MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Applying and Improving a Publicly Available Medication NER Pipeline in a Clinical Cancer EMR

Dr Meg Stevens

Student,

The University of New South Wales (UNSW)

Background and Objectives

Larger Project (CaVa)

- Oral systemic therapy usage not captured in current database
- Aim to reliably extract oral systemic therapy usage from clinical notes using NLP
- Avoids the need for linkage to administrative dispensing data to capture this information

Current Study

- Identify all medication information targets in clinical notes using NLP
- Not specific to cancer therapies or oral agents
- Foundation towards developing the NLP infrastructure for the larger project

Clinical NLP Advancements

Pre-trained language models on biomedical and/or clinical text increasingly available

- less so pre-trained complete pipelines for specific clinical NLP tasks

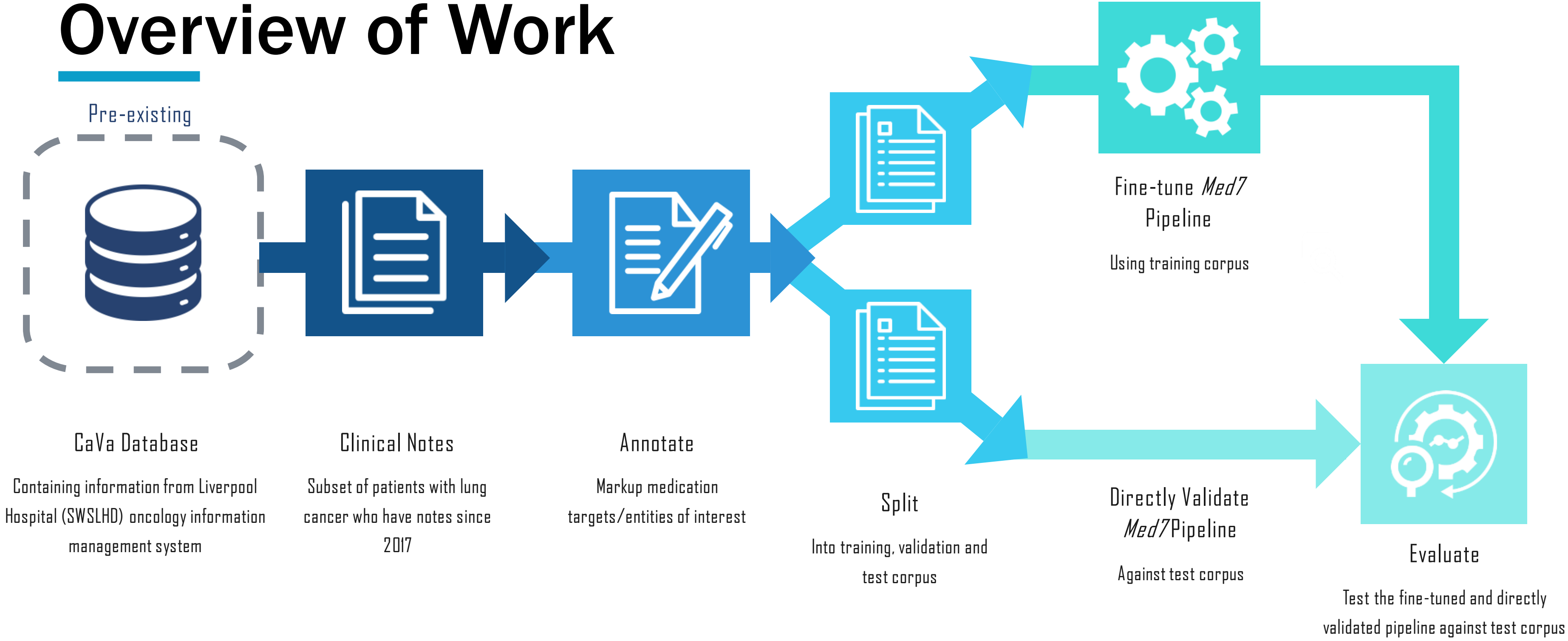
Improved success of transfer learning with Transformer architecture

Downstream tasks (e.g. NER, classification) still require annotated text

We examined *Med7* - one of the few complete publicly available, pre-trained pipelines (with RoBERTa and NER weights) for medication targets in clinical text, with and without further fine-tuning on a small annotated corpus.

