Applying Artificial Intelligence in Structured Real-World Data for Pharmacovigilance Purposes: A Systematic Literature Review

Stella Dimitsaki, Pantelis Natsiavas, Marie-Christine Jaulent

Submitted to: Journal of Medical Internet Research
on: February 28, 2024

Table of Contents

Applying Artificial Intelligence in Structured Real-World Data for Pharmacovigilance Purposes: A Systematic Literature Review

Stella Dimitsaki[1]; Pantelis Natsiavas[2]; Marie-Christine Jaulent[1]

[1]Sorbonne Université Inserm, Université Sorbonne Paris-Nord Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé - LIMICS Paris FR
[2]Institute of Applied Biosciences Centre for Research and Development Hellas Thessaloniki GR

Corresponding Author:
Stella Dimitsaki
Sorbonne Université
Inserm, Université Sorbonne Paris-Nord
Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé - LIMICS
15 Rue de l'École de Médecine
Paris
FR

Abstract

Background: Artificial Intelligence (AI) on real-world data (RWD) (e.g., Electronic Healthcare Records – EHR) has been identified as a potentially promising technical paradigm for the pharmacovigilance (PV) field. There are several applications of AI approaches on RWD, however, most of the studies focus on unstructured RWD, i.e., conducting Natural Language Processing (NLP) on various data sources (e.g., clinical notes, social media, blogs, etc.). Hence, it is essential to investigate how AI is already applied to structured RWD in PV and how new approaches could enrich the existing methodology.

Objective: This manuscript provides a Systematic Literature Review (SLR) depicting the emerging use of AI upon structured RWD for PV purposes to identify relevant trends and potential research gaps.

Methods: The presented SLR methodology is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology/rationale. Relevant scientific manuscripts were retrieved by PubMed on January 31, 2024. The included studies were "mapped" against a set of evaluation criteria, including applied AI approaches, code availability, description of data preprocessing pipeline, implementation of trustworthy AI criteria, and clinical validation of AI models.

Results: The systematic literature review finally yielded 36 studies. There has been a significant increase in studies after 2019. Most of the articles focus on Adverse Drug Reaction (ADR) detection procedures (64%) for specific adverse effects. Furthermore, a significant number of studies (>90%) used non-symbolic AI approaches (Machine Learning – ML and Deep Learning - DL) emphasizing classification tasks. Random forest is the most popular ML approach in this review (47%). The most common RWD sources used are the EHRs (78%). Typically, these data are not available in a widely acknowledged data model to facilitate interoperability and they come from proprietary databases; thus, they are not available to reproduce results. Based on the evaluation criteria classification, 10% of the studies published their code in public registries, 16% of them tested their AI models in clinical environments and 36% of them provided information about the data preprocessing pipeline. Additionally, in terms of trustworthy AI, 89% of the articles follow at least half of the FUTURE AI initiative guidelines.

Conclusions: Artificial intelligence, along with structured real-world data, constitutes a new and promising line of work for drug safety and PV. However, in terms of AI, some approaches haven't been examined extensively in this field (like Explainable AI and Causal AI). Moreover, it would be helpful to have a data preprocessing protocol for real-world data to support pharmacovigilance processes. Finally, because of personal data sensitivity, evaluation procedures have to be investigated further.

(JMIR Preprints 28/02/2024:57824)
DOI: https://doi.org/10.2196/preprints.57824

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓

Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Applying Artificial Intelligence in Structured Real-World Data for Pharmacovigilance Purposes: A Systematic Literature Review

Abstract

Background: Artificial Intelligence (AI) on real-world data (RWD) (e.g., Electronic Healthcare Records – EHR) has been identified as a potentially promising technical paradigm for the pharmacovigilance (PV) field. There are several applications of AI approaches on RWD, however, most of the studies focus on unstructured RWD, i.e., conducting Natural Language Processing (NLP) on various data sources (e.g., clinical notes, social media, blogs, etc.). Hence, it is essential to investigate how AI is already applied to structured RWD in PV and how new approaches could enrich the existing methodology.
Objective: This manuscript provides a Systematic Literature Review (SLR) depicting the emerging use of AI upon structured RWD for PV purposes to identify relevant trends and potential research gaps.

Methods: The presented SLR methodology is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology/rationale. Relevant scientific manuscripts were retrieved by PubMed on January 31, 2024. The included studies were "mapped" against a set of evaluation criteria, including applied AI approaches, code availability, description of data preprocessing pipeline, implementation of trustworthy AI criteria, and clinical validation of AI models.
Results: The systematic literature review finally yielded 36 studies. There has been a significant increase in studies after 2019. Most of the articles focus on Adverse Drug Reaction (ADR) detection procedures (64%) for specific adverse effects. Furthermore, a significant number of studies (>90%) used non-symbolic AI approaches (Machine Learning – ML and Deep Learning - DL) emphasizing classification tasks. Random forest is the most popular ML approach in this review (47%). The most common RWD sources used are the EHRs (78%). Typically, these data are not available in a widely acknowledged data model to facilitate interoperability and they come from proprietary databases; thus, they are not available to reproduce results. Based on the evaluation criteria classification, 10% of the studies published their code in public registries, 16% of them tested their AI models in clinical environments and 36% of them provided information about the data preprocessing pipeline. Additionally, in terms of trustworthy AI, 89% of the articles follow at least half of the FUTURE AI initiative guidelines.
Conclusions: Artificial intelligence, along with structured real-world data, constitutes a new and promising line of work for drug safety and PV. However, in terms of AI, some approaches haven't been examined extensively in this field (like Explainable AI and Causal AI). Moreover, it would be helpful to have a data preprocessing protocol for real-world data to support pharmacovigilance processes. Finally, because of personal data sensitivity, evaluation procedures have to be investigated further.

Keywords: pharmacovigilance; drug safety; artificial intelligence; machine learning; real-world data; systematic review

Introduction

Pharmacovigilance (PV) is defined by the World Health Organization as *"the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem"* [1]. PV plays a crucial role in ensuring the safety of medications and

protecting the health of patients as it mostly focuses on the identification of potential Adverse Drug Reactions (ADRs) after medicinal products have been licensed and released to the public.

ADRs can range from mild and tolerable side effects to severe and life-threatening events. Their impact in terms of public health is significant as there are estimates concluding that ADRs can cause an increase in the duration of hospitalization stays for 9.2 ± 0.2 days for outpatient settings and 6.1 ± 2.3 days for inpatient settings [2]. Typically, PV professionals analyze data from Individual Case Safety Report (ICSRs) databases (like FAERS, the database maintained by the U.S. Food and Drug Administration – FDA) to identify potential PV signals, namely potential causal relationships between an ADR and a drug. ICSRs are typically submitted either from patients or from healthcare/PV professionals and they are the main data source used today for PV. However, ICSR databases are subject to many biases and underreporting has also been identified as a huge issue [3]. Moreover, such databases frequently lack information that could make a significant difference in the examination of a potential signal (e.g., patients' medical history). Hence, the early detection of potential PV signals by collecting and analyzing data from various sources is critical to prevent serious effects as soon as possible.

The term “Real-world data” (RWD) refers to data collected outside the controlled environment of clinical trials, such as electronic health records (EHRs), patient registries, insurance claim databases, ePrescription systems, etc. There is a growing interest in using RWD for PV signal management to facilitate faster and more efficient post-marketing surveillance [4]. The significance of RWD in PV lies in its potential for representing longitudinal real-world patient experiences and healthcare practices that can provide insights into drug safety under real-life conditions. Analyzing RWD could also enrich and consolidate the already existing knowledge on ADRs (e.g., by detecting new cofounders). Indicatively, the use of a federated RWD network has been recently used to validate their value in terms of PV signal management [5].
To this end, the European Medicines Agency (EMA) and the US FDA introduced infrastructures for the leverage of the RWD for drug safety purposes, called DARWIN[a] and Sentinel Initiative[b], respectively. RWD are also actively investigated for purposes beyond drug safety (e.g., epidemiology) [6]. It should be noted that although this type of data could in principle give a good overview of patients’ clinical course, still, two major challenges are preventing the use of RWD: (a) these datasets typically come with significant data quality risks, usually contain a lot of null values and errors, and (b) because of legal, ethical, and regulatory issues (e.g., patient privacy issues) it is difficult to access these data sources.

Rationale

Artificial Intelligence (AI) is widely acknowledged as a potentially very useful technical breakthrough that could be used to support decisions in healthcare (e.g., clinical decision support systems) due to its ability to process efficiently big data to seek useful information. AI could be used to identify patterns and associations at large amounts of data such as RWD where the traditional statistical methods of data analysis may struggle to extract because of the amount and complexity (e.g., nonlinear relations between variables, etc.) of data. AI has been widely investigated regarding its applications in healthcare (e.g. personalized medicine) with promising results [7,8], however, it is not yet widely applied in clinical practice. Along these lines, AI could potentially support multiple aspects of PV, e.g. the identification of patient subpopulations that may be more vulnerable to specific ADRs, contributing to the vision of personalized drug safety management.

[a] https://www.darwin-eu.org/
[b] https://www.sentinelinitiative.org/

Objectives

The objective of this systematic literature review (SLR) is to identify and characterize the current research trends regarding the use of AI upon structured RWD for PV and identify relevant gaps. Sequentially, our purpose is to detect innovative ideas, spot existing limitations, and propose possible directions for future work in this field.

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[c] methodology/rationale is applied. The PRISMA Statement is a road map for authors to describe more precisely the state of the art, the findings of the literature search, and discuss the results.

Eligibility criteria

Journal and conference articles were selected focusing on PV, written in English, reporting the use of symbolic and non-symbolic AI approaches upon RWD, focusing on EHRs, insurance claims' databases, and administrative health data.
Review and opinion articles were excluded from the final manuscript selection. Furthermore, research articles focusing on image and text data (e.g., social media, clinical notes) were also excluded. Therefore, AI methods focusing on the use of Natural Language Processing (NLP), Natural Language Understanding (NLU), image processing, or object detection were considered out of this work's scope.
A key issue that came up during this SLR was the lack of a clear distinction between plain statistical methods and Machine Learning (ML) approaches as these two domains frequently overlap and these two terms are sometimes used interchangeably. In this manuscript, we acknowledge that the difference between AI and statistical methods is that AI creates models that can *learn* from data during iterative training processes while statistical methods are dealing with finding relationships between variables. Thus, we considered the iterative "learning" part of an algorithm as the key feature to classify the algorithm as AI/ML. We excluded papers that were based on algorithms with no iterative "learning" scheme as we considered them to be part of the "plain statistics" approaches. Finally, we excluded papers that focus on adverse events from medical devices.

Search strategy

A PubMed query was formed and executed on January 31, 2024, to include research articles from 2010 to 2024. Error: Reference source not found presents the query structure.

Table 1 Query structure – (Pharmacovigilance terms with OR) AND (AI terms with OR) AND (RWD terms with OR)

Main concepts	Keywords
Pharmacovigilance (Keywords relevant to known ADR categories, synonyms of drug safety, PV terminology, known ICSR databases)	PV OR "pharmacovigil*" OR "pharmaco-vigil*" OR "side effect*" OR "adverse reaction*" OR "Product Surveillance" OR "postmarket*" OR pharmacoepidemiol* OR pharmaco-epidemiol* OR "drug safety" OR "drug event*" OR "toxicit*" OR "drug reaction*" OR "adverse drug*" OR "allerg*" OR "post-market*" OR "post market*" OR vaccinovigil* OR vaccino-vigil* OR eudravigilance OR

[c] http://prisma-statement.org/

	“individual case safety report*” OR ICSR OR VAERS OR FAERS OR AERS OR vigibase OR "adverse effect*" OR "adverse event*" OR hypersensitiv* OR “spontaneous report*” OR “yellow card” OR “yellow-card” OR ADR OR "personalized pharmacovigilance" OR "precision pharmacovigilance" OR "pharmacosurveillance" OR "pharmaco-surveillance"
Artificial Intelligence (Categories of AI, terms that are used in the development of an AI model, explainable/interpretable AI methods, different AI architectures)	"artificial intelligence" OR AI OR "machine learning" OR ML OR "neural network*" OR NN* OR "deep learning" OR DL OR ontolog* OR "knowledge engineering" OR KE OR reasoning OR inference OR “semantic web” OR “OWL” OR “Web Ontology Language” OR SWRL OR “RDF” OR “Resource Description Framework” OR “prediction” OR “estimation” OR “XAI” OR “SHAP” OR “Shapley value” OR “LIME” OR “Local Interpretable Model-agnostic Explanations” OR “DeepSHAP” OR “DeepLIFT” OR “CXplain” OR “Explainable Artificial Intelligence” OR “Explainable machine learning” OR “Interpretable artificial intelligence” OR “Interpretable machine learning”
Real World Data/Real World Evidence (Categories of RWD, data models that are used to store RWD)	"Real World Evidence" OR "Real World Data" OR RWE OR RWD OR “Observational Medical Outcomes Partnership” OR "OMOP" OR "Electronic Healthcare Record*" OR "EHR" OR "Electronic Medical Record*" OR "EMR*" OR EHDEN OR OHDSI OR i2b2 OR Sentinel OR DARWIN OR “Data Analysis and Real World Interrogation Network” OR administrative OR claim* OR “Observational Health Data Sciences and Informatics” OR “European Health Data Evidence Network” OR “multimodal data” OR “multimodal drug data” OR “multidimensional data” OR “multidimensional drug data” OR “multi-modal data” OR “multi-modal drug data” OR “multi-dimensional data” OR “multi-dimensional drug data”

Selection process

The initial phase (phase 1) focused on screening the titles and abstracts of the articles retrieved from Table 1 query to map those that potentially meet our inclusion criteria and exclude the irrelevant studies using the Rayyan tool[d]. Rayyan is an AI tool that is used for the remote collaboration of researchers on systematic literature reviews. The platform gathers the titles and abstract of all the articles in the study and the reviewers can evaluate the eligibility (include, exclude, maybe) of every article based on their review's objectives in blind mode, i.e., each reviewer values the articles without prior knowledge of the other reviewers' decisions. Conflicts were resolved during consensus meetings among all reviewers.
The second phase focused on the full-text review of the papers selected after phase 1 to decide on the papers that would be finally included in this study. In the full-text review of the selected studies based on titles and abstracts, we excluded research papers that did not meet one or more of the inclusion criteria (i.e., strong focus on AI, RWD, and PV) and studies that met the exclusion criteria (e.g., studies related to image and text data, or they are following only statistical approaches).

Data collection process and mapping

The selected studies were further elaborated and mapped against evaluation criteria using a spreadsheet. The mapping criteria main categories are the following: pharmacovigilance objective(s) (Drug safety core activities, Drug safety special topics), data provenance (Data sources categories, Data sources), country/ies of origin, AI algorithms' categories, data preprocessing methods, usage of explainable AI methods, code availability, use of models in clinical practice, ethical AI, etc. Error: Reference source not found presents an external description of the mapping criteria.

Study risk of bias assessment

The selection of studies written in English and the exclusion of AI studies based on text-mining/NLP, image processing, and statistical analysis could be identified as potential risks for bias. Furthermore, the selection of papers only from PubMed could also be identified as a potential bias risk as it could lead to the exclusion of papers from other databases.

