AI-Powered NLP for Comprehensive IBD and CKD Phenotyping Entity and Relationship Extraction of Clinical Features and Biomarkers from EHR

Shashi Khan, Vikash Verma, Louis Brooks Jr, Marissa Seligman, Abhimanyu Roy, Abhinav Nayyar, Ankit Arora, Mahainn Somani, Prathyusha Joshi, Shruti Chaudhary, Ankita Gupta, Riddhi Markan, Anuj Gupta, Vishan Khatavkar, Sudhanshu Chawla, Ram Mishra

Background

Electronic Health Records (EHRs) are largely unstructured, with critical phenotypic information for chronic diseases such as Inflammatory Bowel Disease (IBD) and Chronic Kidney Disease (CKD) embedded in clinical notes. Deep learning–based natural language processing (NLP) enables scalable feature extraction to support robust real-world evidence generation.

Objective

To develop and validate an AI-based NLP pipeline that integrates Named Entity Recognition (NER) and relationship extraction to automate phenotyping across IBD and CKD by capturing clinically relevant entities and their contextual relationships.

Methodology

Data Source: Unstructured clinical notes from EHR databases for patients with (IBD) and (CKD).

Study Design: Retrospective observational analyses were conducted using large-scale electronic health record (EHR) corpora.

Study Period: IBD cohort: January 2007 – December 2024

CKD cohort: January 2014 – December 2025. The study included patients aged 18 years or older.

NLP Model Development: Deep-learning–based Named Entity Recognition (NER) models using a Character-level Convolutional Neural Network with Bidirectional Long Short-Term Memory and Conditional Random Field layers (CharCNN-biLSTM-CRF) were developed to extract diagnoses, disease characteristics and events, disease severity, procedures, and key biomarkers including C-reactive protein (CRP), fecal calprotectin, fecal lactoferrin.

Relationship Extraction: A relationship extraction module was applied to link related clinical entities, such as diagnosis with severity, and biomarkers with corresponding test results.

Model Evaluation: NER performance was evaluated using precision, recall, and F1-score, reviewed against manually annotated notes.

Procedure: These features were identified across multiple diagnostic categories, including colonoscopy, endoscopy, CT imaging, biopsy, and Magnetic Resonance Enterography (MRE).

Procedure Findings: These included inflammation (localized or generalized), skip lesions, polyps or pseudopolyps, cobblestone appearance, strictures, and fistulas.

Disease Event: Disease events were classified as active or inactive and IBD flares were captured.

Microscopic features: This included glomerulosclerosis, nephrosclerosis, interstitial fibrosis, tubular atrophy, and nephrocalcinosis.

Renal Biopsy: This included kidney biopsy via electron microscopy, light microscopy, and immunofluorescence microscopy.

