●●○ Verizon 4:25 PM
100% 17:31
UP SEP 13
12:31 (CEST)
You moved
14,962 Steps
Active time
2h 27m
Longest active
48m
Longest idle
4h 30m
May
TODAY
10:25 PM
Runtastic Me 5:50
Wednesday
10.DECEMBER
STEPS
●●●○○ T-Mobile 5:12 PM
Nov 30
Calories
249
Active Time
1h 19m
Miles
2.8
Steps
6842
Goal: 10000
Level: Active
Do More
950
475
06:00
12:00
18:00
00:00
Me
Trends
Goals
Groups
History
Achievements
●●●○○ T-Mobile 5:15 PM
Close
STEPS
Goal: 10000 Steps
Wednesday, Nov 30, 2016
6842
1:10:14
Active Time
2.83 mi
Distance
218
Cal
5:39 AM
11:59 PM
Stats
Comments
Share
6 PM
12 AM

Robust models via stacked encoders



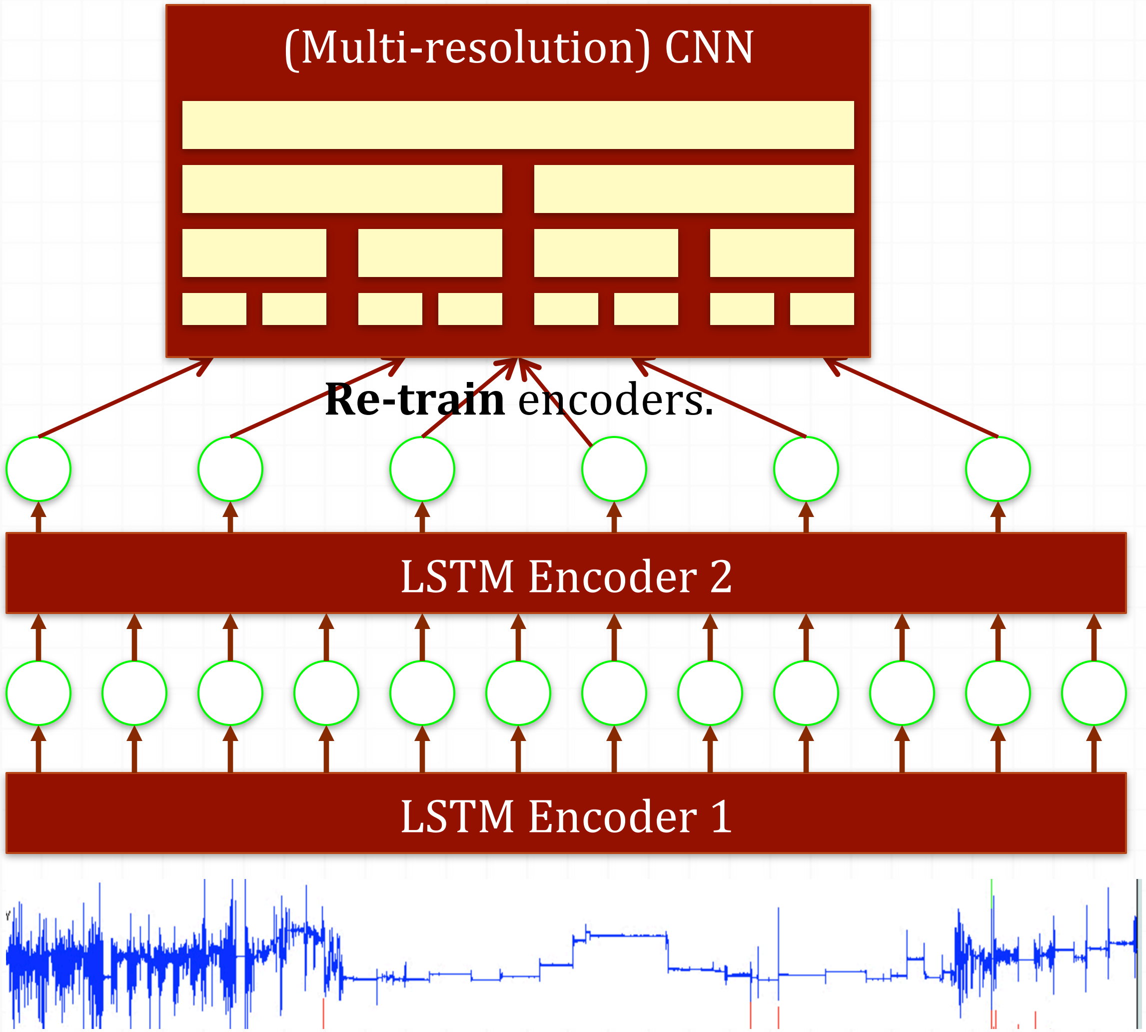
Robust models via stacked encoders



Pre-trained encoders improve performance

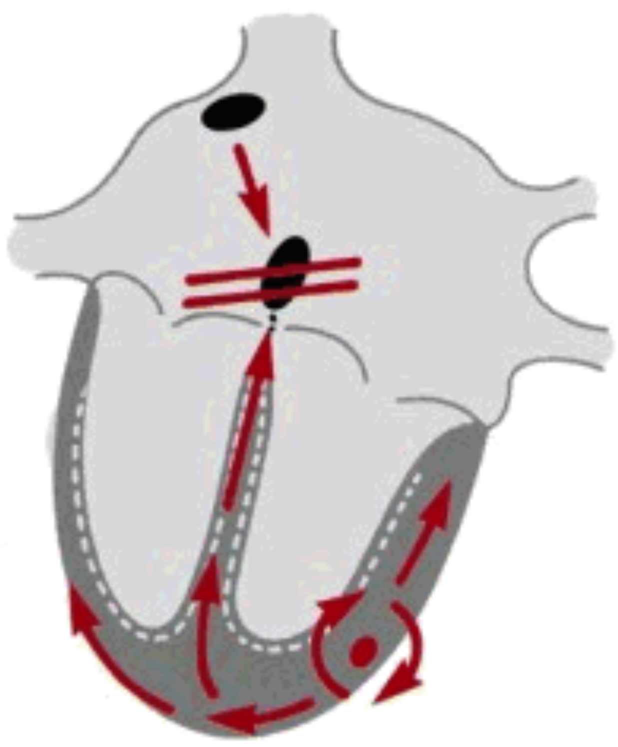
Robust models via stacked encoders

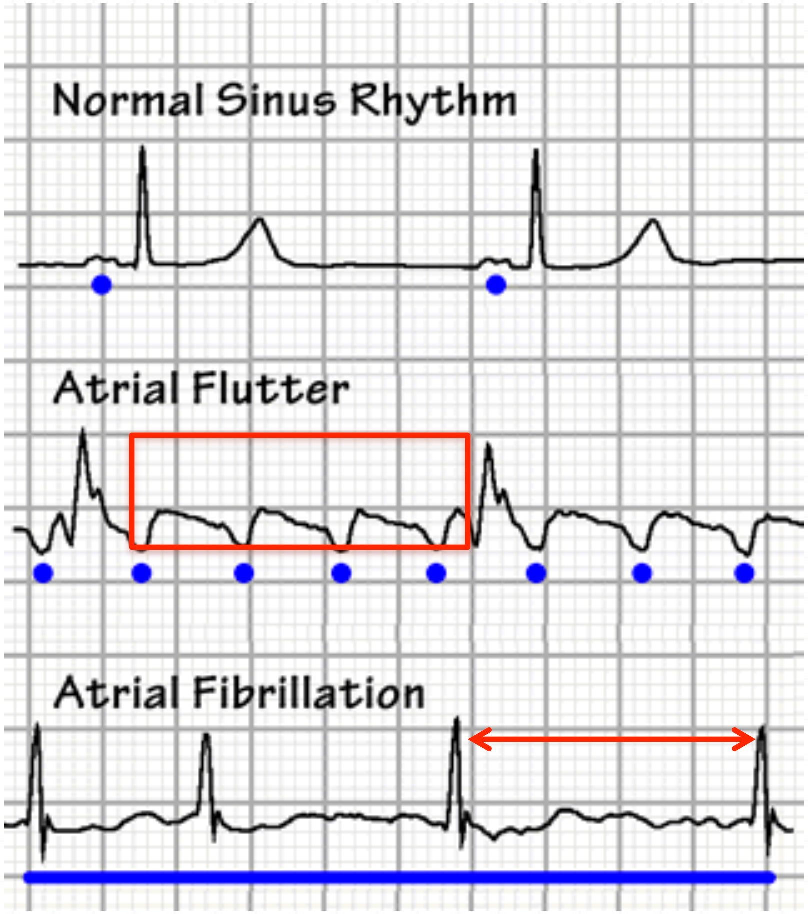
Keep higher-level logic.



