

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

170,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Machine Learning Applications in Pharmacovigilance: Scoping Review

Hager Ali Saleh

Abstract

Background: Pharmacovigilance (PV) is the activity to identify comprehensive information on the safety characteristics of the drug after its marketing. The PV data sources are dynamic, large, structured, and unstructured; therefore, the automation of data processing is essential. **Purpose:** This review aims to identify the machine learning applications in PV activities. **Methods:** Nine (9) studies that were published within the period from 2016 to 2020 were reviewed. The studies were extracted from two databases; PubMed and web of science. The review and analysis were done in December 2020. **Results:** The supervised and semi-supervised learning techniques are applied in the main three PV group activities; adverse drug reactions (ADRs) and signal detection, individual case safety reports (ICSRs) identification, and ADRs prediction. Future research is needed to identify the applicability of unsupervised learning in PV and to formulate the legal framework of the false positive predicted data.

Keywords: machine learning, pharmacovigilance, supervised learning, semi-supervised learning, unsupervised learning

1. Introduction

The World Health Organization's (WHO) definition of pharmacovigilance (PV) is "the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem" [1]. It is difficult to get comprehensive safety characteristics of the drug during the drug development phase because the clinical trials are conducted in a controlled environment in a limited patients number and for a specific duration, however, after the drug marketing, it will be prescribed to thousands of patients in different age groups, therefore, it is obligatory that "safety of all medicines to be monitored throughout their use" [2].

In 2018, the WHO global database of individual case safety reports (VigiBase) has 17 million ADRs reports [3] and the Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) has more than 10 million of which 5 million are serious ADRs and one million caused the death [4]. These databases use spontaneous reporting to collect ADRs, nevertheless, the known criticisms of spontaneous

reporting are under-reporting and uncertainty of the causality assessment¹ [5], therefore, there is a need to find other methods to predict ADRs and to efficiently analyze the available data not only from the structured data from spontaneous reporting databases (SRS) but also from other data sources, such as electronic health records (EHR), clinical narratives, medical literature, social media, and health forums [6].

The PV data sources are dynamic, diverse, structured, and unstructured, accordingly, the manual detection of ADRs and processing of PV data are time-consuming, therefore, the automation of ADRs/signal detection and reports processing will be efficient [7].

Machine learning (ML) is a robust data analysis technique that has statistical and probabilistic techniques to develop models that automatically learn from data and consequently help to accurately identify and predict the source data [8]. ML algorithms are supervised, unsupervised, and semi-supervised learning. In supervised learning, a known label is used to train a model to predict labels from new data, while the unsupervised mathematical methods are used to cluster data, and semi-supervised uses models based on both [8].

This scoping review aims to explore the current applications of machine learning techniques on pharmacovigilance (PV) activities; therefore, the research questions are:

- What are the PV activities and data sources for which the machine learning techniques are currently applied?
- What are the machine learning methods used?

2. Methods

The scoping review was considered to explore the available publications regarding the current applications of machine learning techniques in pharmacovigilance activities. The literature search was performed in December 2020.

2.1 Sources

PubMed and web of science were considered to identify relevant publications related to machine learning and pharmacovigilance. PubMed focuses on the life and the biomedical sciences, while the web of science covers medical and computing and information technology. Boolean operators were used to define the relationship between keyword and Wildcard symbols that were used to expand the scope of the search [9, 10].

2.2 Search criteria

- *Inclusion criteria:* Journal articles in the English language and articles published between 2016 and 2020.
- *Keywords:* The keywords for PubMed are (“machine learning”[MeSH Terms] OR (“machine”[All Fields] AND “learning”[All Fields]) OR “machine learning”[All

¹ Causality assessment of the ADRs is “method used for estimating the strength of relationship between drug exposure and occurrence of adverse reaction(s)” [5].

Fields]) AND (“pharmacovigilance”[MeSH Terms] OR “pharmacovigilance”[All Fields]). While The keywords for the web of science are TOPIC: (machine learning) AND TOPIC: (pharmacovigilance). The number of hits in each database and the total number of hits obtained after applying the filters are shown in **Table 1**.

2.3 The articles selection

All the articles found in the two bibliography databases were reviewed, the duplicate check was done, and 21 duplicates were detected and removed. After that, the remaining

Keywords	PubMed	Web of Science
machine learning AND pharmacovigilance	93	84
<i>After Applied Filters</i>		
Filters: in the last 5 years, Humans, English	50	—
Languages: (English) And Document Types: (Article) Timespan: Last 5 years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	—	46
Total		96

Table 1.
Shows the number of hits per database.

