NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

IMPACT OF SOCIAL MEDIA DATA ON POST-MARKET SAFETY EVALUATION OF MEDICINES

A literature review on automatic mining initiatives of adverse drug reactions

Ana Catarina Simões Martins Mendes

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

IMPACT OF SOCIAL MEDIA DATA ON POST-MARKET SAFETY EVALUATION OF MEDICINES

by

Ana Catarina Simões Martins Mendes

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence

Advisor: Professor Doutor Guilherme Martins Victorino

March 2021

ABSTRACT

Introduction/Background: The impact of adverse drug reactions (ADRs) on public health and national healthcare systems is substantial. The current pharmacovigilance method is time-consuming, incomplete and prone to data loss. Also, due to characteristics inherent to the reporting process, a great portion of ADRs are never reported. Social media (SM) data, due to its volume and immediacy, shows promise for a patient centered way of reporting, and has received increasing attention over the last few years.

Objectives/Methodology: In this research project the author proposes to evaluate how can ADR automatic detection from social media contribute for pharmacovigilance, through a systematic literature review. The review included articles published over the last five years, accounting to 33 publications that were retrieved and reviewed in detail.

Discussion: Several aspects have proven to be critical when developing SM based ADR mining - the main purpose of the analysis (detection of posts containing ADRs and the extraction of specific ADR-drug pairs), the approach (lexicon or machine learning based), and the type of platform used (health-focused or general use). The studies have shown a prevalence of machine learning (ML) based approaches, from which supervised learning is the most popular method, despite the rising trend against the need for costly and time-consuming annotation of data. Mixed approaches have often been used as they seem to derive better performance, whether in combining data sources from general platforms and disease forums, or using distinct sources of annotated data sets, such as biomedical corpus to increase algorithms strength, and even the combination of ML approaches with lexicon based features.

Conclusions/Limitations: The end goal of ADR mining from social media is to be able to identify drugs that are either frequently related to ADRs, or those that are associated with previously unknown ADRs. Combining data from multiple sources will contribute to prevent the impact of serious or previously unknown ADRs, focusing on the issues most pertinent to patients, and will provide a broader safety profile of any medication, with benefits for patients, health systems, companies and regulatory agencies. SM data comes with its specificities (informal language, semantic confusion and ambiguity) that lead to analysis hurdles; hence the method and approach used must be adapted to the purpose of investigation and resources available. Norms and agreed practices to guide these efforts are needed, considering ethical issues, data quality and governance. The progress in information technology and the need to consider patient experience should motivate future research on social media surveillance for complementing conventional pharmacovigilance with patient centric and real time ways of reporting.

KEYWORDS

Social Media; Real World Data; Pharmacovigilance; Medicines; Machine Learning; Adverse drug reactions

INDEX

1. Introduction 1
   1.1. Objectives 2
2. Background 4
   2.1. Pharmacovigilance and data collection – the need for a new paradigm 4
   2.2. Big data in pharmacovigilance 4
   2.3. Social media - From consumer-generated data to ADR data 5
3. Methods 7
   3.1. Data search and selection 7
   3.2. Data abstraction 8
4. Results 9
   4.1. Data collection results 9
   4.2. Dimensions of characterization 10
   4.3. A summary of methodologies and resources 17
      4.3.1. Lexicons and knowledge bases 17
      4.3.2. Automatic classification of ADR containing user posts 18
      4.3.3. ADR mention extraction 19
   4.4. A summary of evaluation techniques and metrics 20
      4.4.1. Qualitative evaluation 20
      4.4.2. Quantitative evaluation 21
5. Discussion 22
6. Limitations and recommendations for future works 26
7. Conclusion 28
8. Bibliography 30

LIST OF FIGURES

Figure 1 - Overview of research objectives 3
Figure 2 - Data search and selection overview 7
Figure 3 - Yearly distribution of publications 9
Figure 4 - Most used social media platform types 12
Figure 5 - Primary ADR detection approach used 12
Figure 6 - Frequence of use of data annotation (supervised approaches) done by experts or other methods 13

LIST OF TABLES

Table 1 - Summary of critical aspects of the studies: publication year, data sources, subdomains of focus (when applicable), number of drugs involved in the studies, the size of the data used, and annotations and availability of the data 11

Table 2 - Summary of the ADR detection/extraction approaches proposed by the studies, their primary research aims, and how the evaluations were designed, to enable an understanding of the success of different approaches for ADR detection/extraction problems. 16

Table 3 - Description of lexicons used in each of the research publications (adapted from Sarker et al., 2015) 18

LIST OF ABBREVIATIONS AND ACRONYMS

ADR	Adverse Drug Reaction
EFPIA	European Federation of Pharmaceutical Industries and Associations
EMA	European Medicines Agency
EU	European Union
FDA	U.S. Food and Drug Administration
HCPs	Health Care Professionals
HMA	Heads of Medicines Agency
ICD	International Classification of Diseases
IMI	Innovative Medicines initiative
PMS	Post-market Surveillance
ML	Machine Learning
NCAs	National competent authorities
NLP	Natural Language Processing
R&D	Research and Development
RCT	Randomized Controlled Trials
RWD	Real World Data
SM	Social Media
WHO	World Health Organization

1. INTRODUCTION

We live in a new era of healthcare, one that is driven by increasing connectivity of health information from several sources (Catell, Chilukuri, & Levy, 2013). Data, in its diverse forms, is accessible and fast, creating countless opportunities for companies in the sector. Nontraditional sources of information are proliferating in healthcare settings – from mobile apps that register drug intake and effects, drugs with built-in sensors that monitor drug adherence, social media platforms that improve Adverse Drug Reaction (ADR) notification rates to build a more realistic safety profile of drugs, medical device software that monitor patient disease progression, wearable devices that track lifestyle options or biometric parameters, and several more (Bate, Reynolds, & Caubel, 2018). The possibilities these data sources enable to sector players are limitless. Other sectors (such as retail or finance) have proven the value that can be obtained from large-scale integration and analysis of heterogeneous data sources (Weber, Mandl, & Kohane, 2014). However, and despite the widespread use of electronic health records, administrative claims, social media and the ubiquity of smart devices, these big data sources have not yet been widely used by healthcare companies (Jarow, LaVange, & Woodcock, 2017). Analysis of this data can generate the scientific evidence that can help understanding diseases, identify new therapeutic intervention points, improve the efficiency of research and development (Wise, Möller, Christie, & Brodsky, 2018) as well as predict treatment effectiveness and safety and understand causal relationships between drugs and health outcomes (Schneeweiss, 2016).

The pharmaceutical/ biotechnological business is built on discovery and development of innovative products. It is a highly competitive market, with a drug-development process known to be extensive, complex and extremely costly: it will normally proceed through four phases of clinical trials over many years (Bate, Juniper, Lawton, & Thwaites, 2016). If the drug successfully passes through phases I, II, and III, it is approved by the national regulatory authorities for use. Phase IV are post-marketing (PM) studies, which aim to reconcile what we know at the end of the clinical trial program with what we do not know, namely: what happens when the medicinal product is used in regular practice? What is its adverse event profile? PM studies are required to continue throughout the product lifecycle (Bere, 2017). Therefore, assessments of the safety, efficacy and appropriate use of new medicines are necessary for treatment development and adoption in clinical practice. Highly controlled randomized clinical trials constitute the gold standard for evaluating the safety and effectiveness of medications, to inform decisions on the approval, coverage and use of a medicine. Researchers and decision makers are becoming increasingly aware that these experimental data alone are insufficient to address those decisions fully (Wise et al., 2018). Real world data (RWD) recorded from routine healthcare delivery helps to provide a more complete representation of effectiveness and safety outcomes (Bate et al., 2016).

Safety signals can be detected from a range of sources, from patient inquiries on websites and search engines to online physician communities, electronic medical records, and consumer-generated media (Catell et al., 2013). For the latter, the mining of social media (SM) has the potential to provide valuable insights into a patient's real-life use of therapeutics (Härmark, Hubert, & Ralph, 2016). The immediacy of these data, along with their widespread availability, can provide insights into adverse effects and noncompliance. Already widely used by commercial organizations to gain brand insight, pharmaceutical companies, regulators, and others are now actively investigating social listening as an

innovative and additional method of pharmacovigilance (Wise et al., 2018). Theoretical benefits are mainly two: reduced costs and accelerated results, i.e., a reduced time until ADR detection (Singh, Schulthess, Hughes, Vannieuwenhuyse, & Kalra, 2018; Stekelenborg et al., 2019). ADR monitoring is therefore moving beyond traditional approaches towards sophisticated methods that identify potential safety signals for unknown adverse events.

The impact of ADRs on public health and national healthcare systems is huge; according to recent data published by the European Commission, ADRs are estimated to cause approximately 5% of all hospital admissions and 197,000 deaths annually throughout the European Union (EU). In addition, 5% of hospitalized patients experience an ADR during their hospital stay (European Commission, 2018). There are clear public health benefits of a more efficient system for evaluation of medicines, for all healthcare parties.

Regulatory agencies FDA (Food and Drug Administration) and EMA (European Medicines Agency) together with pharmaceutical companies are developing joint electronic-surveillance systems to track real-world experiences in dealing with the disease and medications (Ghosh & Lewis, 2015; Jarow et al., 2017; Stekelenborg et al., 2019). One of them is FDA's *MedWatcher Social* - an exploratory data mining tool that monitors several social media channels (including Facebook and Twitter) for ADRs related to 1,400 drugs (Harpaz & Dumochel, 2016). Also, the European social listening initiative *WEB-RADR: Recognizing Adverse Drug Reactions*, is a project developed by Innovative Medicines Initiative (IMI) and European Federation of Pharmaceutical Industries and Associations (EFPIA) set out to provide access to social media data via a visualization platform to allow signal identification and confirmation as well as to develop analytical tools for the analysis of social media content for pharmacovigilance purposes (Ghosh & Lewis, 2015). In parallel, The HMA (Heads of Medicines Agencies) and EMA Joint Big Data taskforce was formed to describe the big data landscape from a regulatory perspective in order to ensure the EU regulatory system has the capability to guide, analyze and interpret these data (European Medicines Agency, 2017). This initiative revealed, for instance that National competent authorities (NCAs) have access and tools to treat big data sets and that they identify, among many applications, signal detection as a priority for big data use (HMA & EMA, 2017). Also, NCAs and pharma recognised there are significant challenges in the use of these data, namely those related to access to the data, understanding their strengths and limitations and the new analytical methods needed to integrate and analyze the datasets in order to generate conclusions needed for decision making (European Medicines Agency, 2017).

1.1. OBJECTIVES

In this research project, the author proposes to evaluate the contributions of automatic ADR detection from social media for pharmacovigilance, through a systematic literature review.

This assessment aims to contribute to guide the development of pharmacovigilance projects using social media data, by identifying priorities, best practices and by uncovering future investigational paths. The focus on SM comes from the volume of users, speed of data generation and possible interactions, making it the most interesting big data source for this purpose. The aim of this research project is to (1) compare results of investigations published over the last few years regarding the main approach used and the type of platform data was gathered from (general use or health-focused SM). According to many authors, it is important to identify the best uses of each data type for pharmacovigilance, including which patient populations, outcomes, or medicines are best suited for

signal detection (Bate et al., 2018; European Medicines Agency, 2017; Price, 2016). Therefore, a second objective of this research is to (2) evaluate critical points on these investigations to assess if any of those (such as evaluating different medications/disease groups) may be related to better results.