Incorporating structured covariates

- Example: how medical history can improve features.





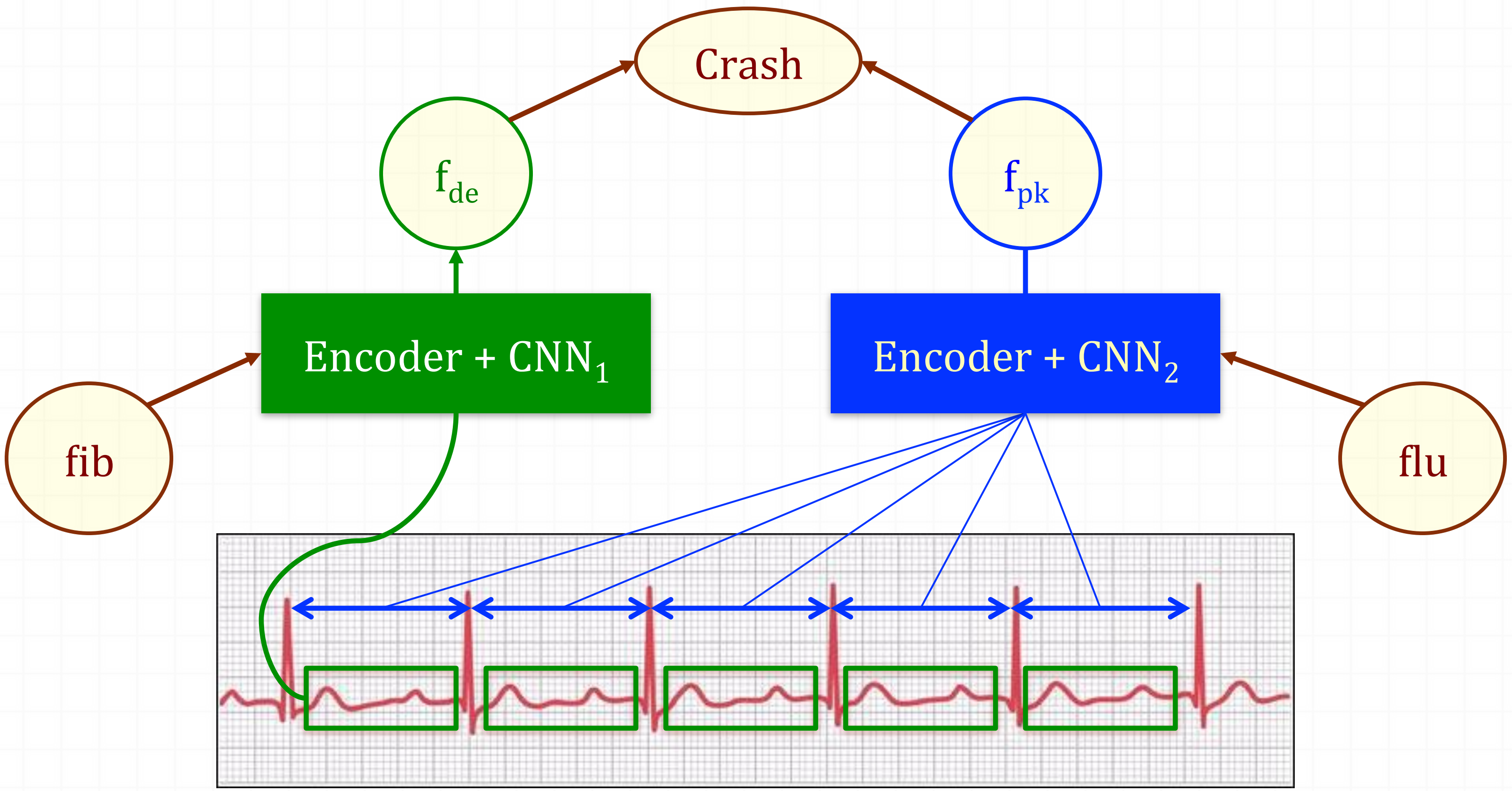
Duty cycle of signal is informative

Peak to peak distance is informative

Source: cardiachealth.org

Incorporating structured covariates

- 2 binary covariates: fibrillation and flutter
- 2 features being learned: $f_{dc\ and}\ f_{pk}$.



