



Big Data Approaches for Conducting Clinical Research: Big Data Analysis Process

심승혁
건국대병원 산부인과

DECLARATION OF INTERESTS

Nothing to declare

부인종양학 연구에서 시도해 볼 만한 주제

* 자궁경부암 환자 추적: 청구자료에서 HPV 검진, 치료 기록, 재발 코드 통합

현재: 고정된 추적 기간 할당

개선: 각 환자의 실제 치료 시작일, 종료일 동적 추정

* 화학요법 독성 식별: 진료 기록과 청구 코드 결합

예: "신독성" 진단 코드 + 신장기능검사 결과 + 항암제 투여 시기 → 인과관계 추정

* 재발 추적: 초회 치료일부터 재발 진단까지의 정확한 interval 계산

문제: 청구자료만으로는 재발 시점 모호 (수술? 화학요법? 추적관찰 중?)

예: 진단 코드 + 수술 코드 + 약물 코드 계층적 해석

Original Investigation | Health Informatics

Assessment of Natural Language Processing of Electronic Health Records to Measure Goals-of-Care Discussions as a Clinical Trial Outcome

Robert Y. Lee, MD, MS^{1,2}; Erin K. Kross, MD^{1,2}; Janaki Torrence, MS^{1,2}; et al

» Author Affiliations | Article Information

RELATED ARTICLES FIGURES SUPPLEMENTAL CONTENT

Close

View Large Download

Figure 1. Training, Prediction, and Validation of the Natural Language Processing (NLP) Model

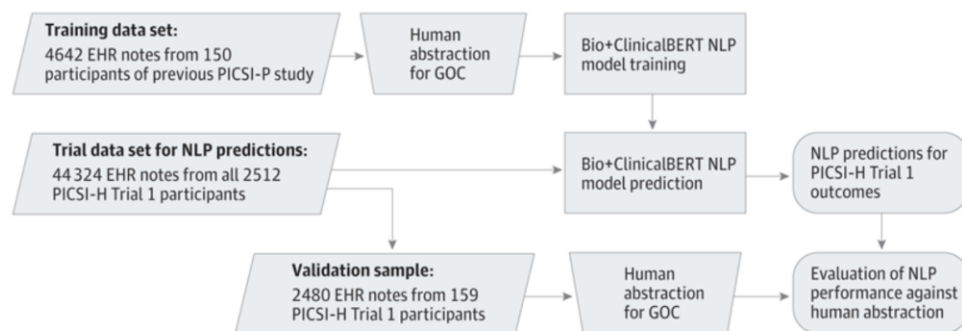
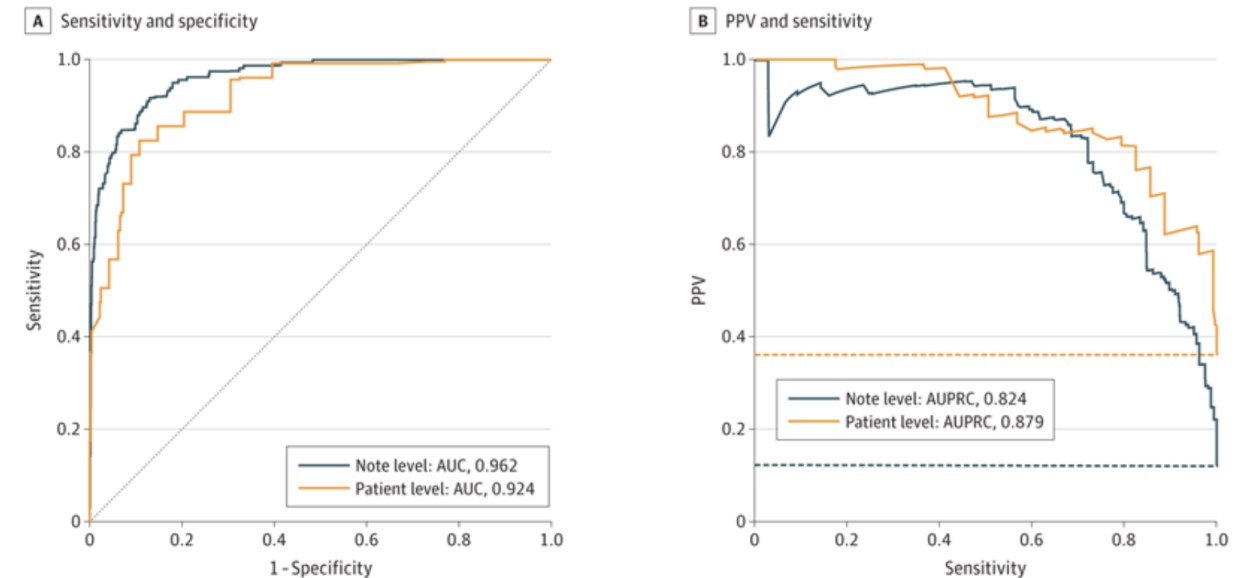


Figure 2. Performance of BERT Natural Language Processing in Classifying 30-Day Documented Goals-of-Care Discussions at Note and Patient Levels in a 159-Patient, 2480-Note Internal Validation Sample



Go to Figure in Article

Representative values of sensitivity, specificity, positive and negative predictive values, and F1 scores from these curves are presented in the Table. B, Dotted lines indicate nondiscriminating classifiers. AUC indicates area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; and PPV, positive predictive value.

Session Title

Q1.

LLM 방식을 사용하면 이런 복잡한 규칙 정의가 필요 없지 않을까?

가능하다면 현재 LLM을 이용한 연구에서 모델의 성능과 구체적 사례는?