Synthesis methods

The mapping strategy was designed based on the three main pillars of the objective and we also included general information about the research papers. Furthermore, we included free-text fields in the mapping Excel file to add significant extra details that cannot be easily classified. These fields are Objective, Methods, Assessment, and interesting results. The criteria contain specific attributes (e.g., Drug safety core activities) which were defined based on previous experience conducting SLR on the field [9] and key interest aspects detected during the review.

Table 2 Mapping criteria architecture

Categories	Criteria	Sub-criteria
General Information	PubMed/Medline ID (Number)	Identification number of articles
	Authors (Text)	List of authors
	Title (Text)	Article's title
	Journal (Text)	Name of journal publication
	Year published	Year of article's publication

[d] https://www.rayyan.ai/

	(Number)	
	Type of organizations (Text)	Type of organizations based on the authors' affiliation. Possible values: Healthcare, Government, Academia/Research, Industry, PV monitoring
	Country (Text)	Country where research conduct based on the authors' affiliation.
Pharmacovigilance	Drug safety core activities (Text)	Possible values: ADE detection, ADE monitoring, ADE prevention, ADE assessment, ADE information collection, ADE reporting
	Drug safety special topics (Text)	Possible values: Comparative drug analysis, Drug interactions, MoA identification/analysis, Personalized drug safety, Signal detection, Specific (class of) disease, Specific (class of) drug(s), Specific adverse effect, Vaccine safety
	Drug (Text)	Drugs being examined in the research papers
	Reaction (Text)	Reactions being examined in the research papers
	Indication (Text)	Indications being examined in the research papers
	Reference terminologies (Text)	Known health informatics terminologies that are detected in the research papers
Artificial Intelligence	Artificial Intelligence categories (Text)	Possible values: Non Symbolic AI, Symbolic AI
	Non-Symbolic AI (Text)	Possible values: Classification, Regression
	Classification (Text)	Possible values: Random Forest, Logistic regression, Artificial neural network, XGBoost, Support Vector Machine, Decision tree, Knowledge graph, K nearest neighbors, Gradient boost, Naive

		Bayes, Random Survival Forest, Extra tree
	Regression (Text)	Possible values: Logistic regression, Linear Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Regularized Cox regression
	Data Preprocessing type (Text)	Possible values: Dimensionality reduction, Feature Engineering, Null imputation, Data cleansing
	Data cleansing (Text)	Possible values: Data normalization, Remove null values
	Feature Engineering (Text)	Possible values: One-Hot Encoding, Binning, Splitting, Calculated Features
	Null imputation (Text)	Possible values: Regression/Classification Imputation
	Explainable AI methods (Text)	Possible values: Local Interpretable Model-Agnostic Explanations (LIME), SHapely Additive exPlanations (SHAP)
	Knowledge representation formalism (Text)	Possible values: OWL, RDF
	Knowledge Engineering core activities (Text)	Possible values: Knowledge extraction, Knowledge integration, Knowledge representation
Real World Data	Data source categories (Text)	Possible values: ADE databases, Clinical narratives, Clinical trials Drug information databases, Drug regulation documentation, EHRs, Genetics and biochemical databases, SRSs, Dispensing records from pharmacies, Administrative Claims data
	Data source(s) (Text)	Possible values: Proprietary closed data sources (e.g. specific hospital EHR), FAERS, SIDER, SMILES,

		UK Biobank, Osteoarthritis Initiative (OAI) dataset, PharmGKB, Twosides, EU-ADR Reference Set, Stockholm EPR Corpus, MIMIC, OMIM, DisGeNet, AEOLUS
	Data Model (Text)	Possible values: OMOP-CDM, Sentinel, Custom
Evaluation criteria	Code availability (Text)	The availability of the code in an open registry. Possible values: Yes, No
	Data preprocessing	Information about the data preprocessing procedures. Possible values: Yes, No
	Clinical use	Information about the evaluation of produced work pipeline in clinical environments. Possible values: Yes, No

Furthermore, in terms of ethical AI, the included studies are evaluated based on trustworthy AI guidelines for solutions in medicine and healthcare from the FUTURE AI initiative [10]. These guidelines are separated into 7 categories (Fairness, Universality, Traceability, Usability, Robustness, Explainability, and a General category). For our evaluation procedure, we included only the highly recommended subcategories from each of the 7 main categories for proof-of-concept (Low technology readiness levels for machine learning models) [11]. Error: Reference source not found presents the selected criteria and their description.

Table 3 FUTURE AI highly recommended, proof-of-concept ML guidelines

	Recommendations	Description
Fairness	Define sources of bias	Identification of possible types and sources of bias for their AI tool during the design phase (e.g. sex, gender, age, ethnicity, socioeconomics, geography, comorbidities or disability of patients, human biases during data labelling).
Universality	Define clinical settings	Specification of the clinical settings in which the AI tool will be applied (e.g. primary healthcare centers, hospitals, home care, low vs. high-resource settings, one or multiple countries)

	Evaluate using external data	Testing of developed AI model to an external dataset with different characteristics from the training set.
Traceability	Provide documentation (e.g. technical, clinical)	Creation of documentations files that provide technical (e.g., public repositories) and clinical information (e.g., bias of the model based on its use)
Usability	Define user requirements	Specification of the model's use from healthcare professionals.
Robustness	Define sources of data variation	Specification of data sources' variation that may impact the AI tool's robustness in the real world (differences in equipment, technical fault of a machine, data heterogeneities during data acquisition or annotation, and/or adversarial attacks).
	Train with representative data	Data for the training process should represent the population based on the case study that AI model is developed for.
	Evaluate & optimize robustness	Risk mitigation measures should be implemented to optimize the robustness of the AI model, such as regularization, data augmentation, data harmonization, or domain adaptation.
Explainability	Define explainability needs	Use of interpretable of explainable models.
General	Engage inter-disciplinary stakeholders throughout the AI lifecycle	
	Implement measures for data privacy and security	
	Define adequate evaluation plan (e.g. datasets, metrics,	

	reference methods)	

Results

Study selection

The PubMed search query originally returned 4264 studies. After the abstract and title screening (phase 1), we selected 93 articles for full-text screening (phase 2). During phase 2, several papers were further rejected, finally including 36 research papers based on the inclusion criteria. PRISMA flowchart presents a detailed overview of the selection procedure in Figure 1.

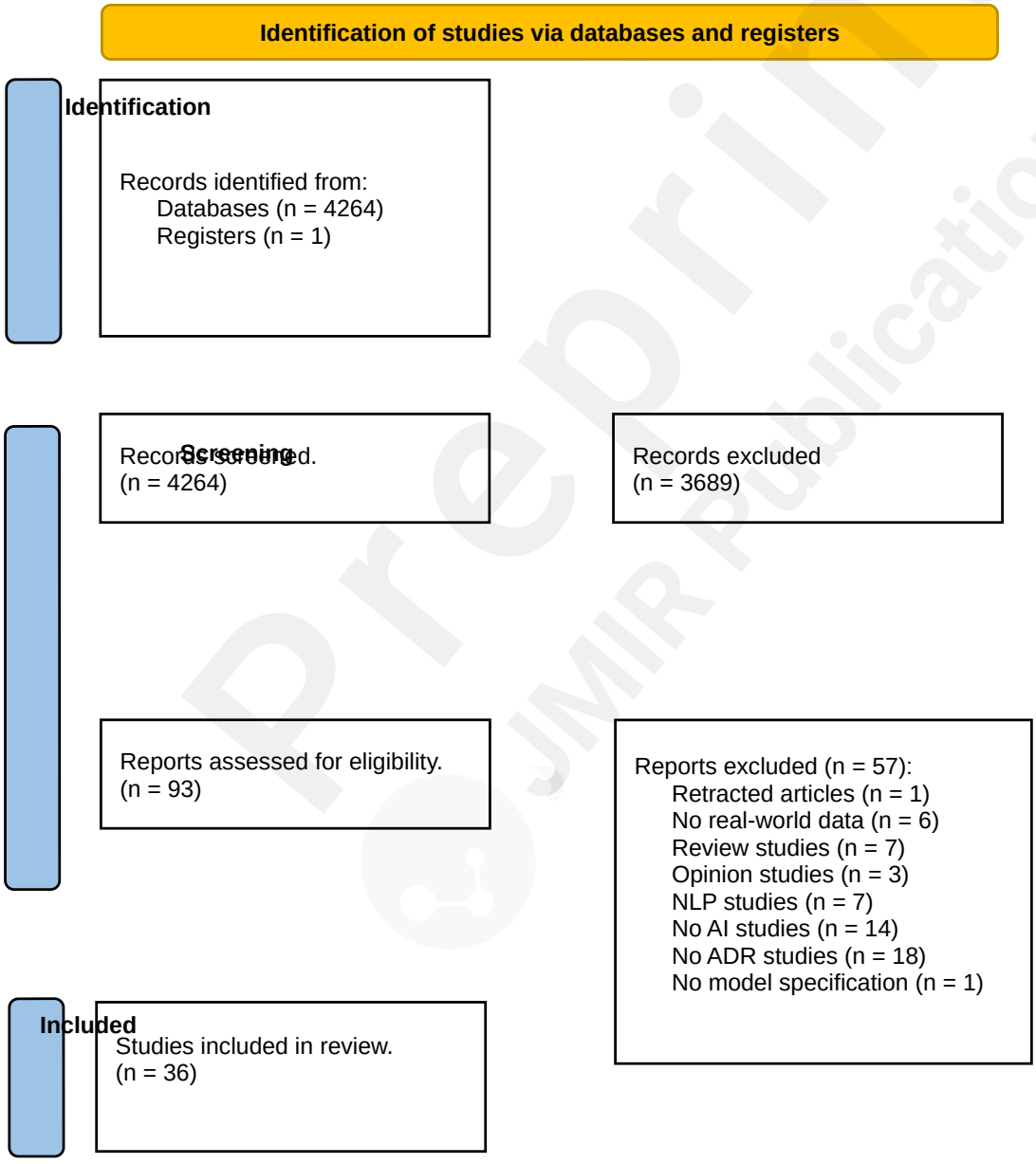


Figure 1 PRISMA flow diagram.

Error: Reference source not found illustrates the connection between the keywords of the included papers in a network. The nodes represent the keywords in the articles and the edges are the

connections between the keywords. The size of the node depends on the frequency of the keyword in the included articles. An interesting result is that AI is mostly represented with the term “machine learning” and RWD with the term “electronic health records”. Contrary, there is a significant variety of terms for describing pharmacovigilance processes.

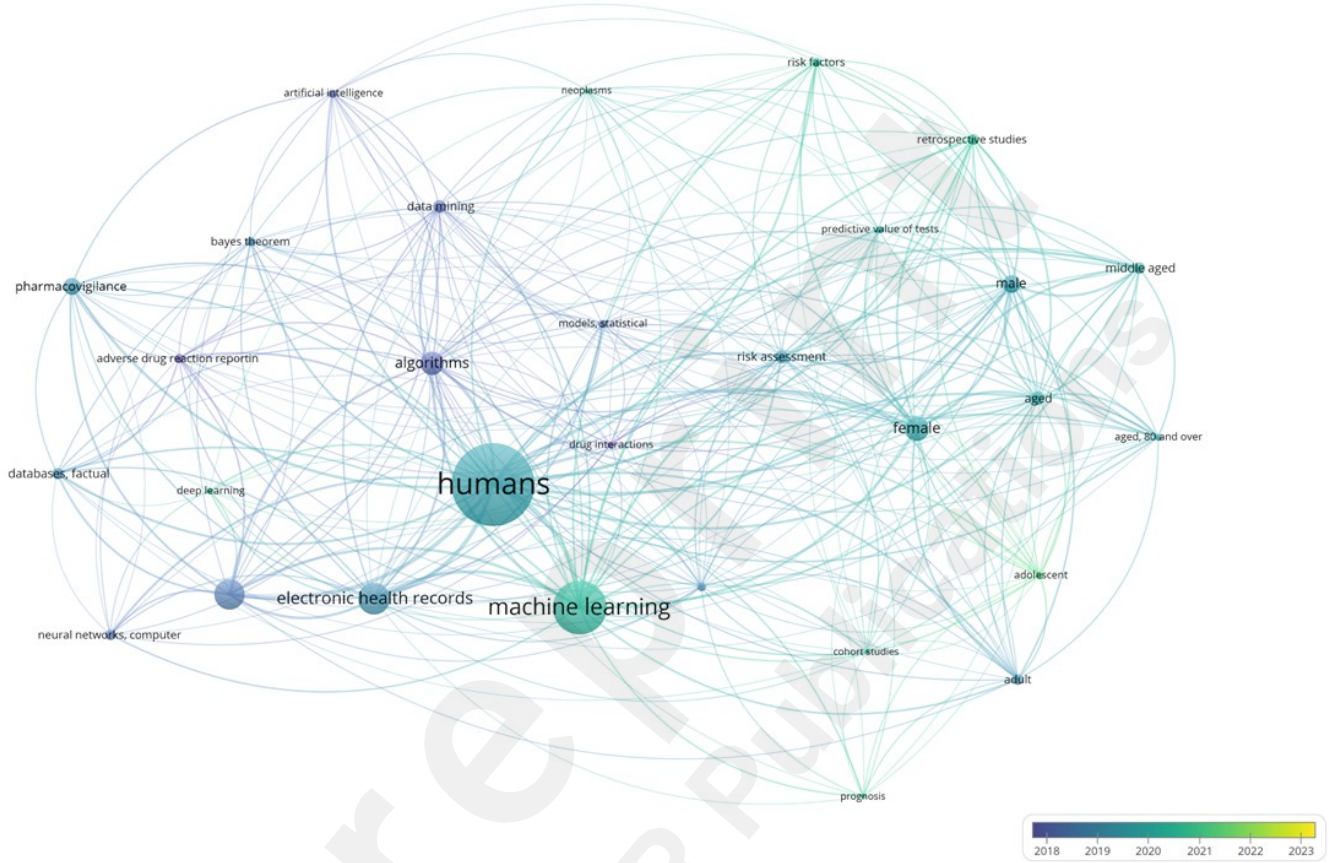


Figure 2 Network visualization of co-occurrence between keywords in included articles in the SLR.

Study characteristics

The included studies were published from 2015 to 2023 where studies are significantly increasing after 2019. Nineteen articles originate from the US, 4 studies from Korea, 4 from the UK, and the rest of the studies are distributed in a variety of other countries (Error: Reference source not found).
In terms of AI, 34 applied only non-symbolic AI, 1 study used only symbolic AI, while 1 study combined the symbolic and non-symbolic AI technical paradigm. Of non-symbolic AI articles, 29 utilized classification tasks whereas 3 of them selected regression algorithms, 3 studies applied causality algorithms (2 of causal inference and 1 of causal discovery) and only one applied association rule mining technique. The association rule mining study [12] follows a mathematical framework called Formal Concept Analysis (FCA) to create association rules between drugs and phenotypes to detect possible adverse drug reactions. Moreover, 6 studies utilized Explainable AI (XAI) techniques, 4 of which used Shapely Additive explanations (SHAP), one Local interpretable model agnostic explanations (LIME) and one tested both approaches.
Regarding RWD (Error: Reference source not found), 28 articles focus on the use of EHRs (from local hospital databases), 4 articles use data from dispensing records from pharmacies, 3 papers employ administrative claims data, 2 focus on patient registries and 1 on insurance claims.

