Learning from heterogeneous sequences of sparse medical data for early prediction of sepsis

Mahbub Ul Alam

Doktorand

Department of Computer and Systems Sciences

Stockholm University

Borgarfjordsgatan 12, Kista

Email: mahbub@dsv.su.se

Sepsis, at a glance

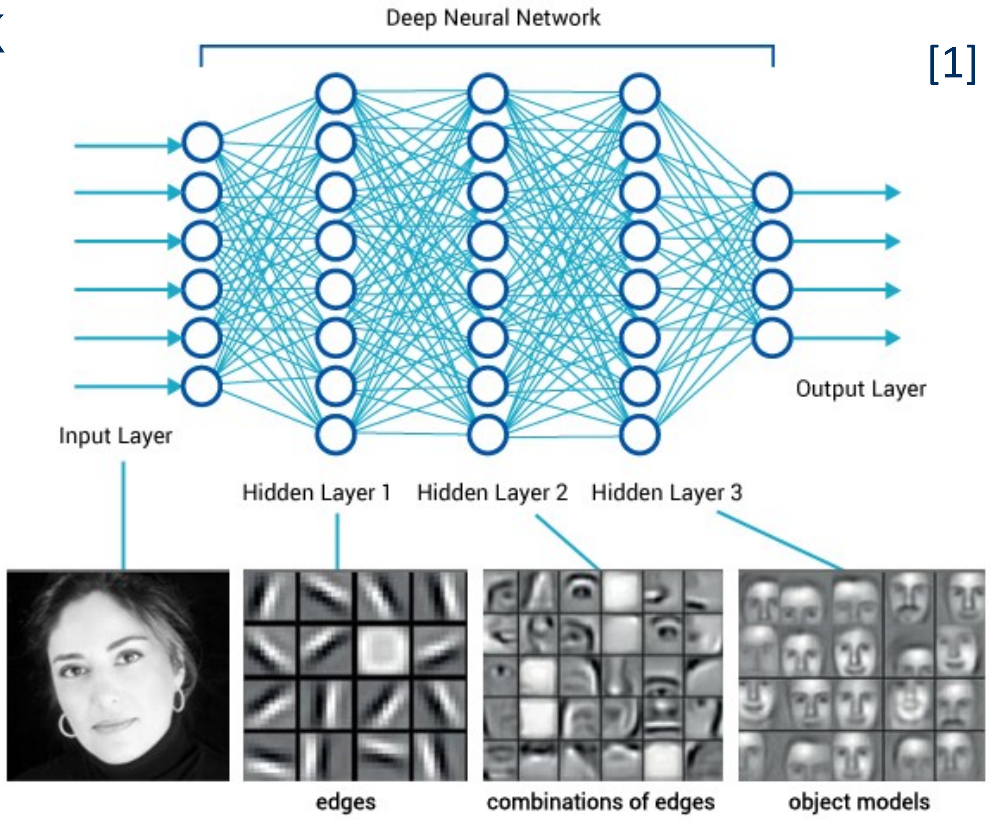
- A life-threatening complication to infections
- A leading cause of hospital morbidity and mortality
- One of the most serious forms of healthcare associated infections
- Survival is dependent on initiating appropriate antimicrobial treatment as early as possible
- Mortality from septic shock increases by 7.6% for every hour that antimicrobial treatment is delayed after the onset

Goal

- Early prediction of sepsis in the non-ICU (intensive care units) setting from electronic health records (EHRs)
- Performance analysis of long short-term memory based recurrent neural network (RNN-LSTM)
- Investigating temporality and sequence length
- Investigating missingness

Deep neural network

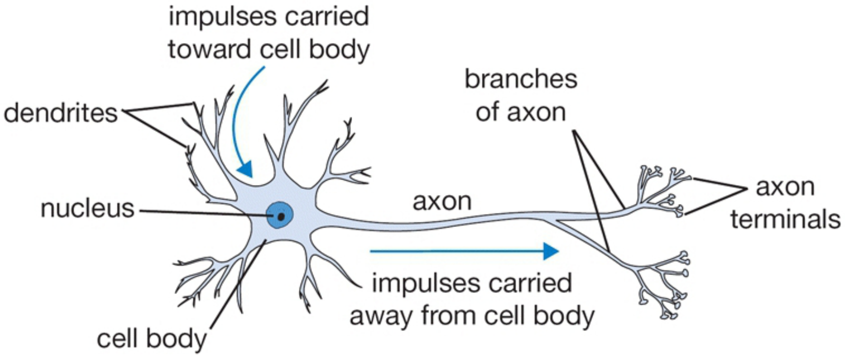
- First learns to detect low-level patterns
- With more data, it might learn to combine these patterns into more complex ones
- With even more data, it might learn to map these higher-level patterns into classes / decisions themselves
- It takes an input, returns an classification output, and in between it learns to represent features through hidden layers of representations.



