Exploring the Generalization of Cancer Clinical Trials Eligibility Classifiers Across Diseases

Yumeng Yang, MS, Ashely Gilliam, MS, Ethan B Ludmir, MD, Kirk Roberts, PhD

May 3rd, 2024

Background

- Clinical trials have been recognized as the gold standard method for determining the safety and efficacy of new treatments
- Finding eligible trials/patients is time-consuming and labor-intensive
- Eligibility Criteria: Describe characteristics that must be shared by all participants
  - Inclusion - Must meet
  - Exclusion - Must **not** meet
- Natural Language Processing (NLP) techniques can help automate screening and criteria matching process

Previous Study

- Develop classifiers to identify common criteria in phase 3 cancer trials
- Fine-tuned on 6 domain-specific BERT-Based models
  - BioBERT
  - BlueBERT
  - ClinicalBERT
  - PubMedBERT
  - SciBERT
  - **ClincalTrialBERT**
- Pretrain domain-specific BERT-based model ClinicalTrialBERT
  - Using 442,370 eligibility criteria sections from ClinicalTrials.gov
- Evaluate results both on trial level and individual criterion level

Common Criteria

- Prior Cancer
- HIV
- Psychiatric Illness
- Hepatitis B Virus (HBV)
- Hepatitis C Virus (HCV)
- Autoimmune Disease (Auto)
- Substance Abuse (Subst)

Previous Study

764 Phase 3 cancer trials (PROTECTOR1)

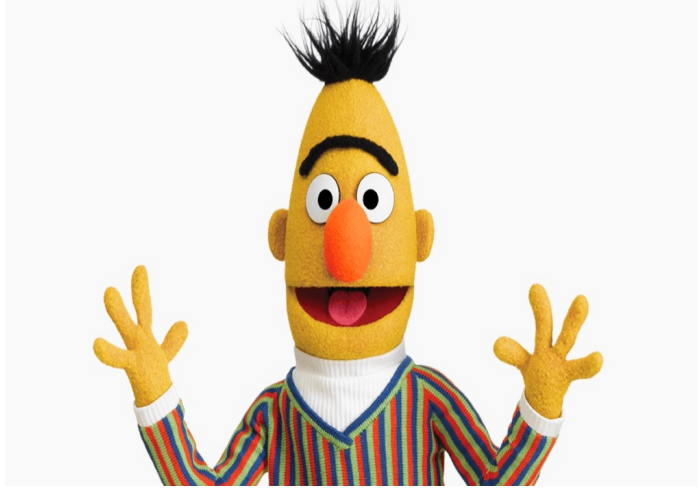
Build keywords list for each criteria

Extract each criterion Annotation

BERT

Generalizability

Exclusion: Prior history of **malignancy** in the past 5 years with the exception of basal cell and squamous cell carcinoma of the skin



Yes

The trial excludes this criteria

No

The trial does not exclude this criteria

Cohorts Selection

- Other phase 3 cancer trials that are not in PROTECTOR1
- Phase 1 and 2 cancer trials
- Heart disease trials
- Type 2 diabetes trials
- Observational trials

Cohorts Retrieval

- Download all trials' XML files from ClinicalTrials.gov
- Build keywords list for each cohort
- Screening condition, phase and intervention sections through all XML files

Data and Annotation

- Randomly select 50 trials from each cohort

- Refined previous keywords list, then match individual criterion with these keywords

	Prior Cancer	HIV	HBV	HCV	Psych	Subst	Auto
Sample Size	523	294	302	307	323	467	274
Cohen's κ	0.92	0.94	0.73	0.88	0.87	0.97	0.85
Agreement Accuracy	0.97	0.97	0.95	0.96	0.94	0.99	0.97

Experiments Design

- Directly make classification with previous fine-tuned ClinicalBERT and ClinicalTrialBERT

- Few-shot transfer learning
  - randomly chose samples in sets of 5, 10, and 15. However, the sets of 5 and 10 are both included in the set of 15 samples.

Results

Criteria	Cohort	ClinicalBERT		ClinicalTrialBERT	
		Criterion	Trial	Criterion	Trial
Prior Malignancy	Phase 3 cancer	0.84	0.91	0.71	0.90
	Phase 1,2 cancer	0.91	0.94	0.69	0.86
	Heart disease	0.75	0.75	0.60	0.61
	Type 2 diabetes	0.76	0.76	0.69	0.72
	Observational	0.50	0.50	0.41	0.48
HIV	Phase 3 cancer	0.92	0.94	0.93	0.95
	Phase 1,2 cancer	0.89	0.91	0.89	0.92
	Heart disease	0.77	0.85	0.79	0.85
	Type 2 diabetes	0.94	0.97	0.94	0.94
	Observational	0.60	0.75	0.62	0.78