With this research work it is intended to answer the following questions:

- What are the potential contributions of social media data for pharmacovigilance?
- Are there performance differences when using different approaches or different platforms types?
- How can we take advantage of each approach in order to maximize success in automatically detecting ADR from SM?

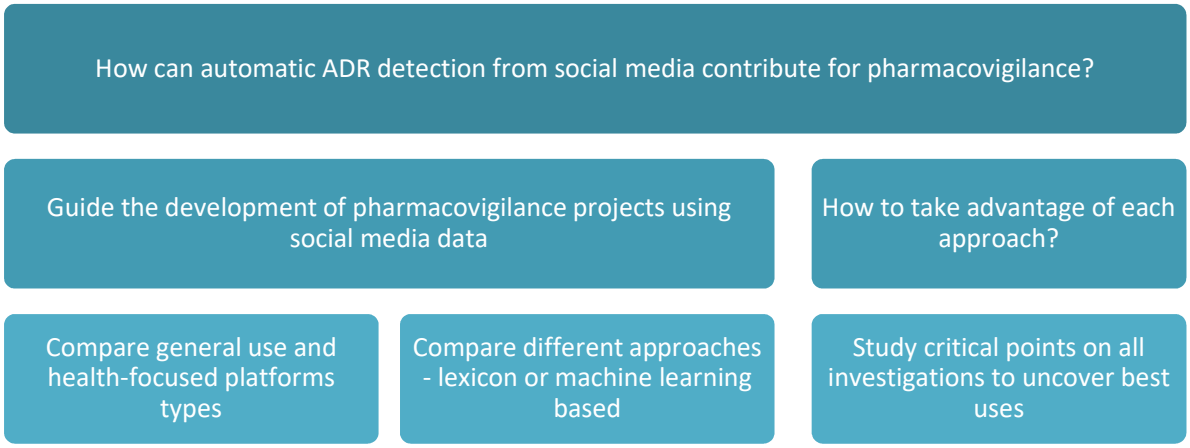


Figure 1 - Overview of research objectives

A background summary is presented on the following section to inform the sample definition and to allow positioning the findings within the current investigation landscape.

The remaining of the dissertation is organized as follows: the *Background* section covers the developments on the Pharmacovigilance field and the use of social media data for ADR detection; the *Methods* section discusses data search, selection and abstraction approaches used for the survey; the *Results* section elaborates on the findings of the survey, including summaries of all the studies that met the inclusion criteria; the *Discussion* section summarizes the main findings while relating to the research objectives; *Limitations and Recommendations for future works* cover some issues found while developing this literature survey and directions for future investigations, and finally the *Conclusions* of the research.

2. BACKGROUND

2.1. PHARMACOVIGILANCE AND DATA COLLECTION – THE NEED FOR A NEW PARADIGM

Pharmacovigilance is defined as the science and activities related to the detection, assessment and prevention of adverse effects or any other drug-related effects. The World Health Organization (WHO) established its Program for International Drug Monitoring in response to the thalidomide calamity in 1961 (Fermont, 2019). Since then, the gold-standard source for ADR detection is the spontaneous suspected ADR reporting system. Spontaneous reporting systems have proven to be vital to post-marketing surveillance, and are effective at detecting many types of ADRs (Harpaz & Dumochel, 2016). The system in place depends upon patients or healthcare professionals to report a potential ADR, which will then be assessed by drug experts and pharmaceutical companies, and the results are then passed on to regulatory agencies. From receiving the reports to the assessment and release, it can take up to several months or even years. This whole process leads to substantial delays and eventual data loss. Another issue with this system is that ADR reporting systems are largely unknown to the public, and among physicians ADR reporting has been declining. The significant delays in detecting ADRs, and the realization that a substantial number of ADRs remain unreported – as Salathé (2016) presented in a recent study made on the United States in which hospital staff did not report 86% of suspected ADRs among patients - have led to the search of complementary approaches for ADR detection (Harpaz & Dumochel, 2016). While patients rarely use official ADR reporting systems, they increasingly use online platforms to investigate and discuss potential ADRs (Salathé, 2016).

In Portugal, the under-notification of ADRs is also a reality, despite its significant impact for public health and health system. In a study of the impact of ADRs in Portuguese patients and health system, data for all hospitalizations that took place between 2000 and 2015 was analyzed, and 5.8% (861 thousand) of these hospitalizations were related to at least one ADR. Hospitalizations with registration of adverse events had a median length of stay of 8 days, median hospitalization costs of 3060.7 Euro, and an in-hospital mortality of 6.7%. The financial impact of these was estimated around 4.8 thousand million euros for all adverse events and related costs (treatments/ hospitalizations). Also, from all cases of ADRs identified in the study, only 17% had been officially notified (Sousa-pinto, Marques, Lopes, & Freitas, 2018).

2.2. BIG DATA IN PHARMACOVIGILANCE

In the last years, 'big data' has become an important asset for several sectors, including healthcare (Schneeweiss, 2016; Schultz, 2013). At first, big data in healthcare referred to the vast volumes of computerized medical information available in the form of electronic health records, administrative or health claims data, disease and drug monitoring registries and some of the data collected through clinical practice, including patients' medical history and drug prescriptions (Jarow et al., 2017; Schneeweiss, 2016). For a long time, these data accumulated without its value being fully recognized and leveraged (Jarow et al., 2017; Trifirò, Sultana, & Bate, 2018). More recently, the notion of big data in healthcare, expanded to include all the information routinely collected as part of patient care, including images and readings from medical equipment, laboratory results as well as the data created by Internet search results and the widespread use of social media and smart devices (Jarow et al., 2017; Min, Osborne, Lynn, & Shakir, 2018). These data can support the assessment of benefits

and risks associated with the use of a medical product and support treatment decisions (Min et al., 2018).

The use of big data for pharmacovigilance has been growing, in an attempt to overcome issues inherent to the traditional spontaneous reporting system. The expanding roles of big data in pharmacovigilance include signal detection, substantiation and validation of drug safety signals (Trifirò et al., 2018; M. Yang, Kiang, & Shang, 2015).

The big data sources, with respect to healthcare, can be divided into electronic health records and patient-generated data. This data can come in the form of coded (structured) information, for example, using International Classification of Diseases (ICD) codification, or in the form of unstructured data, such as physicians' notes from their practice or discharge files. The two data types have complementary strengths—high veracity in the data from traditional sources and high velocity and variety in patient-generated data. However, this last category also comes with unique challenges - patient-generated data is unstructured and context dependent (Salathé, 2016).

Regarding the tools needed to analyze this huge amounts of data, machine learning algorithms (supervised and unsupervised) have been used in pharmacovigilance for automated signal generation. Unsupervised algorithms have been used to identify complex drug safety signals and drug utilization patterns whereas supervised algorithms - such as Natural Language Processing (NLP) - have been extensively used for the correct identification of an ADR from free text (on clinical notes) to find associations between drug mentions and potential ADRs. NLP techniques have also been applied to free text from social media websites, where patients share information on their experience of potential ADRs (Leaman et al., 2010; Stekelenborg et al., 2019; Trifirò et al., 2018).

Several multi-database initiatives are shaping all over the world and have generated the need for creating big data infrastructures adapted for post-marketing drug monitoring. Two examples of such are FDA's program 'Sentinel' and European 'EU-ADR', which have provided access to big data sources for post-marketing pharmacovigilance. Both initiatives have shown to have the power to allow a faster marketing authorization withdrawal for drugs identified with serious safety issues and to increase the power of studies in rare outcomes or special populations (Jarow et al., 2017; Trifirò et al., 2018).

2.3. SOCIAL MEDIA - FROM CONSUMER-GENERATED DATA TO ADR DATA

The term social listening has been used to refer the mining of social media "as an innovative and additional method of pharmacovigilance" (Wise et al., 2018, p.8). Because social media posts usually appear in proximity to the time an event occurs, the use of social media offers opportunities for surveillance that are near real-time, and therefore allow early detection of issues (Bousquet, Dahamna, Guillemin-lanne, & Darmoni, 2017). Another advantage of analyzing patient narratives in discussion forums or blogs is the ability to explore patient-centered issues that may not be detected in existing sources generated by health care professionals (Bousquet et al., 2017). Furthermore, proactively sharing an experience of drug use online could elicit instant feedback from other affected patients within a community of drug users (Price, 2016) and therefore scale-up the data available for each potential drug-ADR pair (Trifirò et al., 2018).

Mostly due to its number of users, social media has become a useful resource for ADR monitoring (Sarker et al., 2015). Being this a patient centric model of reporting, the perception of patients on the effectiveness of treatments is of great importance for drug developers, healthcare providers, payers and other healthcare entities (Bate et al., 2018). However, the findings generated using social media data require clinical judgment, therefore, the process of drug safety signal detection and validation cannot be completely automated without expert assessment (Trifirò et al., 2018). The value of these new activities should be evaluated to understand to what extent they improve signal detection and assessment.

Despite the obvious potential, several issues have been raised in the literature regarding the use of social media for pharmacovigilance, such as feasibility of monitoring the massive volumes of data, concerns regarding the quality and reliability of the data as well as ethical issues (Harpaz & Dumochel, 2016). On quality of the data, the issues can be summarized in lack of specificity, verification difficulties (the data does not meet the basic regulatory definitions of an individual case safety report and lack sufficient details to conduct a medical assessment), low validity (since symptoms are not verified by an healthcare professional) and bias (data mainly from subpopulations who are most likely to share their experiences despite privacy issues and who have technical ability and means of access, which may assume exclusion of the elderly). In addition, these huge datasets contain enormous numbers of potential intervariable correlations, risking the discovery of many false positive correlations (Price, 2016).

3. METHODS

3.1. DATA SEARCH AND SELECTION

Figure 2 - Data search and selection overview

Pharmacovigilance using electronic data is a relatively recent research topic. The use of social media data for this purpose has received significant and growing research attention over the last few years. For this reason, this survey focuses on articles published over the last five years only (2015-2020), gathered using database SCOPUS and advanced search options to obtain relevant citations. When searching, besides enforcing the constraint associated with the year of publication and the language (English), several keyword-based constraints were also added. The aim was to obtain publications that mentioned both Pharmacovigilance and Social Media, and after preliminary searches to browse for the most appropriate keywords, several queries using related words were introduced in the database to uncover the combination that would achieve the most comprehensive set of publications of interest.

In the end, the selected keywords were “Pharmacovigilance” and “Social Media” for title, abstract or keywords only, since the preliminary search showed that their presence in the text was often unrelated to the topic of interest. The data search resulted in 152 publications, and following a preliminary analysis of the information present on titles and abstracts, an initial selection of 67 articles were retrieved for manual review. The inclusion criteria involved studies that presented original data, utilized any social media resource (e.g., forums, message boards, social networks), and indicated the use of automatic algorithms for ADR detection. The preliminary set of articles included studies for which it could not be immediately determined whether the detection methods were automatic or manual. The exclusion criteria included studies that utilized data sources other than patient-generated data. Studies were also excluded if they focused exclusively on drug-drug interactions, detected ADRs in randomized controlled trials, in drug labels, or were not published in English. Studies focused on pre-treatment stages of data for ADR identification and detection (such as creating drug synonyms lexicons for social media) were also excluded.

