



**HAL**  
open science

# A Multiagent System for Integrated Detection of Pharmacovigilance Signals

Vassilis Koutkias, Marie-Christine Jaulent

► **To cite this version:**

Vassilis Koutkias, Marie-Christine Jaulent. A Multiagent System for Integrated Detection of Pharmacovigilance Signals. *Journal of Medical Systems*, 2016, 40 (2), pp.37. 10.1007/s10916-015-0378-0 . hal-01234025

**HAL Id: hal-01234025**

**<https://hal.sorbonne-universite.fr/hal-01234025>**

Submitted on 26 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Multiagent System for Integrated Detection of Pharmacovigilance Signals

Vassilis Koutkias · Marie-Christine Jaulent

Received: date / Accepted: date

**Abstract** Pharmacovigilance is the scientific discipline that copes with the continuous assessment of the safety profile of marketed drugs. This assessment relies on diverse data sources, which are routinely analysed to identify the so-called “signals”, i.e. potential associations between drugs and adverse effects, that are unknown or incompletely documented. Various computational methods have been proposed to support domain experts in signal detection. However, recent comparative studies illustrated that current methods exhibit high false-positive rates, significantly variable performance across different datasets used for analysis and events of interest, but also complementarity in their outcomes. In this regard, in order to reinforce accurate and timely signal detection, we elaborated through an agent-based approach towards systematic, joint exploitation of multiple heterogeneous signal detection methods, data sources and other drug-related resources under a common, integrated framework. The approach relies on a multiagent system operating based on a collaborative agent interaction protocol, aiming to implement a comprehensive workflow that comprises of method selection and execution, as well as outcomes’ aggregation, filtering, ranking and annotation. This paper presents the design of the proposed multiagent system, discusses implementation issues and demonstrates the applicability of the proposed solution in an example

signal detection scenario. This work constitutes a step towards large-scale, integrated and knowledge-intensive computational signal detection.

**Keywords** Pharmacovigilance · Computational Signal Detection Methods · Heterogeneous Data Sources · Multiagent System · Aggregation and Reasoning Scheme

## 1 Introduction

The importance of drug safety is widely recognised [1]. Pharmacovigilance encompasses all data gathering and processing activities related with the detection, assessment, understanding and prevention of adverse effects throughout the entire lifecycle of drugs [2].

One of the most important aspects in the monitoring of marketed drugs is the identification and analysis of new, medically important findings, so-called *signals* that might influence the use of a medicine. Signals have been defined<sup>1</sup> by the Council for International Organisations of Medical Sciences (CIOMS) VIII Working Group as “information that arises from one or multiple sources (including observations and experiments), which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events, either adverse or beneficial, that is judged to be of sufficient likelihood to justify verificatory action” [4]. In the scope of this work, we focus on adverse events.

In the pre-market stage, clinical trials of newly developed drugs constitute the main procedure for identi-

---

V. Koutkias (✉) · Marie-Christine Jaulent  
INSERM, U1142, LIMICS, F-75006, Paris, France;  
Sorbonne Universités, UPMC Univ Paris 06, UMR\_S 1142,  
LIMICS, F-75006, Paris, France;  
Université Paris 13, Sorbonne Paris Cité, LIMICS, UMR\_S  
1142, F-93430, Villetaneuse, France  
E-mail: vasileios.koutkias@inserm.fr

M.-C. Jaulent  
E-mail: marie-christine.jaulent@inserm.fr

---

<sup>1</sup> For a list of comprehensive definitions in the domain of pharmacovigilance the reader is referred to [3].

fying Adverse Drug Reactions<sup>2</sup> (ADRs) resulting from their use. However, due to time constraints, the limited sample population size as well as potential bias, clinical trials do not enable the detection of all possible ADRs. Consequently, post-marketing surveillance is necessary to identify new or incompletely documented ADRs throughout the time a drug is actively prescribed [5]. Active and systematic surveillance requires constant assessment of all drugs for any outcome [2].

Given the necessity to identify as early as possible signals, it has become evident that all the available data sources have to be explored [6]. In the post-marketing context, these sources vary, spanning from spontaneous reports, electronic health records [7], scientific literature [8], and even social media [9]. The availability of this data deluge dictates the need to introduce high-throughput computational methods that will enable efficient knowledge extraction and management, compensating the underlying heterogeneity and complexity.

To this end, various computational signal detection methods have been proposed to explore the above types of data sources [10]. However, these methods are designed for application in a single data source and typically demonstrate high false-positive rates in the provided results. In addition, the task of assessing the outcomes of analysis methods is currently performed manually and requires significant effort. Overall, accurate, timely and evidence-based signal detection remains a challenge [11].

The current work concerns *integrated signal detection*, i.e. a systematic, joint exploitation of multiple heterogeneous signal detection methods, data and other drug-related resources under a common framework, addressing methods' selection and execution, as well as outcomes' aggregation, filtering, ranking and annotation. Heterogeneity refers both to the methods' underlying computational models and the data that they can analyse. This paper presents an agent-based approach for the construction of a knowledge-intensive platform to accommodate these requirements, relying on a multiagent system that operates based on a collaborative agent interaction protocol.

The paper is structured as follows: In section 2, we provide background information concerning computational signal detection and comment on the motivation of this work. In section 3, we describe the challenges that this work implies and we justify the adoption of the agent paradigm. We then present the architecture of the proposed multiagent system, including agent classes

and the overall integrated signal detection strategy. In section 4, we provide details for our prototype implementation as regards the material and the development tools that have been employed. Equally important, we also illustrate the functionality of the multiagent system through an example signal detection scenario. In section 5, we discuss our contribution in comparison with related works and highlight current limitations and future work plans. Finally, section 6 concludes the paper.

## 2 Computational Signal Detection: Background and Motivation

Computational analysis methods constitute a well established support tool for signal detection [10]. Computationally extracted signals can be considered as potential causal relations that have to be verified. As such, they are perceived as early indications informing drug safety experts for appraisal actions, which require the review/analysis of scientific literature, clinical trial data, biological properties of drugs, etc., before concluding at causality.

The output of computational signal detection methods is a list of drug-event pairs, which shall be ideally ranked based on significance factors (statistical or probabilistic), in order to drive the focus of appraisal actions. Since these actions are typically performed manually, they require significant effort and, given the large amount of generated indications, they result in a bottleneck in the pharmacovigilance process [12]. Thus, besides highly accurate detection methods, filtering mechanisms are necessary, in order to exclude: (a) known ADRs, but also (b) associations containing conditions corresponding to the drugs' indicated use, which can be also generated as potential signals, according to the data being explored.

As the volume of data for signal detection is constantly increasing with new sources being considered for analysis (like patient-generated content in social media platforms), the need for advanced computational signal detection frameworks becomes imperative [13]. Based on their nature and the computational analysis that they require, the main data sources for signal detection can be categorised into: a) spontaneous reporting systems, b) structured, longitudinal observational healthcare databases, and c) unstructured / free-text data sources.

In particular, spontaneous reporting systems (SRSs) constitute the dominant source of signals through which suspected cases are voluntarily reported by healthcare professionals (and in some cases by patients) to regulatory authorities and other bodies. Prominent SRSs are the Food and Drug Administration (FDA) Adverse

<sup>2</sup> ADR: response that is noxious and unintended, and that occurs at doses normally used in humans for the prophylaxis, diagnosis, or therapy of disease, or for the modification of physiological function [3].

Event Reporting System (FAERS [14]) in US, EudraVigilance maintained by the European Medicines Agency [15], and VigiBase<sup>®</sup> maintained by the Uppsala Monitoring Centre [16]. The main type of signal detection methods for analysing SRS data relies on the statistical investigation of “disproportionality” (i.e. assessing whether a specific event is reported more frequently for a drug than would be expected if randomly occurring) [17]. Another category of methods exploiting SRS data are the so-called “multivariate” [10], such as logistic regression-based and unsupervised machine learning methods, which are able to analyse more complex drug safety incidents compared to disproportionality-based methods.

Structured observational healthcare data originate mostly from Electronic Health Record (EHR) and administrative claim systems. These data sources have been considered in terms of secondary use for the identification of potential causal relations between drugs and adverse conditions [7], [18]. Compared to spontaneous reports, observational data contain more information, such as diagnosis and lab examination results. The corresponding signal detection methods typically involve the application of data mining techniques in large datasets [19]. A number of methods have been also proposed having their origin from statistical epidemiology, relying on case-control design [20], cohort design [21], self-controlled case series design [22], self-controlled cohort design [23], etc. Notably, the Observational Medical Outcomes Partnership (OMOP) gathered an open-source library of such methods [24].

