Artificial Intelligence Based on Machine Learning in Pharmacovigilance: A Scoping Review

Benjamin Kompa[1,2], M.Phil; Joe B. Hakim[3], BS; Anil Palepu[3], BS; BS; Kathryn Grace Kompa[4], BS; Michael Smith[5], BS; Paul A. Bain[6], PhD; Stephen Wolozynek[7], MD/PhD; Jeffery L. Painter[8], MS/JD; Andrew Bate[9, 10,11], PhD; Andrew L. Beam[1,2,5], PhD

[1] Department of Biomedical Informatics, Harvard Medical School, Boston, MA USA

[2] CausaLAB, Harvard T.H. Chan School of Public Health, Boston, MA USA

[3] Harvard-MIT Department of Health Sciences and Technology, Cambridge, MA

[4] Tufts Medical Center, Boston, MA USA

[5] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA USA

[6] Countway Library of Medicine, Harvard Medical School, Boston, MA USA

[7] Beth Israel Deaconess Medical Center, Boston, MA USA

[8] GlaxoSmithKline, Research Triangle Park, US

[9] GlaxoSmithKline, West London, UK

[10] Department of Non-communicable disease Epidemiology, London School of Hygiene and Tropical Medicine, University of London, United Kingdom

[11] Department of Medicine at NYU Grossman School of Medicine, New York, USA
Corresponding author: Andrew_beam@hms.harvard.edu

Supplementary Information

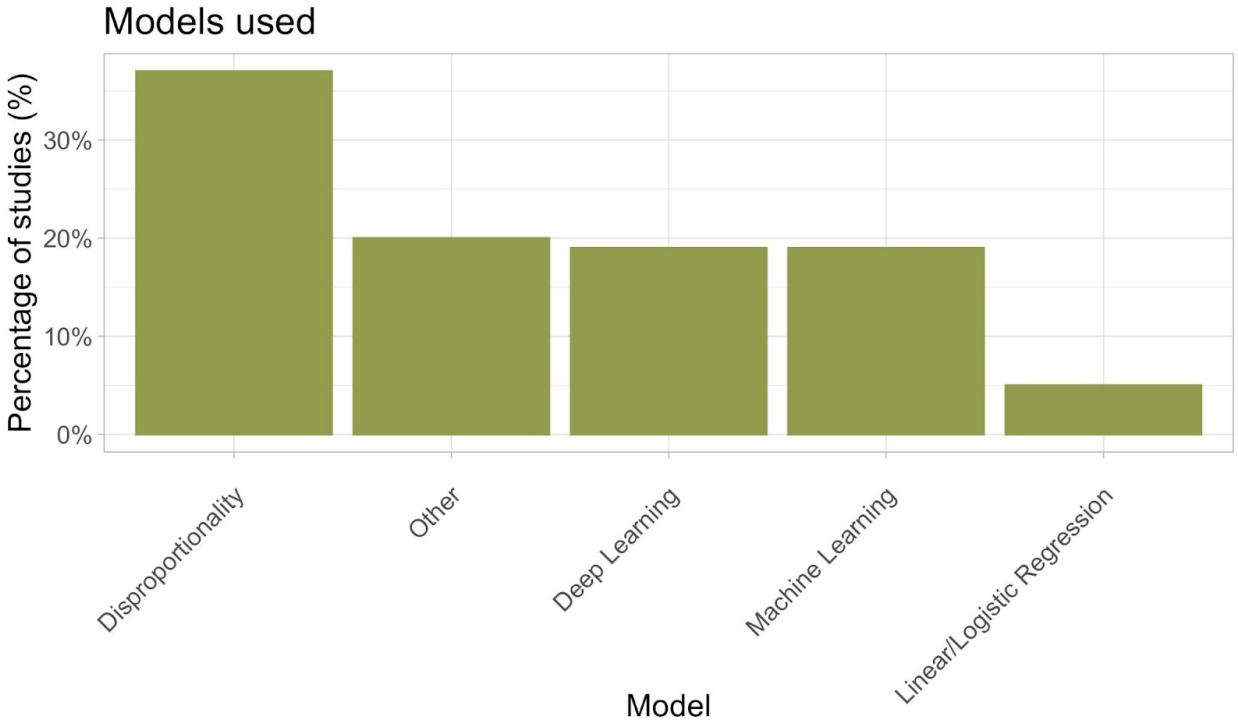

Figure S1: A summary view of the models presented in Figure 2B. “Deep learning” refers to neural network based methods, while “machine learning” contains methods such as random forests, SVMs, etc.