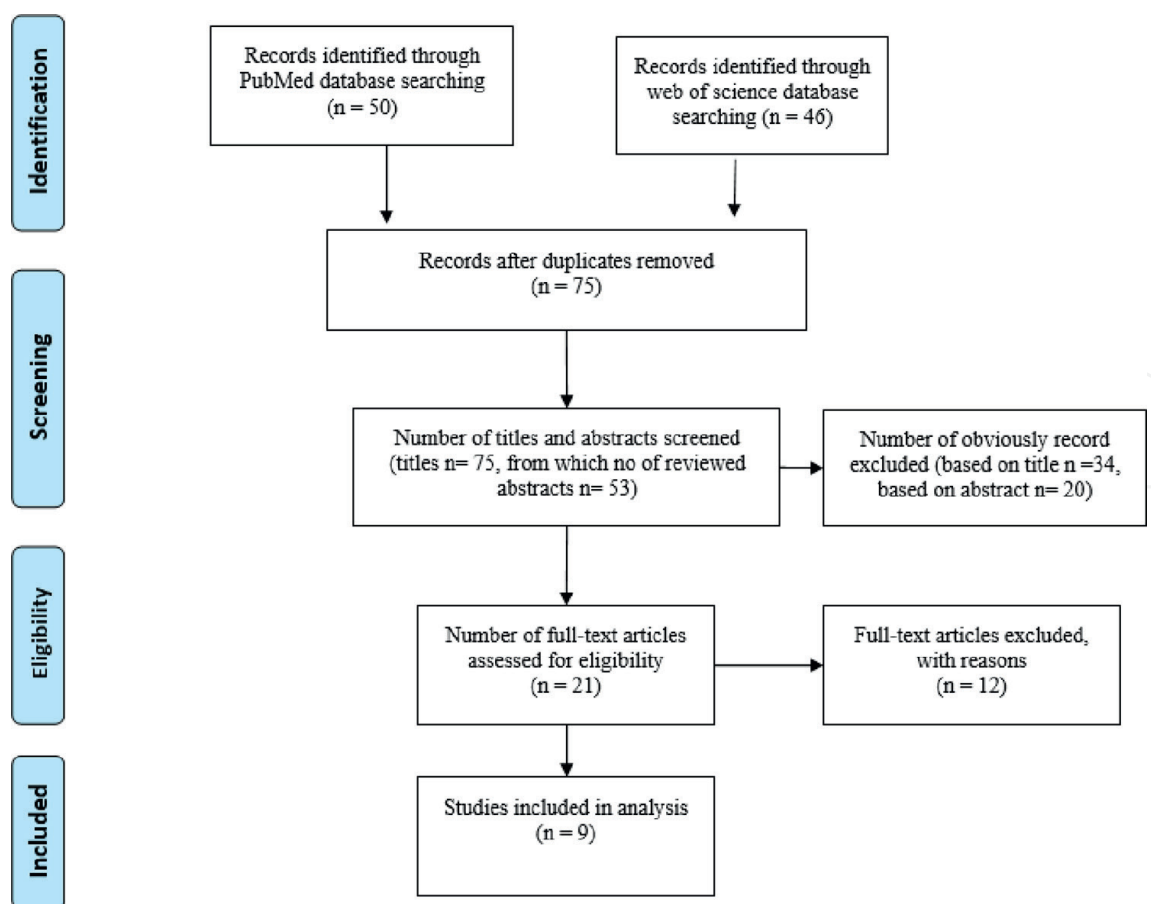


Figure 1.
PRISMA diagram for the articles' selection process.

hits were assessed. The inclusion criteria were peer-reviewed and relevant articles, the relevance means articles that clearly addressed the ML application in PV activities, while the exclusion criteria were articles that addressed PV alone, articles addressed the use of ML in drug-drug interaction (DDI) detection because DDI is not the focus of PV activities, and articles focused on considering more data sources rather than ML Applications.

The hits assessment process was done in three phases firstly assessment of the title, then an assessment of the information provided by the abstracts, eventually assessment of the full text. At each phase, articles were retained or excluded for analysis, based on the inclusion and exclusion criteria. PRISMA flow diagram was used to illustrate the selection process (**Figure 1**) [11].

3. Results

3.1 Overview of articles characteristics

A total of 96 articles were identified of which nine articles met the inclusion criteria, seven were research articles and two reviews. **Table 2** summarizes the retrieved articles according to the year of publication, the first author, the country where the author affiliation is located, the PV activities and data sources, ML technique, and the main findings.

