

# 机器学习在药物警戒领域应用的文献计量分析



李丽敏, 吴文字, 魏芬芳, 唐碧雨, 吴建茹

深圳市药物警戒和风险管理研究院 (广东深圳 518000)

**【摘要】目的** 探讨世界范围内机器学习在药物警戒领域的应用现状和发展趋势, 为开展机器学习在药物警戒领域的应用相关研究提供参考。**方法** 在 Web of Science 文献库中以“机器学习”“药物警戒”等为主题词检索相关文献, 检索时限为建库至 2023 年 3 月 1 日, 利用 R 语言等软件定量分析该领域文献数据, 对年度发文量、机构、国家、关键词等方面的特征开展聚类、共现和突现的可视化分析。**结果** 共纳入 904 篇文献, 文献发表量自 1994 年以来呈现波动上升趋势。合作网络机构之间存在跨领域、跨地区、跨结构的合作。发文量前 5 位的国家为美国、中国、日本、韩国、印度, 中美在该领域的合作相对较密切。信号检测、社交媒体、电子健康记录是该领域的高频关键词, 聚类和关联规则分析显示, 机器学习在该领域主要围绕信号识别、非结构化文本挖掘分析和电子医疗信息的处理分析三方面开展研究, 目前在信号识别、社交媒体信息挖掘、电子医疗信息非结构化的文本处理等方面取得了显著进展, 拓宽了药物警戒的数据来源, 提高了对药品不良反应的实时监测能力。**结论** 大数据和人工智能技术的飞速发展使得机器学习与药物警戒领域融合日益密切, 技术交流合作和学科间的交叉融合日益频繁, 应先优化各种机器学习算法, 提高其在药物警戒领域中的准确性和稳定性; 需要加强对数据隐私和安全性的保护措施, 确保患者信息的安全; 整合医学、数据科学、统计学等领域的专业知识, 以期推动药物警戒领域的技术进步。

**【关键词】** 机器学习; 药物警戒; 文献计量学

## Bibliometric analysis of the application of machine learning in pharmacovigilance

LI Limin, WU Wenyu, WEI Fenfang, Tang Biyu, WU Jianru

Shenzhen Institute of Pharmacovigilance and Risk Management, Shenzhen 518000, Guangdong Province, China

Corresponding author: WU Jianru, Email: wujr@mail.amr.sz.gov.cn

**【Abstract】Objective** To explore the application status and development trend of machine learning in the field of pharmacovigilance worldwide, and to provide reference for the research on the application of machine learning in the field of pharmacovigilance. **Methods** Relevant literature was searched in the Web of Science with the key words of "machine learning" and "pharmacovigilance" from the inception to March 1, 2023. R

DOI: 10.12173/j.issn.1005-0698.202309079

基金项目: 广东省基础与应用基础研究基金自然科学基金项目 (2023A1515011495); 广东省药品监督管理局科技创新项目 (2021ZDB01); 广东省药品监督管理局科技创新项目 (2022TDB16)

通信作者: 吴建茹, 博士, 副主任药师, Email: wujr@mail.amr.sz.gov.cn

language and other software were used to quantitatively analyze the literature data in this field. The clustering, co-occurrence and emergence visual analysis were carried out on the characteristics of annual published papers, institutions, countries, keywords and other aspects. **Results** A total of 904 literature were included. The number of literature published showed a fluctuating upward trend since 1994. There was cross-regional, cross-regional and cross-agency cooperation among the cooperative network institutions. The top 5 countries in the number of publications were the United States, China, Japan, South Korea and India, China and the United States had relatively close cooperation in this field. Signal detection, social media and electronic health records were high-frequency keywords in this field. Clustering and association rule analysis showed that this field focused on three aspects signal recognition, unstructured text mining and analysis, and processing and analysis of electronic medical information. At present, machine learning has made significant progress in signal recognition, social media information mining, and unstructured text processing of electronic medical information, which broaden the data sources of pharmacovigilance, improve the real-time monitoring ability of adverse drug reactions, bringing innovation impetus to the field of pharmacovigilance. **Conclusion** The rapid development of big data and artificial intelligence technologies has led to an increasing integration of machine learning into the field of pharmacovigilance, which promotes technical exchanges and cooperation and cross-disciplinary integration. It is necessary to optimize each machine learning algorithm to improve its accuracy and stability in pharmacovigilance, strengthen the protection measures of data privacy and security to ensure the safety of patient information. Integrating expertise in the fields of science, medicine, and data statistics with a view to promoting technological progress in the field of pharmacovigilance.

**【Keywords】** Machine learning; Pharmacovigilance; Bibliometrics

世界卫生组织（WHO）将药物警戒定义为有关不良反应或任何其他可能与药物相关问题的发现、评估、理解与预防的科学和活动，贯穿于药品全生命周期，是制定药品审评和监管决策的重要技术支撑<sup>[1]</sup>。该项活动需从临床前毒理学数据、自发报告、电子索赔和医疗记录等来源收集大量数据<sup>[2]</sup>并进行管理和分析。目前，机器学习在包括医学在内的许多科学领域中应用广泛<sup>[3]</sup>。而文献计量学是一门结合统计学对文献进行量化的学科，能够有效地识别研究领域的知识发展脉络、主要研究机构和国家、合作网络以及研究主题的演变，为研究领域学者提供一个领域研究的宏观视角。基于此，本文通过对 Web of Science 数据库中的相关文献进行定量分析，旨在系统地检索和分析机器学习在药物警戒领域的应用现状、研究热点和存在的不足，以期为研究者获取该领域相关研究的动态、发展趋势提供参考。