To ensure this literature review included all recent and relevant publications, alerts were set for new publications occurring since the main survey was performed. These alerts were set on SCOPUS following the same criteria used on the main survey. An additional alert was set for journals of interest: the Journal of Biomedical Informatics and Drug Safety were the publications where most of the articles were published. A last alert was set for the highest contributing author for this subject (Abeed Sarker) on database Google Scholar, in order to attempt to obtain additional articles to the ones on Scopus database. The results of the alerts were evaluated from the first literature search in June 2020 up until November 2020, and the relevant papers were included in the overall results. The final set consisted of 33 publications describing automatic methods for ADR detection/extraction from user posted data on social media. These are reviewed in detail over the following sections.

3.2. DATA ABSTRACTION

For all included studies, data was abstracted on characteristics such as study size, research aim(s), primary ADR identification/extraction approach, data source, availability of data, and the type of evaluation performed. For the study size factor, both data size and number of drugs were included. It is also indicated for each study if it focuses on a specific sub-domain of drugs (e.g., heart disease) or includes a general set of drugs. The classification of the primary identification/extraction technique focuses on the general approach that was employed at the final stage of detection. Some techniques relied on ADR lexicons, while another set of techniques relied on detecting linguistic patterns (ML based) for the ADR detection task. The data sources were classified based on the social network type used (health focused or general use platforms). In terms of availability of data, it was based on whether the data used for the study were publicly available for research purposes or not. For each study it is also indicated whether it utilizes annotated data (for supervised machine learning techniques), and if this annotation has been done by domain experts. Regarding the evaluation, articles were categorized based on the type of evaluation used to assess performance (qualitative or quantitative) and specific evaluation approaches and metrics used.

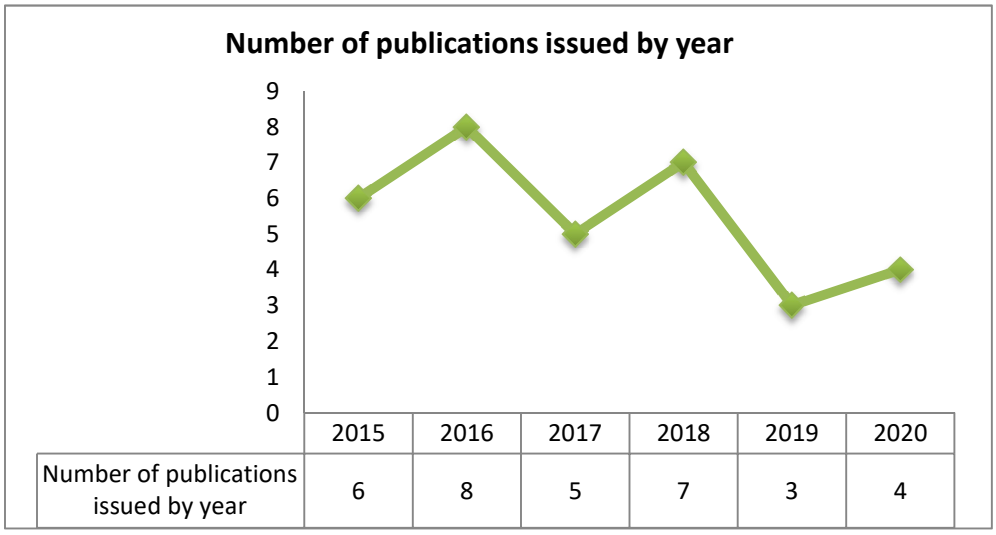
4. RESULTS

This section details the results obtained in the systematic literature survey. On sections 4.1 and 4.2 it is briefly presented the overall data collection process, some statistics and summary tables of the selected literature using the criteria mentioned above. The summary of methodologies and evaluations used are discussed in more detail over sections 4.3 and 4.4.

4.1. DATA COLLECTION RESULTS

Overall, 33 publications were selected for describing automatic methods for ADR detection/extraction from user posted data on social media. This set consists of journal articles and conference proceedings. The articles not included in the final set focused on topics such as creation of ADR lexicons for SM mining, pharmacovigilance studies focusing on data other than patient generated data, reviews and opinion articles. Exploratory studies using manual extraction of ADRs and works detecting personal experience posts on social media were also excluded.

This topic has maintained attention over the last 5 years, with 14 publications from the last three years.



	2015	2016	2017	2018	2019	2020
Number of publications issued by year	6	8	5	7	3	4

Figure 3 - Yearly distribution of publications

The earliest works (Cocos, Fiks, & Masino, 2017) identified are from 2015, and employ machine learning or mixed approaches using NLP based features and manually annotated data for evaluation. The works led by Abeed Sarker (Emadzadeh, Sarker, Nikfarjam, & Gonzalez, 2017; Korkontzelos et al., 2016; Nikfarjam, Sarker, Connor, Ginn, & Gonzalez, 2015; Sarker et al., 2015; Sarker & Gonzalez, 2015) are a cornerstone on the topic, as well as Jing Liu, both having done several publications over the last five years. Regarding journals of interest, most of these works have been published on Journal of biomedical informatics.

4.2. DIMENSIONS OF CHARACTERIZATION

This section presents the characterization of the studies using two summary tables with key information for each study reviewed. These illustrate how research on pharmacovigilance using social media has evolved over the last 5 years.

Title/authors	Source	Sub-domain	# instances	Annotation	Available	# drugs
(Nikfarjam et al., 2015)	Twitter (TW), dailystrenght (DS)	Newest and most used medications	6279 DS; 1784 TW	Yes(E)[1]	Yes	81
(Sarker & Gonzalez, 2015)	Dailystrenght and twitter	Chronic diseases; high prevalence of use	10822 TW; 10617 DS	Yes(E)	Yes	130
(M. Yang et al., 2015)	Medhelp	Most discussed - Biaxin, Lansoprazole, and Luvox	3500 threads	Yes(E)	Yes	3
(Korkontzelos et al., 2016)	Dailystrenght and twitter	Newest and most used medications	6279 DS; 1784 TW	Yes(E)	Yes	81
(Cocos et al., 2017)	Twitter	Newest and most used medications	1784 TW	Yes(E)	Yes	81
(Lee et al., 2017)	Twitter	–	7100 tweets	Yes	Yes	–
(X. Liu & Chen, 2015)	AMA[2] forum, Diabetes Forum, Medhelp Heart Disease discussion boards	Diabetes and heart disease medications	± one million postings	Yes(E)	No	–
(Tutubalina & Nikolenko, 2017)	Askapatient	–	1250 posts	Yes	No	–
(Segura-Bedmar, Martínez, Revert, & Moreno-schneider, 2015)	Forum clinic	–	84000 comments	Yes(E)	Yes	–
(J. Liu, Zhao, & Zhang, 2016)	Medhelp (M), diabetes forum, Twitter	Diabetes and heart disease, others	2200557(M) 61226(forum) 815(TW)	Yes(E)	Yes	–
(Segura-Bedmar et al., 2015)	Forum clinic	–	84000 posts	No	Yes	–
(Comfort, Perera, Hudson, & Dorrell, 2018)	Twitter, FB, tumblr and news media blogs	_	311189 posts	Yes(E)	No	–
(Eshleman & Singh, 2016)	Twitter	Most prescribed	1500 tweets	Yes	Yes	200
(Gupta, Pawar, Ramrakhiyani, Palshikar, & Varma, 2018)	Twitter	–	950 tweets	Yes	Yes	10
(J. Liu, Zhao, & Wang, 2018)	Medhelp	Heart disease	261464 posts	Yes	Yes	–
(Katragadda, Benton, Karnati, Pusala, & Raghavan, 2015)	Twitter	Medications on US drugs list	172800 tweets	Yes	Yes	1949
(Alimova & Tutubalina, 2018)	CADEC[3] and twitter corpus	_	6926 TW; 1248 CADEC	Yes	Yes	–
(Plachouras, Leidner, & Garrow, 2016)	Twitter	Most common, for hypertension	10000 tweets	Yes	No	518

[1] E – Studies that used data annotated by domain experts

[2] American Diabetes Association

[3] Corpus from askapatient

(Morlane-Hondère, Grouin, & Zweigenbaum, 2016)	Meamedica.fr	_	12440 reviews	Yes	Yes	–
(C. Yang & Yang, 2015)	Medhelp	Most discussed	16344 threads	No	No	20
(Mishra, 2015)	Webmd	Oncology drugs	1933 reviews	Yes	No	–
(Chen et al., 2018)	5 online foruns	Methylphenidate	21 million posts	Yes(E)	Yes	1
(Xia, Wang, & Fan, 2017)	Webmd	Diabetes, hypertension	6400 reviews	Yes	Yes	–
(Dai & Wang, 2019)	Twitter	–	32670 tweets	Yes	Yes	–
(Chen, Deldossi, Aboukhamis, Faviez, & Dahamna, 2017)	Health french foruns	Most discussed, most used	325425 messages	Yes(E)	No	12
(J. Liu & Wang, 2018)	Heart disease foruns	Heart disease	261464 posts	Yes(E)	Yes	–
(Fan, Fan, Smith, & Skip, 2020)	Webmd, drugs.com	–	421348 reviews (10000 labeled)	Yes(E)	Yes	–
(Wu, 2019)	Twitter	–	25000 tweets	No	No	–
(Yadav, Ekbal, Saha, & Bhattacharyya, 2019)	Twitter, CADEC, and Medline	–	575 TW; 1248 CADEC; 20967 medline	Yes	Yes	–
(J. Liu, Wang, & Chen, 2019a)	Medhelp	–	2200557	Yes	Yes	–
(Gupta, B, Varma, & Pawar, 2018)	Twitter	–	15131 labeled; 461522 unlabeled	Yes	Yes	–
(Li, Yang, Luo, Xiang, & Lin, 2020)	CADEC, twitter, TwiMed[4]	–	16000	Yes	Yes	–
(Zhang, Cui, & Gao, 2020)	Twitter and dailystrenght	Chronic conditions on DS	5076 TW; 3705 DS	Yes(E)	Yes	8

Table 1 - Summary of critical aspects of the studies: publication year, data sources, subdomains of focus (when applicable), number of drugs on the studies, data size, annotations and availability of the data.

Table 1 suggests that Twitter is the most popular data source, used by 18 studies, followed by Medhelp and Dailystrength used on 6 and 4 publications, respectively.

Regarding the type of platforms used, eight studies use a combination of health focused social media and general use platforms. Ten studies draw their data from general-use social media (mainly twitter) and 15 studies use health focused platforms for ADR mining tasks.

[4] Curated from twiter and pubmed

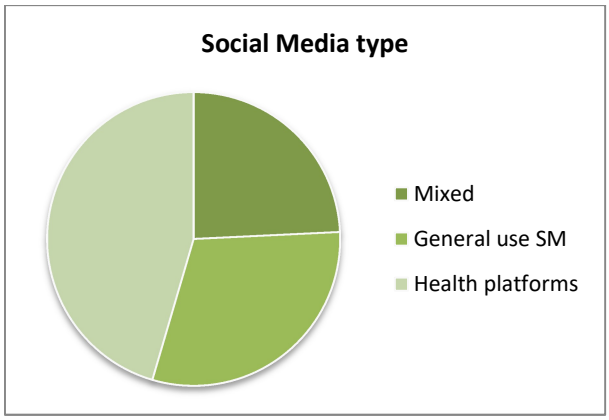


Figure 4 - Most used social media platform types

There is a predominance of machine learning approaches that build on lexicon features to boost detection accuracy.

