Recent Advances in AI-Based Order Recommender Systems for Electronic Health Records

Jonathan Chiang

INTRODUCTION

Recent breakthroughs in data-driven clinical decision support (CDS) underscore the potential of artificial intelligence (AI) to aid clinicians in selecting diagnostic tests, imaging studies, and treatments. As health systems grapple with rising patient complexity and clinician shortages, automated recommender systems can help streamline the process of identifying "next steps" or recommended orders within electronic health record (EHR) workflows. Here, we review the latest progress in EHR-based order recommender systems, focusing on emerging graph neural network (GNN) strategies and large language model (LLM)-enabled interpretability.

RECENT PROGRESS IN CLINICAL ORDER RECOMMENDER SYSTEMS

Early Collaborative Filtering Methods

Initial data-driven approaches employed collaborative filtering or association rules to identify correlations among frequently ordered items. For instance, OrderRex leveraged historical EHR patterns to generate procedure suggestions, demonstrating early proof-of-concept for real-world adoption without adversely affecting care efficiency 1,21,2. Subsequent refinements incorporated multi-layer neural networks, matrix factorization, and autoencoders to address the clinical rationale ("who, what, and why") behind proposed orders 33.

Shift to Graph Neural Networks (GNNs)

Traditional methods often face difficulties in assimilating varied EHR data—such as diagnoses, lab results, clinical notes, and imaging metadata—or in managing "missing labels" (i.e., tests never ordered because they were deemed low-yield) 44. GNN-based models were devised to overcome these challenges by representing patients, orders, and clinical findings as nodes within heterogeneous graphs. Leveraging link prediction, these networks predict subsequent orders with greater robustness to incomplete or missing data 55. In addition, weighting or selection strategies within GNNs enable finer handling of non-ordered tests, preserving model fidelity in complex clinical environments.

Fouladvand et al. Example

A representative GNN-based system by Fouladvand and colleagues 66 encoded pre-referral data for endocrinology and hematology patients in a heterogeneous graph. Compared with collaborative filtering and conventional neural networks, their approach yielded gains of up to 8% in area under the receiver operating characteristic curve (ROC-AUC) for endocrinology referral orders. Notably, it also surpassed human-constructed checklists in both precision and recall. Recent work suggests that coupling these GNN frameworks with large language models can further enhance interpretability, generating guideline-referenced (e.g., UpToDate) explanations for suggested orders 2,52,5.

KEY THEMES: LABEL SELECTION, FEEDBACK LOOPS, AND LARGE LANGUAGE MODELS

Label Selection / Missingness

One persistent issue is that when a recommended order is not carried out, direct outcome labels become unavailable. Studies now employ partial randomization or inverse probability weighting to more accurately capture performance and mitigate bias 44. These methods keep models calibrated, even as the system modifies the underlying data environment.

LLM-Aided Interpretability

The advent of large language models, particularly those fine-tuned on medical corpora, offers a new avenue for explaining GNN outputs. Specifically, LLMs can translate "black-box" predictions into concise bullet points referencing established guidelines 2,52,5. Such transparency has been shown to bolster clinician trust and drive greater adoption of recommender systems.

CONCLUSION AND OUTLOOK

The evolution from purely correlation-based order recommendations to graph-enhanced, LLM-assisted clinical decision support aligns with a broad trend toward more robust and transparent AI solutions in health care. Emerging research highlights three key imperatives for future development: (1) managing partial label missingness, (2) delivering user-friendly explanations via LLM-driven summaries, and (3) embedding real-time feedback to learn from user acceptance and changing evidence.

By leveraging structured EHR data alongside unstructured clinical notes and authoritative knowledge corpora, future recommender systems can provide clinically relevant and guideline-concordant suggestions that adapt dynamically to new information. The synergy between graph neural networks, meticulous label-handling strategies, and large language models represents a promising frontier for enhancing both patient outcomes and clinician efficiency.

REFERENCES

1. Chen JH, Podchiyska T, Altman RB. OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records. *J Am Med Inform Assoc.* 2016;23(2):339-348.
2. Kumar A, Aikens RC, Hom J, et al. OrderRex clinical user testing: a randomized trial of recommender system decision support on simulated cases. *J Am Med Inform Assoc.* 2020;27(12):1850-1859.
3. Noshad M, Jankovic I, Chen JH. Clinical recommender algorithms to simulate digital specialty consultations. *Pac Symp Biocomput.* 2022:290-300.
4. Hunter-Zinck HS, Peck JS, Strout TD, Gaehde SA. Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay. *J Am Med Inform Assoc.* 2019;26(12):1427-1436.

5. Li R, Wang JK, Sharp C, Chen JH. When order sets do not align with clinician workflow: assessing practice patterns in the electronic health record. *BMJ Qual Saf.* 2019;28(12):987-996.
6. Fouladvand S, Reyes Gomez F, Nilforoshan H, et al. Graph-based clinical recommender: predicting specialists procedure orders using graph representation learning. *J Biomed Inform.* 2023;143:104407.