In the category of unstructured free-text data, typical sources include clinical narratives, scientific literature and patient self-reports. Clinical narratives remain a significant part of EHR systems, in which information concerning diagnosis and patient treatment (including drugs) is documented [25]. Scientific literature may also provide indications for pharmacovigilance signals [12] as well as evidence to support the evaluation of findings obtained by other sources. Patient self-reports about drug concerns/problems that are shared among networked communities using social media are an emerging source of signals [26]. However, the subjective nature and the lack of quality control in the generated data constitute important limitations for reliable signal detection. Independently of the data source that is being employed, computational signal detection using unstructured text relies on information extraction and natural language processing (NLP) techniques [27].

Interestingly, recent comparative studies of computational signal detection methods have highlighted [28]-[31]: (a) high false-positive rates, (b) significantly variable performance across different datasets used for anal-

ysis and events of interest, but also (c) complementarity among methods’ outcomes. These findings reinforce the argument for the joint exploration of as many data sources as possible through diverse computational signal detection methods. Another argument for combined signal detection is that potential signals replicated in multiple data sources may indicate increased confidence in establishing a causal relation among drugs and adverse effects and, thus, they are more likely to be true. For example, this hypothesis has been successfully evaluated in [32], combining EHR and SRS data.

To this end, the motivation of this work concerns the design and development of a systematic framework enabling large-scale combinatorial signal detection as described in the following section.

### 3 An Agent-based Approach for Integrated Signal Detection

#### 3.1 Challenges and Adoption of the Agent Paradigm

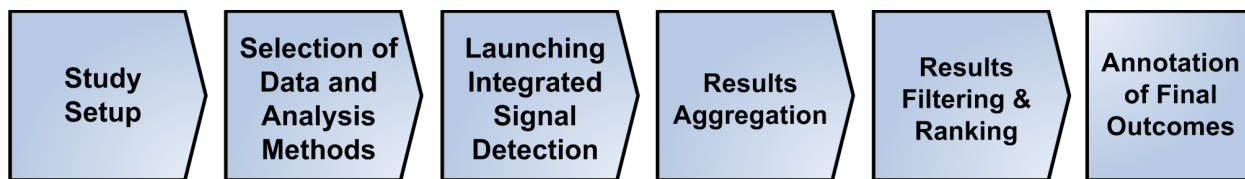
Our view for computational signal detection concerns an integrated framework in which joint exploitation of multiple heterogeneous drug safety data sources can be performed through relevant computational signal detection methods, as well as other drug information sources. However, this broader exploitation increases the complexity of the signal detection procedure for the end-users and requires a technical framework to be set.

More specifically, from the users’ perspective, the technical complexity caused for example by many parameters offered by some detection models<sup>3</sup>, has to be addressed. Similarly, users could be interested in exploring only specific data sources, e.g. based on their origin, coverage and / or content, thus, selection shall be facilitated. In this respect, support in study setup and selection of the resources that will be used is important.

Since the focus in signal detection concerns new ADRs, noisy indications (e.g. known ADRs or associations linking the drug with its indicated use) have to be filtered by accessing reference knowledge. Given the fact that computational signal detection methods typically generate many potential indications for ADRs, the remaining findings shall be ranked to prioritise their subsequent assessment by drug safety experts. In addition, supportive information on the findings would increase their comprehension, beyond just providing statistical measures on each drug-event pair. Such informa-

---

<sup>3</sup> An indicative example concerning the variety of analysis options is illustrated in [33], in which 14 signal detection methods were tested and the number of different combinations of analysis parameters varied from 48 to 162.



**Fig. 1** Elaborated workflow for integrated signal detection.

tion may constitute, for example, the toxicity profile of the drug of interest, recent relevant published studies, as well as information about potentially relevant clinical trials. At a lower level, technical challenges include the development of efficient mechanisms for interfacing with data sources and signal detection methods, as well as aggregating and synthesising method outcomes.

In this regard, Figure 1 depicts the overall signal detection workflow elaborated in this work, starting from study setup and concluding with outcomes annotation and enrichment with supportive information. For the implementation of this workflow, we elaborate on a software platform, in order to address two major aspects. First, the semantic harmonisation of key concepts / information models of signal detection methods available through the platform. Second, the aggregation of the methods’ outcomes and the assessment of the findings’ strength. For the first part we developed the Pharmacovigilance Signal Detectors Ontology (PV-SDO), which has been presented in [34], while the second part is addressed via a multiagent system that is presented in the current paper.

We adopted the “software agent” paradigm, as it offers a favourable way to consider complex systems with multiple distinct and independent components. Agents enable the aggregation of different functionalities in a conceptually embodied and situated whole, while at the same time may exhibit autonomous, social, reactive and/or proactive features [35]. This offers appropriate abstractions for the design and development of large systems of individual agents and of ways in which agents may interact, while taking into account organisational or macro-level issues. Especially through task decomposition and coordination, software agents can contribute in the accomplishment of complex goals and tasks. In the current work, agents provided us with a way of structuring a system around autonomous communicative elements [36], through which we develop mechanisms for automating and improving signal detection in an integrated framework, supporting drug safety experts in accomplishing this task.

Since our ultimate goal is the development of an extensible platform in which new methods, raw data and other relevant data/information sources can be incorporated, the modularity and reusability offered by the

agent paradigm [37] enable us to design a sustainable and scalable architecture comprising of diverse components as presented in the following subsection.

### 3.2 System Architecture: Resources and Agent Classes

In order to implement the workflow presented in Fig. 1, we defined a system architecture comprising of software agents, non-agent software components, data / information sources, and a knowledge component. Non-agent software components correspond to implementations of signal detection methods, while data sources provide the raw data employed for signal detection (e.g. spontaneous reports, EHR data, etc.). Information sources provide reference data on known ADRs and other drug-related information, while the knowledge component corresponds to an ontology-based Semantic Registry describing all the resources of the framework based on PV-SDO [34].

We formulated a collaborative agent team, which operates according to an agent interaction protocol that we defined [38], corresponding to the implementation of each step of the integrated signal detection workflow of Fig. 1. In this regard, agents exploit the resources available in the system, so that signal detection methods are appropriately selected and launched, using the Semantic Registry, while their results are aggregated and post-processed via filtering (to discard noisy indications) and ranking (based on their significance) mechanisms, using the reference ADR and drug information sources, respectively.

Appropriate agent classes were defined to support this functionality. Initially, the main principle for defining these classes was task decomposition and distribution [39], so that each agent is delegated a specific and simple task to accomplish, and the computational burden underlying integrated signal detection is distributed among agents. In the overall analysis and design phase, we followed the principles of MOBMAS [40], a comprehensive methodology that covers the main activities of developing multiagent systems (analysis, multiagent organisation design, agent internal design, agent interaction design, and agent architecture), which is relevant with the development of ontology-based multiagent systems. Based on the collaborative agent interac-