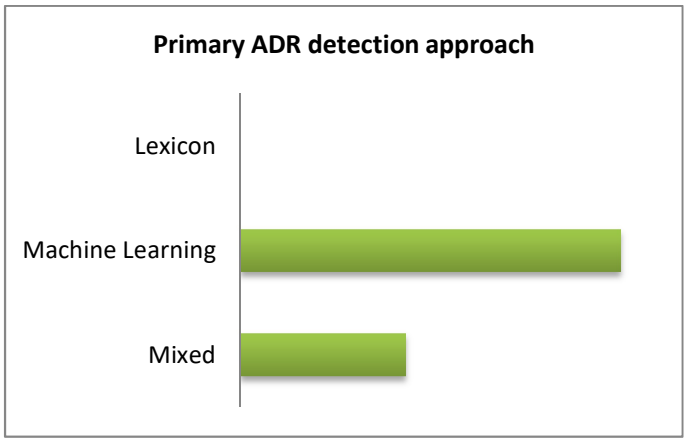


Figure 5 - Primary ADR detection approach used

The table also suggests that only a minority of the studies focused on a small number of drugs for ADR investigation. On most cases, researchers go beyond investigating ADRs associated with a small set of drugs. Furthermore, while some studies focused on specific domains of drugs (e.g., heart disease, diabetes, etc.), most studies, particularly recent ones, tend to concentrate on a range of drugs not specific to a domain.

In terms of data size, the studies can be divided into two important categories—large data sets without expert annotations, and relatively small data sets which contain expert annotations. Expert annotation has been replaced in many cases by less expensive and time-consuming options (as crowdsourcing annotation (Mishra, 2015)). Of the 33 publications included in this review, 14 utilized expert annotated data (Figure 6).

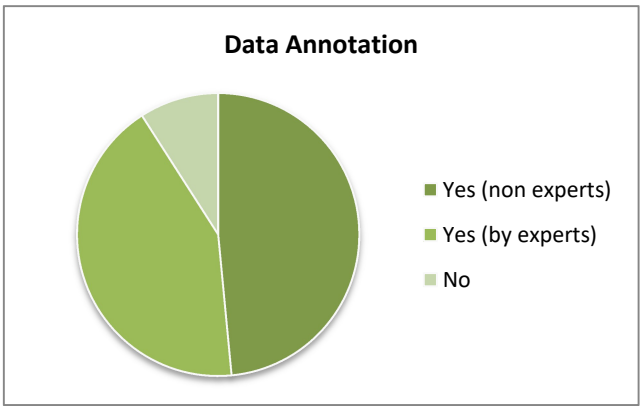


Figure 6 - Frequence of use of data annotation (supervised approaches) done by experts or other methods

Almost all studies exploit annotated data (on supervised approaches) and evaluate the results against gold standards prepared by experts. The availability of annotated data is a concern, as well as the time and cost associated with this approach. However, most data sets have been made publicly available. The results indicate an increasing trend towards diminishing the need for using sets of annotated data for ADR detection. Some of the more recent studies utilize adversarial training as a way to leverage very large volumes of unlabeled data present on social media.

Study	Research aim/concept	Primary detection/ extraction approach	Evaluation type	Main approach
(Nikfarjam et al., 2015)	Design a ML-based approach to extract mentions of ADRs, combining training data from different corpora to boost classification accuracy	*Extraction* Machine learning-based concept extraction system that uses CRF[5]	Quantitative: precision, recall, F-score; compared to expert annotated gold standard	ML/ lexicon
(Sarker & Gonzalez, 2015)	Explore NLP generated features to use in optimized ML algorithms for automatic classification of ADR presence in text segments; investigate combination of training data from distinct corpora to improve classification accuracy	*Detection* Supervised classification approach using NLP based features	Quantitative: F-score, accuracy; compared o benchmark corpus	ML/ lexicon
(M. Yang et al., 2015)	Automated filtering of ADR-containing posts using text classification methods	*Detection* SVM[6] based classifier	Quantitative: F-score; compared to 4 benchmark methods	ML
(Korkontzelos et al., 2016)	Investigate the effect of sentiment analysis features in locating ADR mentions	*Detection* Machine learning-based concept extraction system that uses conditional random fields and sentiment analysis features	Quantitative: precision, recall, F-score; compared to ADRmine	ML/ lexicon
(Cocos et al., 2017)	Develop a scalable, deep-learning approach to improve ADR detection performance	*Detection* Recurrent neural networks based classifiers	Quantitative: precision, recall, F-score; compared to baseline lexicon system and state-of-	ML

[5] Conditional Random Fields
[6] Support Vector Machine

			the-art CRF model	
(Lee et al., 2017)	Build several semi-supervised CNN[7] models for ADR classification in tweets, leveraging different types of unlabeled data	*Detection* Semi-supervised CNN framework	Quantitative: precision, recall, F-score; compared to state-of-the-art supervised classification model	ML
(X. Liu & Chen, 2015)	Develop an integrated information extraction framework for patient reports of ADRs in health social media	*Detection/ Extraction* Shortest dependency path kernel based ML with semantic filtering and report source classification	Quantitative: precision, recall and F-score; compared to baseline co-occurrence method	ML
(Tutubalina & Nikolenko, 2017)	Develop a joint model that combines CRF and RNN[8] to model the sequence of labels for ADR extraction	*Extraction* Supervised model combining RNN and CRF	Quantitative: precision, recall, F-score; compared to baseline methods	ML
(Martínez et al., 2016)	Uses NER[9] and Relation Extraction tasks (drugs, diseases, symptoms and adverse effects) for ADR extraction	*Extraction* Combination of ML processes with NLP supported by lexicons	Quantitative: precision and recall; compared to expert annotated Spanish ADR gold standard	ML/ lexicon
(J. Liu et al., 2016)	Develop a relation extraction system that uses NLP techniques to distinguish between ADR and non-ADR	*Extraction* Information-gain-based feature selection - contribution of each feature; kernel/ ensemble based combination approaches	Quantitative: accuracy, F-score and AUC[10]; compared to co-occurrence and a hybrid method	ML/ lexicon
(Segura-Bedmar et al., 2015)	System for detecting drug effects using relation extraction tasks	*Extraction* Distant supervised - shallow linguistic kernel method	Quantitative: recall, F-score, precision; compared to gold-standard corpus of Spanish SM	ML
(Comfort et al., 2018)	Develop rule-based and ML models for classifying ADR presence from SM and compare their performance to pharmacovigilance experts	*Detection* Rule based and ML based ADR annotator - SVM algorithm	Qualitative/quantitative: Agreement with pharmacovigilance experts by calculating Gwet AC1 statistic (gkappa) + time test	ML
(Eshleman & Singh, 2016)	Construction of two Drug-effect graphs , *degsider* and *degtwitter*, that are used in conjunction to identify ADRs in Twitter-streams	*Detection/ Relation extraction* Drug-effect graph integration model using supervised classifier (random forests) to classify relations into adverse or non-adverse	Quantitative: precision, recall, F-score	ML/ lexicon
(Gupta, Pawar, et al., 2018)	Tackle the problem of labeled data scarcity for ADR mention extraction from social media	*Extraction* Semi-supervised based RNN model, which can	Quantitative: precision, recall, F-score; compared to	ML

[7] Convolutional Neural Network
[8] Recurrent Neural Network
[9] Named Entities Recognition
[10] Area Under the Curve

		leverage unlabeled data on social media	baseline RNN model	
(J. Liu et al., 2018)	Develop a framework for ADR relation extraction using patient-generated content in social media with enhanced performance	*Relation Extraction* Semi-supervised ensemble learning framework, *SSEL-ADE*	Quantitative: AUC; compared to baseline methods	ML/ lexicon
(Katragadda et al., 2015)	Identify ADRs from tweets by modeling it as a link classification problem in graphs to identify negative edges i.e. adverse drug effects.	*Extraction/ Relation Extraction* From twitter drug mentioning graph - decision tree classification	Quantitative: precision, recall, F-score, accuracy; compared to baseline	ML
(Alimova & Tutubalina, 2018)	Develop two ML models with entity-level and context-level sets of features	*Detection* ML models based on Linear SVM and Logistic Regression	Quantitative: precision, recall, F-score; compared to SVM and CNN baseline models	ML
(Plachouras et al., 2016)	A large-scale, near real-time system for computational pharmacovigilance	*Detection* Set of cascaded filters, followed by a supervised ML classifier (SVM)	Quantitative: F-score, precision, recall; compared to baseline model	ML
(Morlane-Hondère et al., 2016)	Identification of medical conditions, based on the concept of post-coordination: first are extracted minimal medical-related entities then these are combined to identify complex ones	*Detection* Two classifiers, the first being based on CRF and the second on SVM	Quantitative: precision, recall, F-score; compared to baseline	ML
(C. Yang & Yang, 2015)	Compare tensor based and matrix based classification for early ADR classification	*Detection* Compare with labeling changes for ADRs and FDA official signaling times	Quantitative/qualitative: confidence, leverage, lift; time compared to official alert	ML
(Mishra, Malviya, & Aggarwal, 2015)	Automated system which takes a drug name as input and outputs information about the drug, from label sections, rating and common issues suffered by the users but not on label	*Extraction* SVM for drug rating and sentiment analysis	Qualitative/quantitative: frequency comparison, agreement with annotators; accuracy	ML
(Chen et al., 2018)	Provide text mining and visualization tools to explore a corpus of posts extracted from social media	*Detection/ Relation extraction* NER and RE	Quantitative: F-score; qualitative: frequency compared to SMPC and Vigibase reports	ML
(Xia et al., 2017)	Research framework using advanced NLP and deep learning for high-performance ADR extraction that can be generally applied to different corpuses	*Extraction* Word embeddings and a deep learning based NER (bi-directional LSTM[11] model)	Quantitative: precision, recall, F-score; compared to baseline lexicon method	ML/ lexicon
(Dai & Wang, 2019)	Propose a novel word embedding-based synthetic minority over-sampling technique	*Detection* Binary classification algorithm (SVM) with imbalanced techniques	Quantitative: F-score, precision, recall; compared to state-of-the-art methods	ML
(Chen et al., 2017)	Recognition of drug names, symptoms, and ADRs in social media texts	*Detection/ Relation extraction* A rule-based approach and a dictionary based	Quantitative: precision, recall, F-score;	Ml/ lexicon