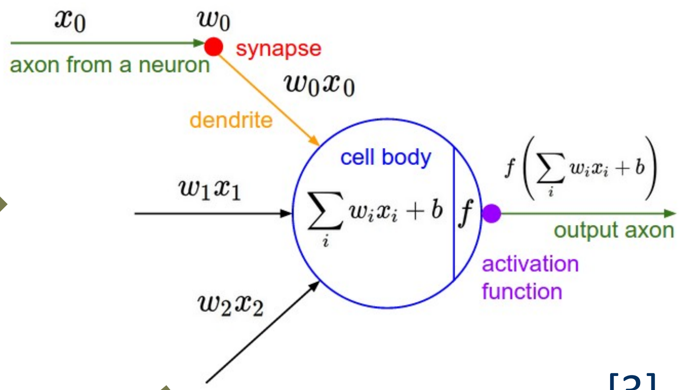
[1]

Stockholm University

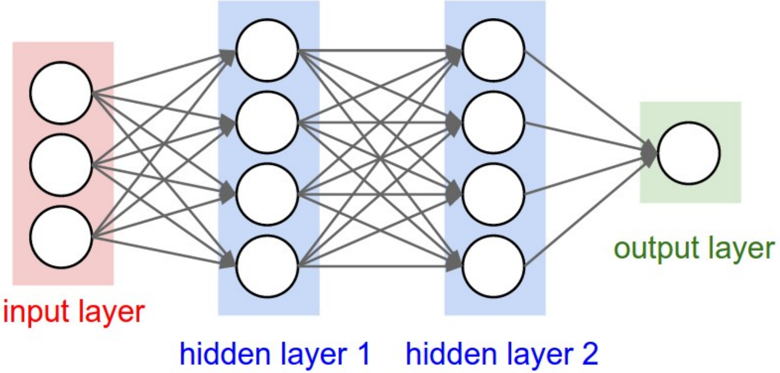
Deep neural network



[2]

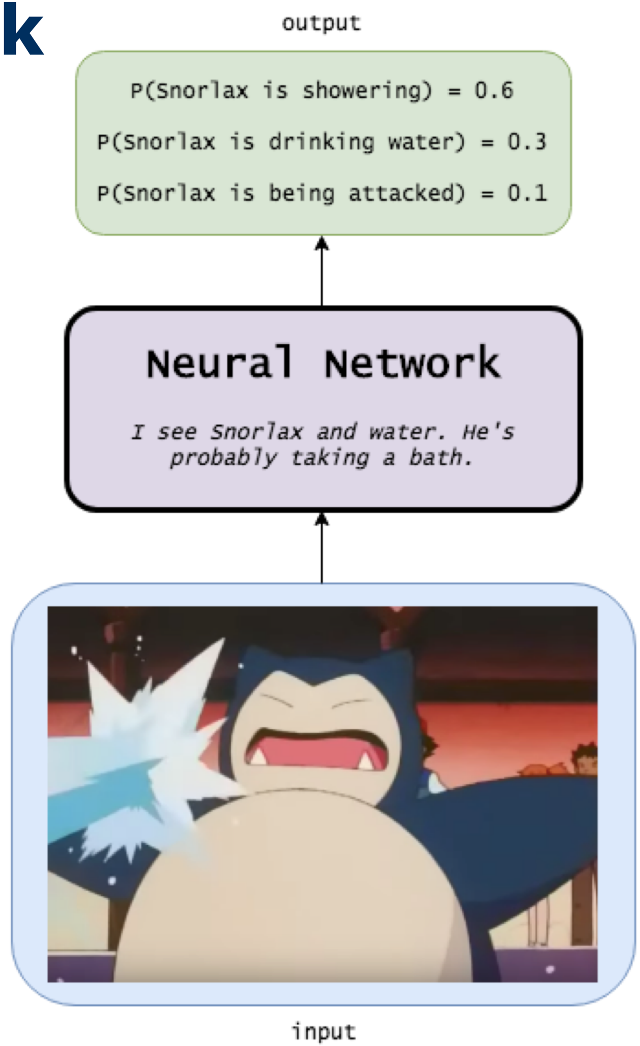


[3]



[4]