## 1 资料与方法

### 1.1 纳入与排除标准

#### 1.1.1 纳入标准

① Web of Science 数据库中已公开发表的与检索主题相关的研究文献；②语种为英文；③论文种类为 article。

#### 1.1.2 排除标准

①发表类型为会议摘要、毕业论文、报告等无全文文献；②作者和年份等信息不全；③重复发表的文献。

### 1.2 文献检索与数据来源

本研究将 Web of Science 作为文献数据源，检索时限为建库至 2023 年 3 月 1 日。构建与机器学习和药物警戒相关的主题词检索式，检索词包括 pharmacovigilance、drug side effects、adverse drug reaction、drug contraindications、drug toxicity、drug safety、drug risk、adverse drug

reaction signal、drug post-marketing surveillance、rational drug use、drug-induced illness、risk-benefit assessment of drug、drug incompatibility、machine learning、deep learning、prediction model、big data、natural language processing、data mining、random forest、k-means clustering、Markov model、support vector machine、neural network、decision trees、classification algorithm、machine learning algorithm、feature extraction 等，检索策略见框 1。

```
#1 TS=(“Pharmacovigilance” OR “Drug Side Effects”
OR “Adverse Drug Event” OR “Adverse Drug Reaction”
OR “Drug Contraindications” OR “Drug Toxicity”
OR “drug safety” OR “drug risk” OR “Adverse drug
reaction signal” OR “Drug post-marketing surveillance”
OR “Rational drug use” OR “Drug-induced illness”
OR “risk-benefit assessment of drug” OR “Drug
Incompatibility” )
#2 TS=( “Machine learning” OR “Deep learning”
OR “Prediction model” OR “Big data” OR “Natural
language processing” OR “Data mining” OR “Random
forest” OR “k-means clustering” OR “Markov model”
OR “Support vector machine” OR “Neural network”
OR “Decision Trees” OR “Classification algorithm” OR
“machine learning algorithm” OR “feature extraction” )
#3 DT=(Article)
#4 #1 AND #2 AND #3
```

框1 检索策略

Box 1. Search strategy

### 1.3 资料提取与统计分析

采用文献计量学方法进行分析，以纯文本格式文件导出文献，每一篇文献包括标题、作者、机构、摘要、关键词、引文等记录。文件导出后，由 3 位人员分别结合排除标准对题目、摘要、全文进行审查并剔除与机器学习和药物警戒主题不相关的文献，如有分歧通过协商讨论确定。再应用 R 语言 stringdist 包、ggraph 包、ggplot 2 包和 Microsoft Excel 2019 等软件对导出的文献数据进行归整、分析和可视化。分析内容包括文献来源、发文章量、机构、地域及关键词等；利用社区发现算法对机构、地区、关键词等指标进行网络可视化分析；利用节点链接图展示国家、作者与关键词之间的层次关系，厘清脉络，挖掘研究热点及方向。

## 2 结果

### 2.1 文献筛选结果

共检索到 1 324 篇文献，筛选排除的文献包括 Book Chapter ( $n=17$ )，Data Paper ( $n=1$ )，Early Access ( $n=5$ )，Proceedings Paper，Correction ( $n=7$ )，Editorial Material ( $n=29$ )，Letter ( $n=9$ )，Meeting Abstract ( $n=26$ )，News Item ( $n=1$ )，Review ( $n=126$ ) 等 420 篇文献。经筛选最终纳入 904 篇文献。

### 2.2 发文章量及趋势

1994—2022 年，机器学习在药物警戒领域应用共发表文献 904 篇，文献发表量呈波动上涨趋势。根据发文章量，机器学习在药物警戒领域中的应用可以分为三个阶段，1994—2004 年为起步阶段、2005—2014 年为平稳发展阶段、2015—2023 年为迅速发展阶段。2004 年以前，研究者一直利用简单的频率统计方法对 WHO 药品不良反应病例报告数据库 (VigiBase) 中的数据进行药物警戒数据挖掘<sup>[4]</sup>。随着大数据、人工智能、社交媒体爆发式的增长和使用，2007 年首篇关于大数据挖掘的药物警戒研究出现，使得机器学习技术逐渐进入药物警戒研究领域。Nikfarjam 等<sup>[5]</sup>创造了一种基于机器学习的序列标记器 ADMine 软件，从社交媒体上的用户帖子中自动提取药品不良反应数据。2015 年起，药品不良反应信号算法越来越复杂化、智能化，使用的新方法和新技术能够更加准确地识别和预测药品不良反应，为药品研发和使用提供了重要的参考和支持，见图 1。

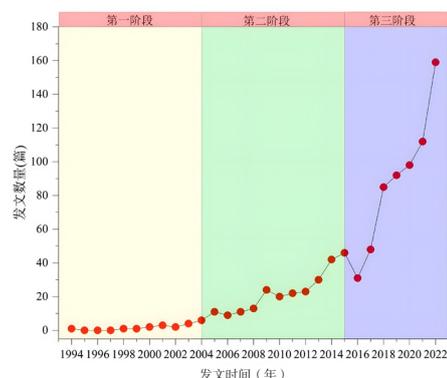


图1 机器学习在药物警戒领域应用的文献年度分布图

Figure 1. Annual distribution of literature on the application of machine learning in pharmacovigilance