Details on Topic Model Analysis

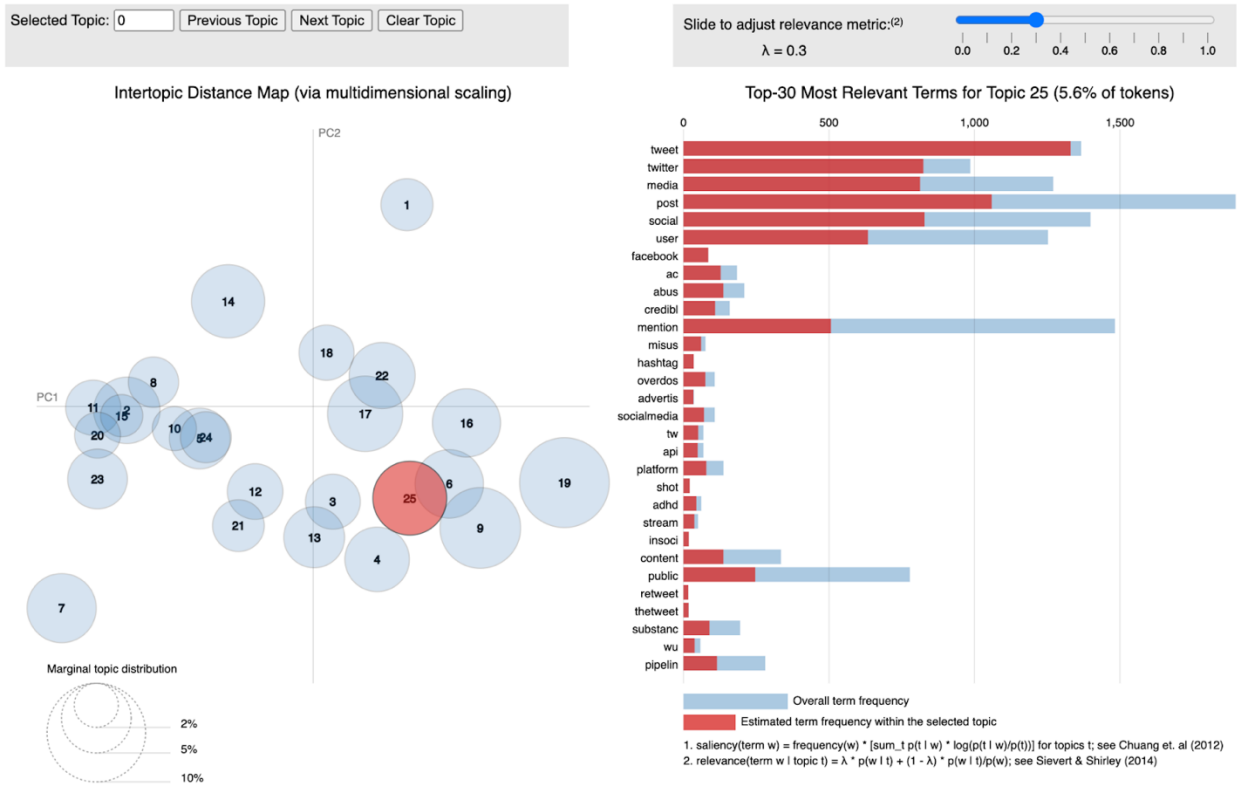

Figure S2: Topic proportions and intertopic distance are plotted on the left with the `LDAvis` package. This package weighs topic proportion by document length. On the right, the top words for topic 25 are shown. By decreasing lambda, one obtains words that are more unique to a topic.

