Machine Learning to Analyse and Reduce Healthcare Harm in New Zealand General Practice

Vithya Yogarajan[1], Sharon Leitch[2], David Reith[3] and Michael Witbrock[1]

[1] Strong AI Lab, School of Computer Science,The University of Auckland, New Zealand
[2] General Practice and Rural Health, Otago Medical School, University of Otago, Dunedin, New Zealand
[3] Office of the Dean, Otago Medical School, University of Otago, Dunedin, New Zealand

Background and Motivation

Definition: Patient Harm
Physical or emotional negative consequences to patients directly arising from health care, beyond the usual consequences of care and not attributable ONLY to the patient's health condition.

Includes: Treatment delays, inconvenience and additional financial costs.

An extensive retrospective review of New Zealand general practice (GP) electronic health records (EHRs) over a three-year period identified preventable patient safety issues [1]. However, this review required extensive time, expert knowledge and resources for collecting, screening and analysing data.

Objectives

This research is an independent study to the retrospective review [1]. However, it makes use of the New Zealand GP EHRs collected in the original study and employs machine learning to identify healthcare harm, harm severity and preventability.

It is a **feasibility study** to determine the viability and limitations of automating identification of healthcare harm, if any, and providing feedback to the GPs.

To develop a prototype to automatically detect patient harm from medical records.

Retrospective Review

New Zealand GP EHRs (de-identified) → Manually annotated by 8 GP researchers

Findings: Identified preventable patient safety issues. Required extensive time, expert knowledge & resources.

Can Artificial Intelligence help?

Current Feasibility Study

Use the data and knowledge from the above study to train, test and evaluate machine learning models.

Selection of a model.

→ Identify opportunities and limitations.

Expected Outcome: Developing a prototype to automatically identify patient harm from EHRs.
If potential harm is predicted, provide GP with feedback.

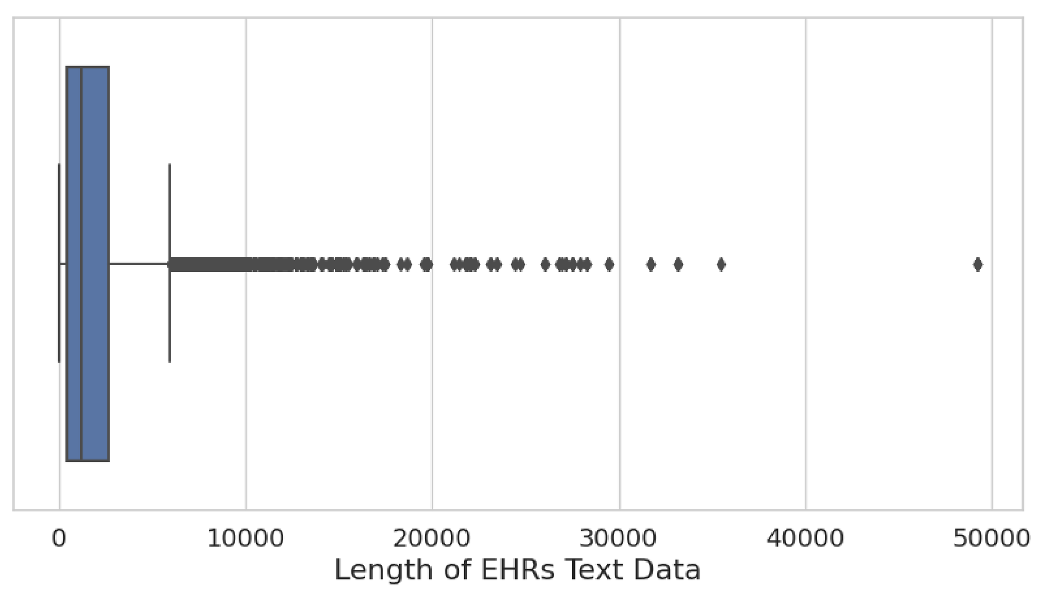
Final Decision: Made by GP using the AI feedback.