期刊是学术和科学报告最重要的来源和指标，通过对期刊分布的分析，可以确定该领域研究的核心方向和研究价值<sup>[6]</sup>。文献数据来源最多的期刊为 *Drug Safety*，共发表相关文章 59 篇，

2022 年的影响因子为 4.2；其次为 *Journal of the American Medical Informatics Association*，发表相关文章 30 篇，2022 年的影响因子为 4.5。文献来源前 10 位的期刊见表 1。

表1 机器学习在药物警戒领域应用的文献来源前10位期刊

Table 1. Top 10 journals for the application of machine learning in the field of pharmacovigilance

序号	期刊	数量	2022年影响因子
1	<i>Drug Safety</i>	59	4.2
2	<i>Journal of Biomedical Informatics</i>	45	4.5
3	<i>Pharmacoepidemiology and Drug Safety</i>	31	2.6
4	<i>Journal of the American Medical Informatics Association</i>	30	6.4
5	<i>Frontiers in Pharmacology</i>	29	5.6
6	<i>PLoS One</i>	19	3.7
7	<i>BMC Bioinformatics</i>	15	3.0
8	<i>Scientific Reports</i>	14	4.6
9	<i>JMIR Medical Informatics</i>	13	3.2
10	<i>BMC Medical Informatics and Decision Making</i>	12	3.5

注：影响因子来源于Web of Science 期刊介绍。

### 2.3 机构合作情况

共有 1 199 家科研机构参与该领域的研究，累计发表文章排名前 5 位的机构包括美国哥伦比亚大学（Columbia University）（发文 20 篇）、瑞典斯德哥尔摩大学（Stockholm University）（发文 16 篇）、美国斯坦福大学（Stanford University）（发文 15 篇）、美国宾夕法尼亚大学（University of Pennsylvania）（发文 14 篇）、日本京都大学

（Kyoto University）（发文 13 篇）。

对研究机构进行聚类与可视化，可分为 9 个群体（图 2）。图 2 提示该领域机构之间存在跨领域、跨地区、跨结构（医院、大学、监管机构、研究机构、企业）的合作，不同机构之间的合作可以各角度探索药品安全性研究，也可以相互弥补研究不足，推动该领域向更高水平发展。

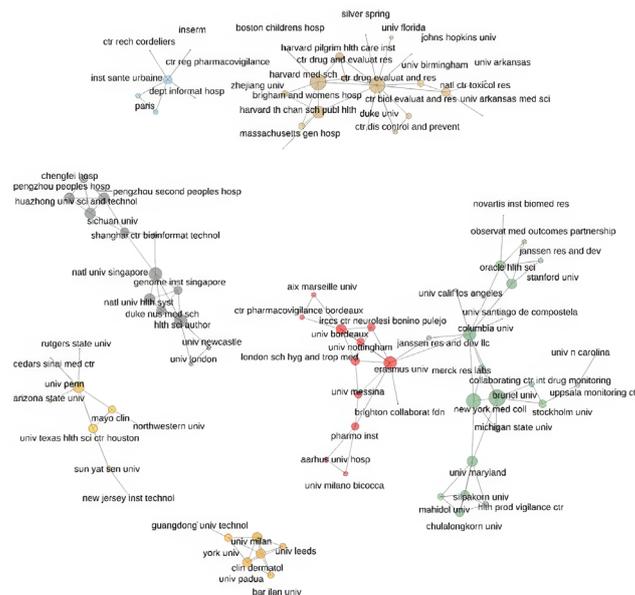


图2 机器学习在药物警戒领域应用的机构合作图

Figure 2. Diagram of institutional cooperation for the application of machine learning in pharmacovigilance

## 2.4 地域合作情况

表 2 显示，该研究领域发表文献排名前五位的国家分别为美国、中国、日本、韩国、印度。中美、美英等国家合作比较频繁，这些国家拥有强大的经济和科技实力，在科研投入、技术创新以及高等教育等方面较为领先，吸引其他发达国家寻求合作，以共享资源、技术和知识<sup>[7]</sup>。

## 2.5 关键词分析

### 2.5.1 关键词频次分析

关键词是一篇论文的核心观点，对某一领域相关文献关键词的分析，有助于挖掘领域的研究热点及核心内容<sup>[8]</sup>。

使用 R 语言 Stringdist 包对该领域文献关键词进行同义词替换和归整，出现的关键词包括药物警戒、药品不良反应、数据挖掘等，见表 3。由于出现频次较高的关键词与检索词相关，过滤掉检索词后，频次前 5 位的关键词包括信号检测、社交媒体、电子健康记录、美国不良反应报告系统、文本挖掘。

由高频关键词可见，目前机器学习在药物警

戒领域的应用以药品不良反应的信号检测研究为主，利用数据挖掘在不同数据来源进行药品不良反应信号检测是该领域的研究重点。

表2 机器学习在药物警戒领域应用的地区发文量及地区合作分布

Table 2. Regional publication volume and regional cooperation distribution of the application of machine learning in pharmacovigilance