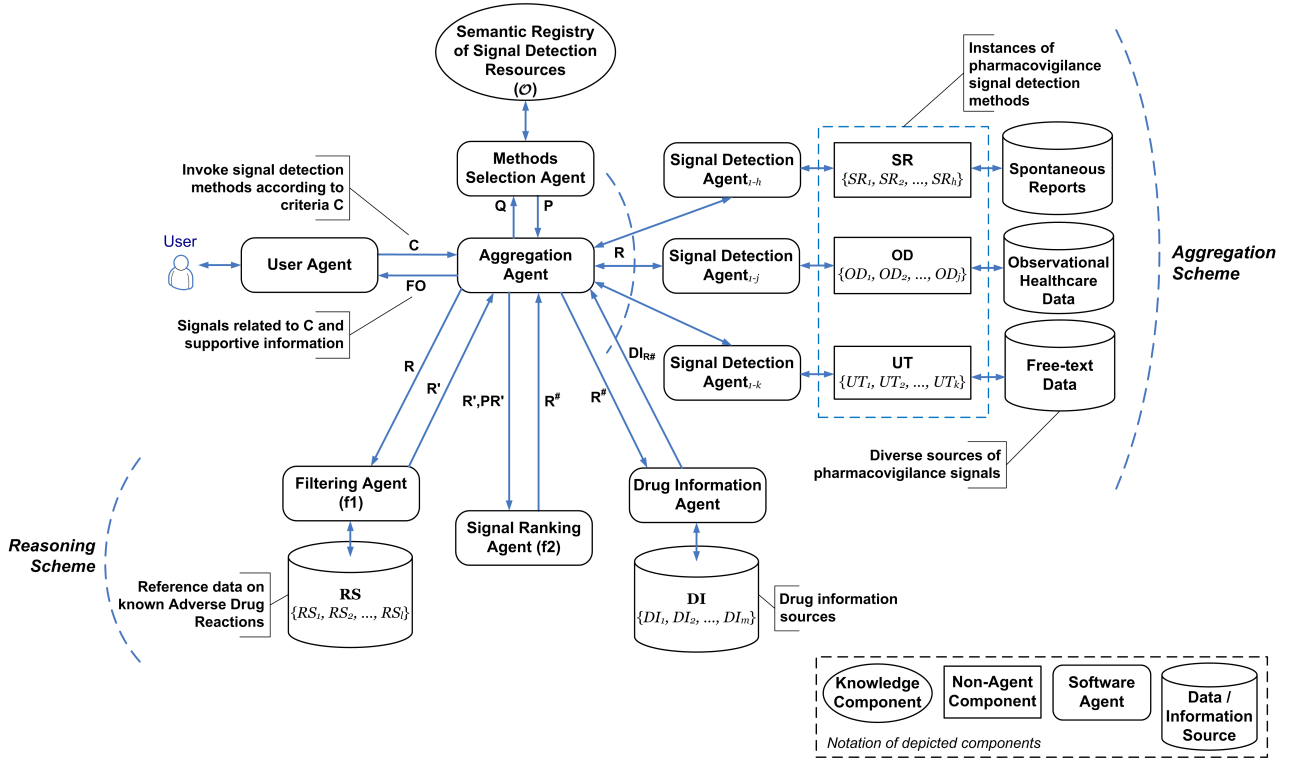


Fig. 2 The proposed system architecture for integrated signal detection.

tion protocol, agent activities are coordinated and result in a multiagent system that performs the proposed integrated signal detection via teamwork.

Figure 2 depicts the proposed system architecture, discriminating between an aggregation scheme and a reasoning scheme. In this respect, let us consider that the architecture comprises of  $h$  SRS-based signal detection methods denoted by the set:

$$\mathbf{SR} = \{SR_1, SR_2, \dots, SR_h\}, \quad (1)$$

$j$  methods appropriate for signal detection in structured longitudinal observational healthcare data sources corresponding to:

$$\mathbf{OD} = \{OD_1, OD_2, \dots, OD_j\}, \quad (2)$$

and  $k$  detection methods for unstructured text sources denoted as:

$$\mathbf{UT} = \{UT_1, UT_2, \dots, UT_k\}. \quad (3)$$

The architecture includes also means to access  $l$  ADR reference data sources:

$$\mathbf{RS} = \{RS_1, RS_2, \dots, RS_l\}, \quad (4)$$

as well as  $m$  drug information sources:

$$\mathbf{DI} = \{DI_1, DI_2, \dots, DI_m\}. \quad (5)$$

$\mathcal{O}$  depicted in Figure 2 denotes the ontology-based *Semantic Registry of Signal Detection Resources* [34].

The following agent classes were defined:

- *User Agent*: an interface agent responsible to support the end-user in defining and conducting an integrated signal detection experiment.
- *Aggregation Agent*: a mediator agent which controls the agent interaction protocol.
- *Methods Selection Agent*: a wrapper agent which queries  $\mathcal{O}$  for appropriate signal detection methods given a specific experiment.
- *Signal Detection Agent*: a wrapper agent capable of receiving requests for signal detection (along with specific analysis parameters), launching the respective signal detection method, and returning the outcomes<sup>4</sup>.
- *Filtering Agent*: a processing agent capable of accessing ADR reference sources, in order to filter the aggregated outcome provided by *Signal Detection Agents*.
- *Signal Ranking Agent*: a processing agent capable of applying diverse criteria/metrics, in order to rank the results provided by the *Filtering Agent*.

<sup>4</sup> Note that for each one of the considered computational methods (1)-(3), a *Signal Detection Agent* is assigned.

- *Drug Information Agent*: an information retrieval agent capable of identifying supportive information for the outcomes of the *Signal Ranking Agent*.

According to the behaviour that the above agent classes have to exhibit, we addressed design issues at the micro-level. For example, the *Aggregation Agent* encapsulates a rule-based logic, in order to implement the agent coordination protocol, while wrapper agents (i.e. the *Methods Selection Agent* and *Signal Detection Agents*) incorporate a translation mechanism capable of converting agent requests to queries in the respective resources and encapsulating the obtained results in an agent message.

Upon a new experiment definition, instances of the above agent classes are launched to undertake the corresponding signal detection workflow<sup>5</sup>.

The behaviour of each agent is presented in the description of the aggregation and reasoning schemes, in subsections 3.3 and 3.4, respectively.

### 3.3 Selection of Methods and Aggregation of Outcomes

The proposed platform aims to support two diverse usage scenarios, i.e.: (a) *signal detection without a specific target* on drugs or health outcomes, which is performed routinely by drug safety authorities, and (b) *targeted signal detection* concerning specific drugs and potential adverse effects, which is particularly interesting for pharmaceutical companies as a mean to identify risks for their marketed drugs.

Upon the definition of a new analysis experiment, in the targeted signal detection scenario the user provides as input the target of the analysis (drug and/or health outcome of interest). In addition, in both scenarios, optional input parameters can be: (a) the resources that will be employed in the experiment (i.e. the sources of pharmacovigilance signals, the reference data sources on known ADRs and the drug information sources), (b) specific values of analysis/configuration parameters offered by signal detection methods, such as the rule for signal generation<sup>6</sup>, the time-at-risk<sup>7</sup>, etc., and (c) ranking criteria to be employed for the prioritisation of the

generated signals. If the above optional input parameters are not defined, then: (a) all the resources available in the platform will be employed, (b) the default values for analysis/configuration parameters defined for each signal detection method will be used, and (c) all the available ranking criteria will be employed, respectively.

Although explicit, the above inputs pose also implicit requirements (e.g. which methods can support the signal generation criterion) for the experiment that will be conducted. Based on these inputs, the selection of the signal detection methods that will be used in a given signal detection experiment is performed by querying the Semantic Registry  $\mathcal{O}$  (Fig. 2).  $\mathcal{O}$  defines the capabilities, interfaces and parameterisation options of each signal detection method, according to expert domain knowledge and the specifications provided by the respective developers. The respective query is forwarded to the *Methods Selection Agent* by the *Aggregation Agent*, given the input parameters provided by the *User Agent*.

In particular, let  $\mathbf{C}$  denote the set of input parameters that the user defined for a given experiment:

$$\mathbf{C} = \{C_1, C_2, \dots, C_q\}, \quad (6)$$

and assume that  $\mathbf{C}$  is met by a set of  $n$  detection methods:

$$\mathbf{P} = \{P_1, P_2, \dots, P_n\}, \quad (7)$$

with  $\mathbf{P}$  denoting a subset of the available ones, i.e.  $\mathbf{P} \subseteq (\mathbf{SR} \cup \mathbf{OD} \cup \mathbf{UT})$ . In particular,  $\mathbf{P}$  is determined by applying a matchmaking query  $\mathbf{Q}$ , between the input parameters  $\mathbf{C}$  and the registered detection methods in  $\mathcal{O}$ , through the *Methods Selection Agent*:

$$\mathbf{Q} : \mathbf{C} \xrightarrow{\mathcal{O}} \mathbf{P}. \quad (8)$$

According to  $\mathbf{P}$ , the corresponding *Signal Detection Agents* are requested by the *Aggregation Agent* to invoke the respective signal detection methods, taking into account potential input parameters defined in (6). The analysis results obtained from each *Signal Detection Agent* are gathered by the *Aggregation Agent*, and correspond to a set  $\mathbf{R}$ , with:

$$\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n\}, \quad (9)$$

where  $\mathbf{R}_i$ ,  $i = \{1, \dots, n\}$  denotes in turn a set of potential signals obtained from the  $P_i$  detection method (7). In addition, elements of  $\mathbf{R}$  may be equal to  $\emptyset$ , in case where no indication for signals is obtained by the respective detection method(s). Next, the *Aggregation Agent* encodes the obtained outcomes  $\mathbf{R}$  into a uniform representation, keeping also provenance information (i.e. which methods and data were employed and with what parameter values) regarding the origin of each  $\mathbf{R}_i$ .

