Downstream Task Performance of BERT Models Pre-Trained Using Automatically De-Identified Clinical Data

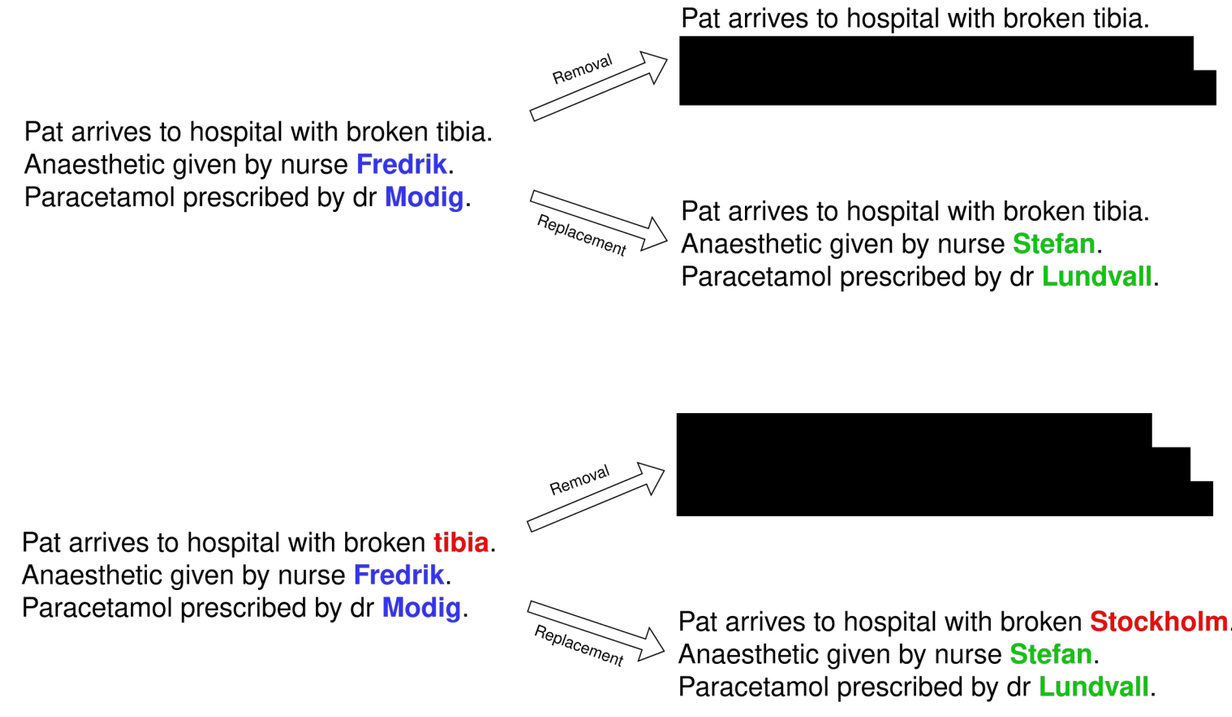
Thomas Vakili, Anastasios Lamproudis, Aron Henriksson & Hercules Dalianis
Department of Computer and Systems Sciences
Clinical Text Mining Group

Continued Pre-Training Using De-Identified Data

Language models for clinical tasks can be improved through **domain adaptation**. However, language models are susceptible to **privacy attacks** which may reveal **sensitive information** about persons in the data. This is especially problematic when training using inherently sensitive **electronic health records** (EHRs).

We suggest lowering the risks to privacy using **automatic de-identification**. This technique uses named entity recognition (NER) to detect sensitive entities such as names and locations. Due to **imperfect precision**, however, data is sometimes **corrupted**.

Previous studies have investigated the impact of this corruption on downstream tasks. This is the first paper examining the **impact of pre-training using de-identified data**.