序号	国家	发文量	合作国家	合作次数
1	美国	224	中国-美国	38
2	中国	221	英国-美国	22
3	日本	60	德国-美国	9
4	韩国	57	德国-英国	8
5	印度	55	意大利-英国	8
6	法国	51	意大利-美国	8
7	英国	49	法国-意大利	7
8	澳大利亚	28	荷兰-西班牙	6
9	意大利	24	瑞典-英国	6
10	西班牙	21	瑞典-美国	6

表3 机器学习在药物警戒领域应用的关键频次统计

Table 3. Statistics of key frequency of application of machine learning in pharmacovigilance

序号	关键词（全部）	频数	关键词（过滤检索词）	频数
1	药物警戒	293	信号检测	79
2	药品不良反应	213	社交媒体	68
3	数据挖掘	193	电子健康记录	46
4	机器学习	158	美国不良反应报告系统	42
5	自然语言处理	125	文本挖掘	42
6	药品安全	87	药物流行病学	36
7	药品不良事件	82	自发呈报系统	34
8	深度学习	82	人工智能	31
9	信号检测	79	药物相互作用	30
10	社交媒体	68	药品研发	22
11	不良事件	54	各类药品	20
12	电子健康记录	46	关系提取	18
13	大数据	43	比例失衡分析	17
14	美国不良反应报告系统	42	预测	17
15	文本挖掘	42	新型冠状病毒感染	16

### 2.5.2 关键词聚类分析

使用社会发现算法对该领域关键词出现次数较高的关键词（过滤掉检索词）进行聚类分析，结果分为三大类。第一类别为信号识别，核心关

键词是信号检测（signal detection），信号检测所用到的数据包括美国食品药品监督管理局不良事件报告系统（FAERS）、自发呈报系统（spontaneous reporting system）、韩国不良反应报告系统数据

库 (KAERS database) 等, 所用到的方法包括比例失衡法 (disproportionality methods)、比例报告比值比 (proportional reporting ratio, PRR) 法、报告比值比 (reporting odds ratio, ROR) 法等。第二类别为对非结构化文本挖掘分析, 主要包括社交媒体 (social media)、Twitter 挖掘 (Twitter mining)、公共健康 (public health) 等基础数据或能够提供基础数据的媒体, 还包括了文本分析所需要的技术手段关键词, 如词嵌入 (word embedding)、情感分析 (sentiment analysis)、文本分类 (text classification) 等。第三类别是对电子医疗信息的处理分析, 主要关键词包括电子健康记录 (electronic health records)、关系提取 (relation extraction)、药物相互作用 (drug-drug interactions), 具体见图 3。

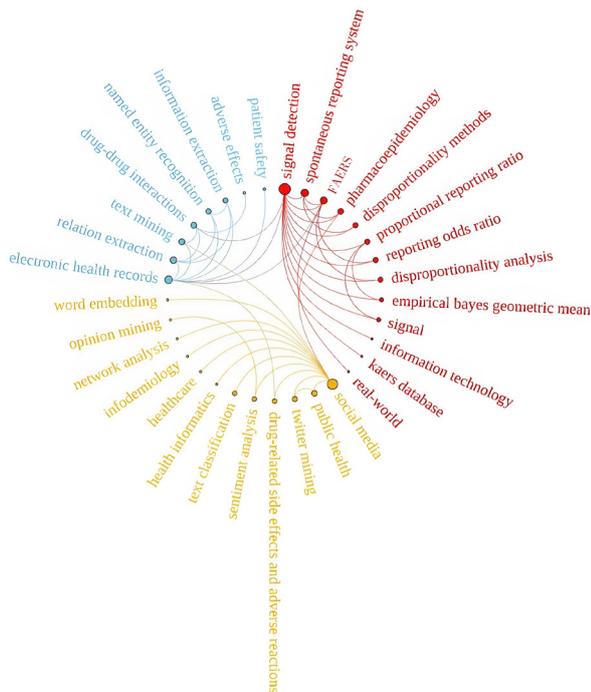


图3 机器学习在药物警戒领域应用的文献关键词复杂网络图

Figure 3. Complex network diagram of literature keywords applied by machine learning in pharmacovigilance

目前信号检测主要使用 FAERS 数据库, 其规范化和开放性便于相关研究人员的使用, 利用比例失衡法、经验贝叶斯几何均值等统计方法对药品与不良反应的关联性进行因果推断; 社交媒体等非结构化文本数据具有非专业、不规范、结构复杂的特点, 需要利用词嵌入、情感分析、

文本分类等方法进行有效信息识别和提取; 另外与社交媒体等非结构化文本不同的是电子医疗信息, 包含了患者基本特征、用药情况、疾病类型等临床信息, 相对规范、专业和完整, 可以利用临床指标关联用药情况, 进行深入挖掘, 还可识别在亚群中出现的药品不良反应。

### 2.5.3 关键词细分

根据上述关键词频次分析和聚类分析结果, 机器学习在药物警戒领域的应用主要在信号识别、社交媒体信息、电子医疗信息 3 个方面的挖掘与运用。将三类细分及相关关键词抽取, 利用社区发现算法和 R 语言 ggraph 包进行聚类关联并绘制网络图。点的大小代表与其相关关键词连接的总次数, 线条的粗细代表两个关键词在同一篇文章中出现的次数。