Deep neural network

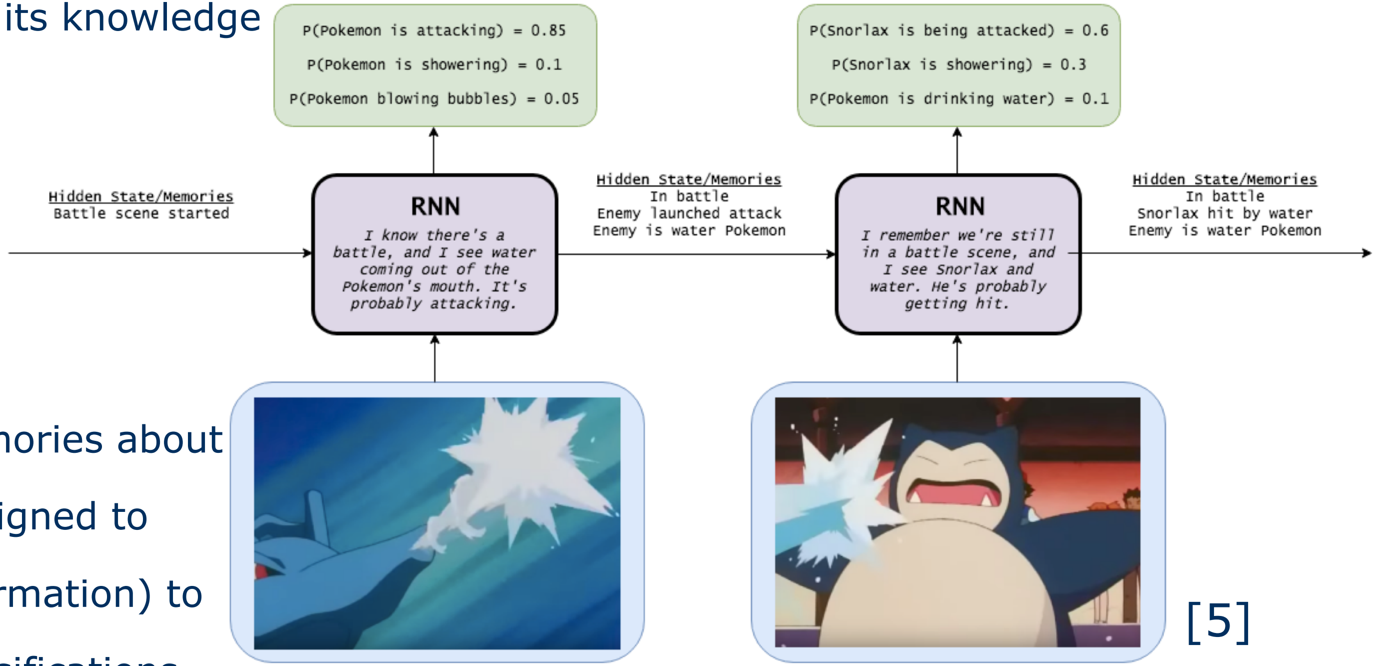


[5]

Stockholm University

Recurrent neural network (RNN)

- After seeing each input, outputs a label and also updates its knowledge of the world



[5]

- Maintains internal memories about the world (weights assigned to different pieces of information) to help perform its classifications.

Stockholm University

Long short-term memory network (LSTM)

- Adding a forgetting mechanism
- ❖ When new inputs come in, it needs to know which beliefs to keep or throw away

- Adding a saving mechanism
- ❖ When new a input comes in, the model first forgets any long-term information it decides it no longer needs

- ❖ Then it learns which parts of the new input are worth using, and saves them into its long-term memory

Stockholm University

Long short-term memory network (LSTM)

- Focusing long-term memory into working memory.

❖ The model needs to learn which parts of its long-term memory are immediately useful.