[11] Long short term memory

		approach		
(J. Liu & Wang, 2018)	Develop a strategy to distinguish ADRs from other semantic types, and determine the drug that an ADE is associated with	*Extraction/ Relation extraction* An improved random subspace method (SSRS[12])	Quantitative: accuracy, AUC; vs baseline	Ml
(Fan et al., 2020)	Detect the presence of ADRs, and for each word classify into three classes: unimportant, drug name, or drug side effect	*Detection/ Extraction* Transformer-based framework for ADR detection and extraction (deep learning LSTM + word and sentence embedding)	Quantitative: AUC, F-score; vs lexicon and ML baselines	Ml/ lexicon
(Wu, 2019)	A neural approach with hierarchical tweet representation and multihead self-attention mechanism to jointly detect tweets mentioning drug names and adverse drug reactions	*Detection* LSTM neural networks, attention mechanisms	Quantitative: precision, recall, F-score; compared to several baseline methods	ML
(Yadav et al., 2019)	Neural network inspired multitask learning framework that can simultaneously extract ADRs from various sources	*Extraction* Adversarial training based multi-task framework	Quantitative: Precision, Recall and F-Score; compared to baselines	ML
(J. Liu, Wang, & Chen, 2019b)	Reduce the reliance on a large number of labeled instances for ADR identification; ADR relation identification. The component takes the sentence that contains drug and event mentions as input, and outputs the class label (ADE or non-ADE)	*Detection/ Relation extraction* A feature weighted-improved disagreement-based semi-supervised learning method, Random subspace, ensemble algorithm	Quantitative: average accuracy and AUC	ML
(Gupta, B, et al., 2018)	Propose a weakly-supervised and a semi-supervised learning based method to generate auxiliary task dataset (detection) and model it in a joint multi-task learning framework	*Detection and Extraction* Multi-task learning based methods - bi-directional; LSTM transducer based	Quantitative: Precision, Recall and F-Score	ML
(Li et al., 2020)	Exploiting a source corpus to improve the performance on small target corpora which only contain hundreds of training instances. Adversarial learning is applied to prevent corpus-specific features from being introduced into shared space	*Detection* Adversarial training based multi-task framework; shared private feature extractor	Quantitative: F-Score	ML
(Zhang et al., 2020)	Investigate whether the details of each ADR word and its corresponding predicates could improve the classification performance. Extract deep and shallow linguistic features for ADR detection and investigate the influence of different feature sets	*Detection* Extraction of linguistic features to train the SVM based predictive model	Quantitative: AUC	ML/ lexicon

Table 2 - Summary of the ADR detection/extraction approaches proposed by the studies, their primary research aims, and how the evaluations were designed

[12] Stratified Sampling based Random Subspace

Table 2 illustrates the two most frequently addressed problems on all publications: detection of comments/ posts discussing ADRs (1), and the extraction of specific ADRs from sentences (2). These two problems are the most relevant for pharmacovigilance initiatives using Social Media. Besides these, one study aims to Map Adverse Drug Reaction Mentions in Tweets to standard Medical Terminology (Emadzadeh et al., 2017). The evaluation approaches used are mostly standard measures such as Recall, Precision and F-score. Later in this section, we discuss some of the evaluation methods mentioned in Table 2.

4.3. A SUMMARY OF METHODOLOGIES AND RESOURCES

4.3.1. Lexicons and knowledge bases

Lexicons are resources that contain lists of ADR mentions collected from diverse sources such as drug labels, clinical trials, or user posts on social media. Significant efforts have been made towards the creation of new knowledge sources adapted for social media, given its own linguistic features. This review of literature revealed that ADR lexicons and knowledge bases have been used as standalone approaches or, more commonly used as added features to ML based approaches. Previous works that attempted to summarize ADR identification and extraction methodologies showed a prevalence in lexicon based efforts (Sarker et al., 2015).

From the set of works included, it becomes clear that recent works diverge from lexicon based methods towards increasingly automatic approaches, using lexicons as one of many features that have proven to improve classification accuracy. With this strategy, the goal is to incorporate domain-specific knowledge into the classification process by performing ADR lexicon matches, using it as a binary feature for presence/absence of ADR mentions, or as a numeric feature computed by frequency of ADR mentions in a text segment. Lexicons remain however, a widely used resource for pharmacovigilance strategies using social media, as a way to explore NLP approaches for generating useful features from text and use them in optimized machine learning algorithms (Sarker & Gonzalez, 2015). From the studies that utilized lexicons, a list of resources containing ADR mentions used is presented below:

Resource	**Description (*adapted from Sarker et al., 2015*)**	**Studies**
COSTART (Coding Symbols for a Thesaurus of Adverse Reaction Terms)	Used for coding, filing, and retrieving post-marketing ADRs.	Alimova & Tutubalina, 2018 Dai & Wang, 2019 Korkontzelos et al., 2016 Nikfarjam et al., 2015 Sarker & Gonzalez, 2015
SIDER(Side Effect Resource)	This is a knowledge base that uses MedDRA and contains ADR information on marketed medicines from public documents and package inserts.	Dai & Wang, 2019 Korkontzelos et al., 2016 J. Liu & Wang, 2018 Nikfarjam et al., 2015 Sarker & Gonzalez, 2015 Zhang et al., 2020
CHV (Consumer Health Vocabulary)	This database was created to map words and phrases used by lay persons to describe ADRs to technical terms used by health professionals.	Dai & Wang, 2019 Fan et al., 2020 Korkontzelos et al., 2016 J. Liu & Wang, 2018

		J. Liu et al., 2018, 2016 X. Liu & Chen, 2015 Plachouras et al., 2016 Sarker et al., 2015 C. Yang & Yang, 2015 Zhang et al., 2020
Canada Drug Adverse Reaction Database (MedEffect)	Contains associations between drugs and adverse reactions.	Sarker & Gonzalez, 2015
FDA AERS	The FDA adverse event reporting system and database designed to support FDA's post-marketing surveillance.	J. Liu et al., 2018, 2016 J. Liu & Wang, 2018 X. Liu & Chen, 2015 Plachouras et al., 2016
UMLS (Unified Medical Language System)	Contains a large collection of biomedical vocabulary. It categorizes medical terms into broad and fine grained categories.	Fan et al., 2020 Martínez et al., 2016 Sarker & Gonzalez, 2015 J. Liu & Wang, 2018 J. Liu et al., 2018, 2016 X. Liu & Chen, 2015 Plachouras et al., 2016 Sarker et al., 2015 Emadzadeh et al., 2017 Zhang et al., 2020
SNOMED CT	A terminology accessible in Spanish containing concepts, descriptions and relationships to represent medical knowledge.	J. Liu & Wang, 2018 Martínez et al., 2016
MedDRA (Medical Dictionary for Regulatory Activities)	A specific, multilingual, standardized medical terminology to enable sharing of regulatory information internationally for medical products.	Chen et al., 2017, 2018 J. Liu & Wang, 2018 Martínez et al., 2016 Zhang et al., 2020

Table 3 - Description of lexicons used in each of the research publications (adapted from Sarker et al., 2015)

Alongside these, drug name lexicons and medical condition lexicons are used in many of the articles. As examples, MeSH[13] and NLM[14] are used for drug names and symptoms (Katragadda et al., 2015), Rxnorm (Dai & Wang, 2019) is also used for drug names, CIMA contains official information about authorized drugs in Spain maintained by the Spanish Agency of Medicines and Health Products, and SpanishDrugEffectDB, a spanish database that contains relations between drugs and effects (Martínez et al., 2016).

4.3.2. Automatic classification of ADR containing user posts

Some of the studies reviewed focused on the automatic classification of user posts to determine whether ADRs are mentioned in the posts or not. This approach is based on the knowledge that most posts on social media are not associated with ADR mentions and thus, filtering out irrelevant posts is critical. ADR data on social media is often characterized as imbalanced - only a very small amount of posts contain ADRs (Sarker et al., 2015). However, on health-related social media websites, the proportion of ADR data is higher (Morlane-Hondère et al., 2016). Several approaches have been pursued and optimized for supervised learning tasks in order to overcome data imbalance in social

[13] Medical Subject Headings
[14] National Library of Medicine

media text (Dai & Wang, 2019; J. Liu & Wang, 2018; Morlane-Hondère et al., 2016; Sarker & Gonzalez, 2015).

Supervised classification approaches were the most commonly used for this classification task, but these require large numbers of manually annotated posts to make reliable evaluations. The lack of large, available annotated data sets is a problem for implementing this approach. A couple of studies (Gupta, Pawar, et al., 2018; Lee et al., 2017) leverage different types of unlabeled data for semi-supervised classification with common machine learning algorithms, such as Convolutional Neural Networks and Recurrent Neural Networks, to tackle the problem of labeled data scarcity.

To overcome data imbalance, some authors employed a number of strategies including the use of weighted classifiers, incorporation of features from other text classification problems, and the combination of multiple corpora for training to augment the minority class. These strategies allow increasing the diversity of classifiers, leading to better recall and overall F-scores for the minority classes. On a recent study, the authors used minority over-sampling and word embeddings to synthesize new training examples from the sentence representation to compensate for large imbalanced datasets in Twitter (Dai & Wang, 2019).

To tackle both data imbalance and lack of annotated datasets, some authors used multi-corpus training strategies, which integrate information from compatible corpora to improve significantly classification performance and reduce the time and costs associated with the annotation of data (Sarker & Gonzalez, 2015). At the same time, as a mean to improve the performance on small target corpora that contain only a few training instances, adversarial training was used (Gupta, B, et al., 2018; Li et al., 2020; J. Liu et al., 2019a; Yadav et al., 2019) to prevent corpus-specific features from being introduced into shared space so that corpora from different sources can be leveraged with minimum extra noises (Li et al., 2020).

The extraction of features from the datasets is a recent and promising strategy since it showed that a careful selection of features can significantly improve classification accuracy. The use of NLP approaches help generating useful features from text and it can include polarity, sentiment, topics, lexicons and many other features (Sarker et al., 2015).

It is worth mentioning a strategy used by some studies, which leveraged sentiment analysis as a mean to identify drug experience posts with negative sentiment association, for identification of ADR containing posts (Korkontzelos et al., 2016; Mishra et al., 2015).

4.3.3. ADR mention extraction

A majority of the studies that met the inclusion criteria focused on identifying specific ADR mentions from user posts and extracting them. About half of the approaches are lexicon-based, meaning that they identify and retrieve ADRs using a list of precompiled ADR mentions (described on Table 3). Applying lexicon-based matching techniques can successfully identify a subset of the ADR mentions in user posts. However, besides only allowing to identify and extract known ADRs (those present in the lexicons used), there are other challenges to using a pure lexicon-based approach to the task (discussed in detail over section 5).

Extracting specific drug-ADR pairs generally follows the preliminary stage of classifying ADR containing posts. In order to pursue mention extraction both the drug mention and related ADR

mention are required. Some studies have focused on differentiating the relationships between ADR and drug pairs (Chen et al., 2018; Eshleman & Singh, 2016; Gupta, Pawar, et al., 2018; Katragadda et al., 2015; C. Yang & Yang, 2015). A class of techniques - association rule mining - is used to identify if a specific drug and ADR are associated. Earlier studies often used co-occurrence techniques (drug and ADR mentions in close proximity indicate a drug-ADR relation), which had low precision. Besides co-occurrence, some studies use rule based (pattern based) methods and statistical learning based approaches. Rule based semantic filtering methods however, suffered from low recall. In recent years, statistical machine learning-based approaches have gained popularity for its increased performance (Nikfarjam et al., 2015).

Eshleman & Singh (2016) and Katragadda et al. (2015) use graph models to perform this task. The first use a drug-effect graph integration model and a supervised classifier (random forests) to classify relations obtained into adverse or non-adverse, and the latter utilizes a twitter drug mentioning graph and a decision tree based model for classification of edges as adverse or non-adverse.

On the study by Chen et al. (2018) signal detection and ADR extraction are based on statistical measures of association describing reporting disproportionality. If the statistical measure crosses certain threshold, which is summarized as a decision rule, the signal is declared positive for a drug X associated with symptom Y.

While most approaches for ADR mention extraction relied on supervised learning methods, which have to deal with the issue of labeled data scarcity, Gupta et al. (2018) adressed this issue with a two phase model based on Recurrent Neural Network algorithm, to leverage unlabeled data present in abundance on social media, to first, predict the drug name given its context in the tweet, and afterwards identify drug-ADR relationships.