Topic	Top 4 words (lambda=0.3)	% of tokens (stm)	% of tokens (LDAvis)
10	liver, hepatox, injuri, cardiomyopathi	2.11%	1.99%
20	statin, cancer, sr, sodium	2.34%	2.14%
15	jader, onset, antibodi, jp	2.60%	1.82%
18	adr, im, classifi, gradient	2.67%	3.11%
21	nsaid, sex, diclofenac, manuscripth	2.80%	2.75%
3	abstract, side, apetit, dri	2.80%	3.09%
1	chines, introduct, china, reaction	2.87%	2.77%
24	fall, anemia, emr, pds	2.93%	2.65%
8	hair, ic, sale, sweden	3.19%	2.60%

16	psuedo, classific, imbalanc, sampl	3.42%	4.77%
12	arthralgia, exposur, icd, scan	3.50%	3.16%
13	guidelin, icsr, cognit, queri	3.63%	3.81%
22	predict, regress, auc, forest	3.75%	4.48%
6	discuss, sentiment, polar, emot	3.83%	4.79%
4	materi, ml, narrat, ehr	4.06%	4.26%
5	conclus, sdrs, sdr, dmas	4.35%	3.78%
17	ddi, ddis, ade, myopathi	4.38%	5.82%
11	cardiac, ebgm, epiderm, hypersensit	4.81%	3.11%
23	regimen, ap, covid, coliti	4.93%	3.65%
25	tweet, twitter, media, post	5.04%	5.60%
2	ae, korea, regul, signal	5.23%	4.52%
9	annot, entiti, ade, sentenc	5.70%	6.67%
14	bcpnn, prr, gps, shrinkag	5.87%	5.49%
7	vaccin, vaer, gbs, influenza	6.17%	4.90%
19	embed, lstm, layer, model	7.01%	8.27%

Table S1: The full table of topics in our LDA with K=25. Different analysis packages weight topics differently. `stm` does not weight by word frequency, but `LDAvis` does.

Queries

PubMed	4644
Embase	4040
Web of Science	2984
IEEE Xplore	121
Total	11,789
After removing duplicate records	7,745

PubMed (NCBI)
20210909
4,644 Records

("Pharmacovigilance"[Mesh] OR "Drug-Related Side Effects and Adverse Reactions"[mesh] OR "Adverse Drug Reaction Reporting Systems"[mesh] OR "Product Surveillance, Postmarketing"[Mesh:NoExp] OR "Drug Overdose"[mesh] OR "Polypharmacy"[mesh] OR pharmacovigil*[tiab] OR pharmacoepidemiol*[tiab] OR drug safety[tiab] OR drug event*[tiab] OR drug toxicit*[tiab] OR drug reaction*[tiab] OR adverse drug*[tiab] OR overdose*[tiab] OR polypharm*[tiab] OR drug aller*[tiab] OR drug eruption*[tiab] OR postmarket surveillance[tiab] OR post market surveillance[tiab] OR postmarketing surveillance[tiab] OR post marketing surveillance[tiab] OR vaccinovigil*[tiab] OR eudravigilance[tiab] OR individual case safety report*[tiab] OR adverse event report*[tiab] OR ICSR[tiab] OR VAERS[tiab] OR FAERS[tiab] OR AERS[tiab] OR vigibase[tiab]
OR (("Pharmaceutical Preparations"[Mesh] OR "Therapeutic Uses"[Mesh] OR "Drug Therapy"[mesh] OR "drug therapy"[sh] OR drug*[tiab] OR pharmaceutical*[tiab] OR medicine*[tiab] OR medication*[tiab])
AND
("adverse effects"[sh] OR "toxicity"[sh] OR "Patient Safety"[Mesh] OR adverse effect*[tiab] OR adverse event*[tiab] OR adverse reaction*[tiab] OR side effect*[tiab] OR hypersensitiv*[tiab] OR cardiotoxi*[tiab] OR toxicit*[tiab])))
AND
("Artificial Intelligence"[mesh] OR "Pattern Recognition, Automated"[mesh] OR artificial intelligence[tiab] OR computational intelligence[tiab] OR machine intelligence[tiab] OR intelligent automation[tiab] OR intelligent system*[tiab] OR machine learning[tiab] OR deep learning[tiab] OR deep network*[tiab] OR supervised learning[tiab] OR natural language process*[tiab] OR neural net*[tiab] OR perceptron*[tiab] OR pattern recognition[tiab] OR random forest*[tiab] OR computer vision[tiab] OR nearest neighbor*[tiab] OR nearest neighbour*[tiab] OR long short term memory[tiab] OR gated recurrent unit[tiab] OR k-mean*[tiab] OR kmean*[tiab] OR ensemble learning[tiab] OR ensemble model*[tiab] OR kernel method*[tiab] OR radial basis function[tiab] OR gaussian process*[tiab] OR transfer learning[tiab] OR pretrained model[tiab] OR graph convolutional network*[tiab] OR graph neural network*[tiab] OR support vector machine*[tiab] OR elasticnet[tiab] OR LASSO[tiab] OR polynomial regression[tiab] OR linear discriminant analysis[tiab] OR quadratic discriminant analysis[tiab] OR decision tree*[tiab] OR boosting algorithm[tiab] OR boosting machine[tiab]