Additionally, a variety of other sources were used along with RWD like drug information databases (n=3), spontaneous reports (n=3), adverse event databases (n=2), ePrescription data (n=2) and genetics and biochemical databases (n=1).
Twenty-three studies used AI for ADR detection and there was a small number of studies that examined ADR assessment (n=4), monitoring (n=2), prevention (n=7) and 2 studies used AI for information collection about ADRs (Error: Reference source not found). The classification studies (Error: Reference source not found) tested several AI techniques with the most used Random Forest algorithm (n=17). However, regression studies (n=3) developed AI models only with XGBoost (n=1) and Logistic Regression (n=2).

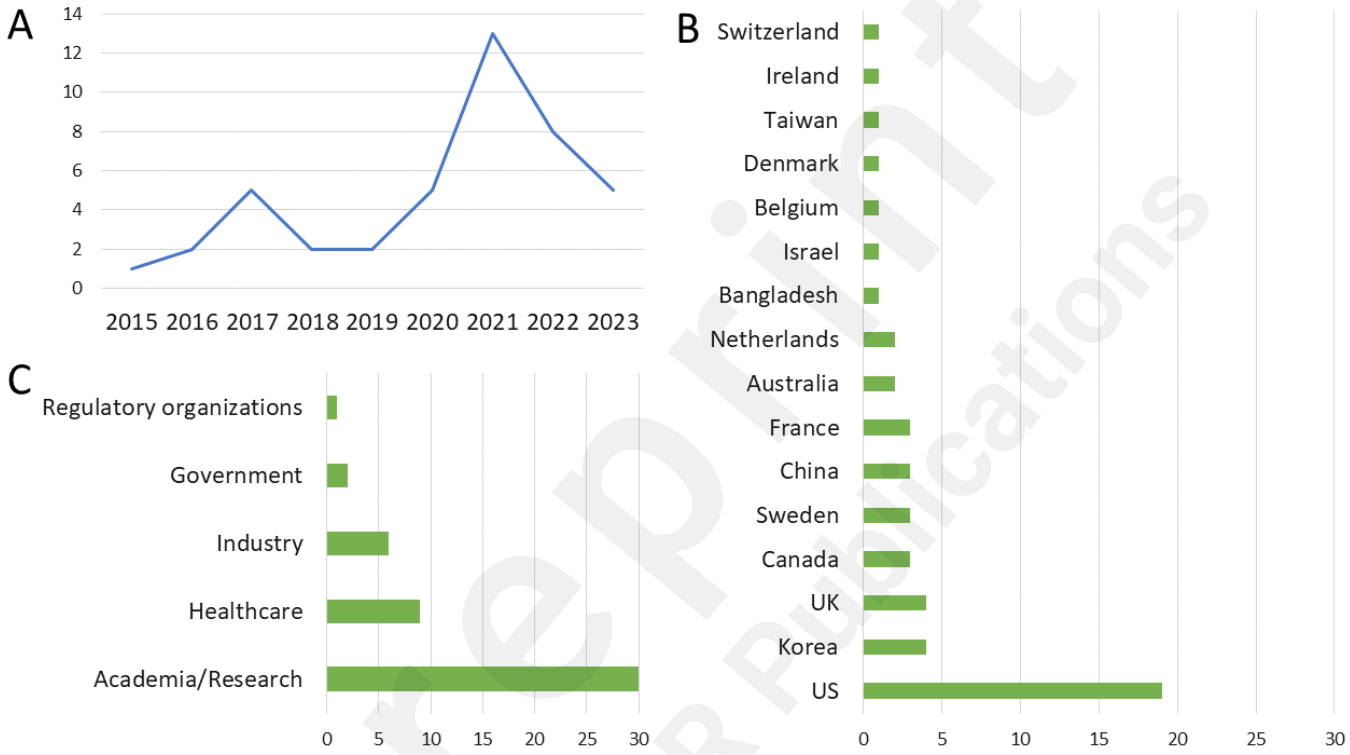


Figure 3 General Bibliographic information about research papers: A. The number of studies through years, B. Country of origin C. Type of organizations that participated in the including studies.

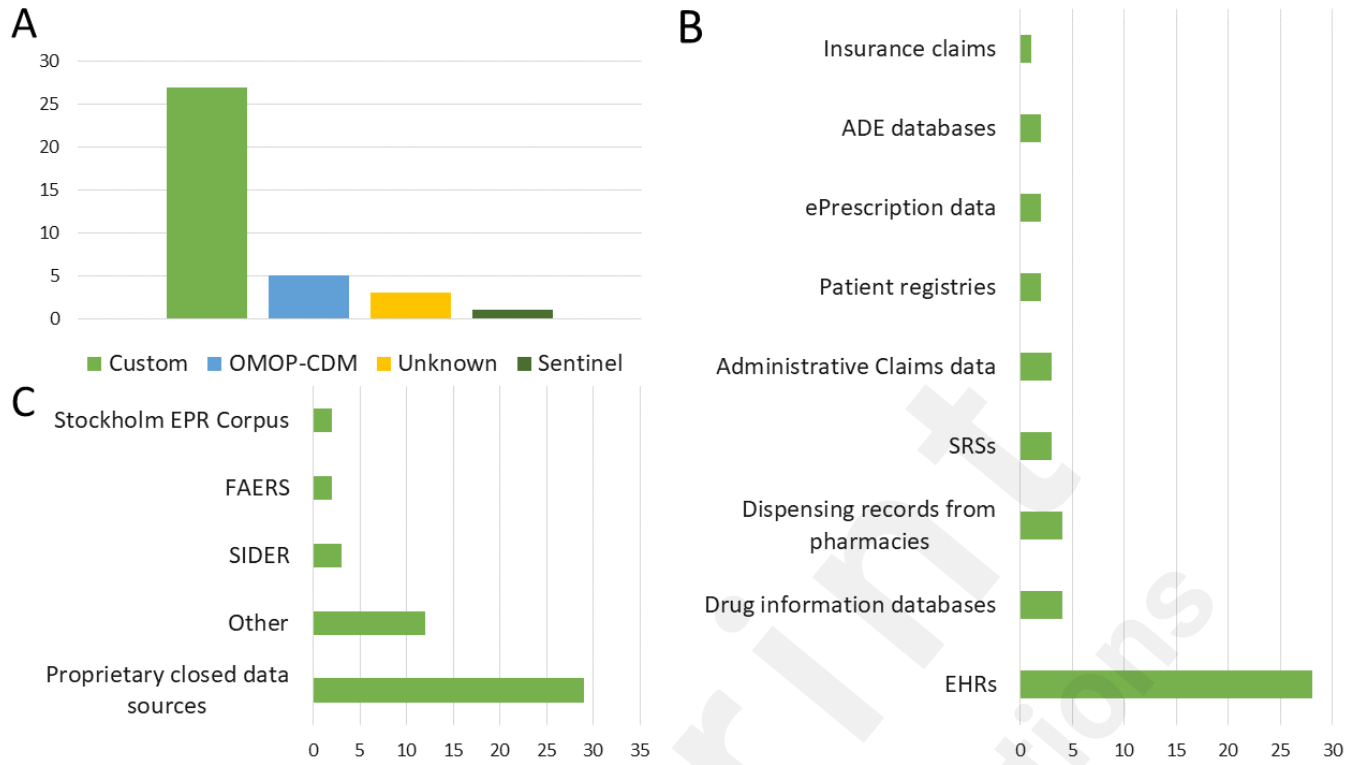


Figure 4 Data description in the included studies A. Studies' distribution based on data models. B. Variety of data sources C. Variety of databases

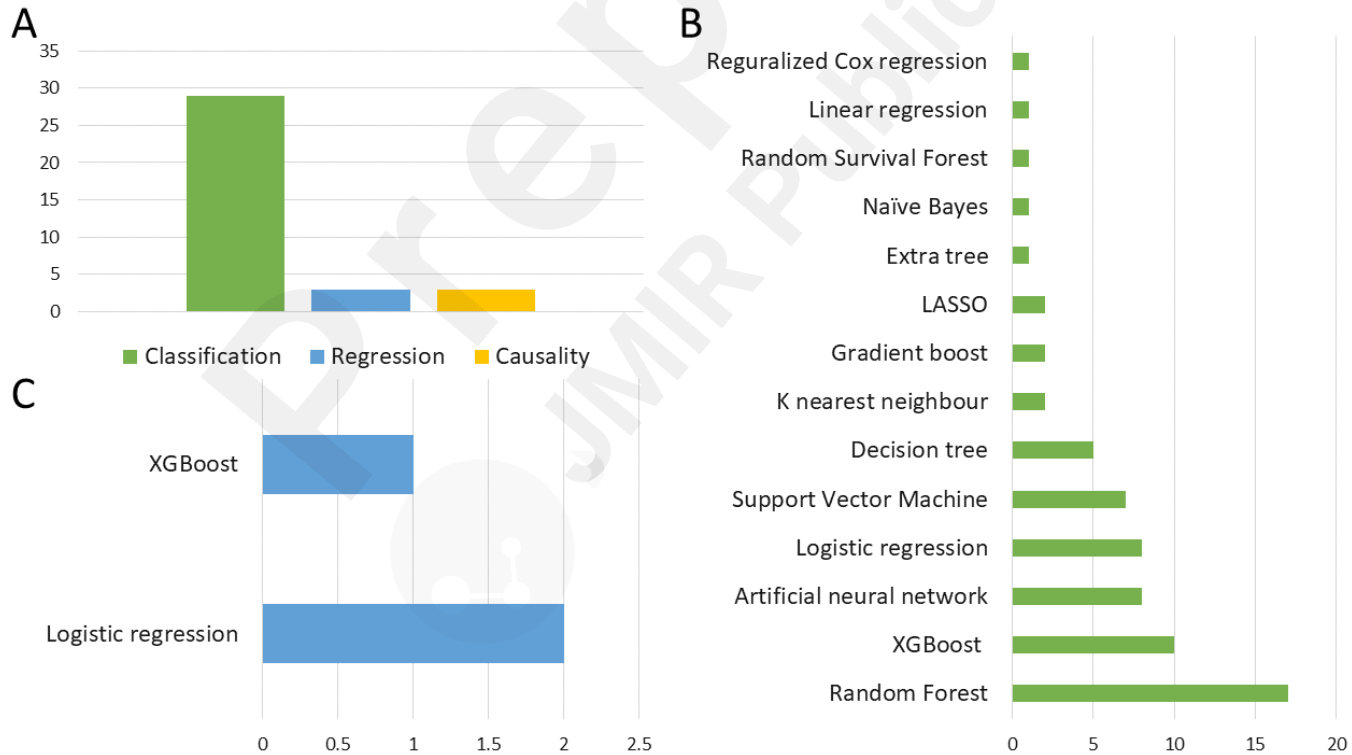


Figure 5 Description of artificial intelligence models A. Distribution of studies based on symbolic AI categories B. Distribution of classification algorithms C. Distribution of regression algorithms.

Thirty-two studies investigated specific drug safety topics, 16 on specific adverse effects, 14 on specific (class of) drug(s), 8 on specific (class of) diseases, 6 on signal detection, 3 on drug interactions, 2 on personalized drug safety, and 1 on vaccine safety (Error: Reference source not found).

Figure 4 illustrates the diversity in data sources that are used in the included studies. Twenty-nine studies select proprietary closed data sources (e.g., specific hospital EHR) for their experiments. Along with EHR data, other data sources were also used (e.g., FAERS and SIDER). Two studies selected the Stockholm EPR (Electronic Patient Record) Corpus. The rest of the RWD data sources (MIMIC, Osteoarthritis Initiative (OAI) dataset) are represented in only 2 studies (1 for every database).

In terms of data models, 27 articles used proprietary data models, 5 articles did not mention any data model, 3 articles used OMOP-CDM, and only 1 the Sentinel model (Error: Reference source not found).

Supplementary figure 1Error: Reference source not found presents the case studies examined from the included articles in the review. As we can see an important number of studies don't work in specific adverse drug reaction case studies. Another significant outcome is that articles contain a variety of case studies, they do not focus on a specific drug, reaction, or indication. We can only observe that chemotherapy drugs and their reactions in different types of cancer are slightly bigger categories in this review.

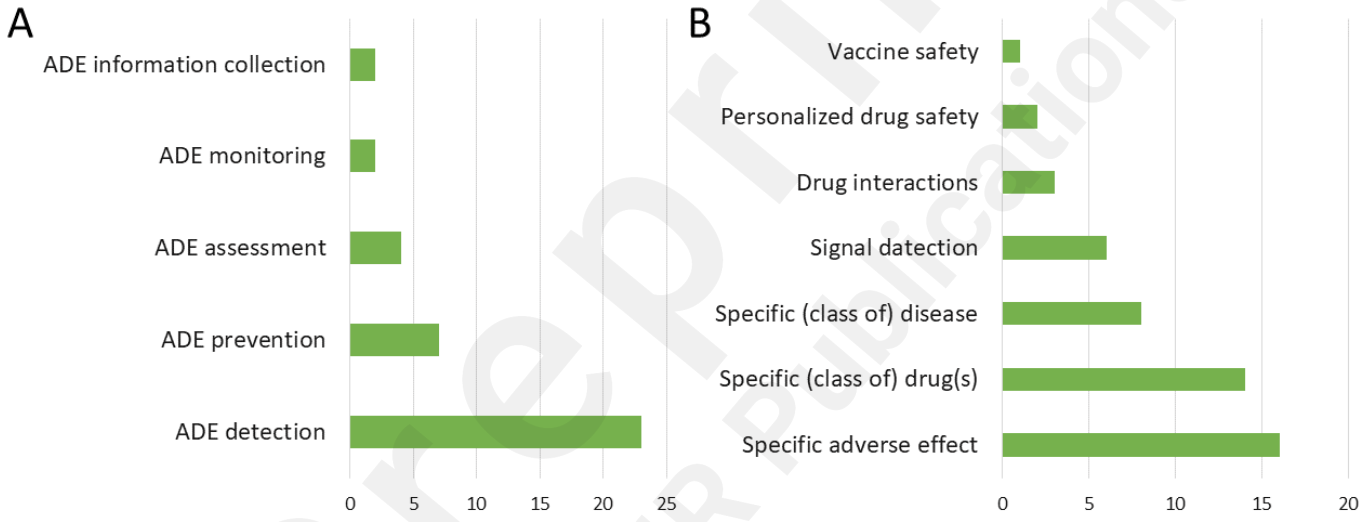


Figure 6 Description of pharmacovigilance processes in the included studies A. Pharmacovigilance core activities B. Specialization of Pharmacovigilance topics.

Although most of the studies used complex AI algorithms (black boxes), like RF (ensemble method) and ANN to construct their prediction models in all ADR categories, there are also many studies that use simple interpretable approaches like logistic regression. Moreover, it is important to highlight that all studies worked on EHR databases except for the ADE assessment category which we also detected one study with a vaccine database.