3.2 The PV activities and ML

Based on analyzed articles ML techniques are used in their PV activities groups. Early detection of ADRs and signal detection, individual case safety reports (ICSRs) identification, and ADRs prediction.

3.3 Early detection of ADRs and signal detection²

3.3.1 Spontaneous reporting systems mining

From the last quarter of 2012 to the second quarter of 2013, 632 722 data were extracted from FAERS reports by using the Apriori algorithm 2933 interacting f drug interaction-adverse event was extracted. The algorithm was effective to detect severe life-threatening and rare ADRs [6].

3.3.2 Electronic health record mining

Discharge summaries mining: The supervised machine learning technique was used to detect the ADRs from discharge notes in a tertiary hospital in Switzerland by using a hybrid method, ML, and rule-based. The manual annotation was used to create the training and testing datasets, while the supervised learning technique is used to classify the discharge notes as positive (had ADRs) or negative (had no ADRs), the automatic detection was efficient compared to the manual one and the accuracy was 0.90 [12]. Furthermore, ML algorithms were used to automate the detection of the relationship between the drug and the ADR from the discharge summaries [14].

² “A signal is defined by WHO as reported information on a possible causal relationship between an adverse event and a drug” [19].

RN	Year	Author	Country	PV activity/ Data source	ML method/ Concept	Main findings
[12]	2020	Vasiliki Foufi	Switzerland	Early detection of ADRs (from Discharge Letters mining)	Supervised learning for the text classification. The following ML algorithm is used: Support Vector Machine (SVM), Naive Bayes Classifier, and Linear Classifier. Where 20% of the dataset is used for testing, while 80% is used for training.	ML algorithms are efficient to detect automatically the ADRs. Naive Bayes Classifier and Linear Classifier have more accuracy than the SVM the accuracy was (0.94, 0.94, and 0.83, respectively)
[13]	2019	Azadeh Nikfarjam	USA	Signal detection (from health forums)	“DeepHealthMiner (DHM), a neural network-based named entity recognition (NER) system” where the supervised learning was used to train the DHM to identify the ADRs	13600 ADRs were detected where the F-measure was 0.738 (0.731 precision and 0.745 recall).
[4]*	2019	Anna O. Basile	USA	Early detection of ADRs (from literature)	the ML models were used to predict ADRs from literature and to detect ADRs from EHRs. Natural Language Processing (NLP) is used to detect ADRs from clinical notes and social media.	The limitation of the SVM is the prediction of unknown ADRs cannot depend on the labeled data.
[14]	2019	Yixuan Tang	Singapore	Early detection of ADRs (From EHR)	A rule-based approach called Readpeer for Active PV (REAP) this structure divided into two steps “named entity recognition (NER) and drug-AE relation extraction” where the ADRs and drug names were recognized then the pairing of ADR-Drug would occur.	The precision and recall of ADRs and drug name detection were 90%. While for detection the relationship between the drug and the ADR the precision was 75% and the recall was 60%.
[6]*	2019	Chun Ye n Lee	Australia	Signal detection (SRS and social media)	using the Apriori algorithm to detect life-threatening ADRs from FAERS reports. SVM classifier to detect from Twitter posts if the users used the drug and to detect the ADRs mentioned in the posts.	The precision of the Apriori algorithm was 85% and sensitivity was 81%. The precision of the SVM classifier was 70%, while the recall was 69%
[15]	2018	Shaun Comfort	USA	ICSR identification (from social media)	Support vector machine (SVM) algorithm used to detect ICSRs from 311,189 social media posts	The ML model spent 48 hr. to finish the task compared to an estimated 44,000 hr. spent by human experts and the accuracy was 74%.

RN	Year	Author	Country	PV activity/ Data source	ML method/ Concept	Main findings
[16]	2018	Shashank Gupta	India	Early detection of ADRs (from social media)	Semi-supervised bidirectional long-short-term-memory (LSTM) where the unsupervised technique is used to train the bidirectional-LSTM model to predict the drug name, and the supervised model to retrain it to predict the label sequence.	The semi-supervised was effective, where the F-score was 0.751.
[17]	2017	Kalpana Raja	USA	Adverse drug reactions prediction (Literature mining)	The researchers used the DDI corpus training data, the following classifier Bayesian network, decision tree, random tree, random forest, and k-nearest neighbors are used to predict ADRs types from the DDI corpus then the performance of each classifier was evaluated using 10- fold cross-validation technique. The random forest showed the best performance (F score = 0.9) After that, the researcher used this ML framework to predict from the literature the ADR types related to psoriasis.	The researchers identified the previously known ADRs (F score =0.9) and predicted the ADRs of psoriasis drugs.
[18]	2016	Vassilis Plachouras	UK	Detection of ADRs (From social media)	SVM classifier identified ADRs based on the surface-textual properties and the known information about drugs' adverse effects.	Accuracy = 74%

RN= reference number, Year = Publication year, Author= First Author, and *= review articles

Table 2.