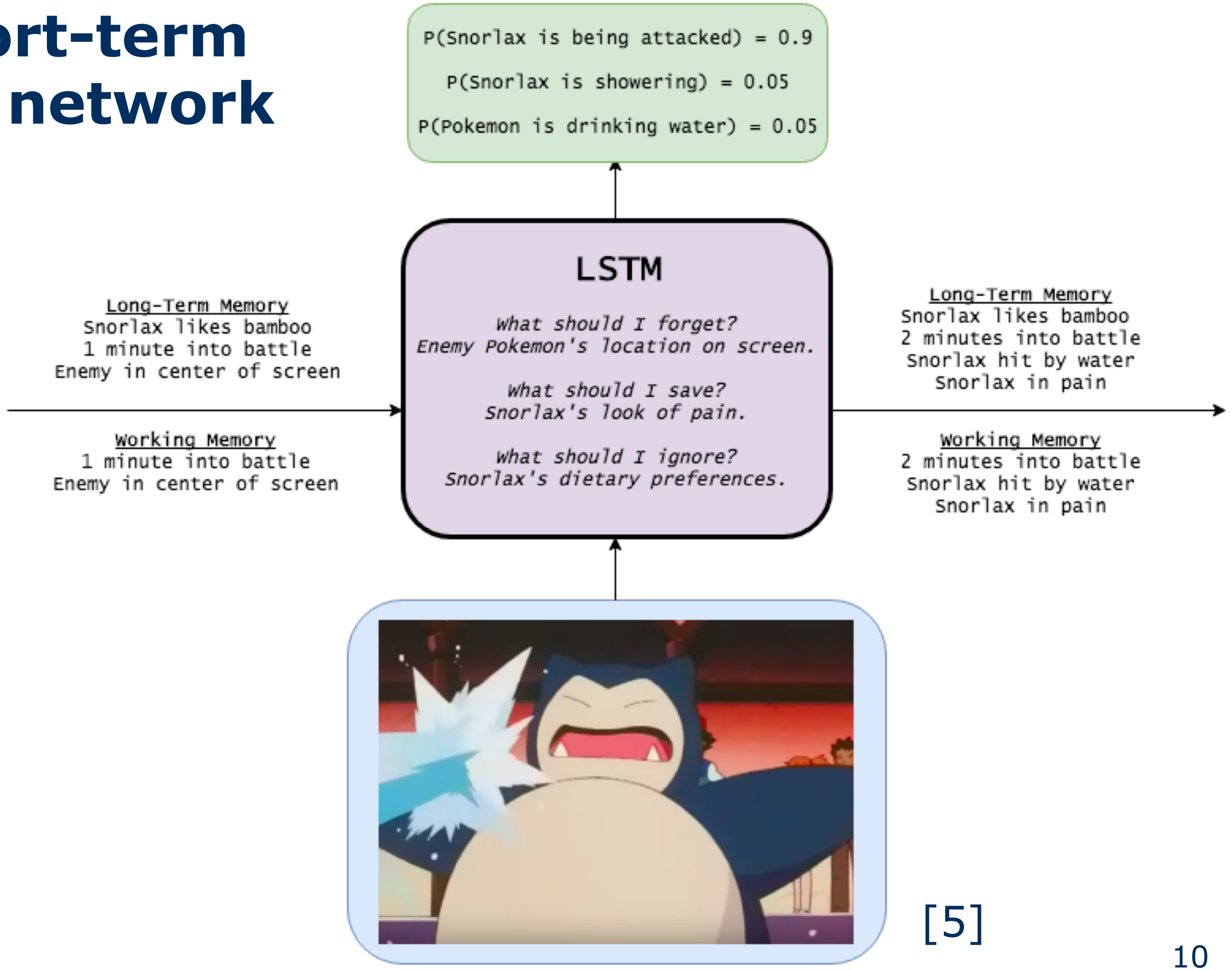
❖ Instead of using the full long-term memory all the time, it learns which parts to focus on instead.

- An RNN can overwrite its memory at each time step in a fairly uncontrolled fashion

- An LSTM transforms its memory in a very precise way
  - By using specific learning mechanisms for which pieces of information to remember, which to update, and which to pay attention to

- This helps it keep track of information over longer periods of time.

Long short-term memory network (LSTM)



[5]

HEALTH BANK - Swedish Health Record Research Bank

- Unique research resource containing a large sets of electronic patient records
- Used in a number of research projects carried out by the Clinical Text Mining Group, Department of Computer and Systems Sciences, Stockholm University
- Contains data from over 512 clinical units from Karolinska University Hospital (2006–2014) over two million patients.
- Structured information contains, a serial number (de-identified) for each patient, age, gender, ICD-10 diagnosis codes, drugs, ab and blood values, admission and dicharge time, and date
- Unstructured data contains text written under different headings

Stockholm University

HEALTH BANK - Swedish Health Record Research Bank

We would like to express our sincere gratitude to Karolinska University Hospital for their contribution with data in HEALTH BANK

Stockholm University

Data

- Patients > 18 years admitted to the hospital between July 2010 and June 2013
- Followed until first sepsis onset, discharge or death
- Excluded if admitted to an obstetric ward and censored during ICU-care
- Encompasses 124,054 patients and 198,638 care episodes
- Sepsis in the cohort is 8.9%

Stockholm University

Care episode

- Constitutes the period between admission and discharge (or death) for a particular patient
- If a patient was admitted via the emergency unit, this arrival time marks the beginning of the episode
- If the time in between the next admission and the previous discharge for the same patient is within 24 hours, the two are considered to be part of the same care episode
- Care episodes may involve stays in several different wards and vary greatly in length, with a median length of around three days

Input data selection

- Defining collection of microbial cultures and tests from all types of body fluids
- Newly administrated antimicrobial treatment is collected based on ATC-codes J01 and J04
- Demographic and physiological for the following 19 parameters: age, body temperature, heart rate, respiratory rate, systolic and diastolic blood pressure, oxygen saturation, supplementary oxygen flow, mental status, leucocyte count, neutrophil count, platelet count, C-reactive protein, lactate, creatinine, albumin, and bilirubin

- The output of the following scoring tools were used: NEWS2, qSOFA and SOFA. The model was only allowed to access data that would be readily available in the EHR - or could be computed from it - at the time of prediction.
- Most of the variables are numeric generally extremely sparse, with a missing rate of more than 90% in some cases

