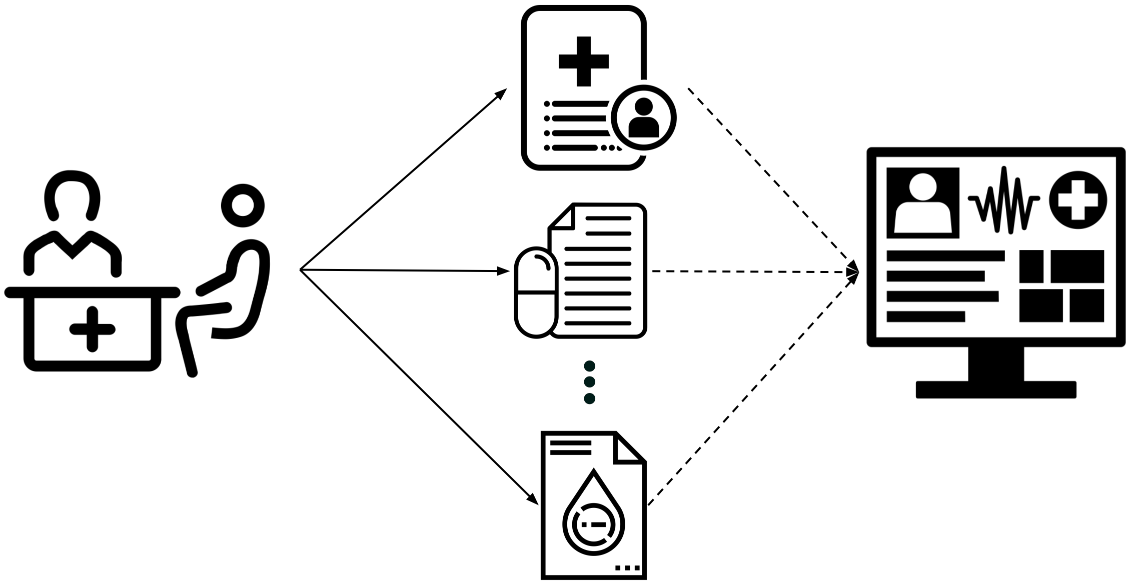
Exploring and Understanding Neural Models for Clinical Tasks

Madhumita Sushil

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

Electronic Health Records

Data in Electronic Health Records



Differential diagnosis prediction

Treatment recommendation

Condition severity estimation

Disease progression estimation

Importance of clinical notes

Unique information present in text, e.g.,

Family history

Signs and symptoms

Reasons for discontinuing treatment

Specific disease phenotypes

Clinical text characteristics

Non-standard terminology and hedging

HTN (hypertension) R/O (rule out)	**Abbreviations and acronyms**
stress+ stress-	**Presence or absence markers**
>stress stress++	**Intensity indicators**
+/-4 weeks +-1 week	**Approximation**
might suggest the possibility of	**Hedging**

Important to use right tools tailored to medical data.

EHR data characteristics: imbalance

Imbalance across ethnicities, age groups, diseases, amount of available data.
Important to ensure models aren't biased before deployment.

Caucasian,
45 years,
Gastrointestinal
disease,
300 records

Caucasian,
60 years,
Cardiovascular
disease,
150 records,

Caucasian,
75 years,
Cardiovascular
disease,
900 records

Asian,
2 years,
Injury and
poisoning,
10 records

Caucasian,
new born,
Perinatal
disease,
1 record

Hispanic,
90 years,
Cardiovascular
disease,
100 records

Research questions

1) How can we develop systems that can make predictions about patient health? (Part 1, Patient Representations)

Research questions

1) How can we develop systems that can make predictions about patient health? (Part 1, Patient Representations)

2) How can we understand what these systems have learned? (Part 2, Model Interpretability)