Shows an overview of the eligible articles listed chronologically.

3.3.3 Social Media and Health forums mining

A combination of supervised and unsupervised ML models (semi-supervised) was used to detect the ADRs mentioned in Twitter posts, where the unsupervised trained model to detect the drug name, while the supervised technique was to retrain the model to detect the ADRs labels [16]. Furthermore, 67172 posts are identified in the health forums, where 13600 ADRs were identified by using the supervised machine learning technique [13].

3.4 ICSR identification

3.4.1 Social media mining

The ICSR to be valid it should have identified the patient, identified the suspect drug, identified ADR, and identified the reporter, so to identify the valid or invalid ICSR from social media posts “ICSR classification framework” was developed by using a support vector machine (SVM) to detect the patient, drug, and ADR, while the reporter was assumed to be the author of the post [15].

3.5 ADRs prediction

3.5.1 System pharmacology

System pharmacology is “the study of drug action using principles from systems biology, considering the effect of the drug on the entire system rather than a single target or metabolizing enzyme.” Its application to PV activities is to focus on “off-target effects and clinical observations of adverse reactions.” [4] An application of this approach was addressed in a published study in 2017, where the researchers evaluated the feasibility of using the “ML models to learn syntactic and semantic information from literature,” to enhance the model prediction the researchers used drug-drug interaction (DDI) information to predict ADRs caused by DDI, and drug-gene interaction (DGI) to predict the ADRs caused by two drugs interaction by the same gene [17].

3.5.2 Event reporting system database mining

The Bayes classifier algorithm was used to predict ADRs from experts’ opinions texts in the ADR case [4].

4. Discussion

Based on the reviewed literature, the benefits of integrating ML with PV activities are the following:

4.1 The data source for post-marketing surveillance

There are two data sources³ structured for example spontaneous reporting systems (SRSs) and unstructured like medical literature, clinical notes, and social media

³ Post marketing surveillance “refers to the process of monitoring the safety of drugs once they reach the market” [20].

posts [6]. The ADRs are collected by regulatory authorities through voluntary reporting to SRSs, therefore, under-reporting is the main drawback of these sources, therefore, it is important to use more data sources to comprehensively collect the safety information [6]. The supervised and semi-supervised machine learning techniques helped in mining other data sources, such as clinical notes, medical literature, and social media [4, 6, 9, 10, 12–14].

4.2 Improve the accuracy and time efficiency

The PV sources are dynamic, which means it is periodically updated over time, these sources become large beside their unstructured characteristics [6], and the accuracy of using ML techniques in the detection or prediction of ADRs was between 74% to 90% [6, 9, 12], the precision was between 0.7 and 0.9 [10, 11], furthermore, the ML model spent 48 hr. to finish the ICSR identification task from social media compared to an estimated 44,000 hr. spent by human experts with accuracy 74% [12].

4.3 ADRs prediction

Predicting ADRs in the early stages will enhance drug safety activities and reduce the financial cost, for example, saving the cost of hospitalization due to the ADRs [21], the ML techniques were used to predict the ADRs from the social media posts, F score=0.9 [14].

4.4 Limitation of the review

Only two databases are considered, the scoping review is not like the systematic review, therefore, it is expected to miss some relevant articles.

5. Conclusion

The supervised and semi-supervised machine learning techniques are applied in the main three PV group activities; detection of adverse drug reactions (ADRs) and signal detection, individual case safety reports (ICSRs) identification, and ADRs prediction. Furthermore, it helps in analyzing large data sources, such as social media and literature, to predict and detect ADRs, accordingly, it complements the drawbacks of spontaneous reporting. Moreover, ML techniques are efficient in terms of accuracy and saving time when compared to human experts.

Knowledge gaps

The supervised learning technique is currently used in PV activities, which has a problem with the scarcity of labeled data [16], so the first knowledge gap is how to apply the unsupervised technique in PV activities.

The second knowledge gap is that PV activities are legally regulated [22], therefore, a regulation should be developed to manage the risk of false-negative detected results.

The third knowledge gap: further research is needed to assess the attitude, knowledge, and practice of PV personnel regarding the applicability of the ML techniques in PV daily practice.