Care episode representation

- Transforming the care episodes into sequences based on a given bin width and experiment with a total of six different window sizes: 1, 2, 3, 4, 6 and 8 hours
- Timestamps associated with clinical events allow to assign values to a given bin
- A variable in a time window can either be missing or have multiple values associated with it
- When multiple values are present in a time window, the "worst" value is chosen
  - defined as the most pathological value for a particular variable and is determined apriori by clinical experts

Handling missing values

- When data is assumed to be missing not at random
  - Imputation is not carried out
  - Missing values are simply assigned an integer value which is not present in the data
  - The idea is that the model may learn to treat missingness as a distinct feature

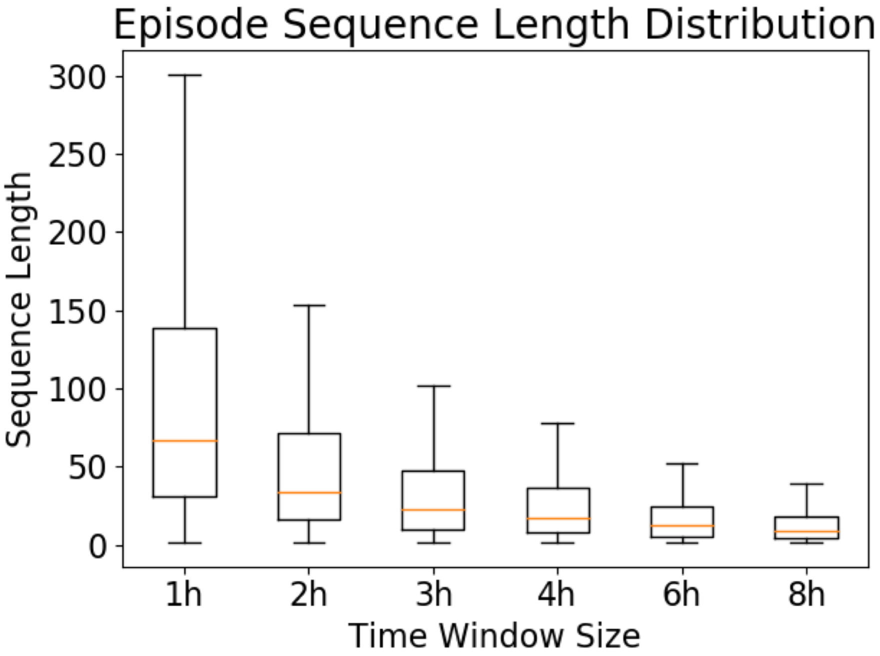
Stockholm University

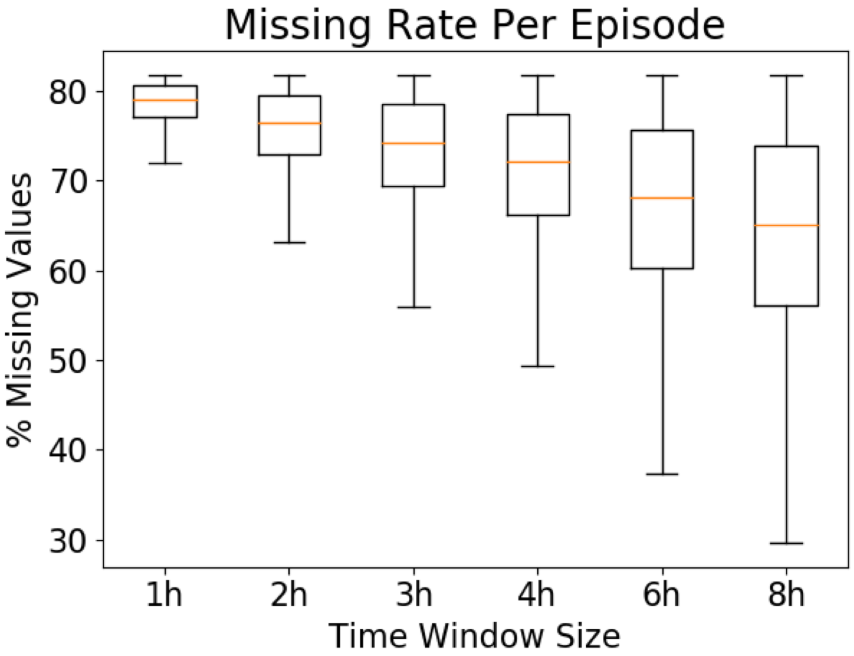
Handling missing values

- When data is assumed to be missing at random
  - When a value exists for a given feature in the care episode, it is carried forward to subsequent windows until another present value is encountered, which is then in turn carried forward and so on
  - When there is no value for a given feature in a care episode, it is imputed globally
  - For categorical features, the most frequent value is chosen, while mean imputation is carried out for numeric features
  - For SOFA, qSOFA, and NEWS2, missing values are not mean-imputed; instead, the score is assumed to be 0 - if missing - at the start of an episode and then carried forward as described above