Incorporating structured covariates

S = structured covariates

1. Introduce S before the last set of convolutions

   LSTM-C

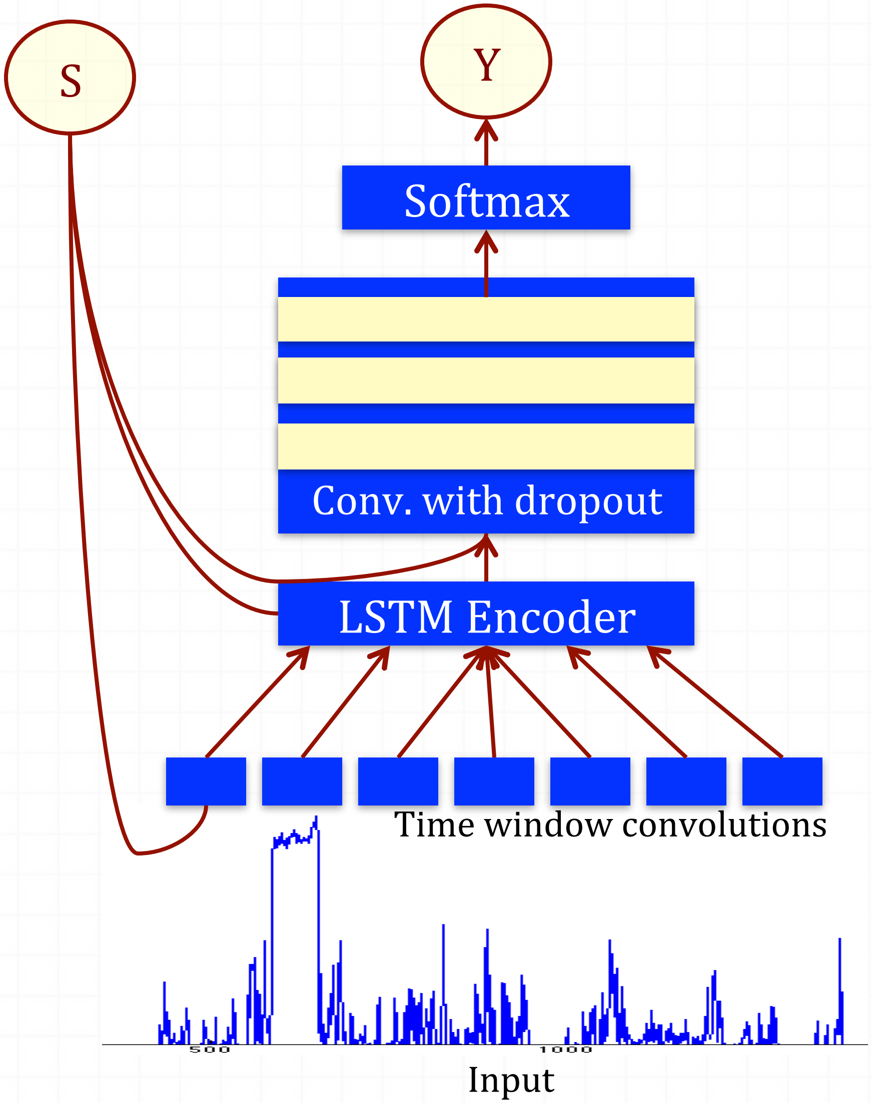
2. Use S in the LSTM structure

   Add terms to nonlinearities of LSTM
   $W_{fs} \bullet s$, $W_{is} \bullet s$, $W_{Cs} \bullet s$, $W_{os} \bullet s$