OR gradient boosting[tiab] OR catboost[tiab] OR semi-supervised learning[tiab] OR self-supervised learning[tiab] OR representation learning[tiab] OR word embedding*[tiab] OR word2vec[tiab] OR glove[tiab] OR neighbor embedding[tiab] OR neighbour embedding[tiab] OR uniform manifold approximation[tiab] OR latent dirichilet allocation[tiab] OR transformer*[tiab] OR masked language[tiab] OR language model*[tiab] OR unsupervised learning[tiab] OR bidirectional encoder representations from transformers[tiab] OR convolutional neural network [tiab] OR reinforcement learning[tiab] OR autoencoder[tiab] OR multilayer perceptron[tiab] OR bayesian additive regression trees[tiab] OR hierarchical cluster[tiab] OR density-based spatial clustering[tiab] OR spectral clustering[tiab] OR topological data analysis [tiab] OR agglomerative clustering[tiab] OR Gaussian mixture model[tiab] OR clustering algorithm[tiab] OR clustering method[tiab] OR bayes classifi*[tiab] OR bayesian classifi*[tiab] OR LSTM[tiab] OR xgboost[tiab] OR t-sne[tiab] OR BERT[tiab] OR BART[tiab] OR DBSCAN[tiab] OR NLP[tiab] OR UMAP[tiab] OR KNN[tiab] OR GRU[tiab] OR RBF[tiab] OR GCN[tiab] OR GNN[tiab] OR SVM[tiab] OR VAE[tiab] OR MLP[tiab] OR TDA[tiab])
AND (english[Filter]) AND (2000:2021[pdat])
AND
("bitter melon"[tiab])

Embase
20210909
4,040 Records
(additional 1,986 conference abstracts saved separately)