<sup>5</sup> This practice has been followed in the current experimental implementation. However, it is also possible to share agent instances across experiments (such as the *Drug Information Agent*, the *Filtering Agent*, the *Signal Ranking Agent*, etc.).

<sup>6</sup> A condition or a set of conditions that needs to be satisfied by a signal detection method, in order to generate a signal.

<sup>7</sup> The period of time a patient is considered to be ‘at-risk’ and, therefore, counting occurrence of conditions as potential events.

### 3.4 Filtering, Prioritisation and Annotation of Outcomes

This part concerns the reasoning scheme, employing the available reference ADR and drug information sources, i.e.  $\mathbf{RS}$  (4),  $\mathbf{DI}$  (5), respectively, in order to: (a) filter  $\mathbf{R}$  (9) for noisy indications and, subsequently, (b) prioritise the remaining novel indications (if any), according to ranking criteria.

In this regard, first, the *Aggregation Agent* provides its results to the *Filtering Agent*, which implements a novelty assessment mechanism  $\mathbf{f1}$ , so that from each  $\mathbf{R}_i$  a subset  $\mathbf{R}'_i$  ( $\mathbf{R}'_i \subseteq \mathbf{R}_i$ ) is produced containing only the novel indications:

$$\mathbf{f1} : \mathbf{R} \xrightarrow{\mathbf{RS}} \mathbf{R}'. \quad (10)$$

As a second step, the *Aggregation Agent* provides  $\mathbf{R}'$  to the *Signal Ranking Agent*, which is in turn capable of applying a prioritisation mechanism  $\mathbf{f2}$  on  $\mathbf{R}'$  based on a set of metrics that is denoted as:

$$\mathbf{PR} = \{PR_1, PR_2, \dots, PR_w\}. \quad (11)$$

In  $\mathbf{PR}$ , we discriminate two types of metrics/criteria:

1. *computational metrics*, e.g. prioritising the common results within the first  $k$  among those provided by the signal detection methods employed in a given analysis experiment, and
2. *domain criteria*, like those proposed for implementing triage schemas in signal detection [41], e.g. ADR seriousness, newer drugs, data originated from multiple countries, etc.

The information to apply domain criteria may be found either in drug information sources (e.g. the drug release date), or in the raw data employed for signal detection. As an example of the latter case, spontaneous reports typically contain a characterisation of the reported event as serious or not, and may further specify seriousness according to the result (e.g. death, life threatening condition, hospitalisation, etc.). Thus, the *Signal Ranking Agent* embodies the required knowledge, in order to interpret such domain criteria.

Let us consider that the end-user defined in  $\mathbf{C}$  (6) a set of ranking metrics  $\mathbf{PR}'$  (with  $\mathbf{PR}' \subseteq \mathbf{PR}$ ) to be applied in a given experiment.

Then,  $\mathbf{R}^\#$  corresponds to the ranking of  $\mathbf{R}'$  based on  $\mathbf{PR}'$ , i.e.  $\mathbf{R}^\# \equiv \text{rank}(\mathbf{R}')|\mathbf{PR}'$ , by applying  $\mathbf{f2}$ , i.e.:

$$\mathbf{f2} : \mathbf{R}' \xrightarrow{\mathbf{PR}'} \mathbf{R}^\#. \quad (12)$$

The final outcome  $\mathbf{FO}$  provided to the end-user via the *User Agent* comprises of  $\mathbf{R}^\#$  along with support

information obtained from the *Drug Information Agent* through  $\mathbf{DI}$  per  $\mathbf{R}^\#$  element, if available, i.e.:

$$\mathbf{FO} = \mathbf{R}^\# \cup \mathbf{DI}_{\mathbf{R}^\#}. \quad (13)$$

Figure 3 illustrates via a sequence diagram the collaborative agent interaction protocol implementing the above signal detection strategy within the integrated framework.

## 4 Implementation and Example Scenario

### 4.1 Material

At the current stage, the development of the proposed integrated signal detection framework relies on publicly available resources. In particular, we employed the following:

- *Sources of raw data for signal detection*: (a) the FAERS spontaneous reporting system; (b) PubMed, the reference bibliographic database in the life sciences, and (c) Twitter, a popular micro-blogging platform.
- *Reference data sources for filtering the outcomes of signal detection methods*: (a) the Side Effect Resource (SIDER) [42], containing information on marketed medicines and their recorded ADRs, which is extracted from public documents and package inserts, and (b) DrugBank [43], since it provides rich information on drug-drug interactions. Both SIDER and DrugBank provide also the indications for each drug's use, which is another type of information employed for filtering.
- *Drug information sources for supporting evidence on novel findings*: (a) ChEMBL [44], which contains 2-D structures, calculated properties and abstracted bioactivities of drugs, but also (b) DrugBank, as it combines drug data (i.e. chemical, pharmacological and pharmaceutical) with comprehensive drug target information (i.e. sequence, structure, and pathway), along with biointeraction information.
- *Signal Detection Methods*: implementations of well-known methods contained in the open-source PhViD  $R$  package [45], and an in-house signal detection method appropriate for unstructured text.
- *Semantic Registry*: PV-SDO [34], an ontology expressed in OWL2 (Web Ontology Language 2) [46], and populated with sample signal detection method implementations, and the reference drug data and information resources mentioned above.

While further resources will be elaborated in subsequent stages of this development, the above material has been considered adequate for a proof-of-concept prototype implementation.



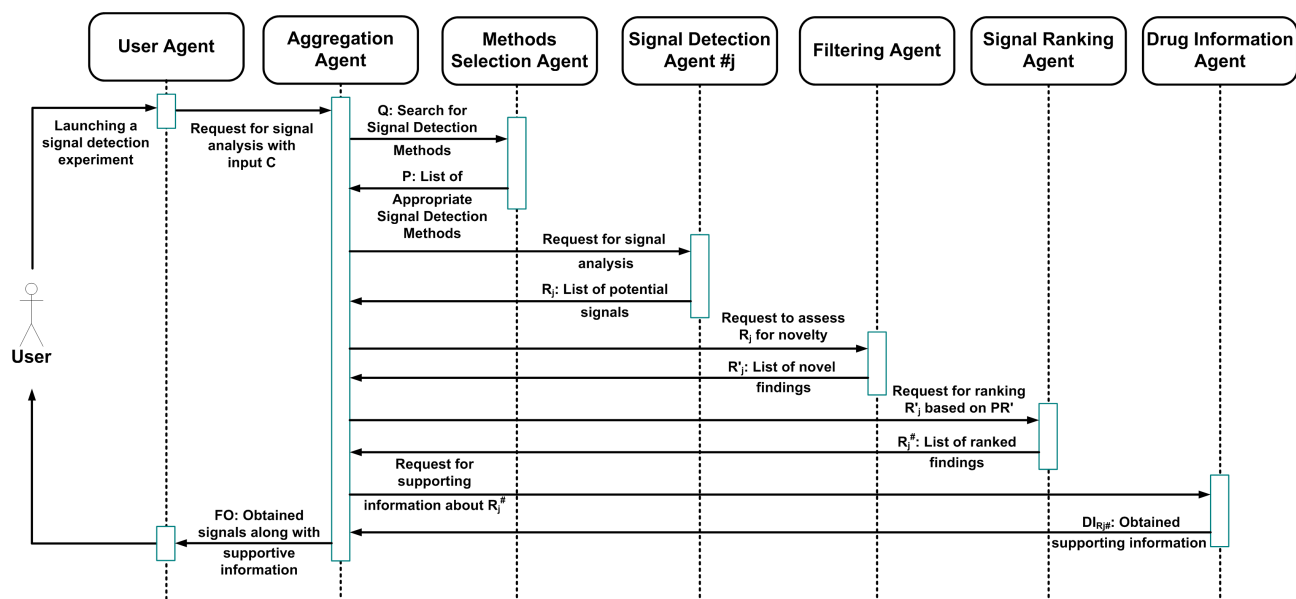


Fig. 3 A partial view of the elaborated agent interaction protocol (for simplicity, only agents are depicted).