RWD databases are also used with other types of data, EHRs are mostly combined with SRSs and Drug information databases, vaccine data with ADE databases, and administrative claims data with SRSs. Furthermore, there are studies where different types of observational types are used for an AI model, Dispensing records from pharmacies with EHRs, and administrative claims data.

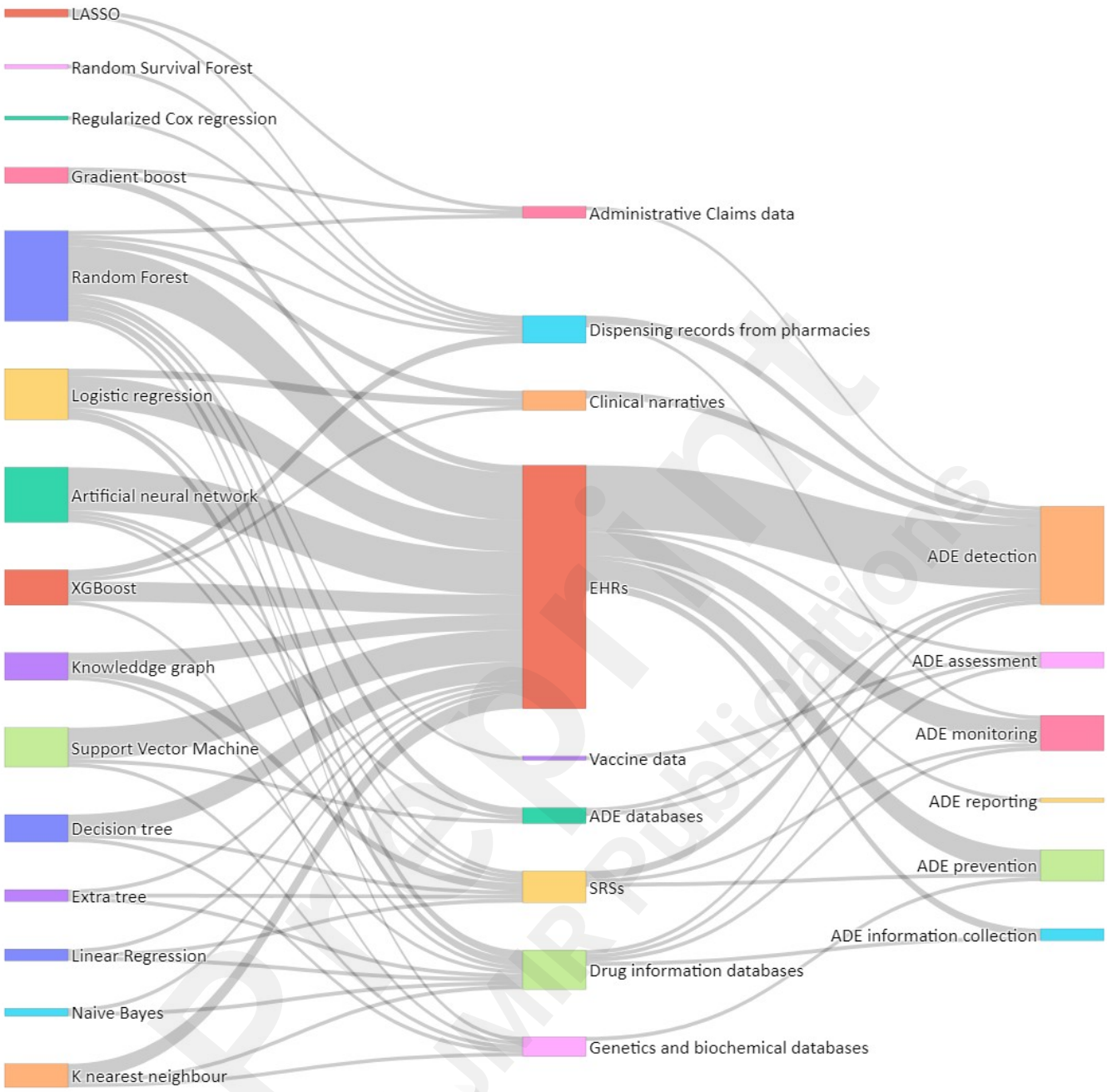


Figure 7 Association between AI models, data sources and drug safety categories

Evaluation results

Only 3 out of 36 studies elaborated in this SLR provide their code openly. Additionally, only 16 studies contain a detailed description of data preprocessing pipelines on real-world data RWD. Moreover, 4 studies evaluate their methodology in the clinical environment (Error: Reference source not found). In terms of trustworthy AI, only 4 studies out of 36 did not reach 50% at the FUTURE AI criteria (Error: Reference source not found). When it comes to the studies that succeeded 75% [13–15], 3 out of 4 studies used external data to evaluate their model which is contained in the Universality criterion (Supplementary table 4).

Table 4 Articles included in the systematic literature review according to the evaluation criteria.

Reference	Year	Code availability	Data Preprocessing	Clinical use	% of FUTURE AI criteria satisfaction

Anastopoulos IN et al. [13]	2021	No	Yes	No	0,75
Ward IR et al. [16]	2021	No	Yes	No	0,67
Zhang W et al. [17]	2020	No	Yes	No	0,50
Kim Y et al. [18]	2021	No	No	No	0,67
Morel M et al. [19]	2020	Yes	No	Yes	0,67
Zou B et al. [20]	2021	No	No	Yes	0,67
Liu et al. [21]	2018	No	No	Yes	0,42
Zhu X et al. [22]	2022	No	Yes	No	0,67
Kidwai-Khan F et al. [23]	2022	No	Yes	No	0,58
Sharma V et al. [24]	2022	No	No	No	0,58
On J et al. [25]	2022	No	Yes	No	0,50
Datta A et al. [26]	2021	No	No	No	0,58
Bagattini F et al. [14]	2019	Yes	Yes	No	0,75
Gibson TB et al. [27]	2021	No	Yes	No	0,58
Jeong E et al. [28]	2018	No	Yes	No	0,67
Zhao J et al. [29]	2015	No	Yes	No	0,58
Zhao J et al. [30]	2016	No	Yes	No	0,42
Segal G et al. [15]	2019	No	No	Yes	0,75
Boland MR et al. [31]	2017	No	No	No	0,67
Wang Z et al. [32]	2021	No	No	No	0,58
Li C et al. [33]	2022	No	No	No	0,58
Jin S et al. [34]	2020	No	No	No	0,67
Hansen PW et al. [35]	2016	No	No	No	0,50
Mosa ASM et al. [36]	2021	No	No	No	0,67
Herrin J et al. [37]	2021	No	No	No	0,50
Pichardo D et al. [38]	2022	No	No	No	0,67
Puzhko S et al. [39]	2021	No	No	No	0,58
Souissi SB et al. [40]	2017	No	No	Yes	0,42
Personeni G et al. [12]	2017	No	No	No	0,42
Zhou Y et al. [41]	2020	Yes	No	No	0,67
Goyal J et al. [42]	2023	No	No	No	0,58

Wang Y et al. [43]	2023	No	Yes	No	0,58
Hughes JH et al. [44]	2023	No	Yes	No	0,58
Sharma V et al. [45]	2023	No	Yes	No	0,67
Akimoto H et al. [46]	2023	No	Yes	No	0,58
Zhang J et al. [47]	2022	No	Yes	No	0,58

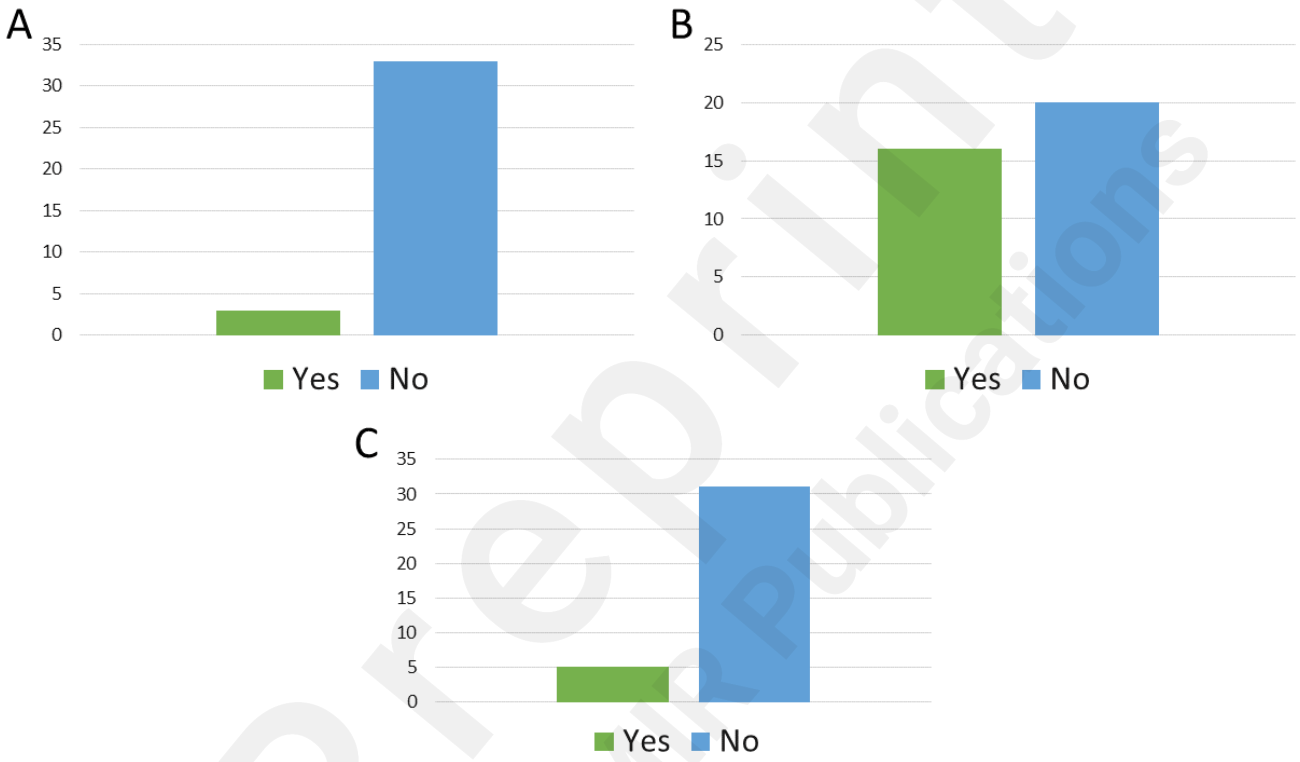


Figure 8 A. Code availability of studies in public registries B. Description of data preprocessing pipeline in included studies C. Clinical validation of the AI models and pipelines.

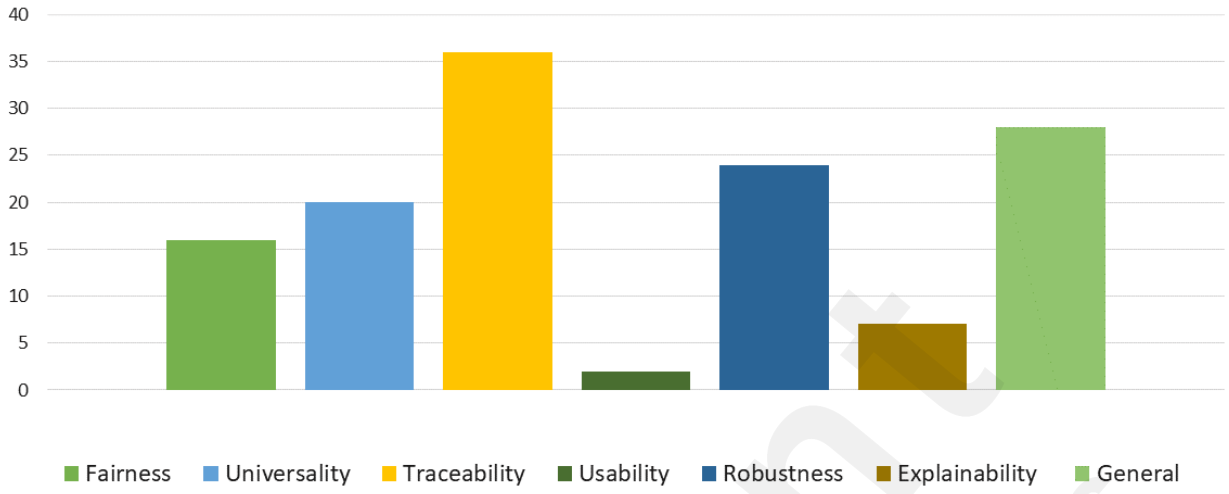


Figure 9 Description of FUTURE AI guidelines.

Interesting results of individual studies

It's important to mention 2 studies that successfully combined AI and self-controlled case series (SCCS) model for ADR detection should be mentioned. Morel et al. [19] introduced the ConvSCCS model where the self-controlled case series (SCCS) model is enriched with a convolutional neural network (CNN). Thus, the ConvSCCS considers a few longitudinal data dimensions (drug exposure) from observational data and can predict a potential ADR without a prior definition of risk windows, which is mandatory in SCCS models. ConvSCCS was tested in glucose-lowering drugs and the risk of bladder cancer case study. Another interesting advantage of the model shown by the results is that the model is useful in high-dimensional data with a reduced amount of data preprocessing. Zhang et al. [17] developed the neural self-controlled case series (NSCCS) to detect probable drug interactions and time-invariant confounders. The NSCCS model was tested in the OMOP reference dataset [48]. The ConvSCCS and NSCCS models outperformed the performances of other SCCS statistical models that are compared. ConvSCCS succeeded in precision and computational speed and NSCCS succeeded AUC score of 0.779.

Furthermore, Kidwai-Khan et al. [23] focused on the prediction improvement of Preventable adverse events (PAE) by integrating an AI decision support tool, EHRs with genetic data (presence/absence of gene contraindicating with a person's medication). It's the only study in this review that combines EHRs with genetic data and one out of the 4 studies that use XAI methods. Additionally, all the AI models achieved high scores at the evaluation metrics (>95%).

There are also a few ADR prediction studies with innovative ideas on feature preprocessing. Jeong et al. [28] which the features of the ML prediction model are calculated from the Comparison of Extreme Laboratory Test results (CERT), Comparison of Extreme Abnormality Ratio (CLEAR), and Prescription pattern Around Clinical Event (PACE) algorithms as inputs and determine whether a drug–laboratory event pair is associated. A different approach is proposed by Wang et al. [32] where the authors try to face the observational data low-quality problems (e.g., missing data) by creating patient embeddings and treating patients as bags with the various number of feature-value pairs, called instances. Thus, the final AI model (AMI-Net3) succeeds with exceptional performance results. Chen et al. [49] also proposed an embedding methodology, called PHASE (PHysiologicAl Signal Embeddings). This study proved that the training of deep embedding models on physiological signals could lead to better forecast of adverse outcomes and this methodology enables data transferability through PHASE models.