1. 'pharmacovigilance'/exp OR 'postmarketing surveillance'/de OR 'adverse drug reaction'/exp OR 'drug toxicity'/exp OR 'drug overdose'/exp OR 'polypharmacy'/exp OR pharmacovigil*:ab,ti,kw OR pharmacoepidemiol*:ab,ti,kw OR 'drug safety':ab,ti,kw OR 'drug event*':ab,ti,kw OR 'drug toxicit*':ab,ti,kw OR 'drug reaction*':ab,ti,kw OR 'adverse drug*':ab,ti,kw OR overdose*:ab,ti,kw OR polypharm*:ab,ti,kw OR 'drug aller*':ab,ti,kw OR 'drug eruption*':ab,ti,kw OR 'postmarket surveillance':ab,ti,kw OR 'post market surveillance':ab,ti,kw OR 'postmarketing surveillance':ab,ti,kw OR 'post marketing surveillance':ab,ti,kw OR vaccinovigil*:ab,ti,kw OR eudravigilance:ab,ti,kw OR 'individual case safety report*':ab,ti,kw OR 'adverse event report*':ab,ti,kw OR ICSR:ab,ti,kw OR VAERS:ab,ti,kw OR FAERS:ab,ti,kw OR AERS:ab,ti,kw OR vigibase:ab,ti,kw
2. (('drug'/exp/mj OR 'drug therapy'/exp/mj) AND ('adverse event'/de/mj OR 'toxicity'/exp/mj OR 'patient safety'/exp/mj))
3. ((drug* OR pharmaceutical* OR medicine* OR medication*) AND ('adverse effect*' OR 'adverse event*' OR 'adverse reaction*' OR 'side effect*' OR hypersensitiv* OR cardiotoxi* OR toxicit*)):ab,ti,kw
4. #1 OR #2 OR #3
5. 'artificial intelligence'/exp OR ('machine learning'/exp NOT 'molecular docking'/de) OR 'artificial intelligence':ab,ti,kw OR 'computational intelligence':ab,ti,kw OR 'machine intelligence':ab,ti,kw OR 'intelligent automation':ab,ti,kw OR 'intelligent system*':ab,ti,kw OR 'machine learning':ab,ti,kw OR 'deep learning':ab,ti,kw OR 'deep network*':ab,ti,kw OR 'supervised learning':ab,ti,kw OR 'natural language process*':ab,ti,kw OR 'neural net*':ab,ti,kw OR perceptron*:ab,ti,kw OR 'pattern recognition':ab,ti,kw OR 'random forest*':ab,ti,kw OR

'computer vision':ab,ti,kw OR 'nearest neighbor*':ab,ti,kw OR 'nearest neighbour*':ab,ti,kw OR 'long short term memory':ab,ti,kw OR 'gated recurrent unit':ab,ti,kw OR 'k-mean*':ab,ti,kw OR kmean*:ab,ti,kw OR 'ensemble learning':ab,ti,kw OR 'ensemble model*':ab,ti,kw OR 'kernel method*':ab,ti,kw OR 'radial basis function':ab,ti,kw OR 'gaussian process*':ab,ti,kw OR 'transfer learning':ab,ti,kw OR 'pretrained model':ab,ti,kw OR 'graph convolutional network*':ab,ti,kw OR 'graph neural network*':ab,ti,kw OR 'support vector machine*':ab,ti,kw OR elasticnet:ab,ti,kw OR LASSO:ab,ti,kw OR 'polynomial regression':ab,ti,kw OR 'linear discriminant analysis':ab,ti,kw OR 'quadratic discriminant analysis':ab,ti,kw OR 'decision tree*':ab,ti,kw OR 'boosting algorithm':ab,ti,kw OR 'boosting machine':ab,ti,kw OR 'gradient boosting':ab,ti,kw OR catboost:ab,ti,kw OR 'semi-supervised learning':ab,ti,kw OR 'self-supervised learning':ab,ti,kw OR 'representation learning':ab,ti,kw OR 'word embedding*':ab,ti,kw OR word2vec:ab,ti,kw OR glove:ab,ti,kw OR 'neighbor embedding':ab,ti,kw OR 'neighbour embedding':ab,ti,kw OR 'uniform manifold approximation':ab,ti,kw OR 'latent dirichilet allocation':ab,ti,kw OR transformer*:ab,ti,kw OR 'masked language':ab,ti,kw OR 'language model*':ab,ti,kw OR 'unsupervised learning':ab,ti,kw OR 'bidirectional encoder representations from transformers':ab,ti,kw OR 'convolutional neural network':ab,ti,kw OR 'reinforcement learning':ab,ti,kw OR autoencoder:ab,ti,kw OR 'multilayer perceptron':ab,ti,kw OR 'bayesian additive regression trees':ab,ti,kw OR 'hierarchical cluster':ab,ti,kw OR 'density-based spatial clustering':ab,ti,kw OR 'spectral clustering':ab,ti,kw OR 'topological data analysis':ab,ti,kw OR 'agglomerative clustering':ab,ti,kw OR 'Gaussian mixture model':ab,ti,kw OR 'clustering algorithm':ab,ti,kw OR 'clustering method':ab,ti,kw OR 'bayes classifi*':ab,ti,kw OR 'bayesian classifi*':ab,ti,kw OR LSTM:ab,ti,kw OR xgboost:ab,ti,kw OR 't-sne':ab,ti,kw OR BERT:ab,ti,kw OR BART:ab,ti,kw OR DBSCAN:ab,ti,kw OR NLP:ab,ti,kw OR UMAP:ab,ti,kw OR KNN:ab,ti,kw OR GRU:ab,ti,kw OR RBF:ab,ti,kw OR GCN:ab,ti,kw OR GNN:ab,ti,kw OR SVM:ab,ti,kw OR VAE:ab,ti,kw OR MLP:ab,ti,kw OR TDA:ab,ti,kw