Research questions

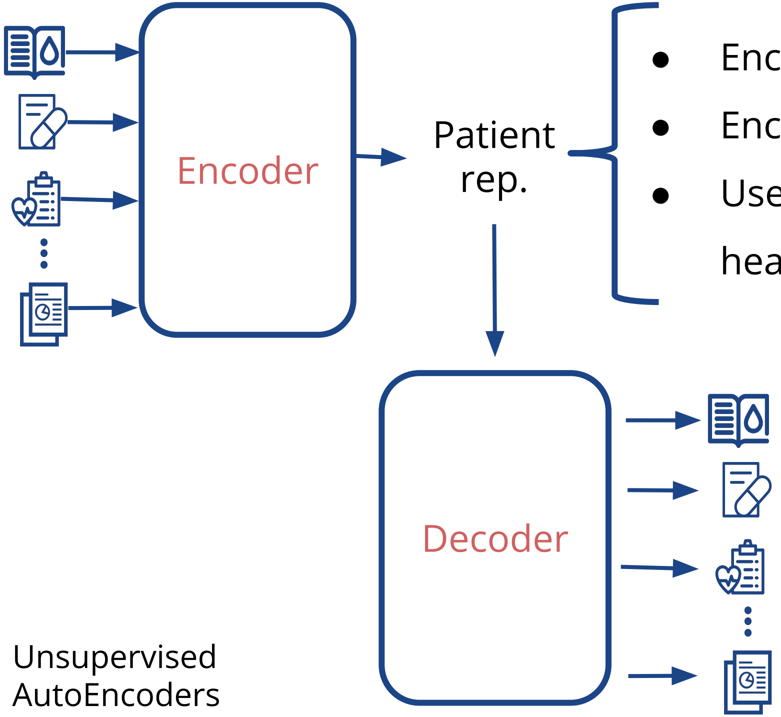
1) How can we develop systems that can make predictions about patient health? (Part 1, Patient Representations)

2) How can we understand what these systems have learned? (Part 2, Model Interpretability)

3) How can we improve these systems using medical domain knowledge? (Part 3, Domain Knowledge Integration)

1. Patient Representations

Task-independent patient representations

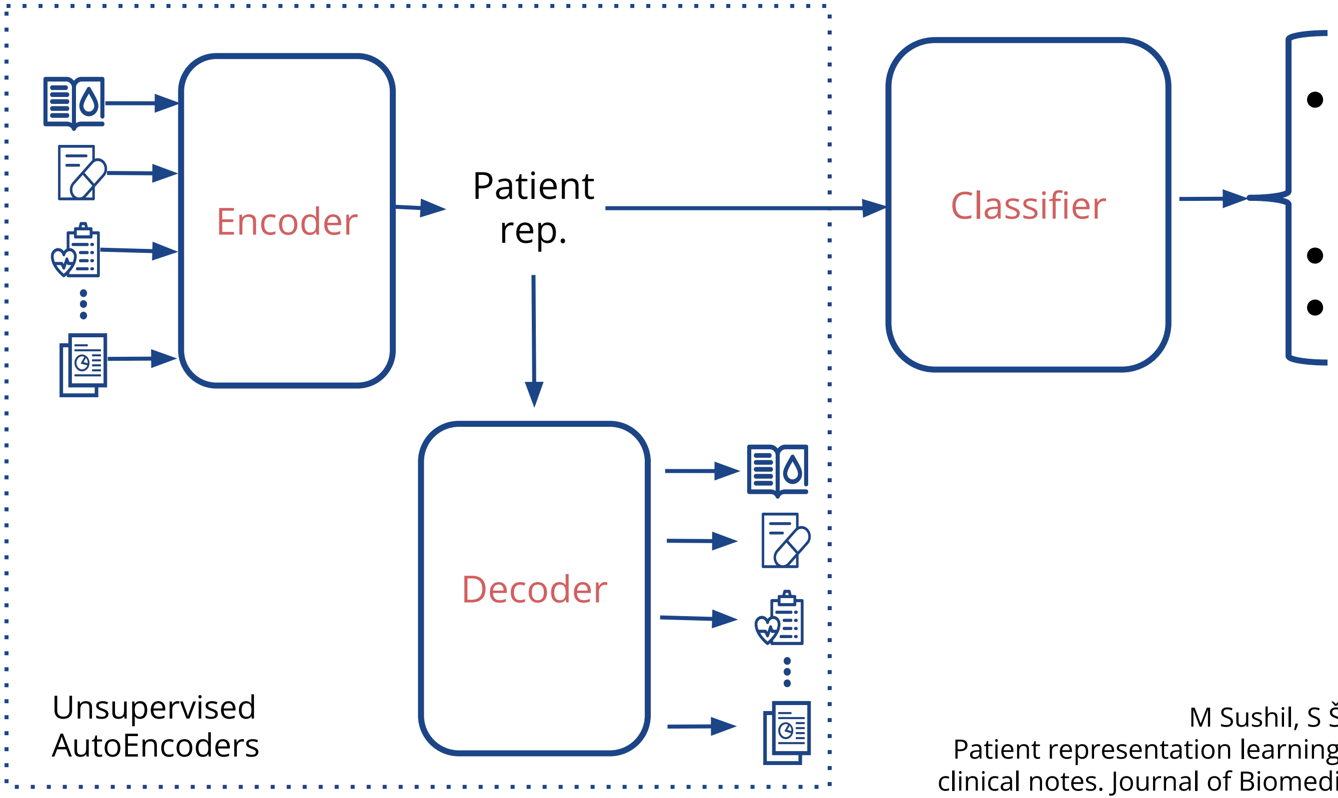


- Encode factual patient info
- Encode text semantics
- Useful for decision-making about patient health