GP Records Review: Study Design and Data

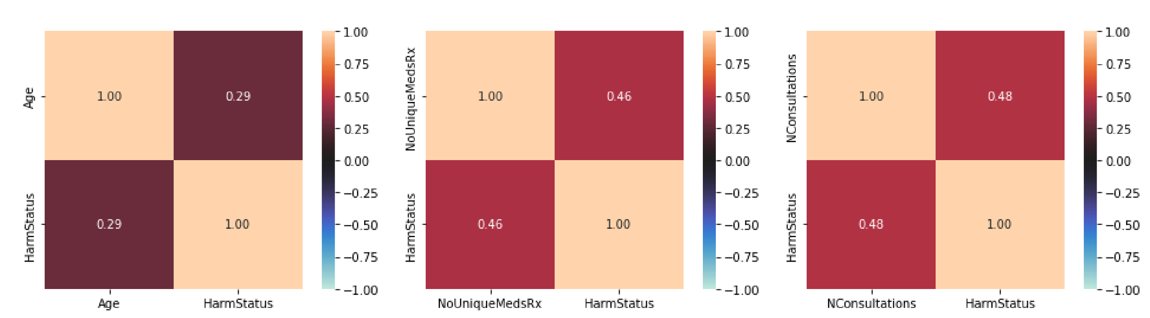
* Primary Health Organization data from the third quarter, 2013.

Tabular Data	Free-Text Data	Labels
age, gender, ethnicity, deprivation (NZDep13), number of consultations, practice size & location	consultations notes, prescriptions, referrals, investigations, discharge summaries	(i) harm or no harm (ii) severity of harm (iii) preventable or not (iv) medical codes

De-identified Text Data
De-identified GP EHRs text length: max = 49,226, average = 2,120



Tabular Data



- Age, number of unique medicines prescribed & number of consultations shows evidence of correlation with patient harm.
- The correlation of patient's gender, deprivation, ethnicity, practice size & location with patient harm is approximately 0.

Preliminary Experiments

Binary classification to predict patient harm or no harm was performed, independently, using both EHRs in the form of free-text and tabular data.

Tabular Data

- Multi-layer perceptrons [3] with categorical variable embeddings was used for the classification task, with the **prediction Sensitivity = 0.85 and Specificity = 0.66**.
- SHAP (SHapley Additive exPlanations) [5] is used to explain the predictions (as shown in examples 1-2).

Example 1- Prediction and True label: Harm. **Example 2- Prediction: Harm; True label: No Harm.**

RED features: increase the prediction of true label; BLUE features: decrease the prediction of true label of the model.

Free-Text EHRs

- PubMedBERT [2] was used for the down-streaming task, with the **prediction Sensitivity = 0.88 and Specificity = 0.7**. EHRs were truncated to 512 tokens due to the model restrictions.
- LIME [4] is used to explain the predictions (as shown in examples 3-4).

Example 3: Prediction and True label: No Harm.

'burning with voids for past days no spasm or other pains no fevers or chills healthy young woman in nad urine blood protein nitrites a cystitis p tmp tabs given from supply return if not resolved scanned document pt transfer

Example 4: Prediction: Harm; True label: No Harm.

'dietnutrition advice given exercise advice given repair in dec r inguinal hernia repair mri of skull for acoustic neuroma normal mri normal fh father aunt of aaa nka ltem classdeidemhbphogtsmoker no diabetes no diabetes bp tchdl cvd risk ltem classdeidemhbphogt twisted or pulled left knee last week while at work complains of daily pain that worsens as the day goes on improved with antiinflammatories also complian of pain posterior right lower extremities believes that it may be due to sciatica also complains of bilateral leg cramps does not drink a lot of water during the day uses a lot of

GREEN: contributes to the prediction. PINK: detract from the final prediction. Shade of color denotes strength.

Concluding Remarks

- Artificial Intelligence working in conjunction with experts moves us further towards reducing healthcare harm in New Zealand general practice.
- Only preliminary experiments of free-text EHRs and tabular text are presented here. However, given the availability of de-identified rich data, with an average length of 2,000 tokens, it is vital to use methods that can handle long sequences.
- Furthermore, incorporating tabular data and text data to use the more recent multi-modal transformers may also improve accuracy of predictions.

References

1. Leitch, S., Dovey, S., Cunningham, W., Wallis, K., Eggleton, K., Lillis, S., ... & Tilyard, M. (2021). Epidemiology of healthcare harm in New Zealand general practice: a retrospective records review study. BMJ Open, 11(7), e048316.

2. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X.,... & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare, 3(1), 1-23.

3. Taud, H., & Mas, J. F. (2018). Multilayer perceptron (MLP). In Geomatic approaches for modeling land change scenarios (pp. 451-455). Springer, Cham.

4. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

5. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Retrospective study [1] was funded by the Health Research Council of New Zealand (HRC-14-185).
Ethical approval for [1] and the current research (using machine learning) was obtained from the University of Otago Human Ethics Committee (HD14/32).