The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

332.THROMBOSIS AND ANTICOAGULATION: CLINICAL AND EPIDEMIOLOGICAL

A Transformer Natural Language Processing Algorithm for Cancer Associated Thrombosis Phenotype

Arash Maghsoudi [1], Emily Zhou [2], Danielle Guffey [3], Shengling Ma, MD [4], Xiangjun Xiao [5], Bo Peng [6], Christopher I Amos [7], Abiodun O. Ouyomi [8], Javad Razjouyan [6], Ang Li, MDMS [9]

[1] Health Services Research Department of Medicine, Baylor College of Medicine, Houston, TX
[2] McGovern Medical School, University of Texas Health Science Center, Houston, TX
[3] Institute for Clinical & Translational Research, Baylor College of Medicine, Houston, TX
[4] Section of Hematology-Oncology, Baylor College of Medicine, HOUSTON, TX
[5] baylor college of medicine, houston
[6] baylor college of medicine, houston, TX
[7] Baylor College of Medicine, Houston, TX
[8] Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX
[9] Section of Hematology-Oncology, Department of Medicine, Baylor College of Medicine, Houston, TX

Introduction: Cancer patients have high risk of venous thromboembolism (VTE). The ability to accurately capture VTE outcome in large electronic health record (EHR) is critical for both surveillance and risk stratification. We previously published a rule-based natural language processing (NLP) algorithm that relied on radiology reports alone with a precision (positive predictive value) of 89% and recall (sensitivity) of 69-78% (sTable 5 PMID36626707). The low recall metric was expected as VTE diagnosis that occur at outside hospital would not be captured by radiology reports. In the current study, we aimed to assess the performance of a machine learning NLP algorithm using both radiology and medical notes to detect cancer associated VTE longitudinally in a large cancer cohort.

Methods: VTE was defined as symptomatic or incidental pulmonary embolism, lower or upper extremity deep vein thrombosis. Unusual thrombosis such as splanchnic vein thrombosis (mostly tumor thrombi) was classified as negative. Figure 1 details the study design: i) patient selection, ii) pre-annotation filtering, iii) gold standard annotation, iv) transformer model development. From a cohort of 9,769 patients with active cancer receiving first-line systemic therapy, we first pre-filtered all clinic progress notes, discharge summaries, and radiology reports to exclude sections with little clinical relevance (e.g. review of systems, physical exam), while maintaining sections with high clinical value (e.g. history of present illness, assessment/plan, hospital problems/course). Then, we kept notes with VTE keyword stems (e.g. VTE, PE, DVT, thromb, embol, clot, filling defect, phlebitis).

We selected 808 patients with at least 1 keyword to determine the gold standard VTE date and location through independent chart review process. This cohort was randomly split into a training and testing cohort at an 7:1 ratio. In the training cohort (n=708), we further annotated all keyword-containing phrases within 30 days (d) from VTE date and from index date until 180d if no VTE (n=~2,000). The ~800 positive and ~1,200 negative keyword-containing phrases were used to train *ClinicalBERT* (bidirectional encoder representation transformer) large language model through split-sample cross validation to finetune the loss function and to determine the prediction threshold. Finally, we applied the entire trained model package to every available notes/reports (n=~2300) in the test cohort (n=100) to assess if the first positively predicted date would correlate with the gold standard VTE date.

Results: The finetuned *ClinicalBERT* model achieved 98% precision and 98% recall on the patient level in the training cohort (n=708). When applied on the 100-patient test cohort, the model accurately predicted first VTE events in 50/54 patients (93% precision) and captured 50/54 of patients with VTE (93% recall sensitivity) (Table 1). Most of the tumor thrombi and superficial venous thrombi were correctly classified as negative. The false positives were related to unusual sentences with both positive and negative findings such as "Thrombosis surrounding the left brachial catheter in basilic vein," "CT chest with PE was done," "Lovenox was initiated for UE DVT, but this was port thrombosis in the basilic vein." The false negatives were related to isolated VTE diagnosis close to death or loss to follow-up. Notably, the precision and recall tradeoff can be further finetuned by changing the minimum number of positively predicted notes from the NLP.

Conclusion: In conclusion, we successfully developed and tested a transformer NLP model to detect cancer associated VTE longitudinally using a combination of clinical notes and radiology reports that achieved a precision of 93% and recall of 93%. This represented an improvement over our original algorithm with radiology report alone. Further validation of the model in external cohorts and finetuning through federated or transferred learning are ongoing to ensure model generalizability and usability.

Disclosures No relevant conflicts of interest to declare.

Figure 1. NLP transformer model development for cancer associated VTE
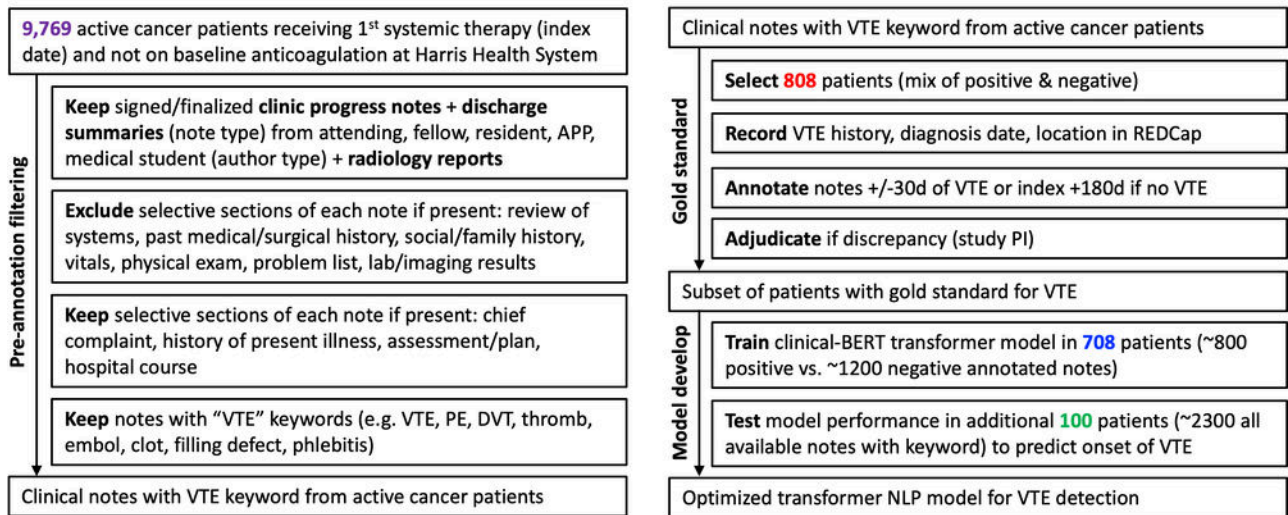
Table 1. Performance of NLP transformer model in training and testing cohorts

Train N=708	VTE + (n=373)	VTE − (n=335)		Test N=100	VTE + (n=54)	VTE − (n=46)	
NLP + (n=374)	366	8	PPV=98%	NLP + (n=54)	50*	4	PPV=93%
NLP − (n=334)	7	327	NPV=98%	NLP − (n=46)	4	42	NPV=91%
	Sn=98%	Sp=98%			Sn=93%	Sp=91%	

PPV = positive predictive value (**precision**), NPV = negative predictive value, Sn = sensitivity (**recall**), Sp = specificity
A patient is considered positive in test cohort if 2+ notes have been predicted + by NLP
* 4 patients in the BERT+ VTE+ stratum had VTE events >30d from true diagnosis date

Figure 1

https://doi.org/10.1182/blood-2023-184756