## 4.2 Development Aspects

Access to the resources mentioned in subsection 4.1 has been realised programmatically. In particular, FAERS data are retrieved through the openFDA Application Programming Interface (API) [47], access to PubMed data relies on the Web services provided by Europe PubMed Central [48], and Twitter posts are retrieved through the available REST (Representational State Transfer) APIs [49]. Access to SIDER, DrugBank and ChEMBL is made through the SPARQL (SPARQL Protocol and RDF Query Language) endpoints [50], available from Bio2RDF [51].

The multiagent system implementation is based on the Java Agent DEvelopment framework (JADE) [52], providing besides open source development tools, the necessary agent execution environment. JADE is compliant with the specifications for interoperable multiagent systems released by the Foundation for Intelligent Physical Agents (FIPA) [53].

*Signal Detection Agents* invoke the respective signal detection methods via Rserve [54], a TCP/IP server allowing for third-party software to use  $R$  from various languages and without the need to initialise  $R$  or link against the  $R$  library. Upon initiation, each *Signal Detection Agent* is registered with the *Directory Facilitator*, a standard agent contained in the FIPA Abstract Architecture [55], which provides “yellow page” services for the agents of the platform. Thus, the *Aggregation Agent* queries the JADE built-in *Directory Facilitator* to obtain the list of relevant *Signal Detection Agents*, according to the analysis scenario.

In the current prototype, knowledge embodied in the *Signal Ranking Agent* for the application and interpretation of domain criteria used for signal ranking, such as seriousness, is represented via a set of simple rules (e.g. to encode that death is more severe than hospitalisation). Although this is a rather simplistic approach, a generic resource providing this kind of knowledge is not currently available.

Agent communication relies on message exchange. Messages are represented in the relevant FIPA standards for agent communication, which set out the encoding, the semantics and the pragmatics of messages. In this regard, messages are structured according to the FIPA ACL Message Structure Specification [56], and are built upon the FIPA Communicative Act Library (e.g. request, inform, agree, etc.) [57]. The messages’ content relies on FIPA-SL [58], and on a basic application ontology that we defined. This JADE-specific ontology contains the respective agent actions and the predicates to facilitate agent communication, according to the methodology proposed in JADE and using the Ontology BeanGenerator tool [59].

While various open-source implementations of signal detection methods are available for SRS and structured observational data, the availability of implementations for unstructured text-based signal detection methods is limited for the moment. Thus, we developed a simple signal detection method relying on the clinical Text Analysis and Knowledge Extraction System (*cTAKES*) [60], and on its built-in Unified Medical Language System (UMLS) [61] Annotator, to identify drug-related terms and effects within PubMed abstracts and

Twitter posts (namely, VigiText). Notably, cTAKES is built upon the Unstructured Information Management Architecture (UIMA) [62], thus, the development of VigiText follows an established standard for managing unstructured data.

We illustrate the functionality of the proposed multiagent system via the following example signal detection scenario.

#### 4.3 Example Scenario

Let us focus on a targeted signal detection scenario and assume that the proposed platform integrates through its multiagent system the following signal detection methods<sup>8</sup>:

- SRS-based detection methods<sup>9</sup>: BCPNN, GPS, PRR, and ROR.
- Observational data detection methods<sup>10</sup>: ICTPD and LGPS.
- Unstructured text detection methods: VigiText.

In this case, (1)-(3) correspond to:

**SR** = {BCPNN, GPS, PRR, ROR},

**OD** = {ICTPD, LGPS}, and

**UT** = {VigiText}, respectively.

Also, the *Filtering Agent* is capable of querying **DrugBank** and **SIDER**, in order to filter the obtained outcomes from known ADRs and findings that correspond to the drug's indicated use, while the *Drug Information Agent* can connect to **ChEMBL** and **DrugBank**, in order to obtain complementary drug information. Thus, in the scope of this scenario, (4)-(5) correspond to:

**RS** = {SIDER, DrugBank} and

**DI** = {ChEMBL, DrugBank}, respectively.

A safety expert working for a pharmaceutical company wishes to search for potential signals concerning **Drug X**, a new drug of the company which has been quite recently made available in the market. Thus, (s)he employs the resources of the platform to analyse data from an observational healthcare database and the reports available in FAERS. The user is prompted by the *User Agent* to provide the input for this analysis (cf. subsection 3.3). Besides **Drug X**, let us consider that

<sup>8</sup> For the sake of simplicity and for illustration purposes, we consider a limited number of detection methods and refer to fictitious data in this example.

<sup>9</sup> Methods contained in [45]. BCPNN: Bayesian Confidence Propagation Neural Network, GPS: Gamma Poisson Shrinkage, PRR: Proportional Reporting Ratio, ROR: Reporting Odds Ratio.

<sup>10</sup> Methods available at [63]. ICTPD: Information Component Temporal Pattern Discovery, LGPS: Longitudinal Gamma Poisson Shrinkage.

(s)he defines for SRS-based detection methods the decision rule for signal generation to be "more than 3 reports", while for the OD-based detection methods (s)he sets the value of the "Time-at-risk" parameter to be "length\_of\_exposure+30days". Thus, (6) can be rewritten as:

$$C = \{Target="Drug-X", Methods="SR, OD",$$

$$SR\_Analysis\_Criterion :$$

$$Signal\_generation\_condition=">3\ reports",$$

$$OD\_Analysis\_Criterion :$$

$$Time-at-risk="length\_of\_exposure+30days"\}.$$

Assuming that the considered ICTPD implementation does not support the defined value for **Time-at-risk**, a fact that is captured in the PV-SDO ontology, ICTPD is implicitly excluded from the analysis experiment as indicated by the *Methods Selection Agent*. Thus, (9) corresponds to:

$$P = \{BCPNN, GPS, PRR, ROR, LGPS\}.$$

Launching a *Signal Detection Agent* for each detection method of **P**, let us assume that the aggregation outcome given by (9) for the targeted **Drug X** is:

$$R = \{R_{BCPNN}, R_{GPS}, R_{PRR}, R_{ROR}, R_{LGPS}\}, \text{ with}^{11}:$$

$$R_{BCPNN} = \{\text{Drug X-MI}, \text{Drug X-Neutropenia}\},$$

$$R_{GPS} = \{\text{Drug X-MI}\},$$

$$R_{PRR} = \emptyset,$$

$$R_{ROR} = \{\text{Drug X-ARF}\} \text{ and}$$

$$R_{LGPS} = \{\text{Drug X-MI}\}.$$

The *Filtering Agent* performs SPARQL queries to **DrugBank** and **SIDER** such as the one presented in Fig. 4, in order to identify which of the above indications shall be discarded. Let us assume that MI is a registered ADR for **Drug X** in both **DrugBank** and **SIDER**. Thus, it is discarded from **R**. As a result, **R'** comprises of:

$$R'_{BCPNN} = \{\text{Drug X-Neutropenia}\} \text{ and}$$

$$R'_{ROR} = \{\text{Drug X-ARF}\}, \text{ while}$$

$$R'_{GPS} = R'_{PRR} = R'_{LGPS} = \emptyset.$$

In order to prioritise the remaining indications, i.e. **Neutropenia** and **ARF**, the user selects to employ the ADR seriousness criterion, i.e. **PR'** = {Seriousness}. Assuming that **ARF** appears to result in more serious effects than **Neutropenia** in the analysed data (e.g. more deaths have been reported with **ARF**), as provided by the *Signal Ranking Agent*, **Drug X-ARF** is ranked prior to **Drug X-Neutropenia**, i.e.:

$$R\# = \{\text{Drug X-ARF}, \text{Drug X-Neutropenia}\}.$$

Finally, the *Drug Information Agent* obtains the entire biochemical profile of **Drug X** and its biointeractions by querying **ChEMBL** and **DrugBank**, respectively.