Only one of the studies developed an application for the prediction of ADRs. Mosa et al. [36] leveraged the interoperability of the decision tree ML model and based on their results designed a rule-based mobile application for the risk of specific adverse drug reactions and indications.
A totally different approach to ADR prediction is introduced by Liu et al. [21]. In this study, authors applied an ML method to develop a prediction model for osteoarthritis ADR in analgetic drugs. Afterwards, through explainability techniques detect patients who might be prescribed analgesic drugs without osteoarthritis ADR risk. The diversity of this study is addressed in a different scope that proposes, instead of predicting ADR based on the patient's history, search for the characteristics that will make the patient suitable for a medication focused on the presence/absence of an ADR.
Recently, the causal machine learning (CML) paradigm has been introduced for PV through the studies of Wang et al. [43] and Zhang et al. [47] who apply causal inference with average treatment effect and causal discovery with directed acyclic graphs, respectively. Wang et al. used CML models to make a representation of a randomized clinical trial (RCT) with EHR data. Their results successfully detected already well-known but also new medications that can cause the suspected ADR of their case study. Furthermore, Zhang et al. create the causal graph of a drug event combination (DEC) and compare the results from two causal discovery algorithms. Their results showcase the causal discovery algorithms' abilities to explore the mechanisms of the suspect drug that could lead to a potential ADR, uncovering unknown causal links.
It should be noted that only a few works focused on the use of symbolic AI [38] compared to the ones focusing on the use of ML. Finally, Pichardo et al. should be highlighted for combining the use of ontologies and ML, namely combining symbolic and non-symbolic AI. The objective of this study is to examine the performance of a clinically informed framework for the prediction of short-term ADRs.
Notably, very few papers focused on the clinical evaluation of the proposed ML approach. Segal et al. present a clinical decision support system that aims to provide medication error alerts to prevent ADRs [15] with significant results as it seems that 40% of the prescriptions were altered due to the respective alerts. Herrin et al. compared the effectiveness of the proposed ML scheme against an already established clinical practice, namely the use of the HAS-BLED approach to evaluate the potential risk for a specific patient to have gastrointestinal bleeding [37].

Discussion

Principal Results

Summarizing the whole picture, the studies about AI methodologies on RWD for pharmacovigilance purposes have significantly increased in the last 5 years, with most studies published after 2019, with the US making the most publications in this field.
Comparing this review study with three recent reviews of the 2022 to 2023 period in the same field we conclude that only Kaas-Hansen et al. [50] could be considered a study with the exact focus with this review. It included only 7 scientific papers as authors selected studies from 2015 to 2021 and only considered studies with more than 1000 patient records. The results of this study are similar to the results of our SLR in terms of dominant AI solutions (classification) and RWD types (EHR). Finally, it is essential to note that this study, like ours, highlights the lack of wide use of common data models, such as OMOP-CDM.

Pharmacovigilance

The reviewed papers focused on the use of ML for detecting ADRs to confirm whether already known ADRs could have been identified using RWD. Another major theme identified is the prediction of one or more ADRs based on the classification of patients with different characteristics, ultimately aiming to support personalized ADR prevention. However, there is a lack of studies that investigate new potential PV signals. Regarding the investigated ADRs, there is a slightly higher interest in chemotherapy drugs for different types of cancer because this type of therapy tends to cause a lot of serious reactions.

Finally, it is important to highlight that only 4 studies in this review were tested in real-world clinical environments, which leads us to the conclusion that there is a lack of generalizability of AI models or a lack of trust in healthcare professionals in AI models. On the other hand, the trustworthy AI evaluation that was followed based on the FUTURE AI guidelines proved that only a few studies don't satisfy half of the criteria, depicting a relatively high research quality.

Real world data

In terms of RWD, EHRs are the most used data source category. EHRs are multidimensional data that could be crucial in detecting post-marketing ADRs. At least in principle, EHRs could be an invaluable data source to investigate potential drug synergies or interactions in diverse populations. Furthermore, the variety of information in patient records could be an advantage in the creation of multimodal datasets, for example, by integrating biological, signaling pathway, and drug information databases.

On the other hand, using EHRs comes with significant burdens as they contain sensitive personal data, resulting in limited access. MIMIC is the only openly available EHR to researchers, but it is not commonly used in PV (only 2 papers out of 36 use it).

Moreover, it's important to mention that RWD preprocessing is challenging due to its complexity and real-world nature (biases, errors, gaps, noise, etc.). To this end, less than half of the articles describe in detail the data preprocessing step in their pipelines.

Another noteworthy outcome is that widely adopted data models like OMOP-CDM, i2b2, and Sentinel appear sporadically in the studies. This could be attributed to the fact that the use of EHR data and AI models is relatively new. It should be noted though that there are initiatives in this direction (e.g., the APPOLO initiative in the context of the OHDSI ecosystem[e]).

Finally, it should be noted that as RWD comes with a very substantial longitudinal dimension and questionable quality (due to gaps, errors, etc.), RWD exploitation for PV purposes requires the development of new approaches focusing on the exploitation of RWD time-related sequential information. While several attempts to exploit this temporal aspect of RWD can already be identified [14,28–30], still, validating AI/ML algorithms focusing on time-series rationale for PV signal detection is a critical issue.

Artificial intelligence

The detection of potential PV signals is a challenging procedure. As a result, the development of ML models that could support the detection of ADR signals could have a significant impact. So far, we can outline two major approaches for ADR signal detection. The first one focuses on creating an AI tool that could discover any unknown relation between a drug and a condition highlighting a potential causal association. The second one emphasizes AI pipelines for specific ADRs, and the input data of the final ML model are preprocessed based on the prior medical knowledge of the specific DEC.

The AI models identified in this review are generally complex. Ensemble methods like Random Forest are the ones mostly used. There is also a significant number of studies that applied artificial neural networks (ANN). RWD contains a lot of diverse information so the relation between different

[e] https://www.ohdsi.org/2023showcase-110/

features may not be linear. Hence, the use of a black-box model (e.g., ensemble, ANN) is essential for discovering more complicated associations in a dataset than linear ones.
Based on the review papers, there is also a lack of use of explainable AI (XAI) models, i.e., LIME and Shap models. Healthcare professionals highlight the necessity to understand the motifs between AI models' tasks to accept the decisions made by the algorithms. This can lead to the biological translation of the results based on existing knowledge but also could reveal new information about a disease, a medication, etc. In terms of PV, explainable AI models on RWD could bring evidence about unknown confounders in an ADR and they could provide more informative results for PV experts about the causes of a potential PV signal. Although XAI methods' results are tested extensively in the healthcare domain, we found only 6 recent studies that were applied in the PV domain [16,18,22,23,49,51] (2 in 2021, 2 in 2022, and 2 in 2023). Another novel approach that is discussed extensively in the explainability field is the newly introduced causal machine/deep learning (CML/CDL). It is a combination of AI and causal inference that uncovers the underlying cause-and-effect relationships among variables. The complexity of RWD is a challenge that CML could potentially solve more efficiently by providing meaningful explanations about causal relationships between variables [52]. These innovative AI models could be a good hypothesis for future work because they seek and present the relationships between different variables in RWD sources with a more informative structure than traditional AI models and they have already been applied efficiently in pharmacological treatment patterns [53]. In this review, there are only three studies that applied CDL to EHR data [20,43,47].
Lastly, a major problem that is identified based on the SLR findings is the lack of code availability. This issue hinders the reproducibility of the model to further testing in different datasets and leads to questioning of developed AI models' robustness and generalizability.

Strengths and Limitations

The strengths of this review include the use of a considerable number of studies (36) which present a thorough knowledge of the specific scientific field. Besides, we compare our review with the last more recent review in the same field and we analyzed the difference between them.

Nevertheless, this systematic review also has several limitations. First, because of the variation in the articles' methodology, we were unable to conduct any kind of meta-analysis on the quantitative results. Finally, it is important to mention the limited number of symbolic AI studies in this review [12,40,54,55]. The construction of a knowledge graph usually requires the use of text mining procedures like NLP and focused on real text like clinical notes. As we exclude NLP from our query, we assume that this led to the narrow number of symbolic AI articles in our review.

Current gaps and potential future work paths

Detecting new PV signals via ML approaches requires evidence of a causal association between the suspect drug and the reaction. To support this, PV professionals can use explainable AI models. To this end, further investigation of CML/CDL approaches could be a very impactful line of research for the identification of PV signals upon RWD.
Another gap identified in this SLR that could indicate future work paths could be multitasking learning approaches. Multitask learning is an ML methodology that takes as input one dataset and can execute more than one prediction task. RWD, like EHRs, are rich data sources that could support more than one task (e.g., PV and pharmacoepidemiology). For instance, a multitask learning model could predict an adverse event, the severity of an adverse event, and the possibility of the same adverse event from other drugs in a patient.
Furthermore, combining ML approaches with symbolic AI is a line of work that offers further potential for exploration. Combining ML with ontologies and automatic reasoning upon Knowledge Graphs (KG) could enable new AI approaches (e.g., neuro-symbolic AI) and provide new insights

based on well-established expert knowledge formed as a KG. Moreover, using ontologies/KG could support the integration with other kinds of data sources (e.g., data sources with low-level biochemical or pharmacokinetics/pharmacodynamics information, signaling pathway information, etc.)
Finally, exploiting the currently formed federated data networks could also be an interesting future area of work. For example, EHDEN is currently setting up a network of more than 180 data partners across Europe using OMOP-CDM as their main data model. The adoption of OMOP-CDM and the potential exploitation of such data networks would add significant prospects to the potential AI models used for PV.

Conclusion

In this paper, we reviewed scientific papers focusing on AI approaches to structured RWD for PV purposes. As a key finding, it should be noted that most models are not designed for PV signal detection but for personalized ADR prediction. Furthermore, explainable AI methods and causal ML/ deep learning are not investigated in-depth. Moreover, there are no identified gold standard methodologies for data preprocessing of structured RWD for PV. Finally, an evaluation of the already developed AI models in external data is difficult because of code unavailability and lack of data access.
Therefore, there is an essential need for more informative and explainable AI models that can be validated on external datasets and for a more detailed description of RWD preprocessing pipelines and methods to examine potential PV signals in clinical practice. Implementing AI approaches on RWD could tackle the problems of PV signaling underreporting and support the vision of personalized ADR management.

Conflicts of Interest

None declared.

Abbreviations

ADE: Adverse events
ADR: Adverse drug reactions
AI: Artificial Intelligence
ANN: Artificial neural network
CDL: Causal deep learning
CML: Causal machine learning
CNN: convolutional neural network
DEC: Drug-Event Combination
EHR: Electronic health records
ML: Machine Learning
NLP: Natural Language Processing
PAE: Preventable adverse events
PHASE: Physiologic Al Signal Embeddings
PV: Pharmacovigilance
RF: Random Forest
RWD: Real-world data
RWE: Real-world evidence
SCCS: self-controlled case series
XAI: Explainable AI

Appendix

Supplementary figure 1 Drugs, reactions, and indications in the included studies A. Drugs B. Indications C. Reactions

Supplementary table 1 Mapping criteria for the studies of systematic literature review

Supplementary table 2 Classification of studies based on the mapping criteria.

Supplementary table 3 FUTURE AI guidelines/criteria description

Supplementary table 4 Classification of studies based on the FUTURE AI criteria.

References

1. World Health Organization (WHO) Pharmacovigilance: Ensuring the Safe Use of Medicines. *World Health* **2004**, 1–6.
2. Formica, D.; Sultana, J.; Cutroneo, P.M.; Lucchesi, S.; Angelica, R.; Crisafulli, S. *The Economic Burden of Preventable Adverse Drug Reactions : A Systematic Review of Observational Studies*; Taylor & Francis, 2018; Vol. 0; ISBN 4201706842.
3. Alatawi, Y.M.; Hansen, R.A. Empirical Estimation of Under-Reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). *Expert Opin. Drug Saf.* **2017**, *16*, 761–767, doi:10.1080/14740338.2017.1323867.
4. Patadia, V.K.; Schuemie, M.J.; Coloma, P.M.; Herings, R.; van der Lei, J.; Sturkenboom, M.; Trifirò, G. Can Electronic Health Records Databases Complement Spontaneous Reporting System Databases? A Historical-Reconstruction of the Association of Rofecoxib and Acute Myocardial Infarction. *Front. Pharmacol.* **2018**, *9*, 1–8, doi:10.3389/fphar.2018.00594.
5. Gauffin, O.; Brand, J.S.; Vidlin, S.H.; Sartori, D.; Asikainen, S.; Català, M.; Chalabi, E.; Dedman, D.; Danilovic, A.; Duarte-Salles, T.; et al. Supporting Pharmacovigilance Signal Validation and Prioritization with Analyses of Routinely Collected Health Data: Lessons Learned from an EHDEN Network Study. *Drug Saf.* **2023**, *46*, 1335–1352, doi:10.1007/s40264-023-01353-w.
6. Pfaff, E.R.; Girvin, A.T.; Bennett, T.D.; Bhatia, A.; Brooks, I.M.; Deer, R.R.; Dekermanjian, J.P.; Jolley, S.E.; Kahn, M.G.; Kostka, K.; et al. Identifying Who Has Long COVID in the USA: A Machine Learning Approach Using N3C Data. *Lancet Digit. Heal.* **2022**, *4*, e532–e541, doi:10.1016/S2589-7500(22)00048-6.
7. Yang, Y.; Yuan, Y.; Zhang, G.; Wang, H.; Chen, Y.C.; Liu, Y.; Tarolli, C.G.; Crepeau, D.; Bukartyk, J.; Junna, M.R.; et al. Artificial Intelligence-Enabled Detection and Assessment of Parkinson's Disease Using Nocturnal Breathing Signals. *Nat. Med. 2022 2810* **2022**, *28*, 2207–2215, doi:10.1038/s41591-022-01932-x.
8. Komorowski, M.; Celi, L.A.; Badawi, O.; Gordon, A.C.; Faisal, A.A. The Artificial Intelligence Clinician Learns Optimal Treatment Strategies for Sepsis in Intensive Care. *Nat. Med. 2018 2411* **2018**, *24*, 1716–1720, doi:10.1038/s41591-018-0213-5.
9. Natsiavas, P.; Malousi, A.; Bousquet, C.; Jaulent, M.C.; Koutkias, V. Computational Advances in Drug Safety: Systematic and Mapping Review of Knowledge Engineering Based Approaches. *Front. Pharmacol.* **2019**, *10*, 415, doi:10.3389/FPHAR.2019.00415/BIBTEX.
10. Lekadir, K.; Feragen, A.; Fofanah, A.J.; Frangi, A.F.; Buyx, A.; Emelie, A.; Lara, A.; Porras, A.R.; Chan, A.-W.; Navarro, A.; et al. FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare. **2023**.
11. Lavin, A.; Gilligan-Lee, C.M.; Visnjic, A.; Ganju, S.; Newman, D.; Ganguly, S.; Lange, D.; Baydin, A.G.; Sharma, A.; Gibson, A.; et al. Technology Readiness Levels for Machine Learning Systems. *Nat. Commun.* **2022**, *13*, doi:10.1038/s41467-022-33128-9.
12. Personeni, G.; Bresso, E.; Devignes, M.D.; Dumontier, M.; Smaïl-Tabbone, M.; Coulet, A.