Stockholm University

Care episode representation





Stockholm University

Experiments

- Experiment 1: different time window sizes
- Experiment 2: handling missing values
- Experiment 3: performance at different time points
- Experiment 4: evaluation of earliness
- Experiment 5: performance with different sequence lengths
- Experiment 6: community onset vs. hospital onset sepsis
- Experiment 7: predicting severe sepsis

Results: predicting severe sepsis

Model	TP			% IHM			Earliness		
	<24h	<48h	All	<24h	<48h	All	<24h	<48h	All
LSTM, w/o imputation, 4h	48	49	56	84.2	86.0	98.2	0.7	1.6	29.7
NEWS2	26	26	29	50.9	50.9	50.9	-0.5	-0.5	14.4
qSOFA	11	11	11	19.3	19.3	19.3	-1.8	-1.8	-1.8

Predictive performance for sepsis cases that led to in-hospital mortality.

TP= Number of true positives

% IHM = Percentage of sepsis-related in-hospital mortality

Earliness = Average prior prediction time to sepsis onset (in hours)

Discussion

- An extensive empirical evaluation is carried out in which six different time window sizes (1, 2, 3, 4, 6 and 8 hours) - affecting missingness and sequence length - are investigated in terms of how this representation impacts on predictive performance and earliness
- The proposed LSTM model, using a 4-hour time window and assuming data is not missing random, clearly outperforms rule-based scoring systems commonly used in healthcare today
- By analyzing the effectiveness of the time window, we can partially infer the diagnosis mechanism implicitly in the model
- It can also be used to investigate the effectiveness of the treatment or the temporal aspect of the physiological markers to identify sepsis earlier

Stockholm University

Future works

- Predicting sepsis should be divided into multiple stages to emulate the actual condition of sepsis
- We will try to incorporate more data including free-text clinical notes
- We will investigate modifications to the neural architecture to make it more task-specific
- We will use additional natural language processing techniques to provide additional important features to our model
- We also plan to investigate the interpretability of the models

Image source

[1] https://www.researchgate.net/figure/Procedure-of-BP-in-training_fig32_326531654

[2] http://cs231n.github.io/assets/nn1/neuron.png

[3] http://cs231n.github.io/assets/nn1/neuron_model.jpeg

[4] http://cs231n.github.io/assets/nn1/neural_net2.jpeg

[5] https://www.topbots.com/exploring-lstm-tutorial-part-1-recurrent-neural-network-deep-learning/

Stockholm University

Questions?