<sup>11</sup> ARF: Acute Renal Failure, MI: Myocardial Infarction.

```

PREFIX dv: <http://bio2rdf.org/drugbank_vocabulary:>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT ?x ?symptom
WHERE {
  ?x ?y ?z .
  # Restrict returned triples based on drug X
  ?x ?s ?drugName .
  ?x rdf:type dv:Drug .
  FILTER regex(?drugName, "drug X")
  FILTER ((?s=rdfs:label || ?s=dct:title || ?s=dct:description || ?s=dv:synonym))
  # Restrict returned triplets based on Myocardial Infarction found in the description of the toxicity resource
  ?x ?t _:a .
  _:a ?u ?effect .
  FILTER regex(?effect, "Myocardial Infarction")
  FILTER ((?t=dv:toxicity || ?t=dv:ddi-interactor-in) && ?u=dct:description)
}

```

Fig. 4 Indicative SPARQL query to assess whether Myocardial Infarction is a known ADR for drug X.

This information is attached to  $\mathbf{R}^\#$  to formulate the final outcome  $\mathbf{FO}$  as in (13), also annotated with provenance information for the experiment.

For illustration purposes, the above scenario refers to fictitious data and it is rather oversimplified, since typically the outcome of a signal detection method returns a long list of drug-event pairs, which shall be further considered for a potential causal relation. Nevertheless, it provides a comprehensive view of the agent interaction protocol and the scope of integrated signal detection that we introduced in this paper.

Notably, taking advantage of the proactive nature of agents [35], the above drug safety screening scenario may be executed as a background process. In this respect, the user is notified promptly as soon as indications for signals are identified. Similarly, in the scenario of screening for signals without a specific target, the process shall be iteratively executed, when new data become available.

## 5 Discussion

### 5.1 Contribution of this Work

This study elaborated on the design and development of advanced IT mechanisms to leverage pharmacovigilance signal detection. Timely and accurate signal detection is of utmost importance in the scope of an active drug surveillance system [64]. A number of quite recent, very significant drug safety issues illustrated that adverse effects of drugs may be detected too late, when millions of patients have already been exposed. To this end, the investigation of diverse data sources via different analysis methods for signal detection can contribute in more timely and accurate findings [6].

In accordance with this perspective, the current work introduced integrated signal detection, i.e. jointly exploiting multiple heterogeneous signal detection methods, data and other drug-related resources under a com-

mon framework, in order to strengthen the reliability and coverage of the analysis outcomes. Besides outcomes' aggregation, the proposed approach included their filtering, ranking and annotation, aiming to support drug safety experts in assessing and prioritising the results. To this end, we take into account both computational metrics and domain criteria in order to prioritise the findings, while we also seek existing knowledge for filtering noisy indications from relevant linked-data based repositories.

Our research was built upon current systematic efforts in the field, as we relied on available implementations of signal detection methods, as well as publicly available drug safety related data and resources. The idea of exploring and comparing different types of pharmacovigilance signal detection methods has been pinpointed in the literature [32], [65]. However, we elaborated on a systematic framework aiming to explore this perspective at large-scale.

The multiagent system that has been proposed provides the basis for developing a platform to conduct integrated signal detection. We defined a collaborative agent integration protocol, in order to coordinate agent activities and implement the workflow for integrated signal detection. Thanks to the modularity and scalability offered by the agent paradigm [37], we are able to replicate our current approach and extend its capabilities by integrating additional detection methods in the platform.

### 5.2 Related Work

In the field of drug safety surveillance, various systematic initiatives have been recently formulated. In particular, the FDA-funded Mini-Sentinel project elaborates on designing safety assessments using multiple existing electronic healthcare data systems, developing and evaluating scientific methods to increase the precision of detection techniques, while also identifying and address-

ing barriers and challenges to build an advanced system for active safety surveillance [66]. Similarly, OMOP implemented and evaluated signal detection methods for longitudinal observational data [24]. More recently, the Observational Health Data Sciences and Informatics (OHDSI) program has been established as a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics [67].

In addition, the EU-ADR Web Platform relies on the exploitation of EHR data collected within a large-scale European initiative [68]. Data are mined for specific drug events, which are then correlated with literature, protein and pathway data, resulting in a rich drug-event dataset. Advanced distributed computing methods based on workflow technologies are tailored to coordinate the execution of data-mining and statistical analysis tasks.

Contrary to the above-mentioned important initiatives, we consider signal detection in an integrative framework that involves multiple heterogeneous data sources and detection methods, rather than focusing on a specific method applied on multiple sources of the same kind, or on diverse methods applied on a specific data source for benchmarking [13]. In this respect, we adopted the agent paradigm as the enabling technology for the development of an integrated platform, due to its proactive, reactive and social nature. These features enabled us to design and implement a novel collaborative multiagent system for signal detection, relying on appropriate mechanisms to automate and improve existing tasks [36], that drug safety experts typically have to undertake manually.

The agent paradigm has been successfully employed in healthcare in various cases [37], with applications spanning from decision support systems [69], disease monitoring [70], telehealth services [71], ambient intelligence and independent living [72], integration of heterogeneous biomedical systems [73], cross-organisational workflow management [74], planning and resource allocation [75], to evaluation of medical systems [76]. Via this work, we illustrate a new application domain of agent technologies in life sciences, i.e. pharmacovigilance.

To the best of our knowledge, an agent-based perspective on the domain of drug safety has been only explored in [77], in which a multiagent framework was proposed for early detection of ADRs using electronic patient data distributed across different sources and locations. The framework relied on intelligent agents employing fuzzy logic to represent, interpret, and compute imprecise and subjective cues that are commonly encountered in ADR identification, as well as to retrieve

prior experiences by evaluating the extent of matching between the current situation and a past experience. Contrary to [77], in the current paper we introduced an agent-based approach for the joint exploitation of multiple, diverse and existing signal detection methods, which are designed to analyse different types of signal sources, under an integrated framework.

From a technical perspective, given that the proposed multiagent system relies on semantic resources, a notable work is the BioSTORM system [78]. BioSTORM elaborated on ontology-based, end-to-end deployment of JADE-based agent systems by exploiting diverse surveillance methods. While BioSTORM shares some architectural similarities with the proposed system, it was created for a different purpose/domain than drug safety that was our focus. In terms of its general rationale, our system shares some common attributes with the DeepQA architecture [79]. DeepQA exploits multiple methods for analysing natural language, identifying sources, finding and generating hypotheses, finding and scoring evidence, as well as merging and ranking hypotheses. Nevertheless, DeepQA is a general-purpose massively parallel probabilistic evidence-based architecture having its focus on unstructured/free-text data based on UIMA, while our work relies on both structured and unstructured resources focusing on a specific application domain.

### 5.3 Challenges and Future Work

Using a first prototype, we are currently elaborating on particular case studies to illustrate the added value of the proposed approach. Defined in collaboration with drug safety experts, these studies concern drugs and adverse effects that are quite rare, as well as new drugs placed in the market, e.g. new anticoagulants, and their association with specific adverse effects. In this regard, the preliminary analysis reveals some “echoing” findings across the diverse data sources explored, i.e. the same signals are discovered in multiple data sources and by multiple detection methods. This may indicate the means of strengthening the outcomes provided by signal detection methods applied separately, since replicated signals obtained from diverse methods and data sources may indicate more probable causal associations between drugs and adverse effects.

Given the uncertainty of the results that signal detection methods provide, approaches for fusing knowledge under uncertainty are being investigated [80]. We are also considering an agent-based voting scheme [81], which will enable us to pursue signal prioritisation by applying multiple ranking criteria and measuring the strength of signals in a comparative fashion. Overall,

combining methods and their results in order to improve accuracy, confidence, or speed such as the approach employed in DeepQA [79], is a part of the developed system that requires further research.

In terms of development, we find the potential to assess whether BioSTORM could be applicable in the particular application domain interesting [78]. In addition, in the current implementation, the *Signal Detection Agents* are handling the *R* output of the respective detection methods. Signal detection methods exposed through standardised, machine readable interfaces, would provide a more systematic solution. To this end, we explore comprehensive frameworks for integrating *R* code in third-party applications, such as DeployR [82]. Enriching and advancing the knowledge according to which the *Signal Ranking Agent* applies domain criteria, requires further research and development.

Up to now, we experimented with signal detection methods that are available in the PhVid *R* package [45]. Further methods will be integrated in the subsequent implementation phase, such as those contained in the OMOP Methods Library [63]. Along this line, we aim to elaborate on additional data sources for signal detection and, particularly work with structured observational data. We are also elaborating on mechanisms for social media screening, since patient self-reports may provide unique insights, especially for drugs used in the treatment of rare diseases, the feedback for which is limited.