Few-shot Result

Criteria	Cohorts	+%	-PROTECTOR1			+ PROTECTOR1		
			5	10	15	5	10	15
Prior Malignancy	Phase 3 cancer	0.16	0.90	0.90	0.90	0.92	0.86	0.84
	Phase 1,2 cancer	0.18	0.93	0.94	0.92	0.89	0.91	0.91
	Heart disease	0.28	0.82	0.69	0.69	0.79	0.72	0.74
	Type 2 diabetes	0.47	0.78	0.71	0.80	0.86	0.80	0.81
	Observational	0.09	0.50	0.50	0.60	0.63	0.42	0.50
HIV	Phase 3 cancer	0.78	0.93	0.94	0.96	0.93	0.96	0.97
	Phase 1,2 cancer	0.79	0.92	0.91	0.91	0.89	0.89	0.91
	Heart disease	0.69	0.77	0.76	0.79	0.84	0.79	0.77
	Type 2 diabetes	0.92	0.94	0.94	0.94	0.94	0.94	0.94
	Observational	0.41	0.70	0.77	0.78	0.77	0.78	0.81

+%: Keywords matching result
- PROTECTOR: Without further training on the entire PROTECOTR1 data
+ PROTECTOR: With further training on the entire PROTECOTR1 data

Conclusions

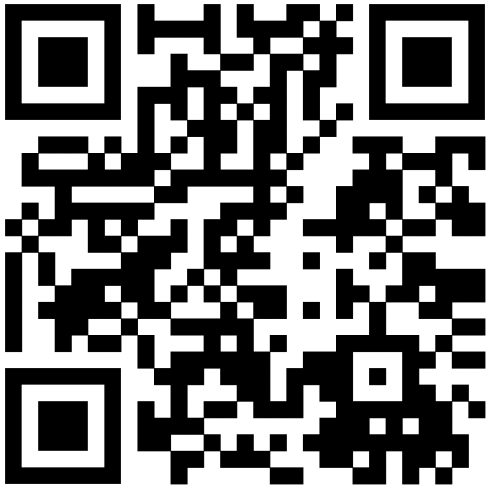
- Emphasizes the need for generalizable eligibility criteria classifiers
- Shows models' robust generalizability to other phase 3 cancer trials
- Highlights the strength of cancer-trained models for disease with similar eligibility requirements
- Notes significant variations in eligibility criteria for different types of trials, like observational trials

Future Work

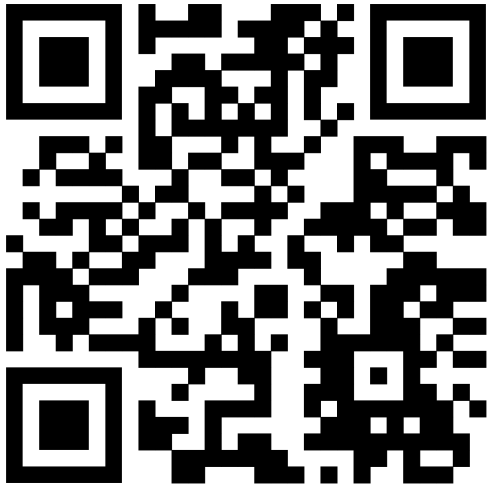
- Cluster criteria for various disease types and phases
- Address class imbalance within training data for specific diseases
- Focus on specific diseases, considering differences based on cancer types, stages, and treatment modalities

Thanks for listening!!

Feel free to reach out if you have any questions at:
yumeng.yang@uth.tmc.edu



Text Classification of Cancer Clinical Trial Eligibility Criteria



Exploring the Generalization of Cancer Clinical Trial Eligibility Classifiers Across Diseases

Keywords list

NCT04676412

Exclusion Criteria:

- Has known untreated central nervous system metastases and/or carcinomatous meningitis
- Has a known history of an additional malignancy, except if the participant has undergone potentially curative therapy with no evidence of that disease recurrence for ≥3 years since initiation of that therapy (Note: The time requirement does not apply to participants who underwent successful definitive resection of basal cell carcinoma of the skin, superficial bladder cancer, squamous cell carcinoma of the skin, in situ cervical cancer, or other in situ cancers.)
- Has radiographic evidence of encasement or invasion of a major blood vessel, or of intratumoral cavitation
- Has an active autoimmune disease that has required systemic treatment in the past 2 years Replacement therapy (e.g. thyroxine, insulin, or physiologic corticosteroid replacement therapy for adrenal or pituitary insufficiency, etc.) is not considered a form of systemic treatment and is allowed
- Has had an allogeneic tissue/solid organ transplant
- Has a known history of human immunodeficiency virus (HIV) infection
- Has a history of (noninfectious) pneumonitis that required systemic steroids or current pneumonitis/interstitial lung disease
- Has a known history of hepatitis B or known active hepatitis C virus infection
- Has a history of a gastrointestinal condition or procedure that in the opinion of the investigator may affect oral study drug absorption

Has a known history of an additional **malignancy**, except if the participant has undergone potentially curative therapy with no evidence of that disease recurrence for ≥3 years since initiation of that therapy (Note: The time requirement does not apply to participants who underwent successful definitive resection of basal cell carcinoma of the skin, superficial bladder cancer, squamous cell carcinoma of the skin, in situ cervical cancer, or other in situ cancers.)