Creating De-Identified Clinical BERT Models

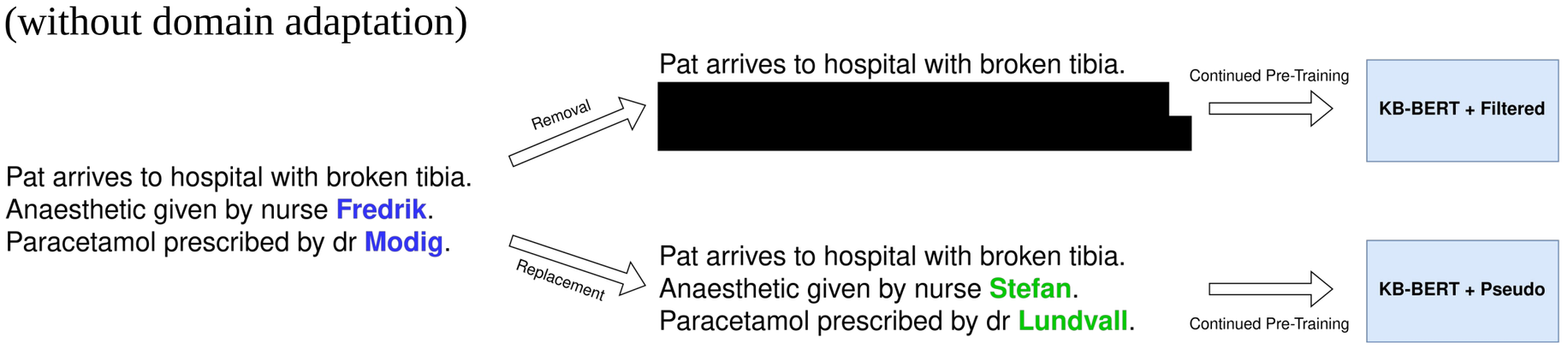
We automatically de-identify **17.9 GB of EHRs** from the *Health Bank* at Stockholm University. These data are used to continuously pre-train **two de-identified models**:

- KB-BERT + Filtered
- KB-BERT + Psuedo

The two de-identified models are compared against evaluations of **two baselines**:

- KB-BERT + Real (trained on unaltered data)
- KB-BERT (without domain adaptation)

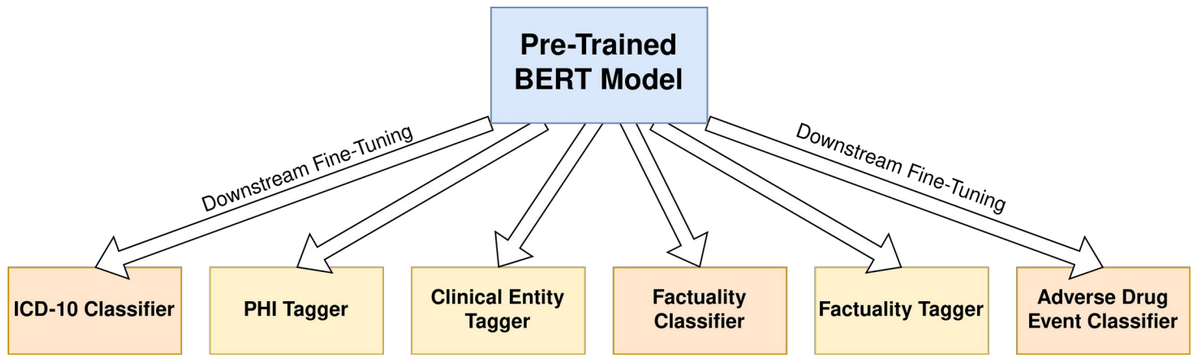
PHI Type	# Predicted Instances	NER Recall	NER Precision
Health Care Unit	19,659,127	80%	87%
Partial Date	19,374,711	83%	94%
Last Name	14,332,309	97%	96%
First Name	12,525,688	97%	98%
Full Date	10,459,935	55%	77%
Location	3,158,031	89%	85%
Age	2,064,111	35%	47%
Organisation	1,078,115	36%	71%
Phone Number	1,262,313	40%	63%



Downstream Performance Doesn't Deteriorate

The models are fine-tuned on **six clinical downstream tasks**. Results show **no noticeable deterioration from training using de-identified data** compared to training using unaltered data.

We will distribute *KB-BERT + Pseudo* as **SweDeClin-BERT** once we have received the necessary ethical permissions to do so.



Model	ICD-10	PHI	Clinical Entity	Factuality	Factuality	ADE
	Classification	NER	NER	Classification	NER	Classification
KB-BERT	0.799	0.91	0.803	0.635	0.630	0.183
KB-BERT + Real	**0.833**	**0.941**	0.858	0.732	0.682	**0.199**
KB-BERT + Filtered	**0.833**	0.929	0.854	0.731	0.672	**0.199**
KB-BERT + Pseudo	0.832	**0.941**	**0.861**	**0.736**	**0.684**	0.191

thomas.vakili@dsv.su.se