Using relation extraction techniques and semantic filter, some works have attempted to distinguish mentions of drug beneficial effects and adverse effects, given that the same term can often be used to describe or an indication or beneficial effect experienced by the user (Fan et al., 2020; J. Liu & Wang, 2018; Martínez et al., 2016).

4.4. A SUMMARY OF EVALUATION TECHNIQUES AND METRICS

The research works designed their own evaluation approaches or used existing ones that are compatible with the proposed methodologies.

The evaluation approaches applied are presented in this section, and they can be divided in two broad categories: qualitative and quantitative.

4.4.1. Qualitative evaluation

The end goal of ADR detection from social media sources is to be able to identify drugs that are either frequently related to ADRs (to estimate frequency) or those that are associated with previously unknown ADRs. Some works merge qualitative and quantitative analysis (Comfort et al., 2018; Chen et al., 2018; Yang & Yang, 2015; Mishra et al., 2015). Chen et al. (2018) compared the frequency of ADRs reported for Methylphenidate on social media with the number of reports for the

same ADRs on SMPC and Vigibase. Mishra et al. (2015) also compared the frequency for each ADR detected versus the frequency notified via traditional reporting systems and evaluated agreement with annotators. On an investigation led by Yang et al. (2015) the timing of social media ADR reporting was compared with that of labeling changes and FDA official signaling times (for a list of specified and known ADRs) to explore the potential of SM reporting for early ADR detection. On the study led by Comfort et al. (2018), the authors used several machine learning models for identification of ADR containing posts and compared their performance and swiftness with that of human pharmacovigilance experts.

4.4.2. Quantitative evaluation

All works used quantitative evaluations, whether by itself or combined with qualitative metrics. Regarding the quantitative metrics used, most tasks used common metrics such as Recall, Precision, F-score and Accuracy. Recall or Sensitivity is the proportion of Real Positive cases that are correctly Predicted Positive. Conversely, Precision or Confidence denotes the proportion of Predicted Positive cases that are correctly Real Positives (Powers, 2011). F-measure is routinely used as a performance metric for different types of prediction problems, including binary classification, and named entity recognition. It enforces a better balance between performance on the minority and the majority class, respectively, and, therefore, it is suitable in the case of imbalanced data. F-measure can be generalized to a weighted harmonic average of recall and precision values, or for simplicity as an unweighted mean, which is often referred to as the F1-score, F-Score or F1-measure (Waegeman, Dembczynski, Cheng, & Hullermeier, 2011). The F-score is a popular measure used in machine learning which trades off precision and recall (to obtain a high F-score, both precision and recall need to be high) (M. Yang et al., 2015). Other metrics for quantitative evaluations have also been used, however less frequently. They include: lift, leverage, confidence (C. Yang & Yang, 2015) proportional reporting ratio (PRR) (Chen et al., 2018; Mishra et al., 2015; Plachouras et al., 2016), Area Under Curve (AUC) (J. Liu & Wang, 2018; J. Liu et al., 2019a, 2018, 2016) and gkappa (Comfort et al., 2018).

Regarding the baseline datasets used for comparisons, over half the studies included used manually annotated data (by experts or not) for the evaluation of the drug-ADR extraction task. Besides the use of annotated data, some authors calculated the performance metrics using various gold standards, such as known adverse reactions from FDA product labels, official regulatory databases and annotated corpora made available by previous studies.

5. DISCUSSION

For pharmacovigilance approaches aiming to leverage big data, several aspects must be considered beforehand. These include the main purpose of the analysis (automatic post filtering for presence of ADR or extracting mentions of ADRs), the approach chosen (lexicon based or machine learning), and the data source type (health forums, twitter-like platforms, or both). In order to make the best use of this data, all these aspects must be evaluated and related to the objective, the specific drug or conditions to study. A roadmap of SM mining for pharmacovigilance is required before starting any approach, in order to assure the best possible performance.

For simplification purposes, this discussion will be divided over critical aspects of SM mining for pharmacovigilance: the main objective of the data mining (ADR detection or extraction), the type of platform used (general use or heath focused), the main approach (lexicon or ML based), and final considerations on best uses of this data.

Automatic post detection and mention extraction tasks

When it comes to SM data harnessing for ADR detection, the two most frequently addressed problems are the detection of comments/ posts discussing ADRs (1) (automatic classification of ADR containing user posts), and the extraction of specific ADR-drug pairs from sentences (2) (ADR mention extraction).

Filtering the posts that contain ADRs is often a preliminary stage of the ADR identification pipeline, following that is the extraction of specific drug-ADR pairs, using relation extraction techniques.

For the detection of posts containing ADRs it is relevant to mention the issue of data imbalance on social media. Several strategies have been applied to tackle this issue and produce better classification results: weighted classifiers, incorporation of features from other text classification tasks, multi-corpus training strategies (combination of multiple corpora for training to augment the minority class and reduce the time and costs associated with the annotation of data) and minority over-sampling (to synthesize new training examples from the sentence representation) (Dai & Wang, 2019; J. Liu & Wang, 2018; Morlane-Hondère et al., 2016; Sarker & Gonzalez, 2015).

While automatic post classification tasks only identify if the posts or comments contain any ADR, mention extraction means we are able to retrieve specific ADRs from posts as well as the drug that is associated with it. In addition to ADR mention extraction several studies also conduct ADR-drug relation extraction to filter out other semantic relations that may exist between drugs and event mentions, like indications or beneficial effects. Association rule mining is used to identify the drug associated with a specific ADR, frequently using machine learning methods (Nikfarjam et al., 2015).

General use platforms vs health forums

In a literature review, Sarker *et al.*(2015) found out that while health-related sources tend to contain higher proportions of relevant ADR data, the volume of data from general social media websites is significantly higher. The issue of imbalanced data is intensified in general social media platforms where health data is scarce when compared with the overall available data (European Medicines Agency, 2017; Ghosh & Lewis, 2015; Sarker & Gonzalez, 2017). Nevertheless, the strategies

employed to counteract it allowed overcoming this issue successfully. However, the effort put into these strategies, must be evaluated with respect to other options available and the value expected to be obtained from general SM data. The use of health forums is an alternative to avoid health data scarcity and might be considered in some cases; using health forums to harness ADR data reduces the work and costs associated with developing data imbalance countering techniques.

From the set of studies evaluated, 7 used a combination of health focused social media and general use platforms; 11 used general social media (most of which twitter); and 15 studies used health focused platforms for ADR mining tasks. Besides the diversity of criteria and approaches in each of the studies, there is no evidence from the data included that one social media platform type may offer better results than others. The studies included showed prevalence in the use of health focused platforms for SM mining, otherwise, a mixed approach was the preferred course of action, taking advantage of the strengths of both types of SM for more comprehensive results.

Machine learning and lexicon based approaches

About half of the studies used lexicon-based approaches, meaning that they identified ADRs using a list of precompiled ADR mentions to identify a subset of the ADR mentions in user posts. This approach only allows identifying and extracting previously known ADRs, present in the resource lexicons used, leaving out potential unknown ADRs. Lexicon based approaches, despite being commonly used, present several challenges when used as a standalone method: consumers use descriptive symptom explanations and idiomatic expressions instead of technical terms found in medical lexicons; even when correctly identified, matched terms are not necessarily adverse effects: the terms used to describe ADRs can also be used for indications or beneficial effects (Sarker et al., 2015). Not acknowledging this will produce an over signalization of potential ADRs and have an impact in the costs associated with the analysis. In the set of studies included, ADR lexicons and knowledge bases have more often been leveraged as added features to ML based approaches.

Regarding ML based approaches, supervised learning was the most popular approach used and has shown promising performances. However, preparation and annotation of data is an expensive process, and especially when there is a need to increase the number of instances for the minority class, large volumes of data require annotation. Nearly all studies included exploit annotated data, whether it was done by experts or other means. There is an apparent trend towards diminishing the need for large annotated datasets for ADR detection. Some of the most recent studies utilize a combination of multiple corpora and adversarial training as a way to leverage labeled data from other sources for supervised classification, or even use a set of techniques aiming to take advantage of the very large volumes of unlabeled data present on social media, as a mean to tackle the problem of labeled data scarcity (Sarker & Gonzalez, 2015; Gupta, B, et al., 2018; Lee et al., 2017; Li et al., 2020; J. Liu et al., 2019; Yadav et al., 2019).

A careful selection of features has shown to significantly improve automatic classification accuracy in ML based approaches. The use of NLP approaches to aid generating useful features from text was used for automatic classification of ADRs. As examples of features used we have polarity, sentiment, topics and lexicon features (Sarker et al., 2015).

The review of the evaluation approaches used shows the absence of a common evaluation methods or common datasets. Regarding evaluation measures, quantitative ones are far more commonly used

than qualitative, particularly F-score, Precision and Recall. The evaluations were mainly made against manually annotated datasets, and some calculated these metrics versus various gold standards, such as known adverse reactions on product labels, official regulatory databases and annotated corpora made available by previous studies.

Best uses

It is important to define the best uses of this data to better take advantage of each approach available and maximize performance while detecting ADR from SM. This could mean directing the efforts towards specific medications, guide the choice of investigational approach or data sources to use on each study case.

For the studies evaluated, most researchers went beyond investigating ADRs associated with a small set of drugs, not setting any drug related restraints. While a minority of studies focused on specific domains of drugs (e.g., medication for heart disease, diabetes, etc.), most studies tend to concentrate on a range of drugs not specific to any domain. In this research project, one of the aim was to assess if some subgroups of ADR data present on social media – referring to any particular medications or diseases – that could be particularly valuable for pharmacovigilance purposes. There was however not enough data on each subset to compare results and reach conclusions regarding best uses. The use of common datasets in the future, as well as standardized evaluation methods could allow comparisons between approaches.

Nevertheless, for any population group or specific medication undergoing safety evaluations, it is important to balance the data that is available and the efforts to acquire and treat it with the potential benefits. There must be a trade-off between the amount of manual screening required by the lower level of SM processing and the potential missing of ADRs by the higher level of SM processing.

Implications for health systems, patients, companies and regulators

Wherever future investigations may lead, current research shows that combining data from multiple sources will provide a broader safety profile for any medication. Reports systematically collected from social media may be able to contribute information on the most pertinent adverse effects to patients (Golder, Scantlebury, & Christmas, 2019), thus creating an innovative patient-centered way of reporting. With this, regulatory decisions (on new drug launches, market renewals and withdrawals) could be based on more complete safety information, one that is gathered from multiple sources, and that is near real time. Limiting the occurrence of serious ADR is one major driver of these investigations, since near real time information would allow for swift withdrawals in case serious ADRs are detected. Throughout the whole healthcare chain, from regulatory appraisals to prescription and intake habits, this untapped information can add value to what is the conventional use of safety data.

Despite the investment necessary to put in place operations to acquire and treat massive amounts of data, for companies and regulatory agencies, being able to benefit from additional and complete safety data that can effectively prevent or diminish the impact of safety issues is a major leap forward. One example of this is the occurrence of major safety issues that could impact marketing authorizations on newly launched medicines; the fact that detection could be made on earlier stages

of development/ launch process could reduce the economic impact for companies. In addition to clear public health benefits of a more efficient system for evaluation of medicines, pharmaceutical companies can use safety data as a competitive advantage (such as value based pricing and reimbursements).