   LSTM-S

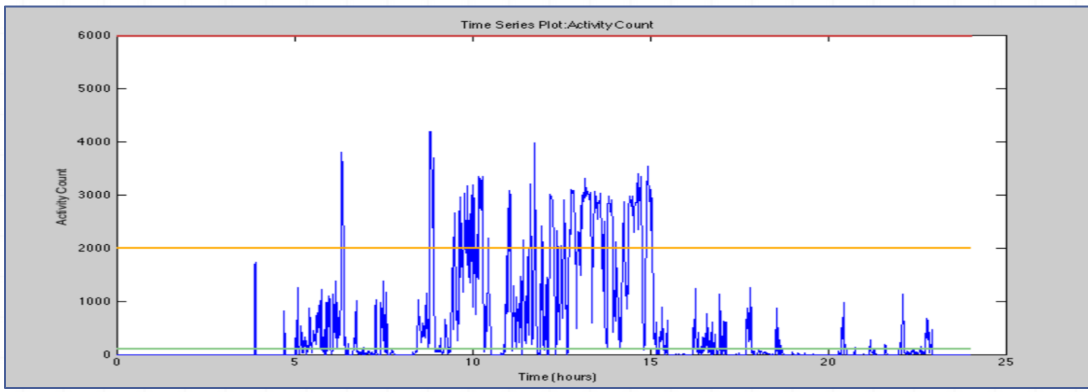
3. Use S as input for time window convolutions

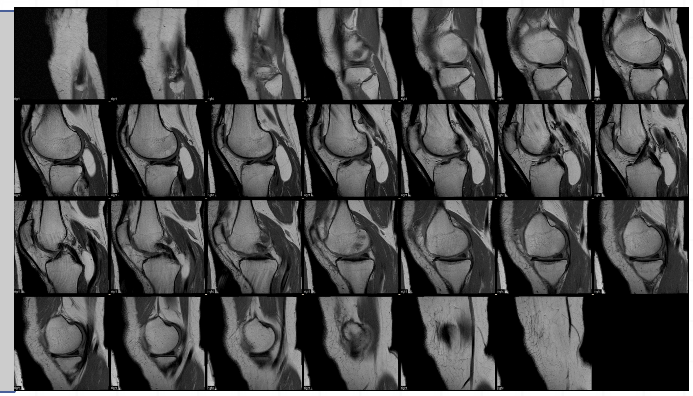
   LSTM-TC



Predicting osteoarthritis progression

- Osteoarthritis Initiative Dataset (OAI) accelerometer data - obtained from 2000 subjects, for a monitoring period of 7 days, expressed as activity counts; 50 structured covariates.
- Predict whether subjects are at risk for OA-related pain.





Accuracy	Histograms + Rand Forests	6 –layer CNN	LSTM-C	LSTM-S	LSTM-TC
Pain / No pain	0.67	0.68	0.70	0.74	0.73

Sample comparisons

- Sinus rhythm. Are these subjects similar?

Sample comparisons

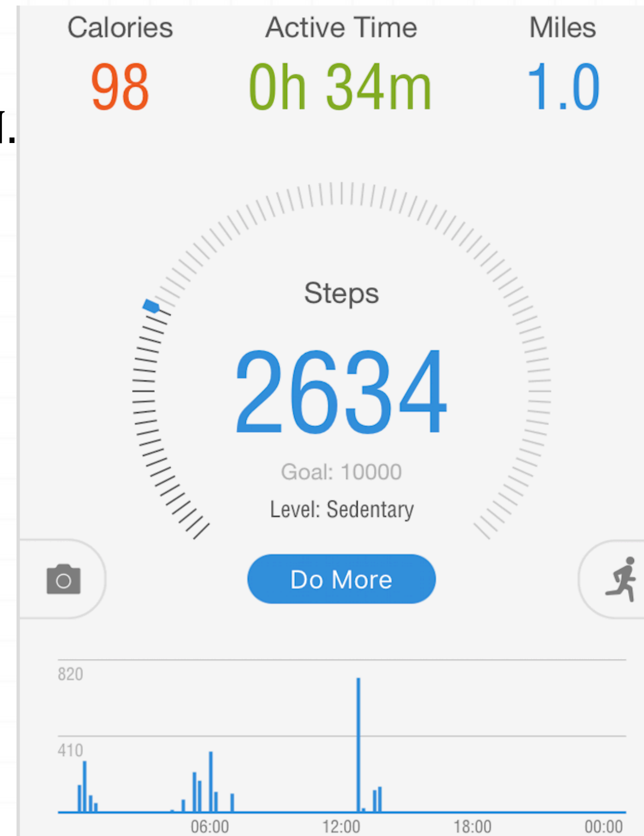
- Sinus rhythm. Are these subjects similar?