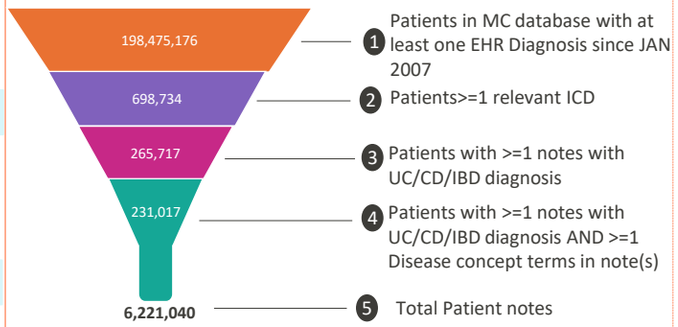

Figure 1. IBD Cohort

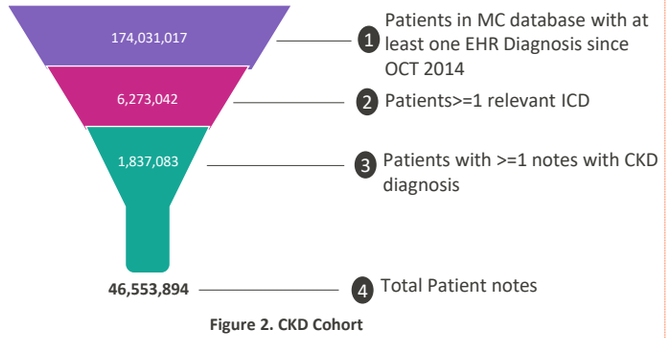

Figure 2. CKD Cohort

Results

- The validated NLP pipeline was applied to over 6.2 million clinical notes for IBD. In the IBD cohort, the Named Entity Recognition (NER) model demonstrated strong performance across clinical concepts. The model achieved F1-scores of 86% for diagnosis, 81% for procedure findings, and 76% for procedure. Disease-related event and severity extraction demonstrated strong performance, achieving F1-scores of 79% for disease events and 80% for disease severity. Biomarker extraction performed robustly, with F1-scores of 82% for biomarkers and 83% for biomarker results.
- The validated NLP pipeline was applied to over 46 million clinical notes for CKD. Extraction of feature severity showed excellent performance, with an F1-score of 96% and Renal Biopsy with an F1-score of 82%. Microscopic feature extraction performed robustly, achieving an F1-score of 88% and diagnosis with an F1-score of 95%.

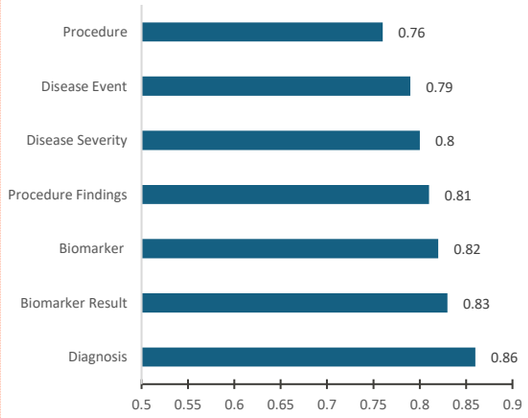

Figure 3. IBD - F1 Score

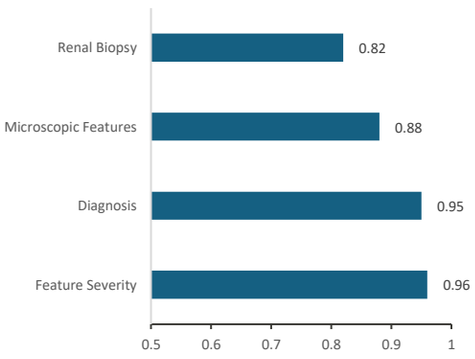

Figure 4. CKD - F1 Score

Conclusions

- This study demonstrates that a deep-learning–based NLP pipeline can accurately and efficiently extract clinically meaningful information from unstructured EHR data across IBD and CKD subtypes.
- The validated NER models achieved strong performance across diagnoses, disease characteristics, events, severity, and biomarkers, renal biopsy, microscopic features, thus, enabling reliable large-scale phenotyping.
- Application of the pipeline to over fifty million across IBD and CKD clinical notes substantially reduced manual chart review burden while preserving clinical detail.
- By enabling scalable, context-aware extraction of real-world clinical data, this approach supports accelerated real-world evidence generation.
- This approach further supports downstream outcomes research and precision medicine applications in IBD and CKD.
- These results demonstrate the replicability of the NLP workflow across therapeutically distinct disease areas, with consistent entity extraction performance observed in both inflammatory bowel disease and chronic kidney disease.

IBD was defined using ICD-10 codes K50*, K51*, and K52.3, and CKD was identified using ICD-10 codes N18.1, N18.2, N18.30, N18.31, N18.32, N18.4, N18.5, N18.6, N18.9, E08.22, E09.22, E10.22, E10.2, E11.22, E11.2, E13.22, I12.9, I12.0, I13, I13.10, and I13.2

Optum