For health systems, establishing the safety profile of medications and having the capability to acquire and interpret near real time safety signals could mean a step forward in preventing ADR caused hospitalizations and related healthcare costs.

6. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The use of social media data for pharmacovigilance has been receiving growing attention over the last few years. Despite that, no guidance or common evaluation framework has yet been defined for ADR detection. Although in this project the author has evaluated strengths and limitations of each approach, at this point, performing a direct comparison of existing approaches is not feasible. Investigational methodologies have progressed in various directions, using distinct datasets (sometimes internal datasets not made available) and processing methods, as well as a variety of complex algorithms, without the development of standard evaluation criteria to allow performance comparisons.

Social media data has some characteristics that make extraction more difficult when compared to other corpora, such as news or biomedical literature; user postings are made in informal language, not respecting grammatical rules, and usually include misspellings, abbreviations, and phrase construction irregularities (Nikfarjam et al., 2015). For extraction tasks, the drug associated with an ADR is not always available or can be ambiguous (it may be implied or the comment may mention multiple drugs) (J. Liu & Wang, 2018). Many symptom and side-effect terms are common in respective dictionaries creating ambiguity and possible misclassification (Rifat, Rashed, Noori, & Hasan, 2019). For these reasons usually surveillance by a domain expert is still necessary whether in preliminary or final stages of the process. Every approach discussed is presented with its own challenges. Lexicon based approaches face specific obstacles when applied to social media linguistic, whereas machine learning based methods require large amounts of data for system development. Supervised classification approaches require manually annotated data, and large numbers of annotated posts are required to make reliable evaluations. The lack of large, available annotated data sets has been a problem for implementing this approach. However the more recent focus on multi-corpus training approaches seem a valid alternative worth pursuing in the future, in order to diminish costs associated with the annotation of data and the need for expert surveillance.

In order to infer the best uses for social media ADR data, in the future, efforts must be directed towards specific groups, rather than universal application, to assess the detection performance of these approaches over different medications or diseases. Although it was out of the scope of this survey, it is crucial for any systematic surveillance system to apply the appropriate preprocessing and post processing modules. For example, the matching of colloquial ADR expressions with official medical terms is of great importance to overcome the previously mentioned language limitations to enter the classification model. Also, the use of sentiment analysis to identify first hand negative experiences of drug intake showed promising results for classification tasks. Up until now these approaches have only been studied as standalone approaches, in the future consolidated end-to-end approaches might give a more accurate perception of the results these methodologies may achieve. There is still a need to assess if SM data tends to include only those inclined to share personal experiences and with technical ability and means of access to social media, which would assume elderly exclusion (Price, 2016).

Finally, norms and agreed practices to guide these efforts are needed, considering the ethical issues like privacy, confidentiality, and follow-up, as well as issues relating to the quality of the information, and more practical ones related to statistical analysis (Audeh et al., 2020). Once these are agreed

upon, industry, regulators and patients can benefit from regulated access and analysis of these data to timely discovery of drug related issues.

7. CONCLUSION

ADRs are among the leading causes of hospitalizations and death in the world. Currently, post-market drug surveillance relies on voluntary reporting systems. With an increase in availability of health related data on social media, interest in medical data mining has been growing rapidly. Information of online discussions about health and medications provide a substantial resource for the identification of ADRs. The end goal of ADR mining from social media is to be able to identify drugs that are either frequently related to ADRs (to estimate frequency) or those that are associated with previously unknown ADRs.

The aim of this research project was to find out how can social media data add value to post-market safety evaluations. For that end a literature review was completed over 33 research works that were conducted over the last five years and used any kind of social media data that was retrieved and treated by automatic means for the purpose of ADR detection. In order to choose the best approach to efficiently take advantage of safety data present on SM, companies and regulatory agencies developing pharmacovigilance roadmaps need to evaluate the state-of-the-art of these initiatives. A secondary objective of this review was to assess if any specific combination of data source, detection approach and population of study would be associated with better performance. However, the lack of common datasets, algorithms and evaluation metrics over these investigations did not allow comparing between approaches.

A transition in research methodologies was evident in the results, towards an increase in the use of supervised ML based approaches, as large annotated data sets are increasingly becoming available. In fact, almost every study published on the topic over the last 5 years relied on machine learning approaches for identifying posts containing ADRs or extracting ADR mentions in text, with lexicon based features being often applied. It is also noteworthy the strong tendency to use combined approaches as they seem to provide better results, whether with combined data sources (general platforms and health-focused forums); mixing different sources of annotated data sets (adding biomedical corpus to increase algorithms strength and diminish the need for costly and time-consuming data annotation); and even the combination of machine learning approaches with lexicon based features. The breakthrough works from Sarker and Gonzalez (2015), followed by works by Korkontzelos et al. (2016), Alimova & Tutubalina, (2018), Li et al., (2020) and Zhang et al. (2020) have shown that added features to ML approaches is another way to improve classification accuracy. The most recent trends in ML based extraction – multitask learning and adversarial learning – use characteristics from several data sources to produce a tailored learning, while reducing the dependency to data annotation.

Social media data can integrate inclusive pharmacovigilance approaches that gather data from several sources in order to achieve broader safety profiles of medications. For its dimension alone, this source has immense potential for uncovering unknown ADRs and issues most pertinent to patients. However, it also comes with its challenges, which is why the decision to integrate this data source and the choice of method and approach to use must always be adapted to the purpose of investigation and resources available.

Future efforts must be directed towards end-to-end projects, studying specific medications or disease groups in order to get a better sense of the best uses for social media data on safety

evaluations. Data privacy, confidentiality, quality of the information, and practical issues related to the analysis are still to be established so that in the future, agreed norms between all parties involved are developed. The progress in information technology and the societal need to consider patients' experiences is undeniable and should motivate future research on social media surveillance for complementing conventional pharmacovigilance with patient centric and real time ways of reporting.

8. BIBLIOGRAPHY

Alimova, I., & Tutubalina, E. (2018). Automated Detection of Adverse Drug Reactions from Social Media Posts. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 3–15). https://doi.org/10.1007/978-3-319-73013-4

Audeh, B., Bellet, F., Noëlle, M., Agnès, B., Le, L., & Cédric, L. (2020). Use of Social Media for Pharmacovigilance Activities : Key Findings and Recommendations from the Vigi4Med Project. *Drug Safety*, *43*, 835–851. https://doi.org/10.1007/s40264-020-00951-2

Bate, A., Juniper, J., Lawton, A., & Thwaites, R. (2016). Designing and incorporating a real world data approach to international drug development and use: what the UK offers. *Drug Discovery Today*, *21*(3), 400–405. https://doi.org/10.1016/j.drudis.2015.12.002

Bate, A., Reynolds, R. F., & Caubel, P. (2018). The hope , hype and reality of Big Data for pharmacovigilance. *Therapeutic Advances in Drug Safety*, *9*(1), 5–11. https://doi.org/10.1177/2042098617736422

Bere, N. (2017). How are medicines evaluated at the EMA. Retrieved January 26, 2020, from https://www.ema.europa.eu/en/documents/presentation/presentation-how-are-medicines-evaluated-european-medicines-agency-part-1_en.pdf

Bousquet, C., Dahamna, B., Guillemin-lanne, S., & Darmoni, S. J. (2017). The Adverse Drug Reactions from Patient Reports in Social Media Project : Five Major Challenges to Overcome to Operationalize Analysis and Efficiently Support Pharmacovigilance Process. *JMIR Research Protocols*, *6*(9). https://doi.org/10.2196/resprot.6463

Catell, J., Chilukuri, S., & Levy, M. (2013). How can Big Data revolutionize pharmaceutical R&D. *McKinsey & Company*, (April). Retrieved from https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d

Chen, X., Deldossi, M., Aboukhamis, R., Faviez, C., & Dahamna, B. (2017). Mining Adverse Drug Reactions in Social Media with Named. *Studies in Health Technology and Informatics*, *245*, 322–326. https://doi.org/10.3233/978-1-61499-830-3-322

Chen, X., Faviez, C., Schuck, S., Agnès Lillo-Le-Louët, A., Texier, N., Dahamna, B., … and Burgun, A. (2018). Mining Patients Narratives in Social Media for Pharmacovigilance : Adverse Effects and Misuse of Methylphenidate. *Frontiers in Pharmacologie*, *9*(541), 1–17. https://doi.org/10.3389/fphar.2018.00541

Cocos, A., Fiks, A. G., & Masino, A. J. (2017). Deep learning for pharmacovigilance : recurrent neural network architectures for labeling adverse drug reactions in Twitter posts, *0*(0), 1–10. https://doi.org/10.1093/jamia/ocw180

Comfort, S., Perera, S., Hudson, Z., & Dorrell, D. (2018). Sorting Through the Safety Data Haystack : Using Machine Learning to Identify Individual Case Safety Reports in Social- Digital Media. *Drug Safety*, *41*(6), 579–590. https://doi.org/10.1007/s40264-018-0641-7

Dai, H., & Wang, C. (2019). Classifying adverse drug reactions from imbalanced twitter data. *International Journal of Medical Informatics*, *129*, 122–132. https://doi.org/10.1016/j.ijmedinf.2019.05.017

Emadzadeh, E., Sarker, A., Nikfarjam, A., & Gonzalez, G. (2017). Hybrid Semantic Analysis for

Mapping Adverse Drug Reaction Mentions in Tweets to Medical Terminology. In *AMIA Annual Symposium Proceedings* (pp. 679–688).

Eshleman, R., & Singh, R. (2016). Leveraging graph topology and semantic context for pharmacovigilance through twitter-streams. *BMC Bioinformatics*, *17*(Suppl 13). https://doi.org/10.1186/s12859-016-1220-5

European Commission. (2018). Proposal for a regulation amending, as regards pharmacovigilance of medicinal products for human use. Regulation (EC) No 726/2004. Impact assessment. Retrieved from http://ec.europa.eu/health/files/pharmacos/pharmpack_12_2008/pharmacovigilance-ia-vol1_en.pdf

European Medicines Agency. (2017). *HMA / EMA Joint Big Data Task Force*. Retrieved from https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf

Fan, B., Fan, W., Smith, C., & Skip, H. (2020). Adverse drug event detection and extraction from open data : A deep learning approach. *Information Processing and Management*, *57*(1), 102131. https://doi.org/10.1016/j.ipm.2019.102131

Fermont, I. (2019). Pharmacovigilance strategy : opportunities for cross-national learning. *Israel Journal of Health Policy Research*, *8*(1), 1–5. https://doi.org/10.1186/s13584-019-0319-3

Ghosh, R., & Lewis, D. (2015). Aims and approaches of Web-RADR : a consortium ensuring reliable ADR reporting via mobile devices and new insights from social media Aims and approaches of Web-RADR : a consortium ensuring reliable ADR reporting via mobile de. *Drug Safety*, *14*(12), 1845–1853. https://doi.org/10.1517/14740338.2015.1096342

Golder, S., Scantlebury, A., & Christmas, H. (2019). Understanding Public Attitudes Toward Researchers Using Social Media for Detecting and Monitoring Adverse Events Data : Multi Methods Study Corresponding Author : Background : *Journal of Medical Internet Res*, *21*(8), 1–13. https://doi.org/10.2196/jmir.7081

Gupta, S., B, M. G., Varma, V., & Pawar, S. (2018). Multi-task Learning for Extraction of Adverse Drug Reaction Mentions, *1*, 59–71. https://doi.org/10.1007/978-3-319-76941-7

Gupta, S., Pawar, S., Ramrakhiyani, N., Palshikar, G. K., & Varma, V. (2018). Semi-Supervised Recurrent Neural Network for Adverse Drug Reaction mention extraction. *BMC Bioinformatics*, *19*(Suppl 8), 1–7.