Sample comparisons

- **Typical** activity traces. Are these people equally healthy?

~~NO.~~

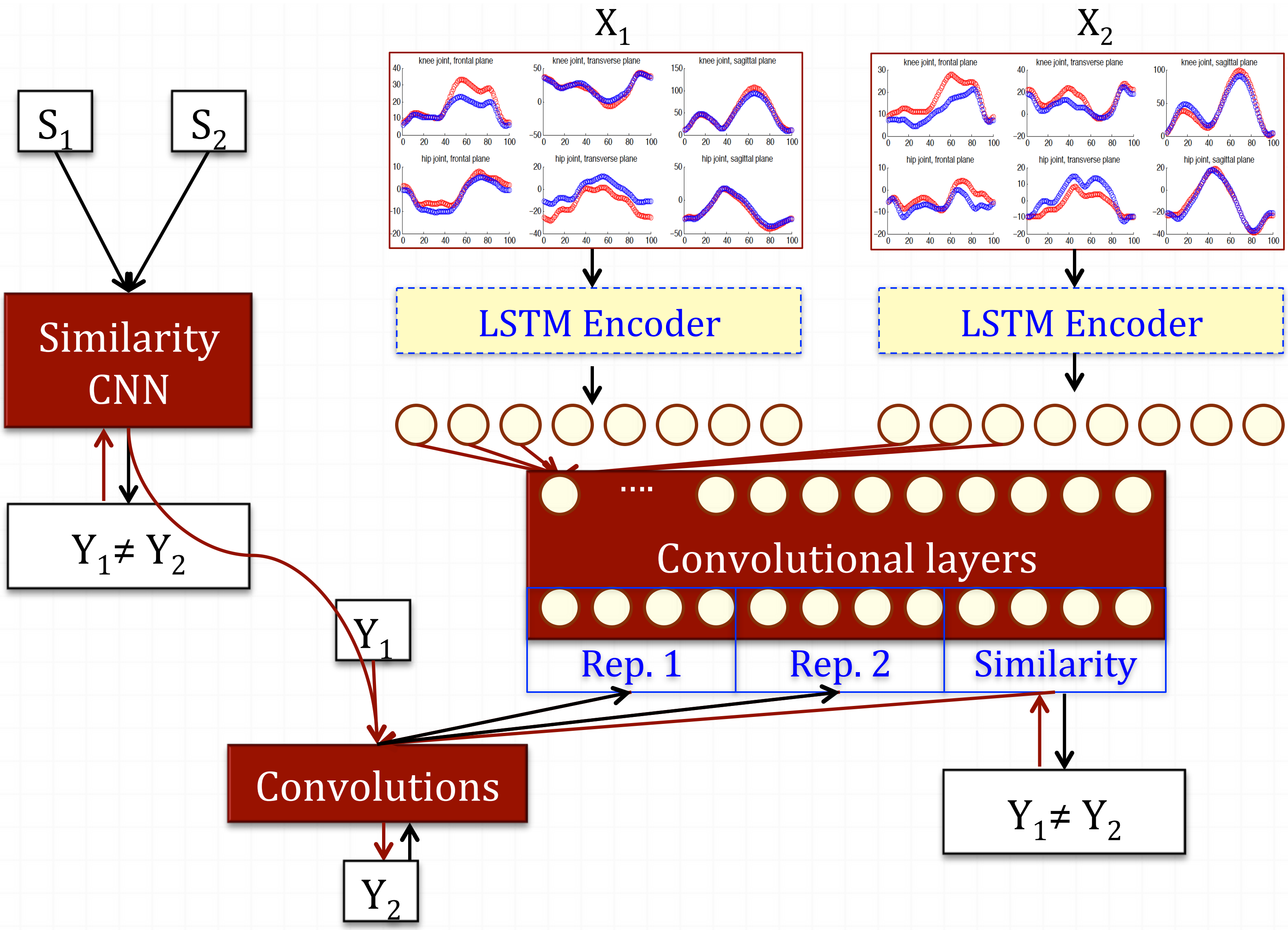
UNKNOWN.