6. #4 AND #5
7. #6 AND [2000-2021]/py NOT 'conference abstract'/it

Web of Science Core Collection (Clarivate Analytics)
Editions = A&HCI, BKCI-SSH, BKCI-S, CCR-EXPANDED, ESCI, IC, CPCI-SSH, CPCI-S, SCI-EXPANDED, SSCI
20210909
2984 Records

Additional 49 Meeting abstracts and 289 unique proceedings papers saved separately.

Limt to year 2000-2021, exclude meeting abstracts

1. TS=("pharmacovigil*" OR "pharmacoepidemiol*" OR "drug safety" OR "drug event*" OR "drug toxicit*" OR "drug reaction*" OR "adverse drug*" OR "overdose*" OR "polypharm*" OR "drug aller*" OR "drug eruption*" OR "postmarket surveillance" OR "post market surveillance" OR "postmarketing surveillance" OR "post marketing surveillance" OR "vaccinovigil*" OR "eudravigilance" OR "individual case safety report*" OR "adverse event report*" OR "ICSR" OR "VAERS" OR "FAERS" OR "AERS" OR "vigibase")

2. TS=(("drug*" OR "pharmaceutical*" OR "medicine*" OR "medication*") AND ("adverse effect*" OR "adverse event*" OR "adverse reaction*" OR "side effect*" OR "hypersensitiv*" OR "cardiotoxi*" OR "toxicit*"))
3. #1 OR #2
4. TS=("artificial intelligence" OR "computational intelligence" OR "machine intelligence" OR "intelligent automation" OR "intelligent system*" OR "machine learning" OR "deep learning" OR "deep network*" OR "supervised learning" OR "natural language process*" OR "neural net*" OR "perceptron*" OR "pattern recognition" OR "random forest*" OR "computer vision" OR "nearest neighbor*" OR "nearest neighbour*" OR "long short term memory" OR "gated recurrent unit" OR "k-mean*" OR "kmean*" OR "ensemble learning" OR "ensemble model*" OR "kernel method*" OR "radial basis function" OR "gaussian process*" OR "transfer learning" OR "pretrained model" OR "graph convolutional network*" OR "graph neural network*" OR "support vector machine*" OR "elasticnet" OR "LASSO" OR "polynomial regression" OR "linear discriminant analysis" OR "quadratic discriminant analysis" OR "decision tree*" OR "boosting algorithm" OR "boosting machine" OR "gradient boosting" OR "catboost" OR "semi-supervised learning" OR "self-supervised learning" OR "representation learning" OR "word embedding*" OR "word2vec" OR "glove" OR "neighbor embedding" OR "neighbour embedding" OR "uniform manifold approximation" OR "latent dirichilet allocation" OR "transformer*" OR "masked language" OR "language model*" OR "unsupervised learning" OR "bidirectional encoder representations from transformers" OR "convolutional neural network" OR "reinforcement learning" OR "autoencoder" OR "multilayer perceptron" OR "bayesian additive regression trees" OR "hierarchical cluster" OR "density-based spatial clustering" OR "spectral clustering" OR "topological data analysis" OR "agglomerative clustering" OR "Gaussian mixture model" OR "clustering algorithm" OR "clustering method" OR "bayes classifi*" OR "bayesian classifi*" OR "LSTM" OR "xgboost" OR "t-sne" OR "BERT" OR "BART" OR "DBSCAN" OR "NLP" OR "UMAP" OR "KNN" OR "GRU" OR "RBF" OR "GCN" OR "GNN" OR "SVM" OR "VAE" OR "MLP" OR "TDA")
5. #3 AND #4