Unsupervised AutoEncoders

M Sushil, S Šuster, K Luyckx, and W Daelemans.
Patient representation learning and interpretable evaluation using clinical notes. Journal of Biomedical Informatics, 84:103 – 113, 2018.

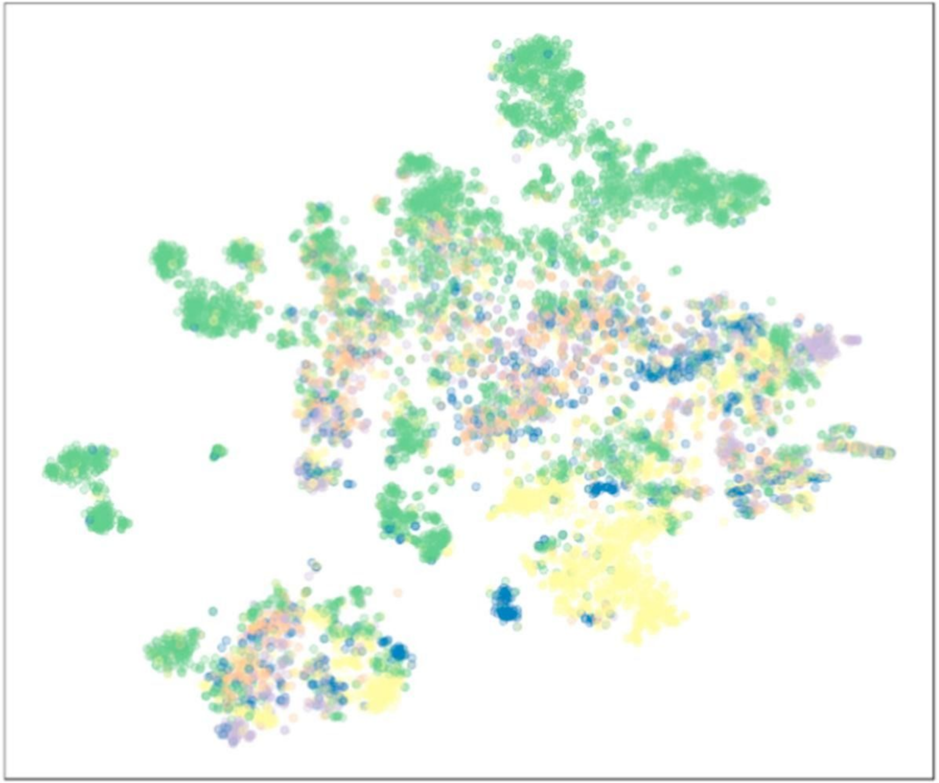
Task-independent patient representations



- In hospital/30 days/1 year patient mortality
- Diagnostic category
- Procedural category

M Sushil, S Šuster, K Luyckx, and W Daelemans. Patient representation learning and interpretable evaluation using clinical notes. Journal of Biomedical Informatics, 84:103 – 113, 2018.

2D visualization of learned representations



Representations learn soft patient cohorts

- Diseases of the circulatory system
- Diseases of the digestive system
- Infectious and parasitic diseases
- Injury and poisoning
- Neoplasms

M Sushil, S Šuster, K Luyckx, and W Daelemans. Patient representation learning and interpretable evaluation using clinical notes. Journal of Biomedical Informatics, 84:103 – 113, 2018.

Summary of Results

Patient representations are significantly better (+20% AUROC) when we need to understand holistic patient conditions from only a few data points.

This advantage is not visible when direct lexical mentions are sufficient for classification.

M Sushil, S Šuster, K Luyckx, and W Daelemans.
Patient representation learning and interpretable evaluation using clinical notes. Journal of Biomedical Informatics, 84:103 – 113, 2018.