AGE: 25

AGE:52

Sample comparisons



Similarity LSTM results

- Predicting OA-related pain and cartilage

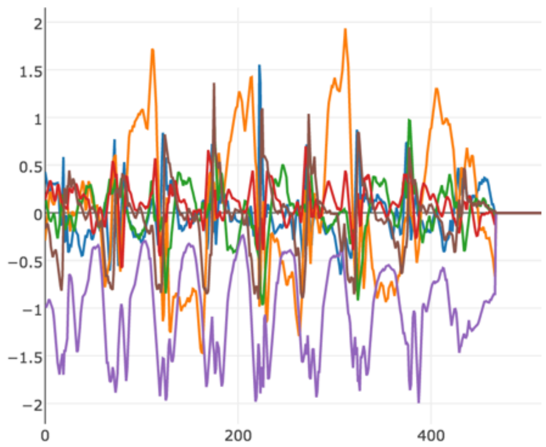
Method Accuracy	Pain / No pain	Joint space narrowing increase
PCA+SVM	0.67	0.70
LSTM-S	0.74	**0.73**
LSTM+SIM	**0.76**	**0.73**
DTW	0.71	0.71
6-layer CNN	0.68	0.72

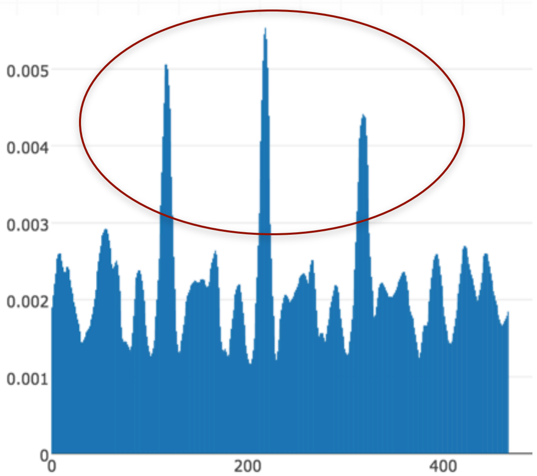
Introducing attention

- Attention allows interpretability of what the network learns
- Can be used to reduce the parameter space
- Sequence is encoded through a bidirectional LSTM
- Last layers: weighted sum of all hidden vectors (attention w)
- Implementation [based on Vinyals 2015]: $u_i = v^T \mathit{tanh}(Wh_i)$

128 sequences (128 x 1028 x 6)	Bi-LSTM with hidden vectors of size 100 (128 x 1028 x 200)	Attention Encoder (128 x 200)	Decoder = FC layer with 2 output classes (128 x 2)	Softmax and Maximum (128 x 1)

Introducing attention

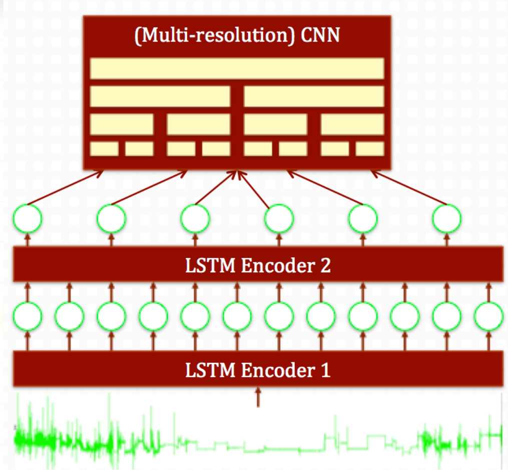


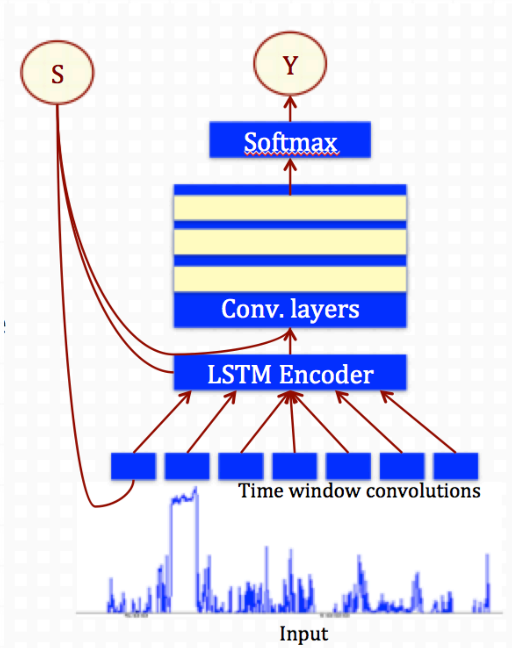


- OU-ISIR Gait Database, 1280 sequences
- 6 times series per patient: triaxial accelerometer/gyroscope
- Classify whether subject as male/female
- Prediction accuracy: 83%

