

Portability of a text mining algorithm for detecting adverse drug reactions in electronic health records across diverse patient groups in two Dutch hospitals



catharina
a santeon hospital

Britt W.M. van de Burgt^{1,2} MSc, Loes F.C. van Dijck³ MSc, Bjorn Dullemond⁴ BSc, Naomi T. Jessurun⁵ PhD, Minou van Seyen⁶ PhD, Prof. Rob J. van Marum^{6,7,8}, Remco J.A. van Wensen⁹ MD, Wai-Yan Liu⁹ PhD, Carolien M.J. van der Linden¹⁰ PhD, Rene J.E. Grouls¹ PhD, Prof. R Arthur. Bouwman^{2,13}, Prof. Erik H.M. Korsten², Prof. Toine C.G. Egberts^{11,12}

Background and importance

Adverse drug reactions (ADRs) are a major healthcare concern, almost always (98%) documented as free-text in electronic health records (EHRs), which complicates their use in clinical decision support (CDS) systems. Text mining (TM) shows promise for extracting relevant data from unstructured text, but its portability across institutions and departments is uncertain.

Aim
This study aims to assess the portability of an earlier developed TM-based ADR identification algorithm by applying it to patient data from a second Dutch hospital that utilizes the same EHR system.

Methods

Study design, setting and population



Jeroen Bosch Hospital & Catharina Hospital
Surgical and non-surgical department
Both HIX



Inclusion:
> 24 h hospitalized (Nov. 2022- Dec. 2023)
n ≥ 60 patients (15 patients per department)



All free-text data, for 1 year prior to the recent discharge

Identification of pADRs

Manual review vs. TM algorithm
2 assessors, for no consensus a third assessor

Possible ADRs in all free-text

Causality: Naranjo score (Lareb)
Serious categorized EMA list

F-score (between 0-1)

Precision = PPV
Recall = sensitivity

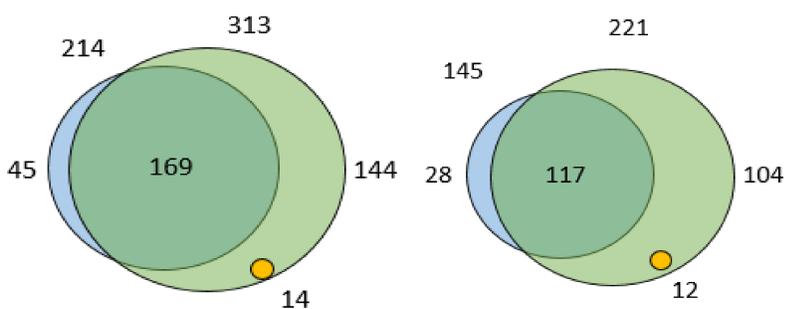
$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results

Total: 25,684 notes & 1,110,987 words

Table 1: Characteristics divided per hospital (CZE vs. JBZ)

Variable	All patients (n=62)		CZE (n=30)		JBZ (n=32)		P-value
Age in years, mean (range)	76	(33-94)	74.6	(33-94)	77.3	(48-92)	0.65
Female, %	50		56.7		43.8		0.31
Charlson Comorbidity Index at last hospitalization, median (range)	4.0 (0-9)		4.0	(0-9)	4.3	(0-7)	0.54
Days of hospitalization, median (range)	8.0	(1-48)	11	(2-48)	4	(1-32)	0.02
Free-text EHR notes per patient, median (range)	347	(70-1,720)	283	(89-1,720)	386	(70-1,231)	0.07
Words per patient, median (range)	12,058	(1,921-104,157)	12,632	(1,921-104,157)	11,472	(2,650-59,949)	0.80
Characters per patient, median (range)	82,248	(14,759-574,699)	86,070	(14,759-574,699)	72,264	(15,360-341,328)	0.47



Venn diagram presenting unique adverse drug reactions (ADRs)
Left CZE, Right JBZ
Green = Text mining algorithm; Blue = Manual review;
Yellow = ADRs found only by the Text mining algorithm

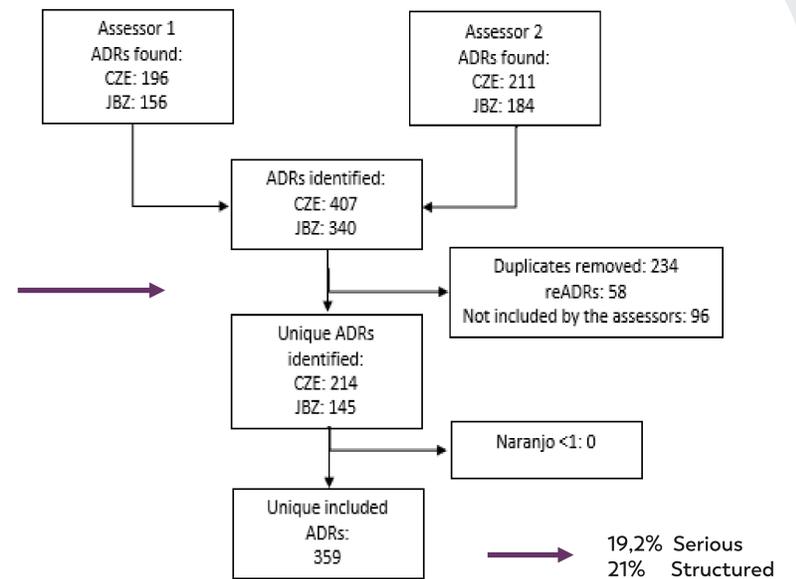


Figure 1: inclusion and exclusion of pADRs

Table 2: TM algorithm versus manual review: the F-score, sensitivity and PPV of total, for both hospitals.

	CZE	JBZ
F-score	0.64	0.64
Sensitivity (%)	169/214 (79)	117/145 (81)
PPV (%)	169/313 (54)	117/221 (53)

Conclusion and Relevance

This study demonstrates that the TM algorithm is portable and can identify pADRs across different hospitals and departments without adaptations. These findings highlight its broader implementation potential for ADR detection in diverse healthcare settings, especially when integrated with CDSS and clinician validation to enhance ADR registration, reduce medication risks, and improve patient safety.

1. Department of Clinical Pharmacy, Catharina Hospital Eindhoven, Eindhoven, the Netherlands.
2. Department of Electrical engineering, signal processing group, Technical University Eindhoven, Eindhoven, the Netherlands
3. Department of Clinical Pharmacy, SJG Weert, Weert, the Netherlands
4. Department of Mathematics and Computer Science, Technical University Eindhoven, Eindhoven.
5. Netherlands Pharmacovigilance Centre LAREB, 's-Hertogenbosch, the Netherlands
6. Department of Clinical Pharmacy, Jeroen Bosch Hospital, 's-Hertogenbosch, the Netherlands
7. Departments of Clinical Pharmacology and Geriatrics, Jeroen Bosch Hospital, 's-Hertogenbosch, the Netherlands.

8. Department of Elderly Care Medicine, Amsterdam Public Health Research Institute, Amsterdam UMC Location VUmc, the Netherlands.
9. Department of Orthopaedic Surgery & Trauma, Catharina Hospital, Eindhoven, the Netherlands.
10. Department of Geriatrics, Catharina Hospital Eindhoven, Eindhoven, the Netherlands.
11. Department of Clinical Pharmacy, University Medical Centre Utrecht, Utrecht, the Netherlands.
12. Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, the Netherlands.
13. Department of Anaesthesiology, Catharina Hospital Eindhoven, Eindhoven, the Netherlands.
contact details: britt.vd.burgt@catharinaziekenhuis.nl