Acknowledgements

This work was done during the author's studies in the Health Informatics Joint Master Program, Karolinska Institute. Therefore, the author expresses her immense gratitude to the program director and to the current research and trends in health informatics course's leaders and teachers, the academic year 2020–2021.

Conflict of interest


The author declares no conflict of interest.

Author details

Hager Ali Saleh
Health Informatics Joint Master Program, Karolinska Institutet, Stockholm, Sweden

*Address all correspondence to: hagerasaleh@yahoo.com

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] World Health Organization. The Importance of Pharmacovigilance [Internet]. 2002. Available from: <https://apps.who.int/iris/handle/10665/42493>. [Accessed: October 6, 2020]
- [2] European Medication Agency. Pharmacovigilance: Overview [Internet]. 2020. Available from: <https://www.ema.europa.eu/en/human-regulatory/overview/pharmacovigilance-overview>. [Accessed: October 7, 2020]
- [3] VigiBase now contains around 17 million ADR reports | SpringerLink [Internet]. 2020. Available from: <https://link.springer.com/article/10.1007/s40278-018-45575-x>. [Accessed: October 7, 2020]
- [4] Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. 2019
- [5] Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions: A systematic review. *Drug Safety*. 2008;**31**(1):21-37
- [6] Lee CY, YPP C. Machine learning on adverse drug reactions for pharmacovigilance. *Drug Discovery Today*. 2019;**24**:1332-1343
- [7] Lee VC. Big data and pharmacovigilance: Data mining for adverse drug events and interactions. *Psychological Medicine*. 2018;**43**(6):340-351
- [8] Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*. 2019;**49**(9):1426-1448
- [9] Massachusetts Institute of Technology. Boolean operators - Database Search Tips - LibGuides at MIT Libraries [Internet]. 2021. Available from: <https://libguides.mit.edu/c.php?g=175963&p=1158594>. [Accessed: March 30, 2021]
- [10] Massachusetts Institute of Technology. Truncation - Database Search Tips - LibGuides at MIT Libraries [Internet]. 2021. Available from: <https://libguides.mit.edu/c.php?g=175963&p=1158679>. [Accessed: March 30, 2021]
- [11] PRISMA. PRISMA Flow Diagram [Internet]. 2021. Available from: <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram>
- [12] Foufi V, Ing Lorenzini K, Goldman JP, Gaudet-Blavignac C, Lovis C, Samer C. Automatic classification of discharge letters to detect adverse drug reactions. *Studies in Health Technology and Informatics*. 2020;**270**:48-52
- [13] Nikfarjam A, Ransohoff JD, Callahan A, Jones E, Loew B, Kwong BY, et al. Early detection of adverse drug reactions in social health networks: A natural language processing pipeline for signal detection. *Journal of Medical Internet Research*. 2019;**21**(6):1-18
- [14] Tang Y, Yang J, Ang PS, Dorajoo SR, Foo B, Soh S, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. *International Journal of Medical Informatics*. 2019;**128**:62-70
- [15] Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting through the safety data haystack: Using machine learning to identify individual case safety reports in social-digital media. *Drug Safety*. 2018;**41**(6):579-590

[16] Gupta S, Pawar S, Ramrakhiyani N, Palshikar GK, Varma V. Semi-supervised recurrent neural network for adverse drug reaction mention extraction. *BMC Bioinformatics*. 2018;**19**(S8):212

[17] Raja K, Patrick M, Elder JT, Tsoi LC. Machine learning workflow to enhance predictions of Adverse Drug Reactions (ADRs) through drug-gene interactions: Application to drugs for cutaneous diseases. *Scientific Reports*. 2017;**7**(1):1-11

[18] Plachouras V, Leidner JL, Garrow AG. Quantifying Self-Reported Adverse Drug Events on Twitter2016. pp. 1-10

[19] Uppsala Mentoring Center. General information about signal [Internet]. 2020. Available from: <https://www.who-umc.org/media/164092/general-information-about-signal-published-in-who-pn.pdf>

[20] Raj N, Fernandes S, Charyulu NR, Dubey A, Hebbar S. Postmarket surveillance: A review on key aspects and measures on the effective functioning in the context of the United Kingdom and Canada. *Therapeutic Advanced Drug Safety*. 2019;**10**:204

[21] Dey S, Luo H, Fokoue A, Hu J, Zhang P. Predicting adverse drug reactions through interpretable deep learning framework. *BMC Bioinformatics*. 2018;**19**(S21):476

[22] European Medicines Agency. Legal framework: Pharmacovigilance. *Expert Review of Clinical Pharmacology*. 2012;**5**:485-488