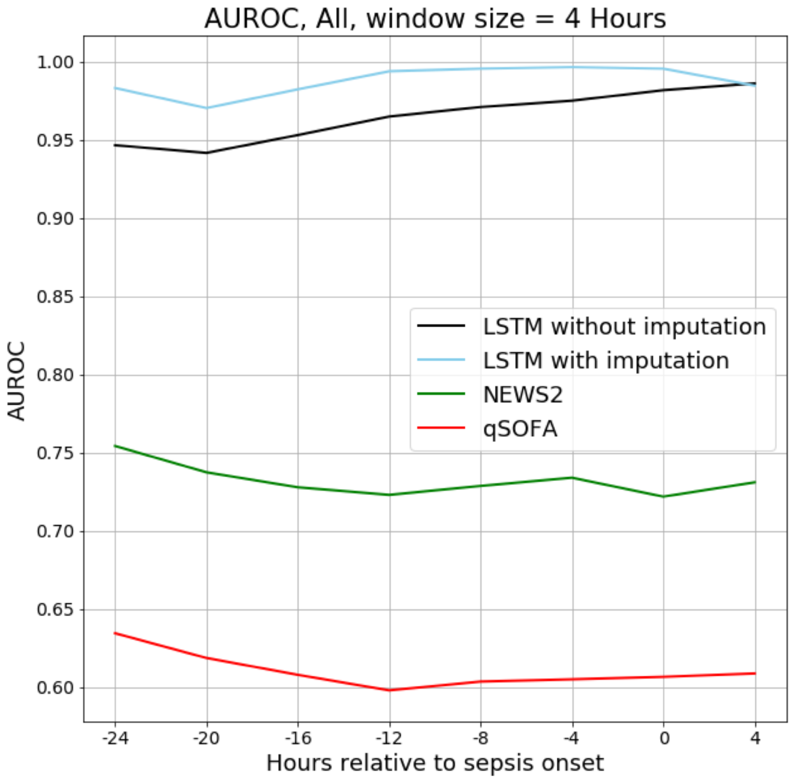
Backup Slides

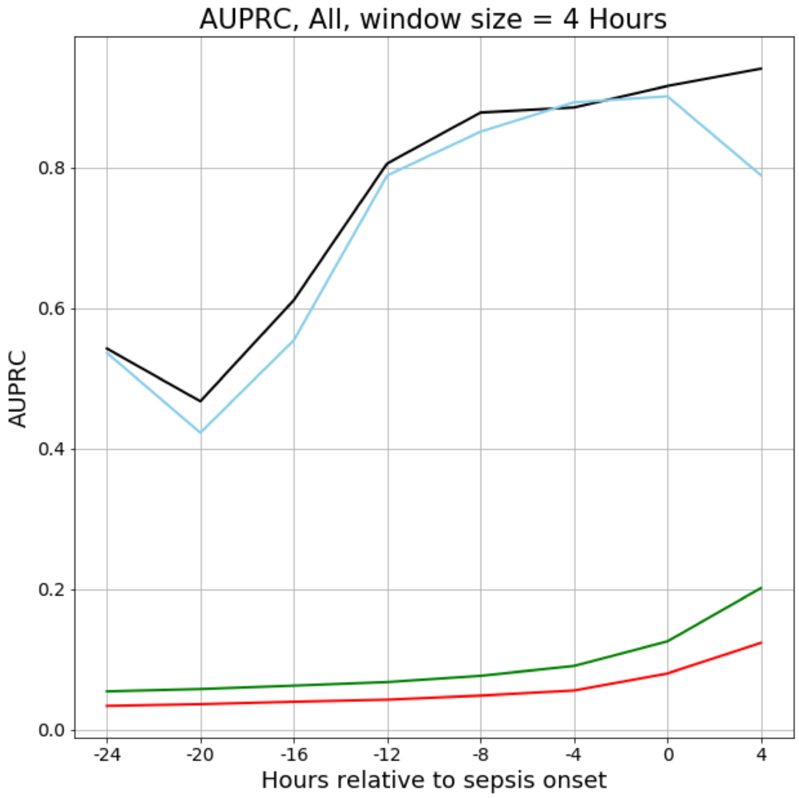
Stockholm University

Results: different time window sizes and handling missing values

W. Size	AUROC	AUPRC	Precision			Recall			F_1			Earliness		
	All	All	<24h	<48h	All	<24h	<48h	All	<24h	<48h	All	<24h	<48h	All
Without Imputation														
1h	**.993**	.907	.745	.759	.781	.823	.887	**1.00**	.782	.819	.877	11.8	**14.2**	28.2
2h	.985	.933	**.869**	**.876**	**.890**	.801	.854	.973	.834	.865	**.930**	4.8	7.2	25.6
3h	.984	.936	.721	.732	.759	.798	.843	.972	.757	.783	.852	3.6	7.9	30.4
4h	.986	**.940**	.812	.819	.838	.811	.850	.971	**.811**	**.898**	.900	3.8	5.9	28.5
6h	.973	.884	.434	.444	.478	.801	.834	.956	.563	.580	.637	1.4	3.1	27.7
8h	.973	.883	.609	.616	.644	.797	.819	.924	.690	.703	.759	-2.0	-0.98	22.4
With Imputation														
1h	.990	.867	.725	.738	.760	.829	**.889**	**1.00**	.773	.807	.864	**11.9**	13.8	27.7
2h	.988	.842	.683	.601	.723	.820	.871	.997	.745	.773	.838	5.5	8.0	27.1
3h	.987	.833	.671	.682	.712	.812	.863	.997	.737	.762	.831	3.8	7.9	30.7
4h	.985	.789	.715	.726	.753	.820	.866	.996	.764	.790	.858	8.3	10.3	**32.8**
6h	.985	.789	.728	.738	.764	.823	.864	.995	.772	.796	.864	2.3	4.2	28.9
8h	.981	.763	.674	.684	.712	**.833**	.870	.994	.745	.766	.830	0.2	1.8	25.8