Pre-training and fine-tuning performed with the open-source Python library spaCy.

Overview of Work



Annotation

5% Pilot
Using annotation guideline

Refine Guideline
Understand points of confusion

Create 'Gold' Annotations
Annotate full sample using refined guideline *(approx. 20 person hours for 3150 notes)*

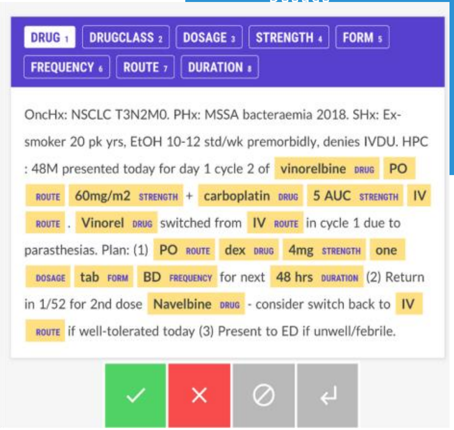
Review Conflicts
Understand systematic differences among annotators

Calculate Agreement
Quantify consistency by comparing overlapping samples

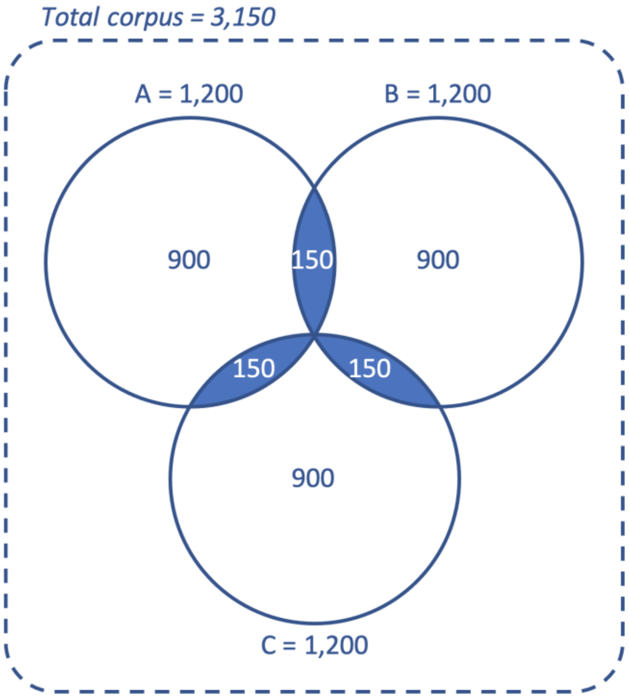
Entities

Drug

Drug Information
- Drug Class
- Strength
- Dosage

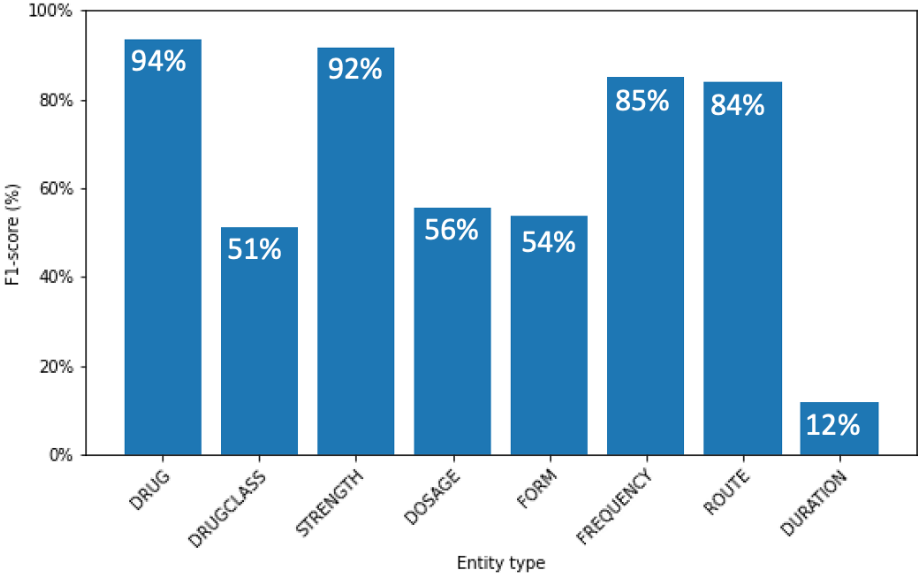


*Fabricated text

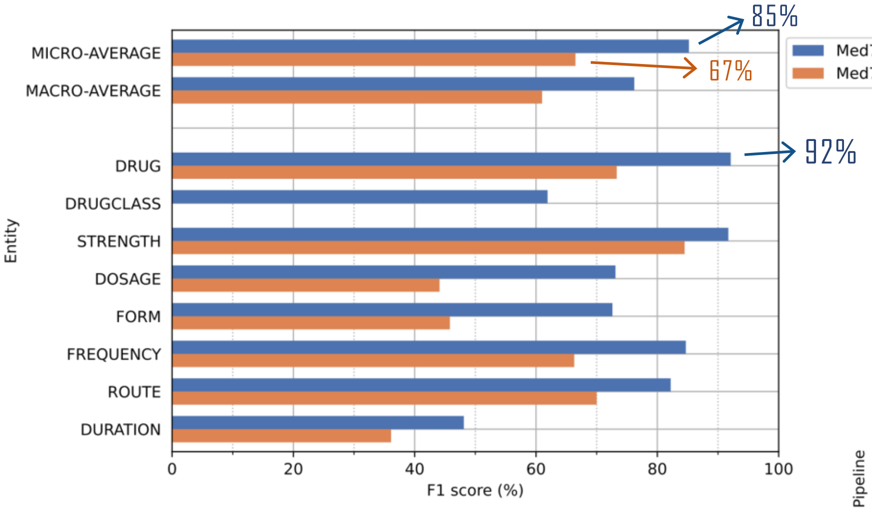


Mean F1-score for all pairs of overlapping samples

- Overall: 82% (strict word and entity matches)
- By entity:



Results



Effects of fine tuning the downstream NER component

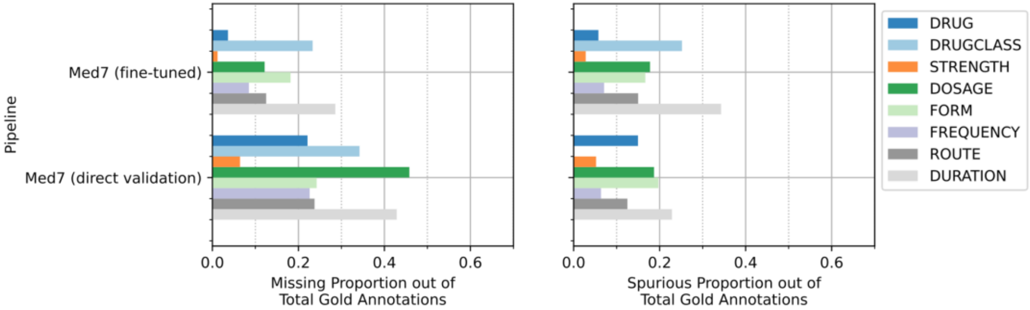
Superior performance per entity and overall

