MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Automating the Identification of Safety Events Involving ML-Enabled Medical Devices

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Dr Ying WANG
Macquarie University
@YingWang_CHI

Dr David LYELL
Macquarie University
@David_Lyell

Prof Enrico COIERA
Macquarie University
@EnricoCOIERA

Prof Farah MAGRABI
Macquarie University
@Farah MAGRABI

@TheInstituteDH #MEDINFO23

AIDH IMIA

Background

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

1. Machine learning (ML) enabled systems have the potential to improve healthcare del
2. With use in clinical settings and by consumers, safety events involving ML-enabled medical devices are emerging.

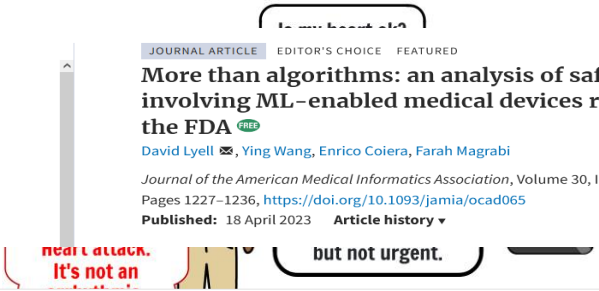
Volume 30, Issue 7
July 2023

JOURNAL ARTICLE EDITOR'S CHOICE FEATURED

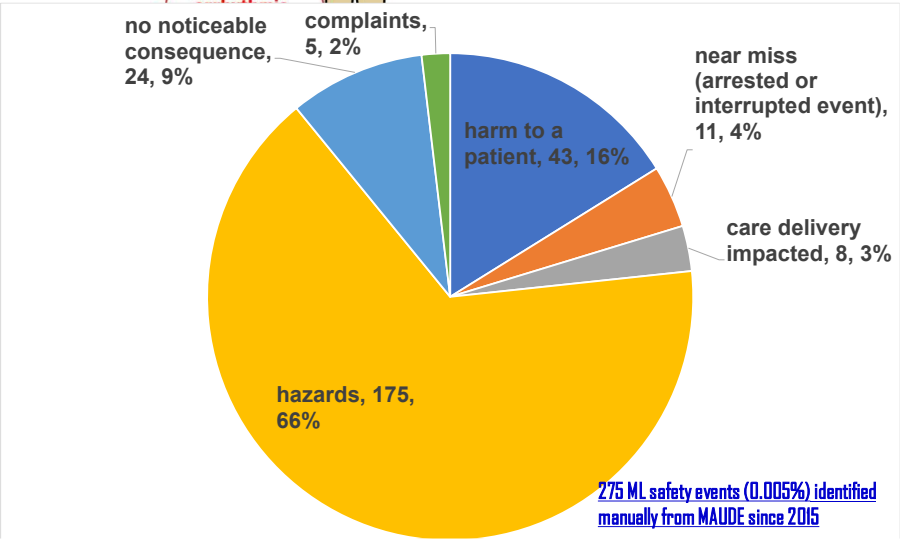
More than algorithms: an analysis of safety events involving ML-enabled medical devices reported to the FDA FREE

David Lyell ✉, Ying Wang, Enrico Coiera, Farah Magrabi

Journal of the American Medical Informatics Association, Volume 30, Issue 7, July 2023, Pages 1227–1236, https://doi.org/10.1093/jamia/ocad065
Published: 18 April 2023 **Article history**



3. FDA MAUDE database is a valuable resource to understand patient safety risks associated with ML medical devices in real worlds.
4. Considering over 2 million events were reported to MAUDE in 2021, manual review on an ongoing basis is not feasible.



Machine Learning driven solution

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Data collection

- Data processing: bag of words
- Numeric representation: term frequency–inverse document frequency (TF-IDF)

Feature representation

- Event reports including updates and follow-up investigations
- Brand names plus reports
- Manufacture names plus reports

Model training and testing

- Binary discriminative classifiers of SVM with radial-basis function kernel
- 10-fold subsampling cross-validation
- Four stratified classifiers with varied feature sets above