图 4 展示信号检测 (signal detection) 与其相关关键词的关联网络, 可以看出与其相关联频次排名前 5 位的关键词有自发呈报系统 (spontaneous reporting system)、FAERS、ROR (reporting odds ratio)、各类数据库 (databases)、PRR (proportional reporting ratio) 等。其中信号检测与自发报告系统关联组线最粗, 提示研究关联性最紧密, 研究最多。研究最深入的网络 (点大且密集) 是以 FAERS、ROR、各类数据库、PRR、经验贝叶斯几何平均值 (empirical bayes geometric mean) 法、信息技术 (information technology)、信号 (signal) 等关键词构成的。表明通过收集和评估个别病例报告来发现安全信号仍然是药物警戒的核心活动, 并且药品不良反应信号识别与监测广泛使用比例失衡分析方法, FAERS 等数据库为此提供了数据基础<sup>[9]</sup>。

机器学习的出现使得从社交媒体中识别和提取药物警戒相关信息成为可能, 拓宽了药物警戒数据来源。图 5 展示在该领域的应用情况, 与社交媒体 (social media) 有关的关键词主要包括 Twitter (社交平台名称)、Facebook (社交平台名称)、生物医学文献 (biomedical literature)、在线健康社区 (online health communities)、电子患者文本 (electronic patient-authored text) 等。通过与这些关键词的关联, 对社交媒体相关的药物警戒研究主要分为三类, 一是 Twitter、Facebook 等通用的社交媒体, 涉及到对此类信息的处理方

法包括情感分析 (sentiment analysis)、无监督预训练 (unsupervised pretraining)、网络分析 (network analysis) 等。二是在线健康论坛 (online health forums)，涉及到的处理方法包括语义映射 (semantic mapping)、带注释的语料库 (annotated corpus)、信息抽取 (information extraction) 等。三是在线健康社区 (online health communities)，对此类信息的处理方法包括异构网络 (heterogeneous networks)、元路径 (Meta path)、可信度分析 (credibility analysis) 等。

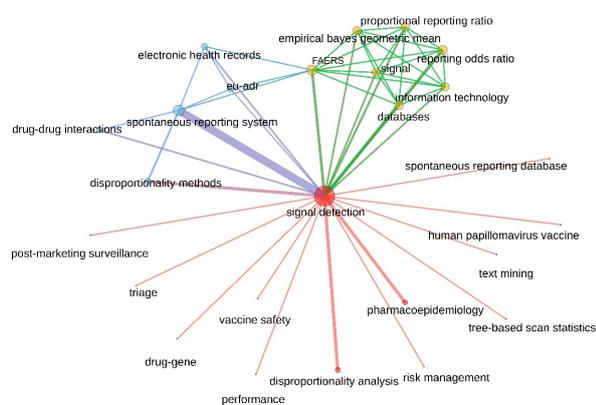


图4 信号检测相关关键词网络图

Figure 4. Signal detection network diagram

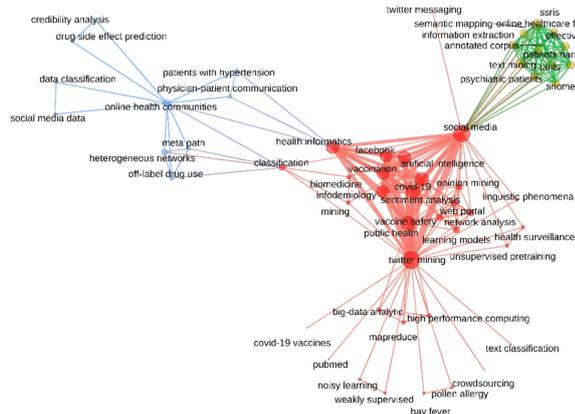


图5 社交媒体相关关键词网络图

Figure 5. Social Media related keywords network graph

Cocos 等<sup>[10]</sup>使用上下文感知模型和由大型未标记数据集形成的单词嵌入等方法，显著提高了社交媒体中的药品不良反应信号检测的可行性，减少了手动数据标记要求，并可扩展到大型社交媒体数据。Bian 等<sup>[11]</sup>通过使用自然语言处理分析来自 Twitter 的 2 亿条推文消息，并

构建支持向量机分类器来查找吸毒者和潜在不良事件的方法。Masino 等<sup>[12]</sup>使用卷积神经网络和词嵌入功能检测 Twitter 上的药品不良反应。在线健康社区主要是指在线用户根据互联网环境中的医疗健康相关信息描述疾病状态的网络社区，如医学知识共享、健康社区成员互动、专家咨询等行为，为医患沟通提供便利，同时也产生了大量的有用数据。利用机器学习挖掘社交媒体中的不良反应信息已经付诸实践，已成为一个研究的热点。但是社交媒体数据存在非结构化、复杂、多噪音、隐私性等特点，为了实现社交媒体带来的好处，许多技术、监管和道德挑战仍有待解决<sup>[13]</sup>。

结果见图 6，药物警戒领域利用机器学习算法对电子医疗信息非结构化的文本数据进行挖掘和运用，主要涉及的关键词包括电子健康记录 (electronic health records, EHR)、电子医疗记录 (electronic medical records, EMR)、医疗记录 (medical records)、健康信息学 (healthcare informatics)、临床药学信息系统 (clinical pharmacy information system) 等。围绕 EHR 和 EMR 进行深入研究，涉及到的算法包括网络分析 (network analysis)、患者分层 (patient stratification)、监管决策支持 (regulatory decision support)、语义网络技术 (semantic web technology)、聚类聚合 (clustering aggregation) 等。现有的观察性研究主要依靠结构化的 EHR 数据来获取药品不良反应信息，然而药品不良反应/事件通常被埋藏在 EHR 叙述中，而不是记录在结构化数据中<sup>[14]</sup>。机器学习的应用促进了 EHR 和 EMR 的深入挖掘，有效提高药品安全监测水平。