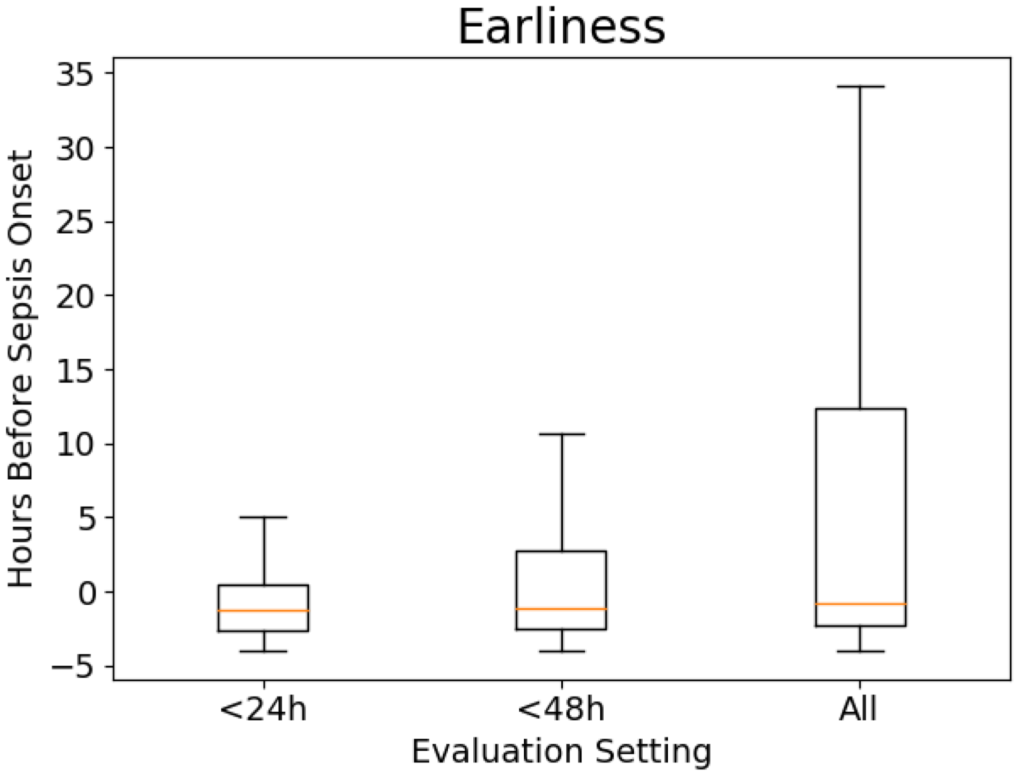
Results: performance at different time points





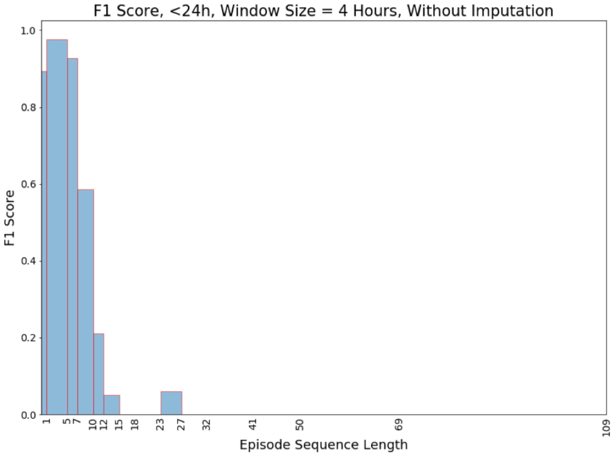
Stockholm University

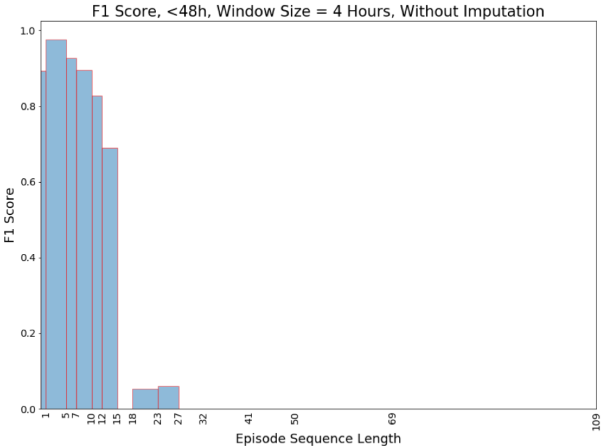
Results: evaluation of early prediction time

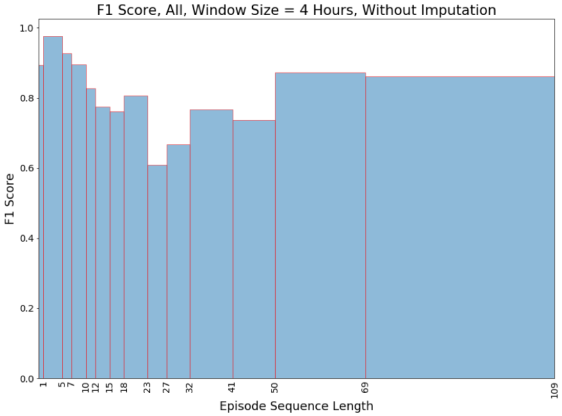


Stockholm University

Results: performance with different sequence lengths







Stockholm University

Results: community onset (CO) vs. hospital onset (HO) sepsis

Metric	CO Sepsis			HO Sepsis		
	<24h	<48h	All	<24h	<48h	All
Precision	.713	.724	.734	1.00	1.00	1.00
Recall	.901	.949	.999	.077	.097	.968
F1	.796	.821	.846	.144	.176	.984
Earliness	1.6	3.7	9.5	4.7	9.9	179.3
Instances	n = 1429			n = 155		
% of Sepsis Episodes	8.17			91.83		