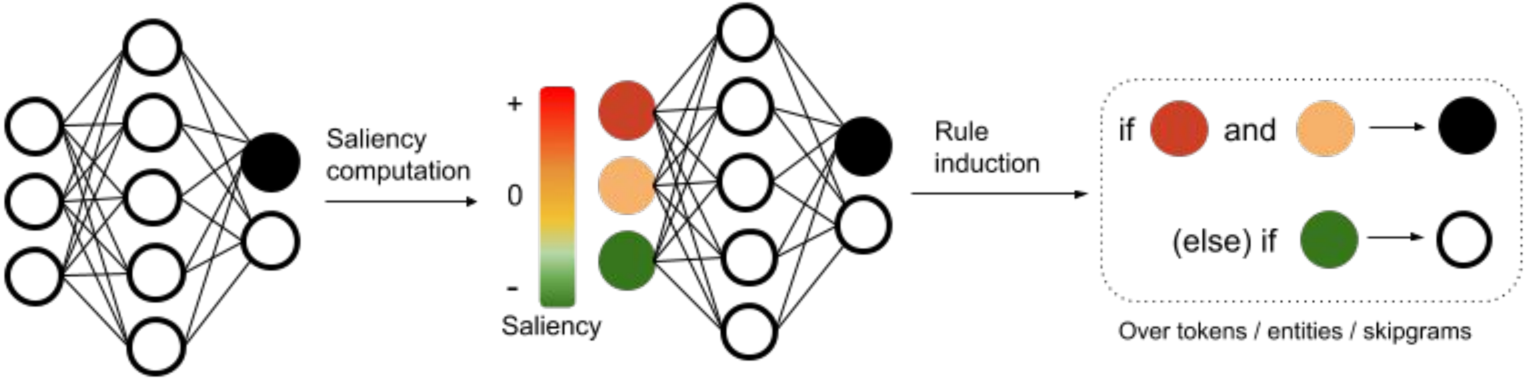
2. Model Interpretability

Need to understand and explore our models

- How can we improve our models?

- Is the model generalized for use across populations?
    - Is it biased towards a specific cohort of patients in one hospital?
    - Is it biased towards properties of the EHR the hospital used?
    - Is it biased towards data pre-processing steps?

Finding explanation rules



What are the most important features?

Frequently related to patient conditions.

Terms absent from a patient's notes used to rule out specific predictions.

Negation terms, absence of numbers and function words. Further context needed for disambiguation.

Encoding longer context in explanations

Combine most important features in the form of *if-then-else* rule lists to mimic neural outputs

M Sushil, S Šuster, and W Daelemans. Rule induction for global explanation of trained models. Workshop on Analyzing and interpreting neural networks for NLP (BlackboxNLP), EMNLP 2018

Encoding longer context in explanations

Combine most important features in the form of *if-then-else* rule lists to mimic neural outputs

if ***<feature1 is imp>*** and ***<feature2 is less imp>*** and … ⇒ **class1**

else if ***<feature1 is absent>*** … ⇒ **class2**

else **class3**

Quantifies interaction between different input features and classes.

M Sushil, S Šuster, and W Daelemans. Rule induction for global explanation of trained models. Workshop on Analyzing and interpreting neural networks for NLP (BlackboxNLP), EMNLP 2018

Explanations for primary diagnostic category prediction (medical entity features)

↑ **Take blood pressure** and
🚫 **Nothing by mouth** and
🚫 **Flagyl**

→**Diseases of the circulatory system** (✔84/90)

Disease-related information is reflected in explanations.

Explanation for in-hospital mortality prediction (medical entity features)

↑ **Physical examination** and

↑ **Pregnancy with medical condition**

→**Dies within hospital** (✔221/222)

The model learns biases in the dataset.

Adding further context in form of skipgrams

*"**no** signs **of infection** were found. "*

M Sushil, S Šuster, and W Daelemans. Distilling neural networks into skipgram level decision lists. Computing Research Repository, 2005.07111, 2020.

Adding further context in form of skipgrams

*"**no** signs **of infection** were found. "*

Importance is computed at skipgram-level instead of individual token-level.

Explanation rule patterns combine important skipgrams to mimic neural outputs.

M Sushil, S Šuster, and W Daelemans. Distilling neural networks into skipgram level decision lists. Computing Research Repository, 2005.07111, 2020.

Explanations for sepsis prediction classifier (Token input features)

↑ **sepsis major surgical** →**septic** (✔209/209)

🚫 **complaint : sepsis** and

↑ **chief hypotension major** →**septic** (✔169/169)

Summary of Results

High precision, lower recall of explanation fidelity (~80% macro F1).

More context is always better!

Explanations can reveal both true patterns and biases in dataset.

Rules can get complex due to several conjunctive clauses.