In the scope of active drug safety surveillance [64], signal detection is a continuous and demanding process. From the computational viewpoint, taking into account the complexity posed by jointly exploiting multiple signal detection methods and drug-related resources under the integrated signal detection approach that we introduced, deploying and executing the proposed multi-agent system in a cloud infrastructure is being considered as a future step [83]. This would accommodate the required power, reliability and scalability for a large-scale deployment accounting for additional signal detection methods and drug safety resources.

## 6 Conclusions

This paper illustrated the application of the multi-agent paradigm for systematically pursuing pharmacovigilance signal detection within an integrated framework. A formal agent collaboration strategy was proposed, taking into account methods' selection, aggregation of methods' outcomes, and subsequently filtering known ADRs and ranking the remaining potential signals for prioritisation. Relying on various data, information and knowledge sources along with a number of

signal detection method implementations, our proof-of-concept development illustrated that the agent-based approach is valid for such an integration. We are currently elaborating to integrate further resources and illustrate the added value in signal detection through concrete case studies. Overall, this work constitutes a step toward large-scale, knowledge-intensive computational signal detection.

**Acknowledgements** This research was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme FP7/2007-2013 under REA grant agreement n° 330422 – the SAFER project.

The authors would like to express their appreciation to the reviewers for their constructive comments and suggestions.

**Conflict of interest** The authors declare that they have no conflicts of interest.

## References

1. WHO, *The Importance of Pharmacovigilance: Safety Monitoring of Medicinal Products*. World Health Organization, Geneva, CH, 2002.
2. Grootheest, A., Richesson, R.. Pharmacovigilance. In Richesson, R., Andrews, J., (Eds.), *Clinical Research Informatics*, Health Informatics, pages 367–387. Springer London, 2012. doi: 10.1007/978-1-84882-448-5\_19
3. Linguist, M., The need for definitions in pharmacovigilance. *Drug Saf* 30(10):825–830, 2007. doi: 10.2165/00002018-200730100-00001
4. Council for International Organizations of Medical Sciences, *Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII*. CIOMS, Geneva, CH, 2010.
5. WHO, *A Practical Handbook on the Pharmacovigilance of Antimalarial Medicines*. World Health Organization, Geneva, CH, 2008.
6. Kelman, C.W. et al., Evaluating medicines: let's use all the evidence. *Med J Aust* 186(5):249–252, 2007.
7. Trifirò, G. et al., Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf* 18(12):1176–1184, 2009. doi: 10.1002/pds.1836
8. Gurulingappa, H., Toldo, L., Rajput, A.M., Kors, J.A., Taweel, A., Tayrouz, Y., Automatic detection of adverse events to predict drug label changes using text and data mining techniques. *Pharmacoepidemiol Drug Saf* 22(11):1189–1194, 2013. doi: 10.1002/pds.3493
9. Sarker, A. et al., Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform* 54:202–212, 2015. doi:10.1016/j.jbi.2015.02.004
10. Harpaz, R. et al., Novel data mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther* 91(6):1010–1021, 2012. doi: 10.1038/clpt.2012.50
11. Hauben, M., Norén, G.N., A decade of data mining and still counting. *Drug Saf* 33(7):527–534, 2010. doi: 10.2165/11532430-000000000-00000
12. Shang, N., Xu, H., Rindflesch, T.C., Cohen, T., Identifying plausible adverse drug reactions using knowledge

- extracted from the literature. *J Biomed Inform* 52:293–310, 2014. doi: 10.1016/j.jbi.2014.07.011
13. Koutkias, V.G., Jaulent, M.-C., Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. *Drug Saf* 38(3):219–232, 2015. doi: 10.1007/s40264-015-0278-8
  14. FDA Adverse Event Reporting System (FAERS). <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/>. Accessed 13 September 2015.
  15. EudraVigilance. <https://eudravigilance.ema.europa.eu/>. Accessed 13 September 2015.
  16. VigiBase<sup>®</sup>. <http://www.unc-products.com/>. Accessed 13 September 2015.
  17. Hauben, M., Bate, A., Decision support methods for the detection of adverse events in post-marketing data. *Drug Discov Today* 14(7-8):343–357, 2009. doi: 10.1016/j.drudis.2008.12.012
  18. Chazard, E., Ficheur, G., Bernonville, S., Luyckx, M., Beuscart, R., Data mining to generate adverse drug events detection rules. *IEEE Trans Inf Technol Biomed* 15(6):823–830, 2011. doi: 10.1109/TITB.2011.2165727
  19. Reys, J., Garibaldi, J., Aickelin, U., Soria, D., Gibson, J., Hubbard, R., Comparison of algorithms that detect drug side effects using electronic healthcare database. *Soft Comput* 17:2381–2397, 2013. doi: 10.1007/s00500-013-1097-4
  20. Madigan, D., Schuemie, M.J., Ryan, P.B., Empirical performance of the case-control design: lessons for developing a risk identification and analysis system. *Drug Saf* 36:S73–S82, 2013. doi: 10.1007/s40264-013-0105-z
  21. Ryan, P.B., Schuemie, M.J., Gruber, S., Zorych, I., Madigan, D., Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. *Drug Saf* 36:S59–S72, 2013. doi: 10.1007/s40264-013-0099-6
  22. Suchard, M.A., Zorych, I., Simpson, S.E., Schuemie, M.J., Ryan, P.B., Madigan, D., Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. *Drug Saf* 36:S83–S93, 2013. doi: 10.1007/s40264-013-0100-4
  23. Norén, N.G., Hopstadius, J., Bate, A., Star, K., Edwards, R.I., Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov* 20:361–387, 2010. doi: 10.1007/s10618-009-0152-3
  24. The Observational Medical Outcomes Partnership. <http://omop.org/>. Accessed 13 September 2015.
  25. LePendou, P. et al., Pharmacovigilance using clinical notes. *Clin Pharmacol Ther* 93:547–555, 2013. doi: 10.1038/clpt.2013.47
  26. Freifeld, C.C. et al., Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf* 37:343–350, 2014. doi: 10.1007/s40264-014-0155-x
  27. Harpaz, R. et al., Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf* 37:777–790, 2014. doi: 10.1007/s40264-014-0218-z
  28. Harpaz, R., DuMouchel, W., LePendou, P., Bauer-Mehren, A., Ryan, P., Shah, N.H., Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin Pharmacol Ther* 93:539–546, 2013. doi: 10.1038/clpt.2013.24
  29. Liu, M. et al., Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. *J Am Med Inform Assoc* 20:420–426, 2013. doi: 10.1136/amiajnl-2012-001119
  30. Schuemie, M.J. et al., Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. *Med Care* 50:890–897, 2012. doi: 10.1097/MLR.0b013e31825f63bf
  31. van Holle, L., Bauchau, V., Signal detection on spontaneous reports of adverse events following immunisation: a comparison of the performance of a disproportionality-based algorithm and a time-to-onset-based algorithm. *Pharmacoevidemiol Drug Saf* 23:178–185, 2014. doi: 10.1002/pds.3502
  32. Harpaz, R. et al., Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 20(3):413–419, 2013. doi: 10.1136/amiajnl-2012-000930
  33. Stang, P. et al., Variation in choice of study design: Findings from the epidemiology design decision inventory and evaluation (EDDIE) survey. *Drug Saf* 36:S15–S25, 2013. doi: 10.1007/s40264-013-0103-1
  34. Koutkias, V.G., Jaulent, M.-C., Leveraging post-marketing drug safety research through semantic technologies: the Pharmacovigilance Signal Detectors Ontology. In *Proceedings of the 7th International Workshop on Semantic Web Applications and Tools for Life Sciences*, CEUR Workshop Proceedings, Vol. 1320, Berlin, Germany, December 9-11, 2014.
  35. Singh, M.P., Huhns, M.N., *Service-Oriented Computing: Semantics, Processes, Agents*. Wiley, 2005.
  36. Luck, M., McBurney, P., Preist, C., A manifesto for agent technology: towards next generation computing. *Auton Agent Multi-Ag* 9(3):203–252, 2004, doi: 10.1023/B:AGNT.0000038027.29035.7c
  37. Isern, D., Sánchez, D., Moreno, A. Agents applied in health care: a review. *Int J Med Inform* 79(3):145–166, 2010. doi: 10.1016/j.ijmedinf.2010.01.003
  38. Klusch, M., Sycara, K., Brokering and matchmaking for coordination of agent societies: a survey. In Omicini, A., Zambonelli, F., Klusch, M., Tolksdorf, R., editors, *Coordination of Internet Agents*, pages 197–224. Springer-Verlag, 2001.
  39. Durfee, E.H., Distributed Problem Solving and Planning. In: *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA, MIT Press, 2000, pp. 121–164.
  40. Tran, Q.-N.N., Low, G., MOBMAS: a methodology for ontology-based multi-agent systems development. *Inform Software Tech* 50(7-8):697–722, 2008. doi:10.1016/j.infsof.2007.07.005
  41. Lindquist, M., Use of triage strategies in the WHO signal-detection process. *Drug Saf* 30(7):635–637, 2007. doi: 10.2165/00002018-200730070-00014
  42. Kuhn, M., Campillos, M., Letunic, I., Juhl Jensen, L., Bork, P., A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 6(1), 2010. doi: 10.1038/msb.2009.98
  43. The DrugBank database. <http://www.drugbank.ca/>. Accessed 13 September 2015.
  44. The ChEMBL database. <https://www.ebi.ac.uk/chembl/>. Accessed 13 September 2015.
  45. Ahmed, I., Poncet, A., *PhViD: An R package for Pharmacovigilance signal Detection*, 2013. R package version 1.0.6.
  46. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax, 2nd Ed. <http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/>. Accessed 13 September 2015.
  47. The openFDA Drug API. <https://open.fda.gov/drug/event/>. Accessed 13 September 2015.