IEEE Xplore (IEEE.org)
20210909
121 Records

Limit to Journal, Early Access, Year2000-2021
All metadata:(pharmacovigilance OR pharmacoepidemiol* OR "drug safety" OR "drug event*" OR "drug toxicitity" OR "drug reaction*" OR "adverse drug" OR overdose* OR polypharmacy OR "drug allergy" OR "drug eruption*" OR "postmarket surveillance" OR "post market surveillance" OR "postmarketing surveillance" OR "post marketing surveillance" OR vaccinovigilance OR eudravigilance OR "individual case safety report" OR "individual case safety reports" OR vigibase OR "adverse effect" OR "adverse event" OR "adverse events" OR "drug hypersensitivity" OR cardioxicity)
AND
All metadata:("artificial intelligence" OR "computational intelligence" OR "machine intelligence" OR "intelligent automation" OR "intelligent system" OR "intelligent systems" OR "machine learning" OR "deep learning" OR "deep network" OR "deep networks" OR "supervised learning" OR "natural language" OR "neural net" OR "neural nets" OR "neural

network" OR "neural networks" OR perceptron OR "pattern recognition" OR "random forest" OR "computer vision" OR "nearest neighbor" OR "nearest neighbour" OR "long short term memory" OR "gated recurrent unit" OR "k-mean*" OR kmean OR kmeans OR "ensemble learning" OR "ensemble model" OR "kernel method" OR "radial basis function" OR "gaussian process*" OR "transfer learning" OR "pretrained model" OR "graph convolutional network" OR "support vector machine" OR elasticnet OR LASSO OR "polynomial regression" OR "linear discriminant analysis" OR "quadratic discriminant analysis" OR "decision tree" OR "decision trees" OR "boosting algorithm" OR "boosting machine" OR "gradient boosting" OR catboost OR "semi-supervised learning" OR "self-supervised learning" OR "representation learning" OR "word embedding" OR word2vec OR glove OR "neighbor embedding" OR "neighbour embedding" OR "uniform manifold approximation" OR "latent dirichilet allocation" OR transformer OR "masked language" OR "language model*"OR "unsupervised learning" OR "convolutional neural network" OR "reinforcement learning" OR autoencoder OR "multilayer perceptron" OR "bayesian additive regression trees" OR "hierarchical cluster" OR "density-based spatial clustering" OR "spectral clustering" OR "topological data analysis" OR "agglomerative clustering" OR "Gaussian mixture model" OR "clustering algorithm" OR "clustering method" OR "bayes classifi*" OR "bayesian classifi*")

Countway Protocol Template

Title

Identify the report as a protocol for the type of review you're conducting.
The impact of artificial intelligence in pharmacovigilance from 2000-2021: a scoping review

Registration Information

If you've registered your protocol, please include the registration number or DOI.

Authors

Who is your research team? List all contributors and their roles. Who is the librarian collaborator working on your project?
Benjamin Kompa – Drafting protocol, reviewing papers, composing draft
Joe Hakim – Query formation, reviewing draft
Anil Palepu - Reviewing papers and draft
Cindy Wang - Reviewing papers and draft
Michael Smith - Reviewing papers and draft
Stephen Woloszynek - Reviewing papers and draft
Paul Bain – Query formation, protocol review, pulling articles, reviewing draft
Jeffery Painter – Pharmacovigilance perspective, reviewing draft
Andrew Bate – Pharmacovigilance perspective, reviewing draft
Andrew Beam – AI perspective, composing draft

Support

Who is the review funder/sponsor? What's their role in the review?

GSK – GSK is funding a research collaboration with the Beam Lab with a focus on using artificial intelligence to improve pharmacovigilance and drug safety. Multiple members of the GSK will provide feedback on our review from a pharmacovigilance perspective.