Discovering Associations between Adverse Drug Events Using Pattern Structures and Ontologies. *J. Biomed. Semantics* **2017**, *8*, doi:10.1186/s13326-017-0137-x.

13. Anastopoulos, I.N.; Herczeg, C.K.; Davis, K.N.; Dixit, A.C. Multi-Drug Featurization and Deep Learning Improve Patient-Specific Predictions of Adverse Events. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1–11, doi:10.3390/ijerph18052600.
14. Bagattini, F.; Karlsson, I.; Rebane, J.; Papapetrou, P. A Classification Framework for Exploiting Sparse Multi-Variate Temporal Features with Application to Adverse Drug Event Detection in Medical Records. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, doi:10.1186/s12911-018-0717-4.
15. Segal, G.; Segev, A.; Brom, A.; Lifshitz, Y.; Wasserstrum, Y.; Zimlichman, E. Reducing Drug Prescription Errors and Adverse Drug Events by Application of a Probabilistic, Machine-Learning Based Clinical Decision Support System in an Inpatient Setting. *J. Am. Med. Informatics Assoc.* **2019**, *26*, 1560–1565, doi:10.1093/jamia/ocz135.
16. Ward, I.R.; Wang, L.; Lu, J.; Bennamoun, M.; Dwivedi, G.; Sanfilippo, F.M. Explainable Artificial Intelligence for Pharmacovigilance: What Features Are Important When Predicting Adverse Outcomes? Explainable Artificial Intelligence for Pharmacovigilance. *Comput. Methods Programs Biomed.* **2021**, *212*, doi:10.1016/j.cmpb.2021.106415.
17. Zhang, W.; Kuang, Z.; Peissig, P.; Page, D. Adverse Drug Reaction Discovery from Electronic Health Records with Deep Neural Networks. *Proc. ACM Conf. Heal. Inference, Learn.* **2020**, *2020*, 30, doi:10.1145/3368555.3384459.
18. Kim, Y.; Jang, J.H.; Park, N.; Jeong, N.Y.; Lim, E.; Kim, S.; Choi, N.K.; Yoon, D. Machine Learning Approach for Active Vaccine Safety Monitoring. *J. Korean Med. Sci.* **2021**, *36*, 1–13, doi:10.3346/jkms.2021.36.e198.
19. Morel, M.; Bacry, E.; Gaïffas, S.; Guilloux, A.; Leroy, F. ConvSCCS: Convolutional Self-Controlled Case Series Model for Lagged Adverse Event Detection. *Biostatistics* **2020**, *21*, 758–774, doi:10.1093/biostatistics/kxz003.
20. Zou, B.; Mi, X.; Tighe, P.J.; Koch, G.G.; Zou, F. On Kernel Machine Learning for Propensity Score Estimation under Complex Confounding Structures. *Pharm. Stat.* **2021**, *20*, 752–764, doi:10.1002/pst.2105.
21. Liu, L.; Yu, Y.; Fei, Z.; Li, M.; Wu, F.X.; Li, H.D.; Pan, Y.; Wang, J. An Interpretable Boosting Model to Predict Side Effects of Analgesics for Osteoarthritis. *BMC Syst. Biol.* **2018**, *12*, doi:10.1186/s12918-018-0624-4.
22. Zhu, X.; Hu, J.; Xiao, T.; Huang, S.; Shang, D.; Wen, Y. Integrating Machine Learning with Electronic Health Record Data to Facilitate Detection of Prolactin Level and Pharmacovigilance Signals in Olanzapine-Treated Patients. *Front. Endocrinol. (Lausanne).* **2022**, *13*, 2576, doi:10.3389/FENDO.2022.1011492/BIBTEX.
23. Kidwai-Khan, F.; Rentsch, C.T.; Pulk, R.; Alcorn, C.; Brandt, C.A.; Justice, A.C. Pharmacogenomics Driven Decision Support Prototype with Machine Learning: A Framework for Improving Patient Care. *Front. Big Data* **2022**, *5*, doi:10.3389/FDATA.2022.1059088.
24. Sharma, V.; Kulkarni, V.; Jess, E.; Gilani, F.; Eurich, D.; Simpson, S.H.; Voaklander, D.; Semenchuk, M.; London, C.; Samanani, S. Development and Validation of a Machine Learning Model to Estimate Risk of Adverse Outcomes Within 30 Days of Opioid Dispensation. *JAMA Netw. Open* **2022**, *5*, E2248559, doi:10.1001/jamanetworkopen.2022.48559.
25. On, J.; Park, H.A.; Yoo, S. Development of a Prediction Models for Chemotherapy-Induced Adverse Drug Reactions: A Retrospective Observational Study Using Electronic Health Records. *Eur. J. Oncol. Nurs.* **2022**, *56*, 102066, doi:10.1016/J.EJON.2021.102066.
26. Datta, A.; Flynn, N.R.; Barnette, D.A.; Woeltje, K.F.; Miller, G.P.; Joshua Swamidass, S. Machine Learning Liver-Injuring Drug Interactions with Non-Steroidal Antiinflammatory

Drugs (NSAIDs) from a Retrospective Electronic Health Record (EHR) Cohort. *PLoS Comput. Biol.* **2021**, *17*, doi:10.1371/journal.pcbi.1009053.

27. Gibson, T.B.; Nguyen, M.D.; Burrell, T.; Yoon, F.; Wong, J.; Dharmarajan, S.; Ouellet-Hellstrom, R.; Hua, W.; Ma, Y.; Baro, E.; et al. Electronic Phenotyping of Health Outcomes of Interest Using a Linked Claims-Electronic Health Record Database: Findings from a Machine Learning Pilot Project. *J. Am. Med. Informatics Assoc.* **2021**, *28*, 1507–1517, doi:10.1093/jamia/ocab036.
28. Jeong, E.; Park, N.; Choi, Y.; Park, R.W.; Yoon, D. Machine Learning Model Combining Features from Algorithms with Different Analytical Methodologies to Detect Laboratory-Event-Related Adverse Drug Reaction Signals. *PLoS One* **2018**, *13*, doi:10.1371/journal.pone.0207749.
29. Zhao, J.; Henriksson, A.; Kvist, M.; Asker, L.; Boström, H. Handling Temporality of Clinical Events for Drug Safety Surveillance. *AMIA Annu. Symp. Proc.* **2015**, *2015*, 1371.
30. Zhao, J.; Henriksson, A. Learning Temporal Weights of Clinical Events Using Variable Importance. *BMC Med. Inform. Decis. Mak.* **2016**, *16*, doi:10.1186/s12911-016-0311-6.
31. Boland, M.R.; Polubriaginof, F.; Tatonetti, N.P. Development of A Machine Learning Algorithm to Classify Drugs of Unknown Fetal Effect. *Sci. Rep.* **2017**, *7*, doi:10.1038/s41598-017-12943-x.
32. Wang, Z.; Poon, J.; Wang, S.; Sun, S.; Poon, S. A Novel Method for Clinical Risk Prediction with Low-Quality Data. *Artif. Intell. Med.* **2021**, *114*, doi:10.1016/j.artmed.2021.102052.
33. Li, C.; Chen, L.; Chou, C.; Ngorsuraches, S.; Qian, J. Using Machine Learning Approaches to Predict Short-Term Risk of Cardiotoxicity Among Patients with Colorectal Cancer After Starting Fluoropyrimidine-Based Chemotherapy. *Cardiovasc. Toxicol.* **2022**, *22*, 130–140, doi:10.1007/s12012-021-09708-4.
34. Jin, S.; Kostka, K.; Posada, J.D.; Kim, Y.; Seo, S.I.; Lee, D.Y.; Shah, N.H.; Roh, S.; Lim, Y.-H.; Chae, S.G.; et al. Prediction of Major Depressive Disorder Following Beta-Blocker Therapy in Patients with Cardiovascular Diseases. *J. Pers. Med.* **2020**, *10*, 1–12, doi:10.3390/jpm10040288.
35. Hansen, P.W.; Clemmensen, L.; Sehested, T.S.G.; Fosbøl, E.L.; Torp-Pedersen, C.; Køber, L.; Gislason, G.H.; Andersson, C. Identifying Drug-Drug Interactions by Data Mining: A Pilot Study of Warfarin-Associated Drug Interactions. *Circ. Cardiovasc. Qual. Outcomes* **2016**, *9*, 621–628, doi:10.1161/CIRCOUTCOMES.116.003055.
36. Mosa, A.S.M.; Rana, M.K.Z.; Islam, H.; Mosharraf Hossain, A.K.M.; Yoo, I. A Smartphone-Based Decision Support Tool for Predicting Patients at Risk of Chemotherapy-Induced Nausea and Vomiting: Retrospective Study on App Development Using Decision Tree Induction. *JMIR mHealth uHealth* **2021**, *9*, doi:10.2196/27024.
37. Herrin, J.; Abraham, N.S.; Yao, X.; Noseworthy, P.A.; Inselman, J.; Shah, N.D.; Ngufor, C. Comparative Effectiveness of Machine Learning Approaches for Predicting Gastrointestinal Bleeds in Patients Receiving Antithrombotic Treatment. *JAMA Netw. Open* **2021**, *4*, doi:10.1001/jamanetworkopen.2021.10703.
38. Pichardo, D.; Michael, R.; Mercer, M.; Korina, N.; Onukwugha, E. Utility of a Clinically Guided Data-Driven Approach for Predicting Breast Cancer Complications: An Application Using a Population-Based Claims Data Set. *JCO Clin. Cancer Informatics* **2022**, doi:10.1200/cci.21.00191.
39. Puzhko, S.; Schuster, T.; Barnett, T.A.; Renoux, C.; Munro, K.; Barber, D.; Bartlett, G. Difference in Patterns of Prescribing Antidepressants Known for Their Weight-Modulating and Cardiovascular Side Effects for Patients with Obesity Compared to Patients with Normal

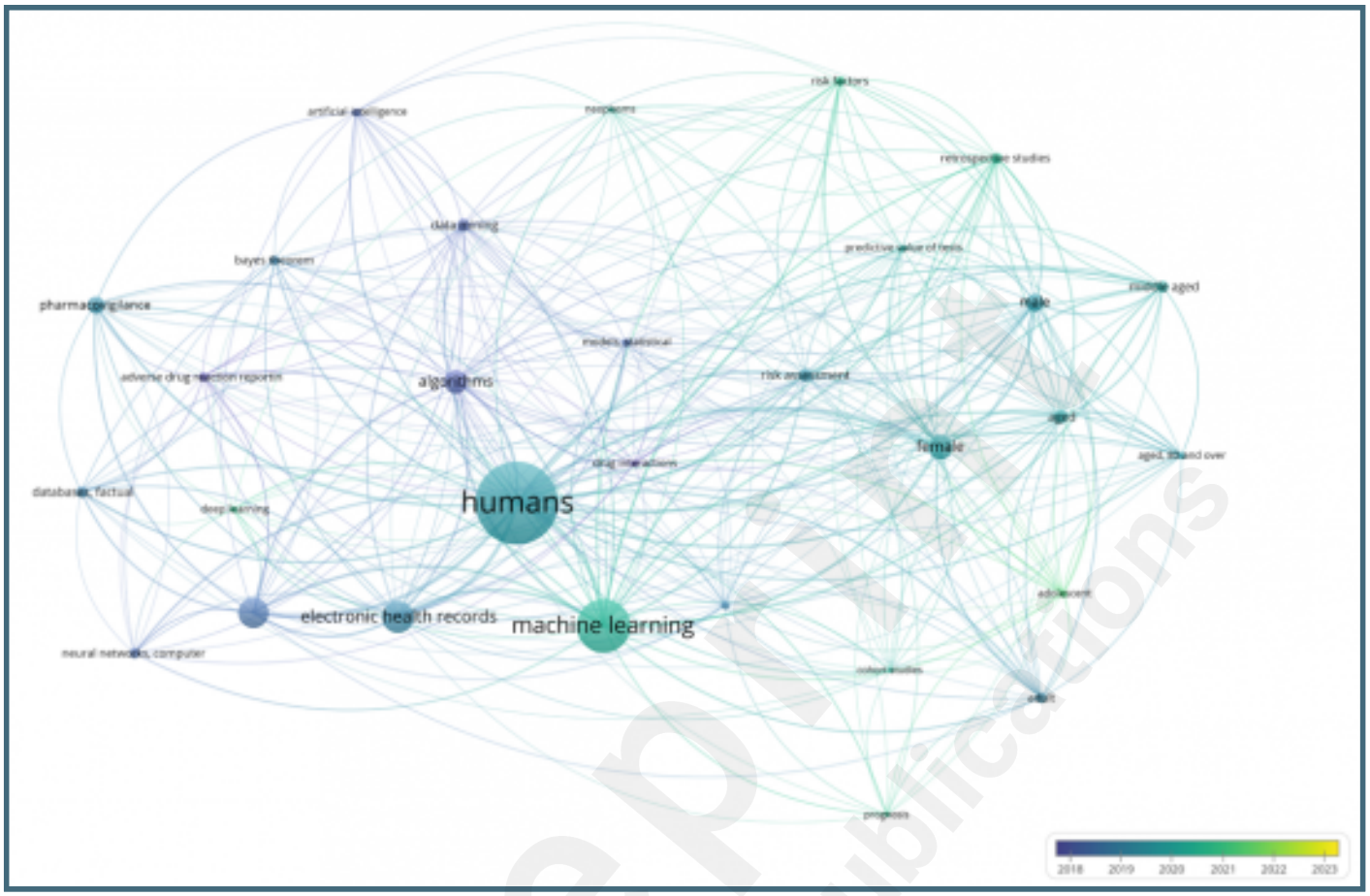
Weight. *J. Affect. Disord.* **2021**, *295*, 1310–1318, doi:10.1016/j.jad.2021.08.018.
40. Souissi, S. Ben; Abed, M.; Elhiki, L.; Fortemps, P.; Pirlot, M. Reducing the Toxicity Risk in Antibiotic Prescriptions by Combining Ontologies with a Multiple Criteria Decision Model. *AMIA Annu. Symp. Proc.* **2017**, *2017*, 1625.
41. Zhou, Y.; Hou, Y.; Hussain, M.; Brown, S.A.; Budd, T.; Wilson Tang, W.H.; Abraham, J.; Xu, B.; Shah, C.; Moudgil, R.; et al. Machine Learning–Based Risk Assessment for Cancer Therapy–Related Cardiac Dysfunction in 4300 Longitudinal Oncology Patients. *J. Am. Heart Assoc.* **2020**, *9*, doi:10.1161/JAHA.120.019628.
42. Goyal, J.; Ng, D.Q.; Zhang, K.; Chan, A.; Lee, J.; Zheng, K.; Hurley-Kim, K.; Nguyen, L.; He, L.; Nguyen, M.; et al. Using Machine Learning to Develop a Clinical Prediction Model for SSRI-Associated Bleeding: A Feasibility Study. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, doi:10.1186/s12911-023-02206-3.
43. Wang, Y.; Ma, J.; Ma, S.; Wang, J.; Li, J. Causal Evaluation of Post-Marketing Drugs for Drug-Induced Liver Injury from Electronic Health Records. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* **2023**, *2023*, 1–4, doi:10.1109/EMBC40787.2023.10340721.
44. Hughes, J.H.; Tong, D.M.H.; Burns, V.; Daly, B.; Razavi, P.; Boelens, J.J.; Goswami, S.; Keizer, R.J. Clinical Decision Support for Chemotherapy-Induced Neutropenia Using a Hybrid Pharmacodynamic/Machine Learning Model. *CPT Pharmacometrics Syst. Pharmacol.* **2023**, *12*, 1764–1776, doi:10.1002/psp4.13019.
45. Sharma, V.; Joon, T.; Kulkarni, V.; Samanani, S.; Simpson, S.H.; Voaklander, D.; Eurich, D. Predicting 30-Day Risk from Benzodiazepine/Z-Drug Dispensations in Older Adults Using Administrative Data: A Prognostic Machine Learning Approach. *Int. J. Med. Inform.* **2023**, *178*, doi:10.1016/j.ijmedinf.2023.105177.
46. Akimoto, H.; Hayakawa, T.; Nagashima, T.; Minagawa, K.; Takahashi, Y.; Asai, S. Detection of Potential Drug-Drug Interactions for Risk of Acute Kidney Injury: A Population-Based Case-Control Study Using Interpretable Machine-Learning Models. *Front. Pharmacol.* **2023**, *14*, doi:10.3389/fphar.2023.1176096.
47. Zhang, J.; Kummerfield, E.; Hultman, G.; Drawz, P.E.; Adam, T.J.; Simon, G.; Melton, G.B. Application of Causal Discovery Algorithms in Studying the Nephrotoxicity of Remdesivir Using Longitudinal Data from the EHR. *AMIA ... Annu. Symp. proceedings. AMIA Symp.* **2022**, *2022*, 1227–1236.
48. Stang, P.E.; Ryan, P.B.; Racoosin, J.A.; Overhage, J.M. Research and Reporting Methods Annals of Internal Medicine Advancing the Science for Active Surveillance : Rationale and Design for the Observational Medical Outcomes Partnership OF. *Ann. Intern. Med.* **2010**, *153*, 600–608.
49. Chen, H.; Lundberg, S.M.; Erion, G.; Kim, J.H.; Lee, S.I. Forecasting Adverse Surgical Events Using Self-Supervised Transfer Learning for Physiological Signals. *npj Digit. Med.* **2021**, *4*, doi:10.1038/s41746-021-00536-y.
50. Kaas-Hansen, B.S.; Gentile, S.; Caioli, A.; Andersen, S.E. Exploratory Pharmacovigilance with Machine Learning in Big Patient Data: A Focused Scoping Review. *Basic Clin. Pharmacol. Toxicol.* **2023**, *132*, 233–241, doi:10.1111/bcpt.13828.
51. Sharma, V.; Kulkarni, V.; Eurich, D.T.; Kumar, L.; Samanani, S. Safe Opioid Prescribing: A Prognostic Machine Learning Approach to Predicting 30-Day Risk after an Opioid Dispensation in Alberta, Canada. *BMJ Open* **2021**, *11*, doi:10.1136/bmjopen-2020-043964.
52. Prosperi, M.; Guo, Y.; Sperrin, M.; Koopman, J.S.; Min, J.S.; He, X.; Rich, S.; Wang, M.; Buchan, I.E.; Bian, J. Causal Inference and Counterfactual Prediction in Machine Learning for

