Fulfilling the promise of Large Language Models: Getting patients the care they need

Christopher Pettengell, BMBCh,[1] Aaron Leibtag, Bcom,[1] Steven Aviv, BBusSci,[1] Viktoriia Mokriak, MA,[1] Aaron Li, BSc,[1]

[1]*Pentavere Research Group Inc., Toronto, ON, Canada*

Presented at AMIA 2024 Clinical Informatics Conference | Minneapolis, MN | May 21–23, 2024 **P18**

BACKGROUND

- Medical error is the 3rd leading cause of death in the United States and in heart failure alone, up to 90% of patients are not receiving foundational therapies according to guidelines[1].
- Limitations in current Electronic Health Records (EHR) systems impair clinicians' ability to unlock value seamlessly and automatically from millions of data points buried in clinical text.
- If medication data are structured, it is often incomplete or out of date. More often they are buried in free-text clinical documentation which can include historical and contemporaneous dictation with varying levels of detail or even conflicts.
- Large language models (LLM) have shown impressive capabilities, with Med-PaLM 2 recently demonstrating expert performance on USMLE-type exam[2], but concerns about their application in a clinical setting remain, in part due to challenges resolving context across multiple clinical documents and hallucinations.
- This research aims to develop an approach to extract accurate medication data at scale and identify gaps in care.

METHODOLOGY

- Employed a two-step process using a LLM (model-L) to summarize and a BERT-based sequence labeling model (model-B) to extract medication information from clinic notes.
- Model-L utilized LoRA with prefix tuning to ensure that it focused on relevant information, resolving conflicts and summarizing evidence from source documents.
- Model-B was trained to extract information from the results obtained from model-L. We leveraged existing manually labeled dermatology data for training to produce sentence-level predictions for the drug as well as its associated status and relations (e.g. dosage, frequency). The final output of model-B is structured data with all brand names normalized to the drug.
- The models were then deployed and assessed for accurate extraction of medications and their doses from patient notes to identify patients not receiving foundational therapies (**Figure 1**).

RESULTS

- ~300 outpatient dermatology patient records were included in the model-L training dataset. These records were unstructured and included multiple different note types (e.g., clinical notes, letters, and pathology reports).
- Model-L provided a compact summary of the status of [Drug A] and conclusion that patient is on [Drug A] resolving longer context conflicts. (**Table 1**).
- The output of Model-L alone without model-B achieved an **F1 0.77**. This was in part due to hallucinations. (**Table 2**).
- The model-B dataset consists of 4,433 clinical text samples, of which 3,310 were used for training and 1,123 for testing.
- Subsequent processing through model-B was able to then extract the specific medication data including dose, for example: "[DRUG A], 40mg, OD."
- The final output of the dual-model approach with model-L and model-B is structured data. (**Table 3**).
- Preliminary comparison to manual review showed the output achieved **F1 ≥0.95**.

Figure 1. **Technical workflow for models.**

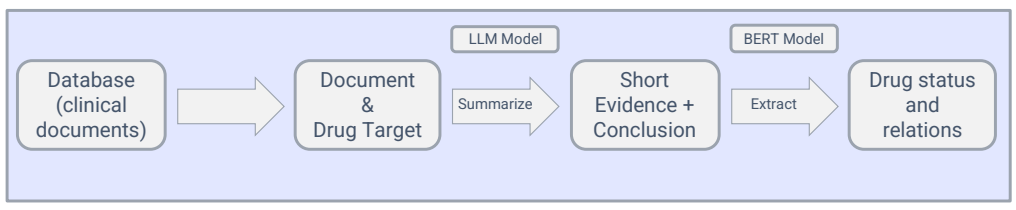

Table 1.

Example of input text[a]	Example Outputs of Model-L
Imp: Mild psoriasis and Crohn's on [Drug A] with fair control, OA of knees…. Plan: Start with 20mg [Drug A], continue last week with 40 mg daily Will send back if need IA cortisone or physio…. Therapy: [Drug B] 50mg twice weekly- switch to [Drug A] and stop 50mg treatment….	**Evidence:** Start with 20 mg [DRUG A], continue last week with 40 mg daily. **Conclusion:** patient is currently on [DRUG A].
Impression: severe psoriasis, edema bilateral lower legs with some eroded areas Plan: continue [Drug A] 90 mg q 8 wks, start Unna's boot for lower leg edema/ erosions	**Evidence:** currently on [Drug A] 90mg every 8 wks. **Conclusion:** patient is currently on [Drug A].

Table 2.

Example of input text[a]	Example Outputs of Model-L Hallucinations
F/U re: psoriasis - currently taking [Drug A] and [DRUG B] Impression: 1.psoriasis Plan: 1.will switch to [DRUG C] Rx given: F/U 2 mos enrollment form faxed into Spectrum	**Evidence:** Current usage of [Drug C]. **Conclusion:** patient is currently on [Drug C]. In the given text input, patient is on [DRUG A].Plan: F/U 2 mos. Imp: Psoriasis. Current therapy: [DRUG C].

Table 3. **Example of the final output after using model-B to extract information obtained from model-L .**

patient	drug	status	dose
1	ustekinumab	ON	45 mg q 12 wks
2	adalimumab	ON	40 mg q 1 wk
3	adalimumab	ON	40 mg q 2 wks
4	ustekinumab	ON	90 mg q 8 wks
5	guselkumab	ON	100 mg q 8 wks

[a] Some context may be hidden or replaced within the examples to protect patient protected health information.

CONCLUSIONS

- Extracting medication information across multiple unstructured notes is challenging, in part, due to context conflicts and hallucinations.
- This study has demonstrated that it is possible to use a dual-model approach of LLM and BERT-based models to enhance the robustness and accuracy of the medication extraction process.
- This data will empower clinicians to review treatment patterns and ensure that patients are getting the treatments they need to improve quality of care, equity, and patient experience.
- Next steps include further validation and trialing this approach in a new pan-Canadian study of heart failure patients to optimize care in the community and demonstrate the impact that LLM can have for patients.

References:
1. Huitema AA, Daoust A, Anderson K, Poon S, Virani S, White M, et al. Optimal usage of Sacubitril/Valsartan for the treatment of heart failure: The Importance of Optimizing Heart Failure Care in Canada. CJC Open. 2020;2(5):321–7. doi:10.1016/j.cjco.2020.03.015
2. Singhal K, Tu T, Gottweis J, Sayres R, et al. Towards Expert-Level Medical Question Answering with Large Language Models. ArXiv. 2023. abs/2305.09617.