## 2.6 国家、作者与关键词

结果见图 7。将发文量前 3 位的国家相应排名前三的第一作者及相应前五位关键词使用节点链接图可视化，第一层为国家，第二层为作者，第三层为作者对应的关键词。美国的作者包括 Ji YQ、Zheng CY、Zolnoori M，其主要研究与机器学习有关的算法包括关联规则 (association rules)、模糊逻辑 (fuzzy logic)、时间关联挖掘 (temporal association mining) 等；中国的作者有 Liu J、Shu YM、Zhu LL，其主要研究内容与机器学习有关的算法包括整体学习 (ensemble learning)、多视角主动学习 (multi-



多的机构分别是 Columbia University (美国哥伦比亚大学)、Stanford University (美国斯坦福大学) 和 Stockholm University (瑞典斯德哥尔摩大学), 均为国际知名院校, 这些机构的跨学科、跨地域合作模式, 不仅促进了知识的交流和技术创新, 也为全球药物警戒研究的网络构建提供了范例。其中 Columbia University 在该领域主要研究使用机器学习算法来分析和挖掘大规模的医疗数据, 以便快速检测和预测药品副作用和不良反应。此外, 他们还研究将机器学习应用到药品开发和临床试验中 (<https://www.dbmi.columbia.edu/>)。Stockholm University 的研究涉及利用机器学习来改进药品不良反应监测和药物安全性评估的方法。如 Stockholm University 的 Zhao 等<sup>[16]</sup>将机器学习用于现有的 EHR 数据以快速检测药品不良反应, 尝试自动检测药品不良反应事件; Henriksson<sup>[17]</sup>将语义空间集合应用于医疗问题, 有助于医疗保健数据的二次使用。另外, 美国食品药品监督管理局 (FDA) 自 2020 年起致力于使用机器学习进行药物安全性预测, 其基于上市药物的药理学靶向不良事件谱预测新药在批准时未标记的不良事件。FDA 的 Daluwatte 等<sup>[18]</sup>通过使用机器学习模型中的其他特征来推进这项研究, 预测了 53 个具有高阳性预测值的严重不良事件, 其中存在明确的靶-事件关系。Schotland 等<sup>[19]</sup>改进了 Daluwatte 等的药理学靶-事件预测模型, 预测批准的药物标签上的不良事件。FDA 的研究成果表明集成的机器学习模型能够预测药品上市后的不良事件, 具有良好的整体性能, 能够更好地了解药物安全性。这些知名院校机构的研究将机器学习应用到药物警戒的各个环节, 为机器学习在药物警戒中的应用和发展提供了新的思路和方向。

### 3.4 关键词分析揭示的研究热点

关键词分析的结果突显了 3 个主要的研究热点: 信号识别、社交媒体挖掘以及电子医疗信息处理。这些热点与药物警戒领域的实际挑战紧密相连, 特别是在药品不良反应的快速识别和大规模医疗数据的有效管理方面。

信号识别在药物警戒中扮演着至关重要的角色。机器学习技术通过结合各类药品安全相关数据, 显著提升了检测、识别、关联和预测药品不良反应/事件的速度和效率<sup>[20-21]</sup>。研究者们利用

自发呈报系统中的个案报告, 应用比例失衡算法 (如 PRR、ROR) 和逻辑回归等方法, 对药品与不良反应的关系进行深入分析, 挖掘药品安全风险信号<sup>[22]</sup>。此外, 机器学习自动化地对不良事件报告中的药品信息进行编码, 与标准化药物字典 (如 WHO Drug Global) 进行匹配, 极大节约了时间和人力资源<sup>[23]</sup>。

社交媒体挖掘作为新兴的数据源, 在药物警戒领域显示出巨大潜力<sup>[10]</sup>。Lian 等<sup>[24]</sup>使用自然语言处理技术分析 Twitter 中的数据, 成功识别了新型冠状病毒疫苗的不良事件。这表明社交媒体数据可以作为传统药物警戒数据的重要补充

电子医疗信息处理方面, 机器学习技术被用于深入挖掘电子健康记录和电子医疗记录, 以识别药品不良反应信息<sup>[14, 25]</sup>。这种方法不仅提高了药品安全监测的水平, 还有助于监管决策支持和患者分层<sup>[26]</sup>。

尽管机器学习在药物警戒领域的应用前景广阔, 但也面临一些挑战。首先, 机器学习的可解释性、稳定性和准确性仍需进一步研究<sup>[27]</sup>。其次, 资源的缺乏, 包括专有的药物警戒系统数据库、人力资源以及政府支持的不足, 限制了机器学习技术的广泛应用。最后, 隐私和安全问题导致数据共享困难, 这也是制约机器学习发展的一个重要因素<sup>[28]</sup>。面对这些挑战, 未来的研究需要集中精力提高机器学习模型的透明度和可信度, 同时寻求跨学科合作, 整合资源, 以推动机器学习在药物警戒领域的进一步发展。