Actionable Healthcare. *Nat. Mach. Intell.* **2020**, *2*, 369–375, doi:10.1038/s42256-020-0197-y.
53. Belthangady, C.; Giampanis, S.; Jankovic, I.; Stedden, W.; Alves, P.; Chong, S.; Knott, C.; Norgeot, B. Causal Deep Learning Reveals the Comparative Effectiveness of Antihyperglycemic Treatments in Poorly Controlled Diabetes. *Nat. Commun.* **2022**, *13*, 1–10, doi:10.1038/s41467-022-33732-9.
54. Jiang, G.; Liu, H.; Solbrig, H.R.; Chute, C.G. Mining Severe Drug-Drug Interaction Adverse Events Using Semantic Web Technologies: A Case Study. *BioData Min.* **2015**, *8*, doi:10.1186/S13040-015-0044-6.
55. Pacaci, A.; Gonul, S.; Sinaci, A.A.; Yuksel, M.; Erturkmen, G.B.L. A Semantic Transformation Methodology for the Secondary Use of Observational Healthcare Data in Postmarketing Safety Studies. *Front. Pharmacol.* **2018**, *9*, 435, doi:10.3389/FPHAR.2018.00435/BIBTEX.

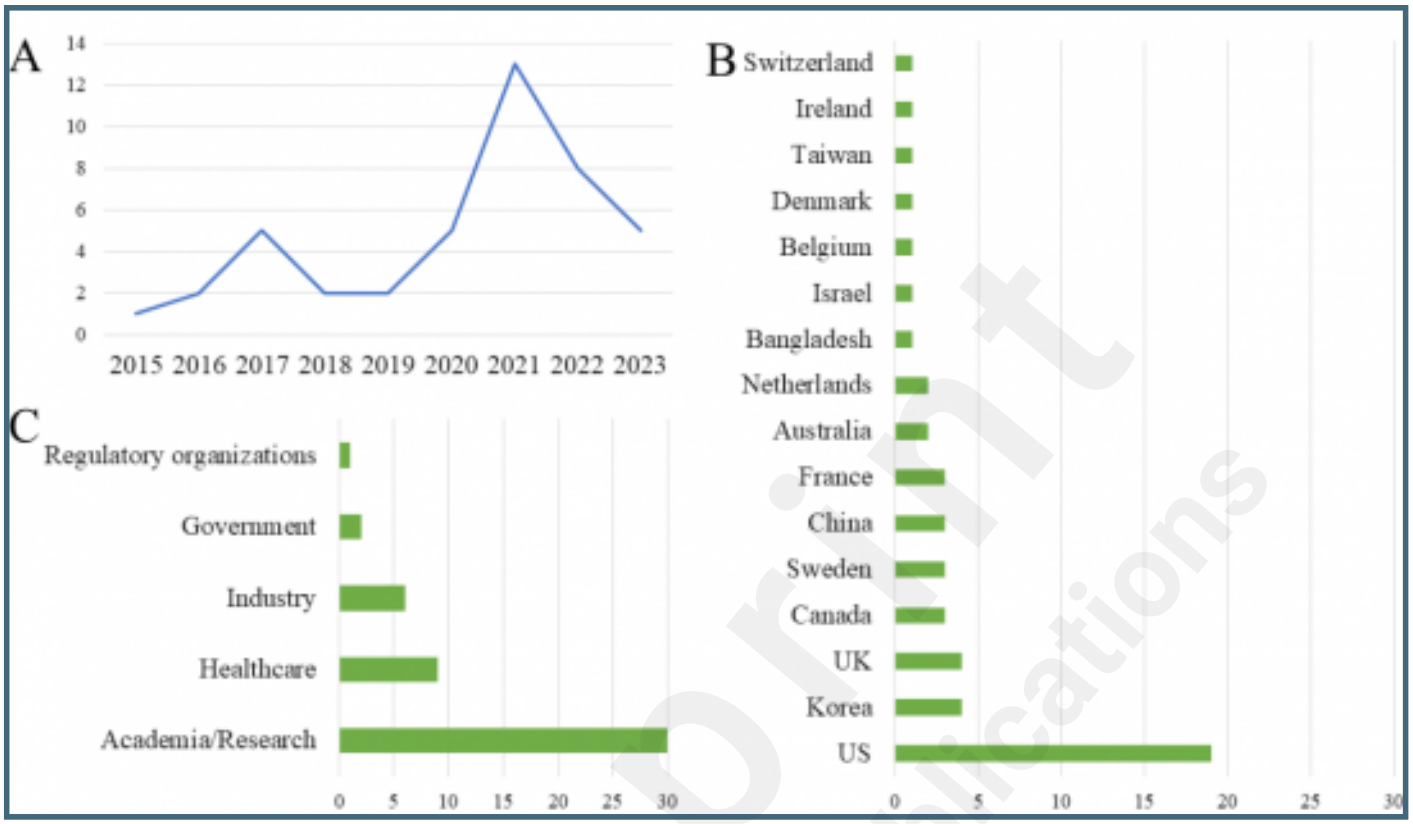
Supplementary Files

Figures

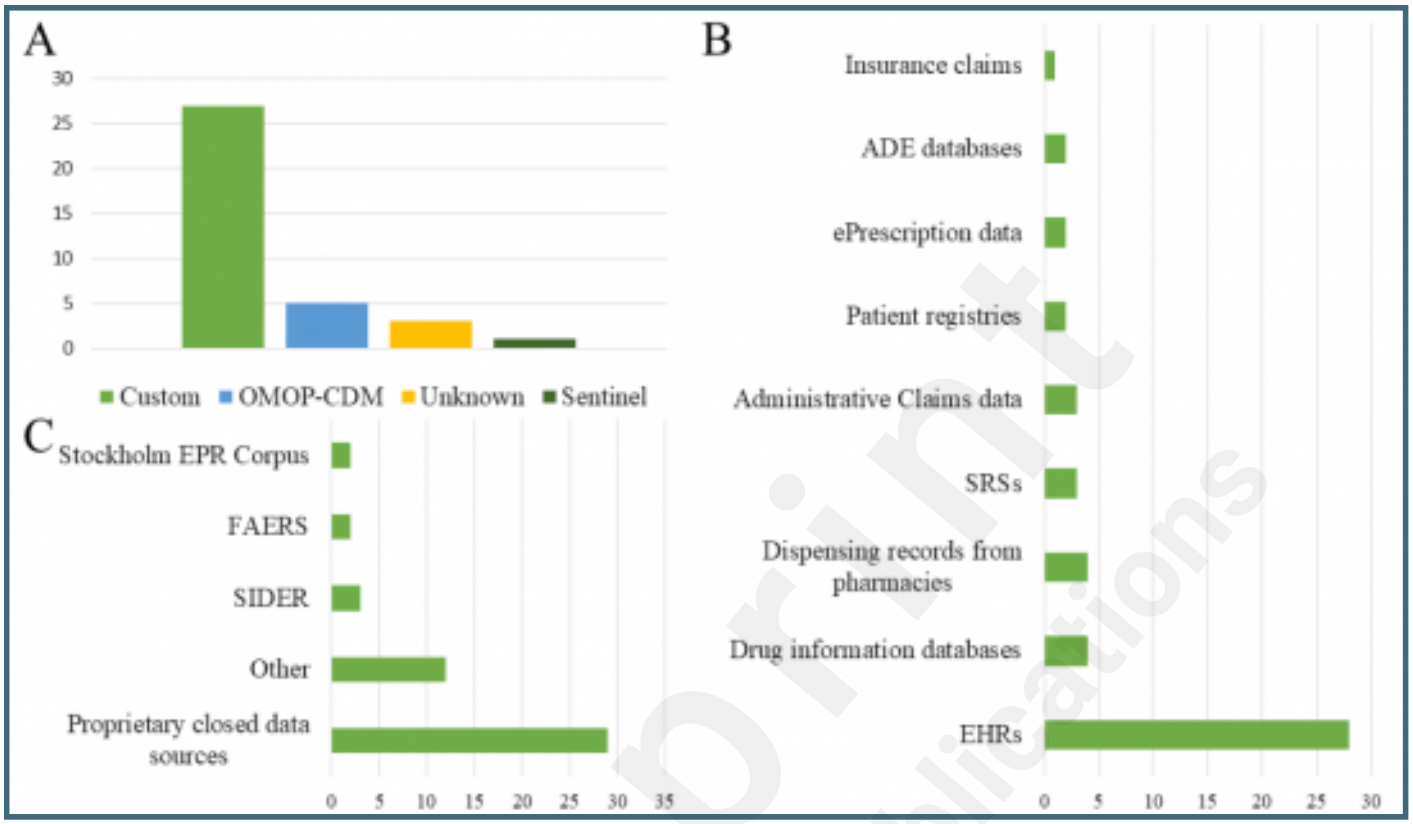
Network visualization of co-occurrence between keywords in included articles in the SLR.



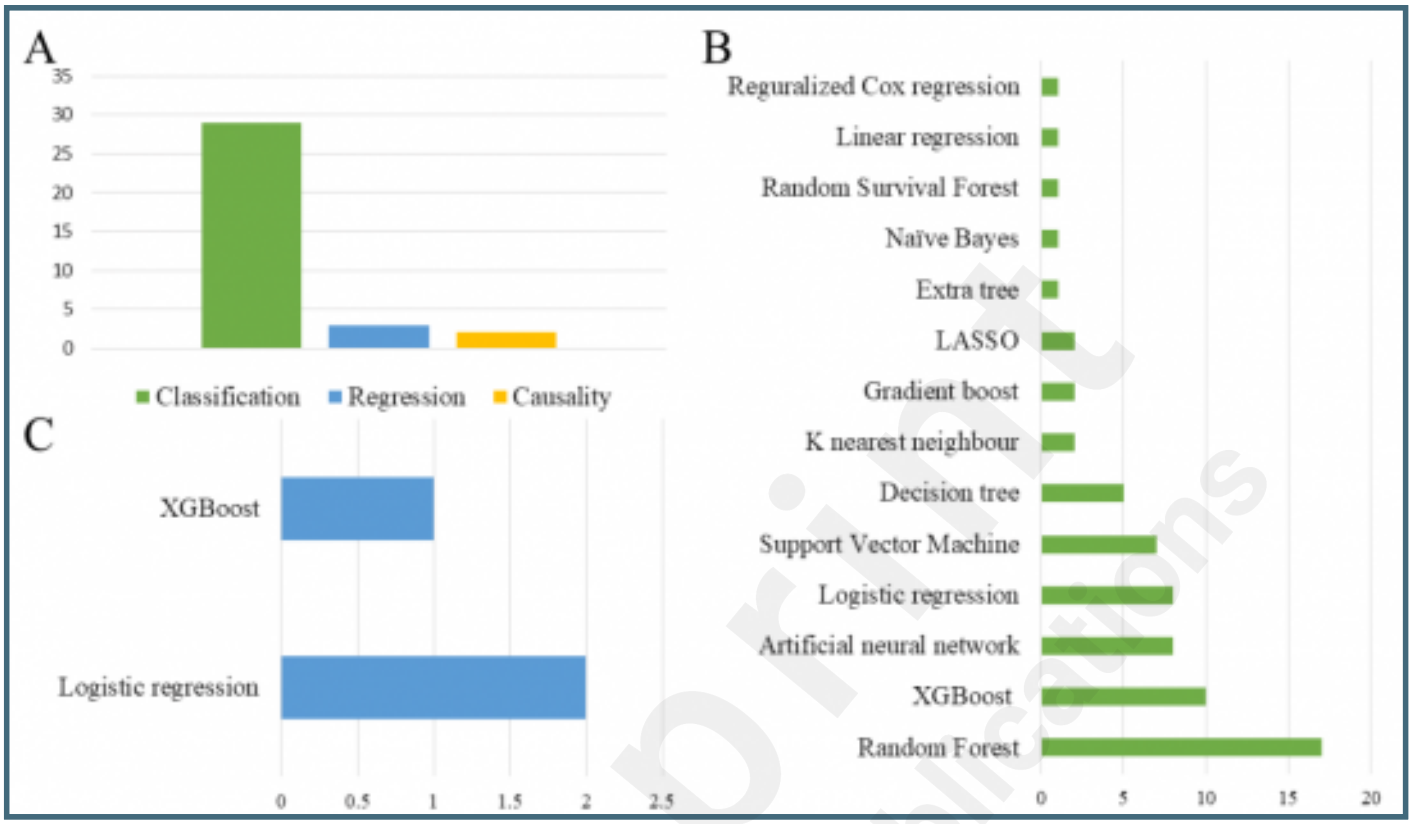
General Bibliographic information about research papers: A. The number of studies through years, B. Country of origin C. Type of organizations that participated in the including studies.