48. Europe PubMed Central RESTful Web Service. <http://europepmc.org/restfulwebservice>. Accessed 13 September 2015.
49. The Twitter REST APIs. <https://dev.twitter.com/rest/public>. Accessed 13 September 2015.
50. SPARQL 1.1 Overview, W3C Recommendation, 21 March 2013. <http://www.w3.org/TR/sparql11-overview/>. Accessed 13 September 2015.
51. bio2RDF. <http://bio2rdf.org/>. Accessed 13 September 2015.
52. Java Agent DEvelopment framework (JADE). <http://jade.tilab.com/>. Accessed 13 September 2015.
53. The Foundation for Intelligent Physical Agents (FIPA). <http://www.fipa.org/>. Accessed 13 September 2015.
54. Rserve - Binary R server. <http://rforge.net/Rserve/>. Accessed 13 September 2015.
55. Foundation for Intelligent Physical Agents, FIPA Abstract Architecture Specification, SC00001L, <http://fipa.org/specs/fipa00001/SC00001L.pdf>. Accessed 13 September 2015.
56. Foundation for Intelligent Physical Agents, FIPA ACL Message Structure Specification, SC00061G, 03/12/2002. <http://www.fipa.org/specs/fipa00061/SC00061G.pdf>. Accessed 13 September 2015.
57. Foundation for Intelligent Physical Agents, FIPA Communicative Act Library Specification, SC00037J, 03/12/2002. <http://www.fipa.org/specs/fipa00037/SC00037J.pdf>. Accessed 13 September 2015.
58. Foundation for Intelligent Physical Agents, FIPA SL Content Language Specification, SC00008I, 03/12/2002. <http://www.fipa.org/specs/fipa00008/SC00008I.pdf>. Accessed 13 September 2015.
59. Caire, G., Cabanillas, D., JADE Tutorial: Application-defined Content Languages and Ontologies. 15 April 2010. <http://jade.tilab.com/doc/tutorials/CLOntoSupport.pdf>. Accessed 13 September 2015.
60. The clinical Text Analysis and Knowledge Extraction System (cTAKES). <http://ctakes.apache.org/>. Accessed 13 September 2015.
61. The Unified Medical Language System (UMLS). <http://www.nlm.nih.gov/research/umls/>. Accessed 13 September 2015.
62. The Unstructured Information Management Architecture (UIMA). <http://uima.apache.org/>. Accessed 13 September 2015.
63. The OMOP Methods Library. <http://omop.org/MethodsLibrary>. Accessed 13 September 2015.
64. Norén, N.G., Edwards, I.R., Modern methods of pharmacovigilance: detecting adverse effects of drugs. *Clinical Medicine* 9(5):486–489, 2009. doi: 10.7861/clinmedicine.9-5-486
65. Liu, M. et al., Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* 19(e1):28–35, 2012. doi: 10.1136/amiajnl-2011-000699
66. The Mini-Sentinel project. <http://www.mini-sentinel.org/>. Accessed 13 September 2015.
67. Observational Health Data Sciences and Informatics (OHDSI) program. <http://www.ohdsi.org/>. Accessed 13 September 2015.
68. Oliveira, J.L. et al., The EU-ADR web platform: delivering advanced pharmacovigilance tools. *Pharmacoepidemiol Drug Saf* 22(5):459–467, 2013. doi: 10.1002/pds.3375
69. González-Vélez, H. et al., Healthagents: distributed multi-agent brain tumor diagnosis and prognosis. *Appl Intell* 30(3):191–202, 2009. doi: 10.1007/s10489-007-0085-8
70. Bromuri, S., Schumacher, M.I., Stathis, K., Ruiz, J., Monitoring Gestational Diabetes Mellitus with Cognitive Agents and Agent Environments. Proc. of IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol. 2, pp. 409–414, Lyon, France, Aug. 22–27, 2011. doi: 10.1109/WI-IAT.2011.37
71. Koutkias, V., Chouvarda, I., Maglaveras, N., A multi-agent system enhancing home-care health services for chronic disease management. *IEEE Trans Inf Technol Biomed* 9(4):528–537, 2005, doi: 10.1007/978-3-662-04401-8\_8
72. Kaluža, B. et al., A multi-agent care system to support independent living. *Int J Artif Intell T* 23(1): 2014. doi:10.1142/S0218213014400016
73. Koutkias, V.G., Malousi, A., Maglaveras, N., Engineering agent-mediated integration of bioinformatics analysis tools. *Multiagent Grid Syst* 3(2):245–258, 2007.
74. Bouzguenda, L., Turki, M., Dynamic medical cross-organizational workflow management system. *J Med Syst* 38:32, 2014. doi: 10.1007/s10916-014-0032-2
75. Cortés, U., Vázquez-Salceda, J., López-Navidad, A., Caballero, F., UCTx: a multi-agent system to assist a transplant coordination unit. *Appl Intell* 20(1):59–70, 2004. doi: 10.1023/B:APIN.0000011142.91514.57
76. Tian, Y., Zhou, T.-S., Yao, Q., Zhang, M., Li, J.-S., Use of an agent-based simulation model to evaluate a mobile-based system for supporting emergency evacuation decision making. *J Med Syst* 38:149, 2014. doi: 10.1007/s10916-014-0149-3
77. Ji, Y. et al., A distributed adverse drug reaction detection system using intelligent agents with a fuzzy recognition-primed decision model. *Int J Intell Syst* 22(8):827–845, 2007. doi: 10.1002/int.20230
78. Nyulas, C.I., O'Connor, M.J., Tu, S.W., Buckeridge, D.L., Okhmatovskaia, A., Musen, M.A., An ontology-driven framework for deploying JADE agent systems. Proc. of IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol. 2, pp. 573–577, Sydney, Australia, Dec. 9–12, 2008. doi: 10.1109/WIAT.2008.25
79. Ferrucci, D. et al., Building Watson: an overview of the DeepQA project. *AI Magazine*, 31(3):59–79, 2010. doi: 10.1609/aimag.v31i3.2303
80. Santos Jr, E., Wilkinson, J.T., Santos, E., Fusing multiple Bayesian knowledge sources. *Int J Approx Reason*, 52(7):935–947, 2011. doi: 10.1016/j.ijar.2011.01.008
81. He, H., Cao, Y., Wen, J., Cheng, S., A boost voting strategy for knowledge integration and decision making. In Sun, F., Zhang, J., Tan, Y., Cao, J., Yu, W., editors, *Advances in Neural Networks - ISNN 2008*, volume 5263 of *Lecture Notes in Computer Science*, pages 472–481. Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-87732-5\_53
82. DeployR. <http://deployr.revolutionanalytics.com/>. Accessed 13 September 2015.
83. Burian, P., Multi-agent systems and cloud computing for controlling and managing chemical and food processes. *J Chem Chem Eng* 6:1121–1135, 2012.