### 3.5 启示与局限性

机器学习技术在药物警戒领域, 特别是在社交媒体和电子医疗信息中提取和识别药品安全信息方面, 已经取得较大进展。如何更有效地确定、收集、确认最有力的证据, 以辅助临床决策和患者选择<sup>[29]</sup>, 是未来研究的重点之一。因此, 应进一步推动跨学科、跨机构和跨地域的交流与合作, 整合证据和优化技术方法, 进一步加强计算机科学、医学、统计学和数学等学科与药物警戒领域的融合研究。

Web of Science 数据库在学术文献收录方面具有权威性和广泛性, 但本研究仅选择该库作为数据来源, 具有一定的局限性。单一数据库可能未能全面覆盖所有相关文献, 尤其是发表在地区性或专业性数据库中的研究。此外, 检索策略可能

受限于关键词选择和逻辑组合,影响文献的全面性和代表性。

针对这些局限性,未来的研究应考虑扩展检索的数据库范围,包括但不限于 PubMed、Scopus 和 Embase 等文献数据库,以提高文献检索的全面性;考虑多语言文献的检索,以增强研究的国际性和包容性。这将有助于更全面地理解机器学习在药物警戒领域的应用,并推动该领域的研究向前发展。

### 参考文献

- 1 杨悦. 我国药物警戒制度实施与 ICH 药物警戒指导原则转化适用 [J]. 中国药物警戒, 2020, 17(2): 65-71. [Yang Y. Implementation of pharmacovigilance system in drug administration law and application of ICH E2 guideline[J]. Chinese Journal of Pharmacovigilance, 2020, 17(2): 65-71.] DOI: 10.19803/j.1672-8629.2020.02.01.
- 2 Hauben M, Hartford CG. Artificial intelligence in pharmacovigilance: scoping points to consider[J]. Clin Ther, 2021, 43(2): 372-379. DOI: 10.1016/j.clinthera.2020.12.014.
- 3 Kompa B, Hakim JB, Palepu A, et al. Artificial intelligence based on machine learning in pharmacovigilance: a scoping review[J]. Drug Saf, 2022, 45(5): 477-491. DOI: 10.1007/s40264-022-01176-1.
- 4 Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation[J]. Eur J Clin Pharmacol, 1998, 54(4): 315-321. DOI: 10.1007/s002280050466.
- 5 Nikfarjam A, Sarker A, O'Connor K, et al. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features[J]. J Am Med Inform Assoc, 2015, 22(3): 671-681. DOI: 10.1093/jamia/ocu041.
- 6 刘娟娟, 陈奕, 张佳玲, 等. 21 世纪全球视域下陇药红芪科学协作与热点前沿知识图谱构建及可视化分析[J]. 中草药, 2023, 54(12): 3932-3948. [Liu JJ, Chen Y, Zhang JL, et al. Construction and visual analysis of scientific collaboration and hot frontier knowledge graph of Longyao *Hedysari Radix* from global perspective in 21st century[J]. Chinese Traditional and Herbal Drugs, 2023, 54(12): 3932-3948.] DOI: 10.7501/j.issn.0253-2670.2023.12.020.
- 7 Leal MM, Sanz MM, Ferrando JRC, et al. A comparative analysis of the pharmacovigilance systems of brazil, spain, the european union and the united states based on the information provided by their regulatory agency websites[J]. Daru, 2019, 27(1): 379-387. DOI: 10.1007/s40199-019-00249-4.
- 8 杨婷, 张晓滕, 张冰, 等. 国内外痛风性关节炎的研究现状—基于 CiteSpace 软件的可视化分析 [J]. 中国实验方剂学杂志, 2020, 26(20): 169-177. [Yang T, Zhang XM, Zhang B, et al. Research status of gouty arthritis at home and abroad—visual analysis based on citespace software[J]. Chinese Journal of Experimental Traditional Medical Formulae, 2020, 26(20): 169-177.] DOI: 10.13422/j.cnki.syfjx.20202018.
- 9 Caster O, Aoki Y, Gattepaille LM, et al. Disproportionality analysis for pharmacovigilance signal detection in small databases or subsets: recommendations for limiting false-positive associations[J]. Drug Saf, 2020, 43(5): 479-487. DOI: 10.1007/s40264-020-00911-w.
- 10 Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts[J]. J Am Med Inform Assoc, 2017, 24(4): 813-821. DOI: 10.1093/jamia/ocw180.
- 11 Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events[J]. SHB12 (2012), 2012, 2012: 25-32. DOI: 10.1145/2389707.2389713.
- 12 Masino AJ, Forsyth D, Fiks AG. Detecting adverse drug reactions on twitter with convolutional neural networks and word embedding features[J]. J Healthc Inform Res, 2018, 2(1-2): 25-43. DOI: 10.1007/s41666-018-0018-9.
- 13 Chen X. Online health communities influence people's health behaviors in the context of COVID-19[J]. PLoS One, 2023, 18(4): e0282368. DOI: 10.1371/journal.pone.0282368.
- 14 Munkhdalai T, Liu F, Yu H. Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning[J]. JMIR Public Health Surveill, 2018, 4(2): e29. DOI: 10.2196/publichealth.9361.
- 15 Zhao Y, Yu Y, Wang H, et al. Machine learning in causal inference: application in pharmacovigilance[J]. Drug Saf, 2022, 45(5): 459-476. DOI: 10.1007/s40264-022-