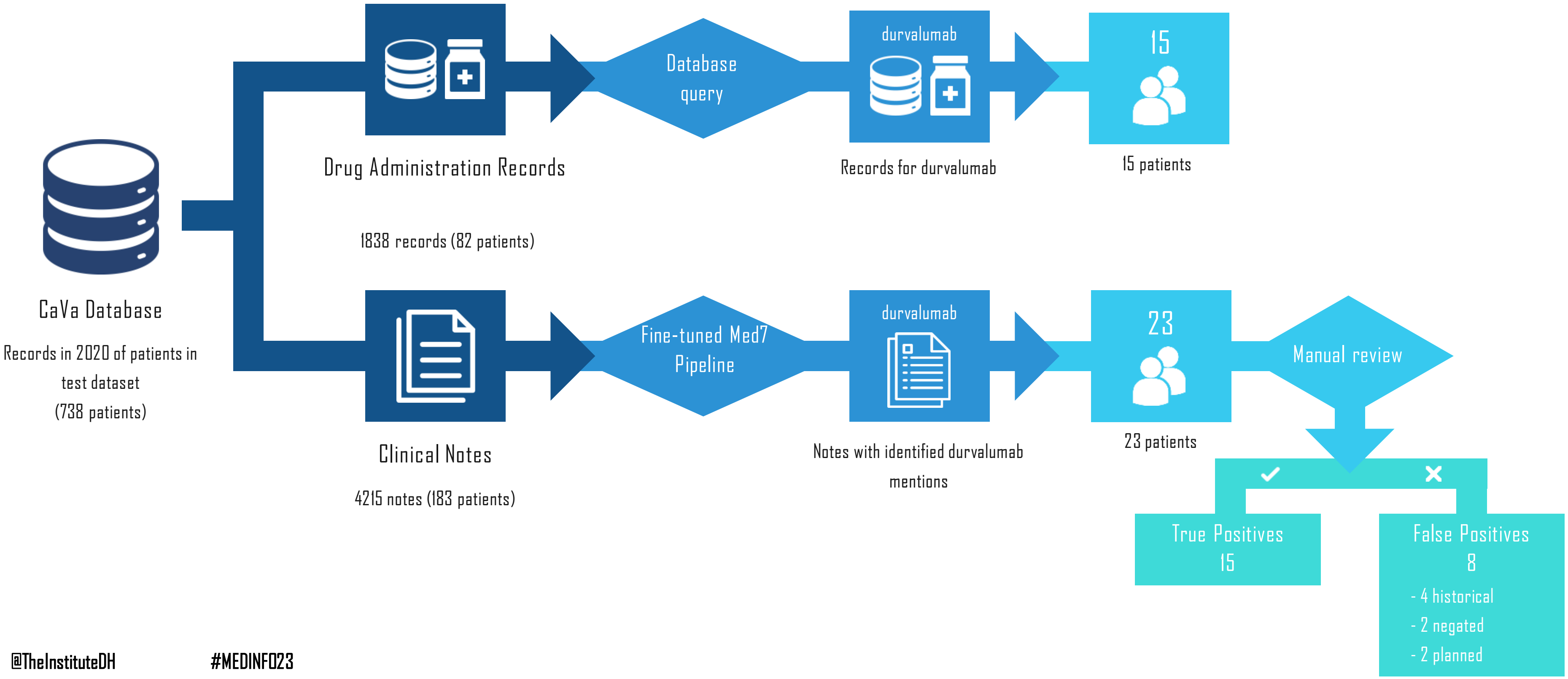
Smallest impact on performance for *strength*

Largest impact on performance for *dosage*

Large improvement on picking up false negative (missing) entities

Less improvement in reducing false positive (spurious) entities




CaVa Database
Records in 2020 of patients in test dataset (738 patients)
Drug Administration Records
1838 records (82 patients)
Database query
durvalumab
Records for durvalumab
15
15 patients
Clinical Notes
4215 notes (183 patients)
Fine-tuned Med7 Pipeline
durvalumab
Notes with identified durvalumab mentions
23
23 patients
Manual review
True Positives
15
False Positives
8
- 4 historical
- 2 negated
- 2 planned

Key Takeaways

Similar training and target data are important for developing high quality clinical NLP solutions

High performance is achievable with low annotation budgets, especially if utilising pre-trained language models

Pre-trained pipelines for clinical targets are not ubiquitous. Pipeline components can be sourced from previous work in the same institution, particularly language models

NER performance was proportional to the quality and quantity of annotated data for that entity

Further dependency/context resolution is vital for clinical utility

Next Steps

Optimisation of Annotation

Learn from error analysis and optimise iteratively

Leverage other strategies e.g. weak supervision with labelling functions, active learning with 'human-in-the-loop' annotation

Dependency Resolution

Context of mention (planned/negated/considered/historical)

Linkage of annotated drug information to drug names

Patient-Level Validation

Against high-quality data e.g. linkage to administrative dispensing data