Härmark, L., Hubert, R., & Ralph, L. (2016). Patient-Reported Safety Information : A Renaissance of Pharmacovigilance ? *Drug Safety*, *39*, 883–890. https://doi.org/10.1007/s40264-016-0441-x

Harpaz, R., & Dumochel, W. (2016). Big Data and Adverse Drug Reaction Detection. *Clinical Pharmacology and Therapeutics*, *99*(3), 268–270. https://doi.org/10.1002/cpt.302

HMA, & EMA. (2017). HMA / EMA Joint Task Force on big data – Surveys results. In *Big Data Meeting* (pp. 1–41). Retrieved from https://www.ema.europa.eu/en/documents/other/hma/ema-joint-task-force-big-data-surveys-results_en.pdf

Jarow, J., LaVange, L., & Woodcock, J. (2017). Multidimensional Evidence Generation and FDA Regulatory Decision Making Defining and Using " Real-World " Data. *JAMA*.

Katragadda, S., Benton, R., Karnati, H., Pusala, M., & Raghavan, V. (2015). Detecting Adverse Drug Effects Using Link Classification on Twitter Data. In *International Conference on Bioinformatics*

and Biomedicine (pp. 2–6). IEEE.

Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., & Gonzalez, G. H. (2016). Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of Biomedical Informatics*, *62*, 148–158. https://doi.org/10.1016/j.jbi.2016.06.007

Leaman, R., Wojtulewicz, L., Sullivan, R., Skariah, A., Yang, J., & Gonzalez, G. (2010). Towards Internet-Age Pharmacovigilance : Extracting Adverse Drug Reactions from User Posts to Health-Related Social Networks. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing* (pp. 117–125).

Lee, K., Qadir, A., Hasan, S. A., Datla, V., Prakash, A., Liu, J., & Farri, O. (2017). Adverse Drug Event Detection in Tweets with Semi-Supervised Convolutional Neural Networks, 705–714.

Li, Z., Yang, Z., Luo, L., Xiang, Y., & Lin, H. (2020). Exploiting adversarial transfer learning for adverse drug reaction detection from texts. *Journal of Biomedical Informatics*, *106*(April), 103431. https://doi.org/10.1016/j.jbi.2020.103431

Liu, J., & Wang, G. (2018). Pharmacovigilance from social media : An improved random subspace method for identifying adverse drug events. *International Journal of Medical Informatics*, *117*, 33–43. https://doi.org/10.1016/j.ijmedinf.2018.06.008

Liu, J., Wang, G., & Chen, G. (2019a). Identifying Adverse Drug Events From Social Media Using An Improved Semisupervised Method. *IEEE Intelligent Systems*, *34*(2), 66–74. https://doi.org/10.1109/MIS.2019.2893158

Liu, J., Wang, G., & Chen, G. (2019b). Identifying Adverse Drug Events From Social Media Using An Improved Semisupervised Method, 66–74.

Liu, J., Zhao, S., & Wang, G. (2018). SSEL-ADE : A semi-supervised ensemble learning framework for extracting adverse drug events from social media. *Artificial Intelligence In Medicine*, *84*, 34–49. https://doi.org/10.1016/j.artmed.2017.10.003

Liu, J., Zhao, S., & Zhang, X. (2016). An ensemble method for extracting adverse drug events from social media. *Artificial Intelligence In Medicine*, *70*, 62–76. https://doi.org/10.1016/j.artmed.2016.05.004

Liu, X., & Chen, H. (2015). A research framework for pharmacovigilance in health social media : Identification and evaluation of patient adverse drug event reports. *Journal of Biomedical Informatics*, *58*, 268–279. https://doi.org/10.1016/j.jbi.2015.10.011

Martínez, P., Segura-bedmar, I., Revert, R., Luna, A., Martínez, J., & Moreno-schneider, J. (2016). Turning user generated health-related content into actionable knowledge through text analytics services. *Computers in Industry*, *78*, 43–56. https://doi.org/10.1016/j.compind.2015.10.006

Min, J., Osborne, V., Lynn, E., & Shakir, S. A. W. (2018). First Conference on Big Data for Pharmacovigilance. *Drug Safety*, *41*(12), 1281–1284. https://doi.org/10.1007/s40264-018-0727-2

Mishra, A. (2015). Towards Automatic Pharmacovigilance : Analysing Patient Reviews and Sentiment on Oncological Drugs, 1402–1409. https://doi.org/10.1109/ICDMW.2015.230

Mishra, A., Malviya, A., & Aggarwal, S. (2015). Towards Automatic Pharmacovigilance : Analysing Patient Reviews and Sentiment on Oncological Drugs. In *International Conference on Data*

Mining Workshop (pp. 1402–1409). IEEE. https://doi.org/10.1109/ICDMW.2015.230

Morlane-Hondère, F., Grouin, C., & Zweigenbaum, P. (2016). Identification of Drug-Related Medical Conditions in Social Media. In *International Conference on Language Resources and Evaluation* (pp. 2022–2028).

Nikfarjam, A., Sarker, A., Connor, K. O., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media : mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of Americal Medical Information Association*, *22*, 671–681. https://doi.org/10.1093/jamia/ocu041

Plachouras, V., Leidner, J. L., & Garrow, A. G. (2016). Quantifying Self-Reported Adverse Drug Events on Twitter : Signal and Topic Analysis. In *International Conference on Social Media & Society* (pp. 1–10). https://doi.org/10.1145/2930971.2930977

Powers, D. (2011). Evaluation: from precision, recall and f-measure to roc , informedness , markedness & correlation. *International Journal of Machine Learning Technology*, *2*(1), 37–63.

Price, J. (2016). What Can Big Data Offer the Pharmacovigilance of Orphan Drugs ? *Clinical Therapeutics*, *38*(12), 2533–2545.

Rifat, J. R., Rashed, S., Noori, H., & Hasan, R. (2019). Pharmacovigilance study of opioid drugs on Twitter and PubMed using artificial intelligence. In *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICCCNT45670.2019.8944668

Salathé, M. (2016). Digital Pharmacovigilance and Disease Surveillance : Combining Traditional and Big-Data Systems for Better Public Health. *The Journal of Infectious Diseases*, *214*(Suppl 4), 399–403. https://doi.org/10.1093/infdis/jiw281

Sarker, A., Ginn, R., Nikfarjam, A., Connor, K. O., Smith, K., Jayaraman, S., … Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance : A review. *Journal of Biomedical Informatics*, *54*, 202–212. https://doi.org/10.1016/j.jbi.2015.02.004

Sarker, A., & Gonzalez, G. (2015). Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of Biomedical Informatics*, *53*, 196–207. https://doi.org/10.1016/j.jbi.2014.11.002

Schneeweiss, S. (2016). Improving Therapeutic Effectiveness and Safety Through Big Healthcare Data. *Clinical Pharmacology and Therapeutics*, *99*(3), 262–265.

Schultz, T. (2013). Turning Healthcare Challenges into Big Data Opportunities : A Use-Case Review Across the Pharmaceutical Development Lifecycle. *Health Informatics*, *39*(5), 34–40.

Segura-Bedmar, I., Martínez, P., Revert, R., & Moreno-schneider, J. (2015). Exploring Spanish health social media for detecting drug effects. *BMC Medical Informatics*, *15*(56-Suppl 2), 1–9.

Singh, G., Schulthess, D., Hughes, N., Vannieuwenhuyse, B., & Kalra, D. (2018). Real world big data for clinical research and drug development. *Drug Discovery Today*, *23*(3), 652–660. https://doi.org/10.1016/j.drudis.2017.12.002

Sousa-pinto, B., Marques, B., Lopes, F., & Freitas, A. (2018). Frequency and Impact of Adverse Events in Inpatients : A Nationwide Analysis of Episodes between 2000 and 2015. *Journal of Medical Systems*, *42*(48), 1–9. https://doi.org/10.1007/s10916-018-0898-5

Stekelenborg, J. Van, Ellenius, J., Maskell, S., Bergvall, T., Caster, O., & Dasgupta, N. (2019). Recommendations for the Use of Social Media in Pharmacovigilance : Lessons from IMI WEB-RADR. *Drug Safety*, *42*(12), 1393–1407. https://doi.org/10.1007/s40264-019-00858-7

Trifirò, G., Sultana, J., & Bate, A. (2018). From Big Data to Smart Data for Pharmacovigilance : The Role of Healthcare Databases and Other Emerging Sources. *Drug Safety*, *41*(2), 143–149. https://doi.org/10.1007/s40264-017-0592-4

Tutubalina, E., & Nikolenko, S. (2017). Combination of Deep Recurrent Neural Networks and Conditional Random Fields for Extracting Adverse Drug Reactions from User Reviews. *Journal of Healthcare Engineering*, *2017*, 1–9. https://doi.org/10.1155/2017/9451342

Waegeman, W., Dembczynski, K., Cheng, W., & Hullermeier, E. (2011). An Exact Algorithm for F-Measure Maximization. In *Advances in Neural Information Processing Systems 24* (pp. 1–9). https://doi.org/10.5555/2986459.2986616

Weber, G., Mandl, K., & Kohane, I. (2014). Finding the Missing Link for Big Biomedical Data. *JAMA*, *311*(24), 2479–2480. https://doi.org/10.1001/jama.2014.4228

Wise, J., Möller, A., Christie, D., & Brodsky, E. (2018). The positive impacts of Real- World Data on the challenges facing the evolution of biopharma. *Drug Discovery Today*, *23*(4), 788–801. https://doi.org/10.1016/j.drudis.2018.01.034

Wu, C. (2019). MSA : Jointly Detecting Drug Name and Adverse Drug Reaction Mentioning Tweets with Multi-Head Self-Attention. In *12th ACM International Conference on Web Search and Data Mining* (pp. 33–41). https://doi.org/10.1145/3289600.3290980

Xia, L., Wang, G. A., & Fan, W. (2017). A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media. In *International Conference on Smart Health* (pp. 237–248). https://doi.org/10.1007/978-3-319-67964-8

Yadav, S., Ekbal, A., Saha, S., & Bhattacharyya, P. (2019). A Unified Multi-task Adversarial Learning Framework for Pharmacovigilance Mining. In *57th Annual Meeting of the Association for Computational Linguistics* (pp. 5234–5245). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1516

Yang, C., & Yang, H. (2015). Exploiting Social Media with Tensor Decomposition for Pharmacovigilance. In *15th Internation Conference on Data Mining Workshops* (pp. 188–195). IEEE. https://doi.org/10.1109/ICDMW.2015.163

Yang, M., Kiang, M., & Shang, W. (2015). Filtering big data from social media – Building an early warning system for adverse drug reactions. *Journal of Biomedical Informatics*, *54*, 230–240. https://doi.org/10.1016/j.jbi.2015.01.011

Zhang, Y., Cui, S., & Gao, H. (2020). Adverse drug reaction detection on social media with deep linguistic features. *Journal of Biomedical Informatics*, *106*(April), 103437. https://doi.org/10.1016/j.jbi.2020.103437