Introduction

Rationale

What are your reasons for producing this review? What are you trying to achieve? What gap in knowledge is this covering?
This review will provide a complementary perspective to the intersection of AI and pharmacovigilance. Our lab group has deep expertise from the AI and healthcare perspective, where there are similar challenges. We would like to survey modern developments in this intersection and identify areas of potential improvement. While there is great promise in applying AI to drug safety, there have yet to be any impactful examples that we are aware of.

Objectives

What specific question is your review addressing? State your hypothesis in a PICO format. Identity the participants, interventions, exposure, issue, comparators, and outcomes. If your question does not fit the PICO structure or is exploratory in nature, you are likely doing a scoping review.
Using the PCC format (population, concept, context) from JBI:
What pharmacovigilance and drug safety applications have artificial intelligence methods between applied to during the period of 2000-2021?

Methods

Eligibility Criteria

What criteria are you using to include and exclude studies from your review? Your eligibility criteria includes your inclusion and exclusion criteria. Include a detailed definition of your independent and dependent variables. These are the elements that are used to include or exclude studies during review: study design (e.g. randomized controlled trials, observation design), setting, time frame, participants characteristics? Is there a well justified date limit (e.g. did something seminal happen in your field that changed practice)? Will you include non-English studies (it is highly recommended that you do)? Will you include meeting abstracts, dissertations, or pre-prints?

Date limit: 2000 to 2021. Artificial intelligence is a rapidly changing field. We are specifically interested in machine learning applications within the broader umbrella of AI and there has been a dramatic increase in the success of these methods in the past two decades as computational power has increased.

Non-English studies: Excluded

Article type: Abstracts, primary research studies, and pre-prints will all be included

Terms: We've curated a list of terms to capture our notion of artificial intelligence and machine learning. We've worked with our collaborators as GSK to do the same for pharmacovigilance and drug safety.

Information Sources

What information sources are you gathering information from? E.g. List the databases, trial registries, and grey literature & the dates of coverage you're searching. Your librarian collaborator will help you choose the appropriate sources.

PubMed
Embase
Web of Science
IEEE Xplore

Study Selection Process

What will your process for selecting studies entail? How many reviewers will be required in the title/abstract screening process? How many votes will be required to move a record from title/abstract screen to the full text screen? How will you reconcile differences? Will you include an analysis of interrater reliability? Describe the process of each phrase of your review.

1) Initial limited review and query formation
    a. We worked with the librarian to do an initial review of PubMed and Web of Science
    b. We've composed a query based on terms relating to AI and pharmacovigilance. We'll search the above databases for matching works.
2) Pulling documents from the above 4 databases
    a. These documents will be imported into Covidence, a software that will manage distributing papers to reviewers and recording their responses.
    b. Two votes will be required to move a record to the full text screen.
    c. For a record with only one vote, we'll have a third reviewer determine whether the piece qualifies for the full text screen.
    d. We have 7 reviewers ready to review roughly 5000 articles.
    e. No plans at this point to assess interrater reliability

Data Extraction & Management

List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationales. How will you obtain & confirm data from investigators? If more than one person is doing extractions, how will you reconcile differences? How will you deal with missing data? How are you going to manage your records & data during your review process (e.g. create data extraction tables in Covidence)?

See above

Risk of Bias and Strength of Evidence

How will you assess risk of bias in individual studies (will this be done at outcome or study level or both). There are formal tools (see link). Will you use GRADE or a different tool for analyzing strength of evidence?
Optional for non-systematic reviews.

Data Synthesis

Describe your plan for analyzing your extracted data. We recommend you consult someone with statistical expertise to make these decisions. Are there any proposed additional analyses (e.g. sensitivity, subgroup analyses, meta-regression)? If not a quantitative synthesis (i.e. qualitative synthesis), describe your method for summarizing data.

We'll qualitative evaluate the relevant studies that we find for several qualities:

- The pharmacovigilance task type (e.g. classification, regression, unsupervised analysis, reinforcement learning tasks)

- The machine learning/AI methods used
- Any obvious issues with their methodology

We're going to synthesize this into a broader discussion on the inductive biases that we can incorporate into our models in order to maximize their success on pharmacovigilance tasks.