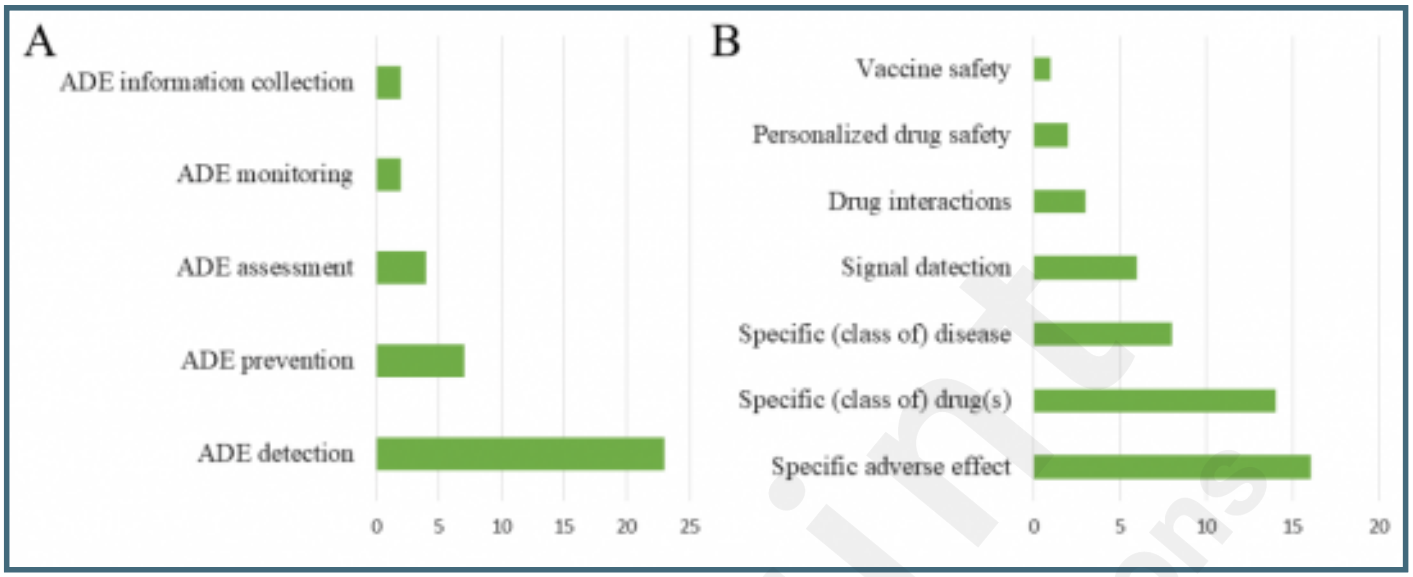
Data description in the included studies A. Studies' distribution based on data models. B. Variety of data sources C. Variety of databases.



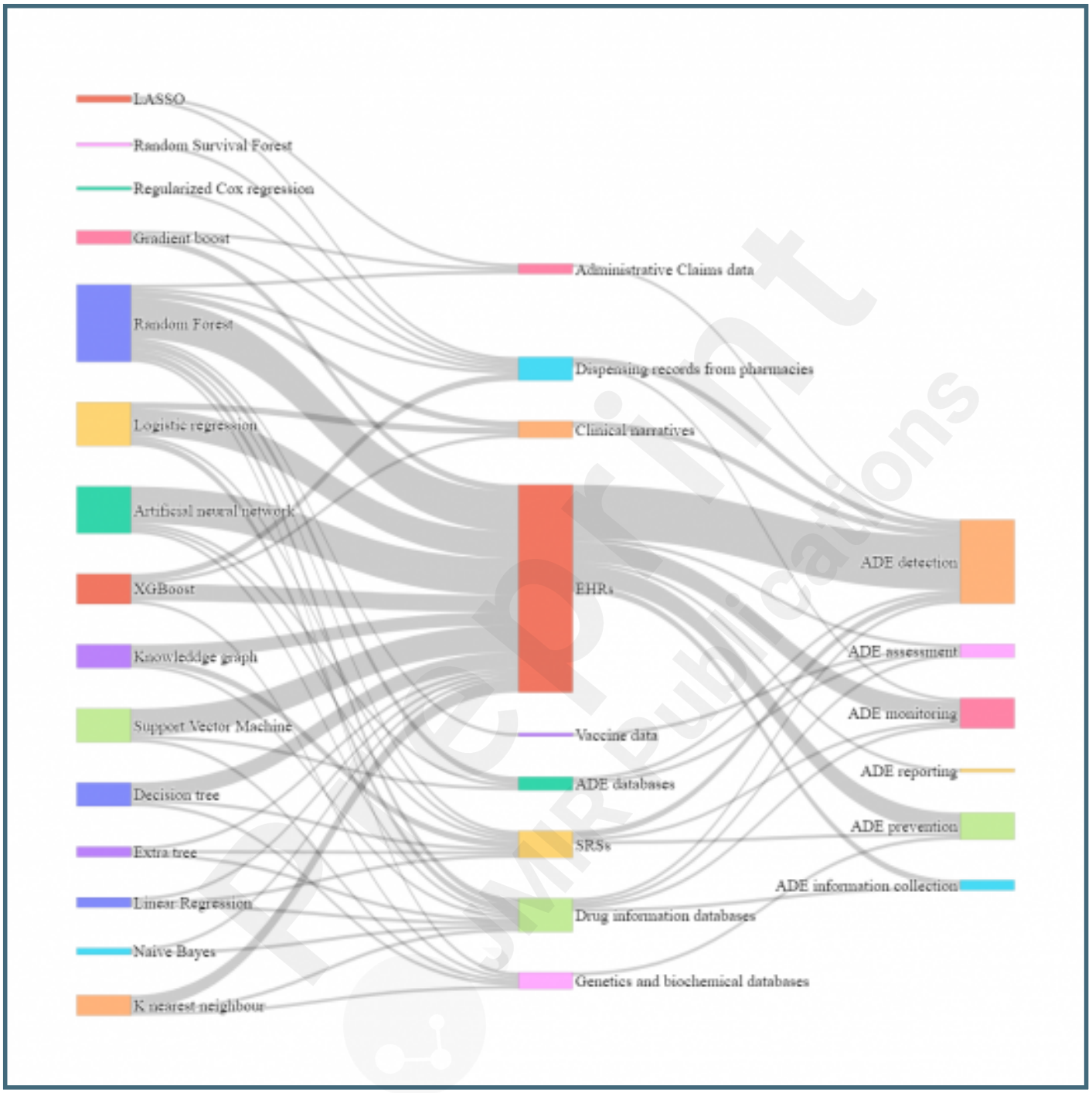
Description of artificial intelligence models A. Distribution of studies based on symbolic AI categories B. Distribution of classification algorithms C. Distribution of regression algorithms.



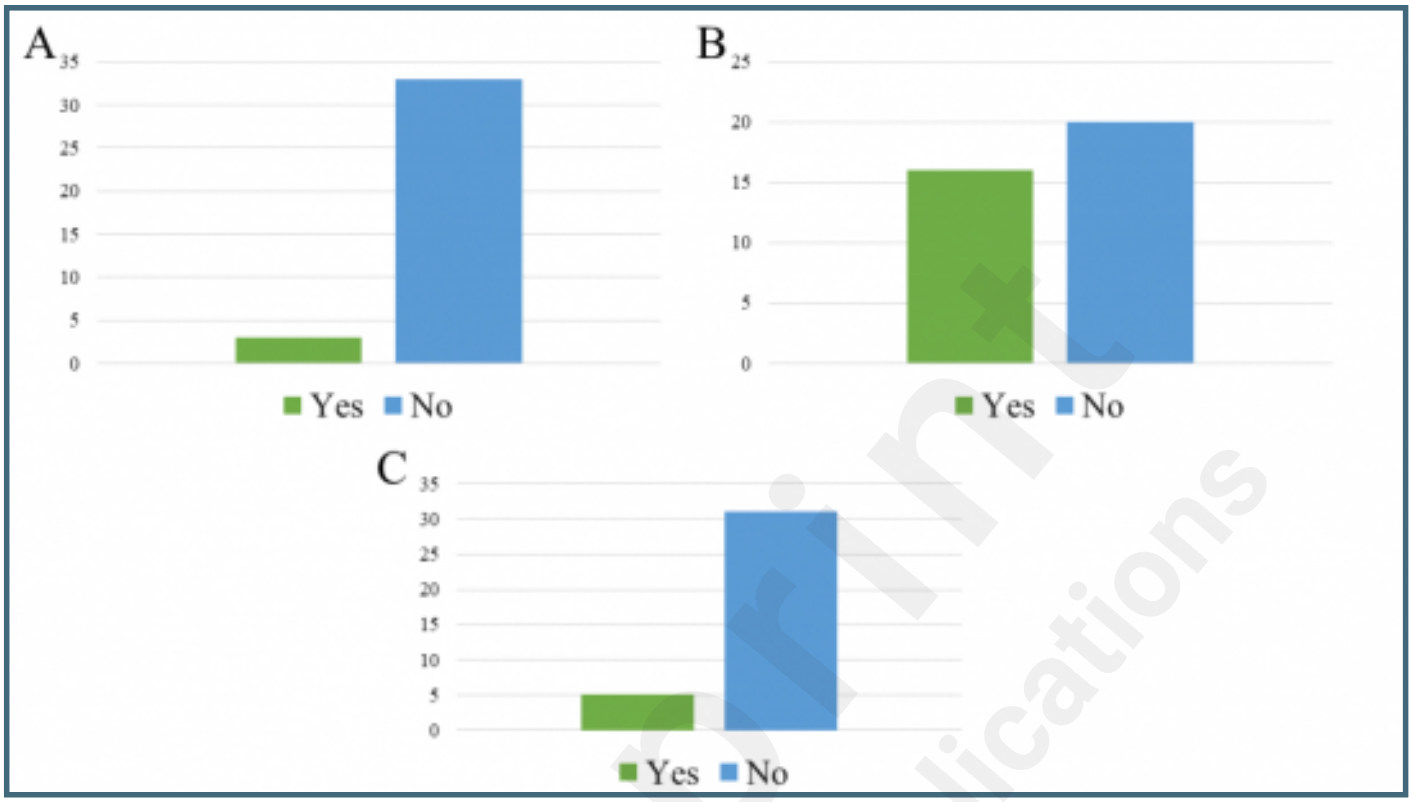
Description of pharmacovigilance processes in the included studies A. Pharmacovigilance core activities B. Specialization of Pharmacovigilance topics.



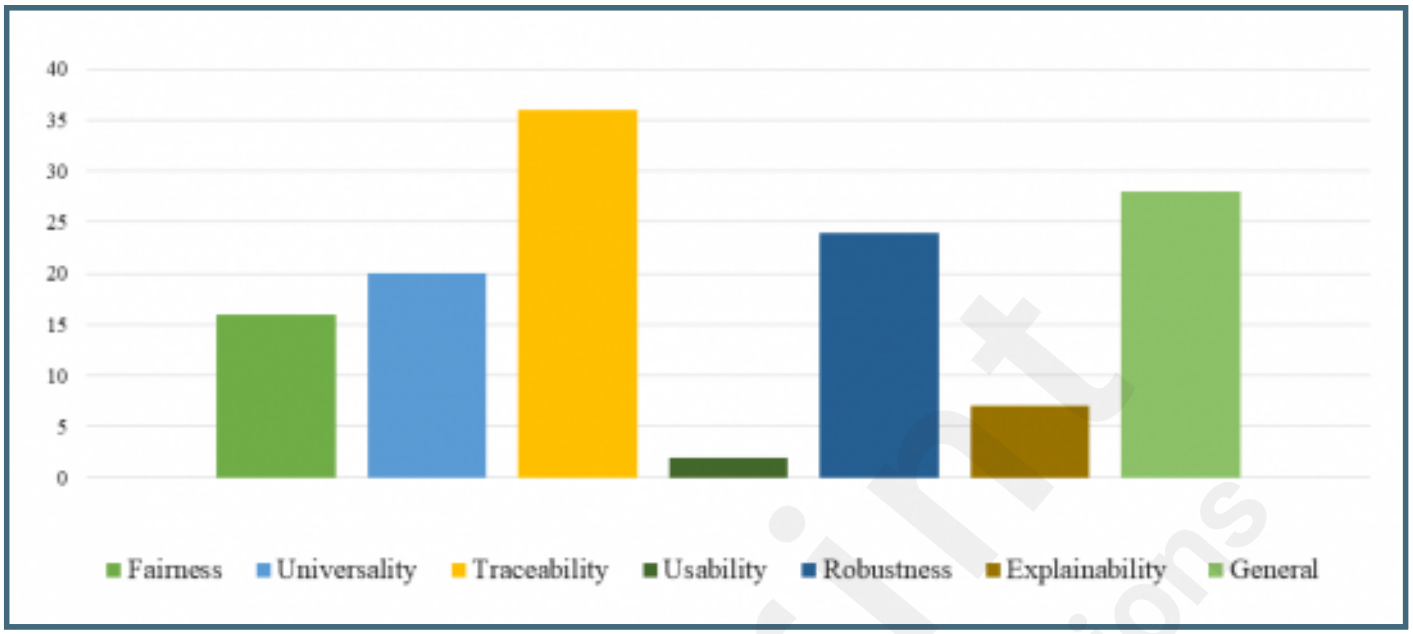
Association between AI models, data sources and drug safety categories.



A. Code availability of studies in public registries B. Description of data preprocessing pipeline in included studies C. Clinical validation of the AI models and pipelines.



Description of FUTURE AI guidelines.



Multimedia Appendixes

Drugs, reactions, and indications in the included studies A. Drugs B. Indications C. Reactions.
URL: http://asset.jmir.pub/assets/3ae18d3de41637b7b5a21d842c82ed68.png

1 Mapping criteria for the studies of systematic literature review 2 Classification of studies based on the mapping criteria. 3 FUTURE AI guidelines/criteria description 4 Classification of studies based on the FUTURE AI criteria.
URL: http://asset.jmir.pub/assets/2f3291256ad634706bf373c8d6300a16.xlsx