Error analysis and verification

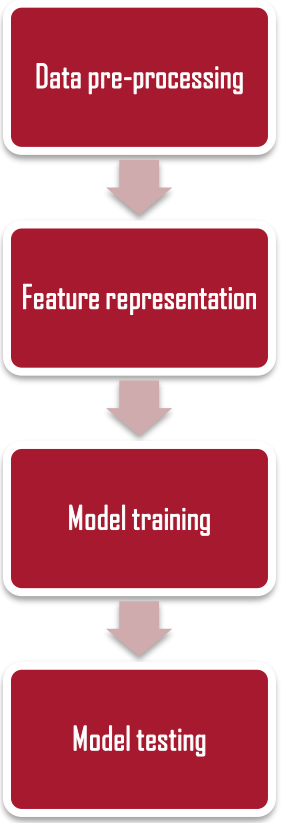
- Performance measures: precision, recall and F1 score
- Verification of classifier identified positives

Training and testing datasets

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Datasets	Event type	Stratified dataset (n=)
Training set (75%) From 1 January 2018 to 31 October 2021	ML safety events	207
	General safety events	4,044,796
	Total	**4,045,003**
Testing set (25%)	ML safety events	68
	General safety events	1,348,266
	Total	**1,348,334**
External testing set From 1 November 2021 to 31 March 2022		895,627



Results: performance of four classifiers

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Dataset		Report classifier	Generic type classifier	Brand name classifier	Combined classifier
Testing set	False negative (n)	27	22	23	4
	False positive (n)	4	2	5	0
	Total (n)	**31**	**24**	**28**	**4**
	Precision (%)	94.4	97.1	93.2	100
	Recall (%)	71.6	75.6	74.7	93.9
	F1 score (%)	81.4	**85**	82.9	**96.9**
External testing set	Classifier identified MLSEs (n)	43	38	38	7,707
Verified as	MLSEs (n)	31	38	38	/
	Data input issues attributed to non-ML	5	0	0	/
	False positives (n)	7	0	0	/

Table: Performance of report classifier, generic type classifier, brand name classifier and combined classifier on testing and external datasets.

Report classifier

Generic type classifier

Brand name classifier

Combined classifier

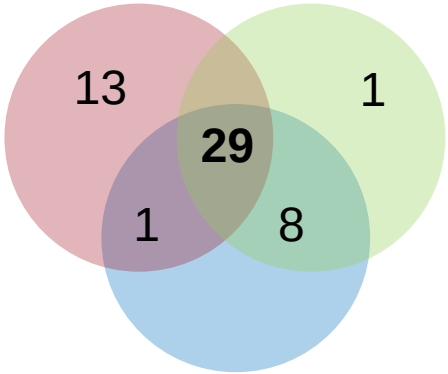
Results: error analysis

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Classifier identifications

- 52 ML safety events identified by three classifiers on external testing set.
- 29 (56%) identified by all three involving known ML devices.
- 9 events identified by two classifiers
  - 8 events by brand name and generic type
  - 1 event by report and generic name
- 13 events identified by report alone and one by brand name



True and false positives

- 40 of 52 events (77%) manually verified as true positives (TP).
- 5 false positives (10%) did not involve ML devices but associated with device input problems, such as poor data acquisition.
- 7 false positives (13%) neither HIT problems nor ML devices.
- All 29 events overlapped by the three classifiers were verified as TPs,
  - Data acquisition problems: 90% (n=26) from imaging systems.
  - Algorithm errors: two TPs (6.9%) with an ECG device designed for detecting normal sinus rhythm and sever arrythmias.

Discussion

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Feasibility

Model performance

- Stratified text classifiers can be used to identify rare ML events from large collections of data.

Generalizability

Real-world use of AI

- Current classifiers don't generalize well to identify rare class events from highly imbalanced dataset.

Device input problems

Common safety risks

- Mostly identified by classifiers and made up the bulk of the training set.

Algorithm errors

Critical safety risk for ML devices

- One user's device suggested atrial fibrillation during heart attack.
- Another user experienced atrial fibrillation, but the device indicated a normal rhythm.

Future work

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Active learning

Data annotation

- iteratively ask safety experts to verify events identified by the classifier and label more ML safety events.

Large Language Models

Generative AI

- pre-trained or fine-tuned on a dataset of safety events to identify ML safety events.

SafeHealth.ai

Knowledge sharing

- Insights on medical AI devices
- Safety risk analysis of ML events

MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

Thank you and comments?

Dr Ying WANG

Macquarie University
@YingWang_CHI
E: ying.wang@mq.edu.au
W: safehealth.ai