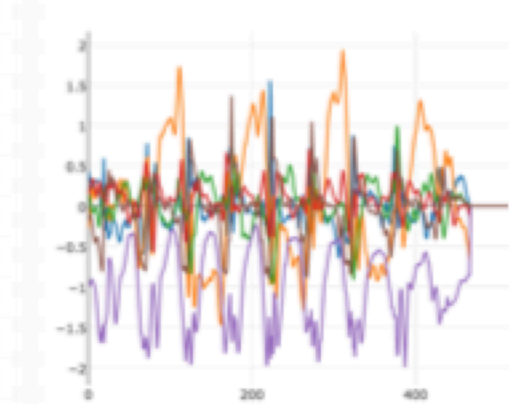
Future work:

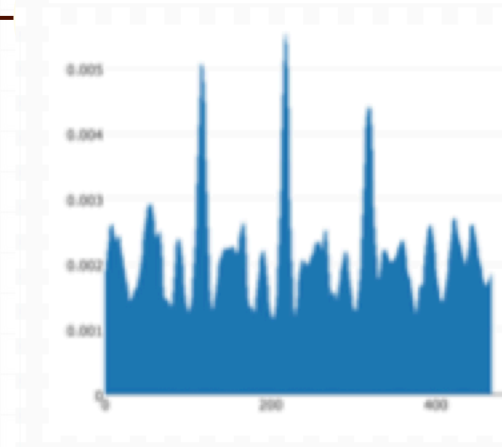
- Improve the model architecture (replace RNNs with CNNs)
- Compute separate attention weights for each dimension of the input

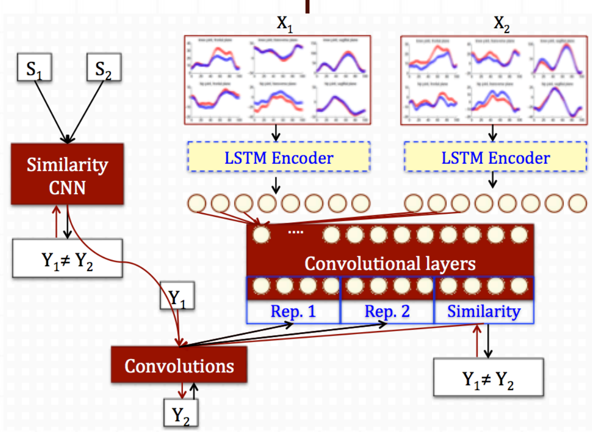

(Multi-resolution) CNN
LSTM Encoder 2
LSTM Encoder 1


S
Y
Softmax
Conv. layers
LSTM Encoder
Time window convolutions
Input

Review






X_1
X_2
S_1
S_2
Similarity CNN
LSTM Encoder
LSTM Encoder
....
Convolutional layers
Rep. 1
Rep. 2
Similarity
$Y_1 \neq Y_2$
Y_1
Convolutions
Y_2
$Y_1 \neq Y_2$

Madalina Fiterau
mfiterau@cs.stanford.edu

Stanford University

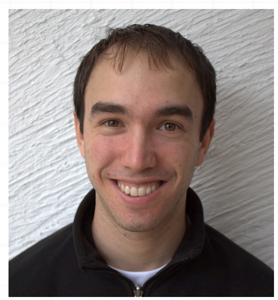
mobilize
Center for Mobility Data Integration to Insight

Thanks!

We acknowledge support from the NIH (U54 EB020405).
We thank Dr. Eni Halilaj for her contributions to the Osteoarthritis Progression experiment.

Chris Ré

Scott Delp



David Sontag

Masters Students:
- Suvrat Booshan
- Nopphon Siranart
- Charles Bornhonesque