- 01155–6.
- 16 Zhao J, Henriksson A, Boström H. Cascading adverse drug event detection in electronic health records[C]. IEEE International Conference on Data Science & Advanced Analytics, 2015. DOI: [10.1109/DSAA.2015.7344869](https://doi.org/10.1109/DSAA.2015.7344869).
  - 17 Henriksson A. Ensembles of semantic spaces: on combining models of distributional semantics with applications in healthcare[EB/OL]. (2015) [2023–07–11]. [https://xueshu.baidu.com/usercenter/paper/show?paperid=582cbcb8f6e8a12212253cae16827812&site=xueshu\\_se&hitarticle=1](https://xueshu.baidu.com/usercenter/paper/show?paperid=582cbcb8f6e8a12212253cae16827812&site=xueshu_se&hitarticle=1).
  - 18 Daluwatte C, Schotland P, Strauss DG, et al. Predicting potential adverse events using safety data from marketed drugs[J]. BMC Bioinformatics, 2020, 21(1): 163. DOI: [10.1186/s12859-020-3509-7](https://doi.org/10.1186/s12859-020-3509-7).
  - 19 Schotland P, Racz R, Jackson DB, et al. Target adverse event profiles for predictive safety in the postmarket setting[J]. Clin Pharmacol Ther, 2021, 109(5): 1232–1243. DOI: [10.1002/cpt.2074](https://doi.org/10.1002/cpt.2074).
  - 20 Edrees H, Song W, Syrowatka A, et al. Intelligent telehealth in pharmacovigilance: a future perspective[J]. Drug Saf, 2022, 45(5): 449–458. DOI: [10.1007/s40264-022-01172-5](https://doi.org/10.1007/s40264-022-01172-5).
  - 21 周虎子威, 张云静, 于玥琳, 等. 机器学习方法在预测麻精药品不合理使用风险中的应用现状和思考[J]. 药物流行病学杂志, 2023, 32(4): 446–457. [Zhou HZW, Zhang YJ, Yu YL, et al. Application of machine learning methods in predicting the risk of irrational use of narcotic and psychotropic drugs: current status and considerations[J]. Chinese Journal of Pharmacoepidemiology, 2023, 32(4): 446–457.] DOI: [10.19960/j.issn.1005-0698.202304010](https://doi.org/10.19960/j.issn.1005-0698.202304010).
  - 22 Vilar S, Friedman C, Hripesak G. Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media[J]. Brief Bioinform, 2018, 19(5): 863–877. DOI: [10.1093/bib/bbx010](https://doi.org/10.1093/bib/bbx010).
  - 23 Meldau EL, Bista S, Rofors E, et al. Automated drug coding using artificial intelligence: an evaluation of WHO drug koda on adverse event reports[J]. Drug Saf, 2022, 45(5): 549–561. DOI: [10.1007/s40264-022-01162-7](https://doi.org/10.1007/s40264-022-01162-7).
  - 24 Lian AT, Du J, Tang L. Using a machine learning approach to monitor COVID–19 vaccine adverse events (VAE) from Twitter data[J]. Vaccines (Basel), 2022, 10(1): 103. DOI: [10.3390/vaccines10010103](https://doi.org/10.3390/vaccines10010103).
  - 25 李海龙, 赵厚宇, 周一帆, 等. 基于电子医疗数据库的药品不良反应信号挖掘方法概述 [J]. 药物流行病学杂志, 2018, 27(8): 541–549. [Li HL, Zhao HY, Zhou YF, et al. Data mining methods for adverse drug reaction signals detection in healthcare databases: a literature review[J]. Chinese Journal of Pharmacoepidemiology, 2018, 27(8): 541–549.] DOI: [10.19960/j.cnki.issn1005-0698.2018.08.012](https://doi.org/10.19960/j.cnki.issn1005-0698.2018.08.012).
  - 26 Xu Q, Ahmadi E, Amini A, et al. Identifying and mitigating potential biases in predicting drug approvals[J]. Drug Saf, 2022, 45(5): 521–533. DOI: [10.1007/s40264-022-01160-9](https://doi.org/10.1007/s40264-022-01160-9).
  - 27 Scabro S, Portelli B, Chersoni E, et al. Increasing adverse drug events extraction robustness on social media: case study on negation and speculation[J]. Exp Biol Med (Maywood), 2022, 247(22): 2003–2014. DOI: [10.1177/15353702221128577](https://doi.org/10.1177/15353702221128577).
  - 28 Bate A, Luo Y. Artificial intelligence and machine learning for safe medicines[J]. Drug Saf, 2022, 45(5): 403–405. DOI: [10.1007/s40264-022-01177-0](https://doi.org/10.1007/s40264-022-01177-0).
  - 29 Rocca E, Copeland S, Ralph Edwards I. Pharmacovigilance as scientific discovery: an argument for trans–disciplinarity[J]. Drug Saf, 2019, 42(10): 1115–1124. DOI: [10.1007/s40264-019-00826-1](https://doi.org/10.1007/s40264-019-00826-1).

收稿日期: 2023 年 09 月 22 日 修回日期: 2024 年 06 月 21 日  
 本文编辑: 周璐敏 杨燕