3. Domain Knowledge Integration

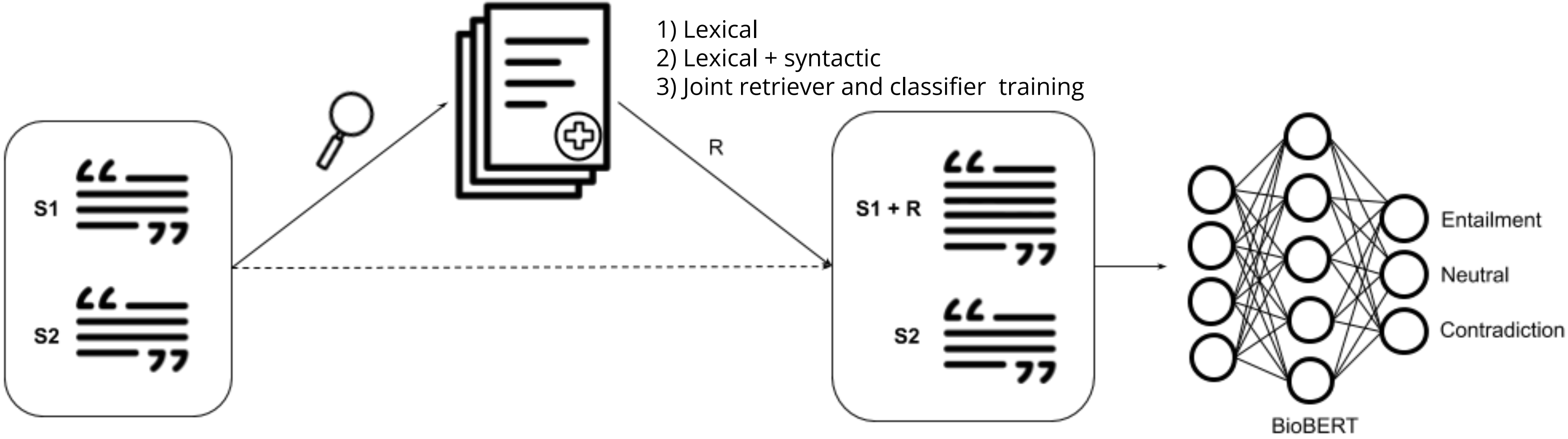
Textual knowledge augmentation - why?

SOTA models are significantly worse in clinical language reasoning than domain experts.

Augmenting domain knowledge can bridge the gap.

Text is more easily available than ontologies.

Natural Language Inference



Is the meaning of S2 entailed in S1?

Summary of Results

Accuracy before and after knowledge integration in (Bio)BERT are comparable (~83%).

Limited success due to the complexity of finding relevant information in an unsupervised setup.

Open question for future research.

Roadmap for the future

Sequence-aware representation learning for long documents (>1k tokens).

Roadmap for the future

Sequence-aware representation learning for long documents (>1k tokens).

Multi-modal clinical decision making.

Roadmap for the future

Sequence-aware representation learning for long documents (>1k tokens).

Multi-modal clinical decision making.

Reducing complexity of explanations (esp. no. of absent phrases) while also retaining high fidelity.

Roadmap for the future

Sequence-aware representation learning for long documents (>1k tokens).

Multi-modal clinical decision making.

Reducing complexity of explanations (esp. no. of absent phrases) while also retaining high fidelity.

Exploring causal explanations for more complex neural architectures.

Roadmap for the future

Sequence-aware representation learning for long documents (>1k tokens).

Multi-modal clinical decision making.

Reducing complexity of explanations (esp. no. of absent phrases) while also retaining high fidelity.

Exploring causal explanations for more complex neural architectures.

Text-based unsupervised knowledge retrieval for reasoning.

Thank you!

References

- E Scheurwegs, M Sushil, S Tulkens, W Daelemans, and K Luyckx. Counting trees in random forests: predicting symptom severity in psychiatric intake reports. Journal of Biomedical Informatics, 75: S112-S119, 2017.
- M Sushil, S Šuster, K Luyckx, and W Daelemans. Patient representation learning and interpretable evaluation using clinical notes. Journal of Biomedical Informatics, 84:103 – 113, 2018.
- M Sushil, S Šuster, and W Daelemans. Rule induction for global explanation of trained models. Workshop on Analyzing and interpreting neural networks for NLP (BlackboxNLP), EMNLP 2018
- M Sushil, S Šuster, and W Daelemans. Distilling neural networks into skipgram level decision lists. Computing Research Repository, 2005.07111, 2020.